CRM Series in Mathematical Physics

M. B. Paranjape · Richard MacKenzie
Zora Thomova · Pavel Winternitz
William Witczak-Krempa *Editors*

Quantum Theory and Symmetries

Proceedings of the 11th International Symposium, Montréal, Canada

Springer

CRM Series in Mathematical Physics

Editorial Board:

Joel S. Feldman
Department of Mathematics
University of British Columbia
Vancouver, BC, Canada
feldman@math.ubc.ca

Duong H. Phong
Department of Mathematics
Columbia University
New York, NY, USA
phong@math.columbia.edu

Yvan Saint-Aubin
Département de Mathématiques et de Statistique
C.P. 6128, Succursale Centre-ville Montréal
Université de Montréal
Montréal, QC, Canada
saint@dms.umontreal.ca

Luc Vinet
Département de Physique
C.P. 6128, Succursale Centre-ville Montréal
CRM, Université de Montréal
Montréal, QC, Canada
vinet@crm.umontreal.ca

The Centre de recherches mathématiques (CRM) was created in 1968 by the Université de Montréal to promote research in the mathematical sciences. It is now a national institute that hosts several groups and hold special theme years, summer schools, workshops, and a postdoctoral program. The focus of its scientific activities ranges from pure to applied mathematics and includes statistics, theoretical computer science, mathematical methods in biology and life sciences, and mathematical and theoretical physics. The CRM also promotes collaboration between mathematicians and industry. It is subsidized by the Natural Sciences and Engineering Research Council in Canada, the Fonds FCAR of the Province of Québec, and the Canadian Institute for Advanced Research and has private endowments. Current activities, fellowships, and annual reports can be found on the CRM Web page at www.CRM.UMontreal.CA.

The CRM Series in Mathematical Physics includes monographs, lecture notes, and proceedings based on research pursued and events held at the Centre de recherches mathématiques.

More information about this series at http://www.springer.com/series/3872

M. B. Paranjape • Richard MacKenzie
Zora Thomova • Pavel Winternitz
William Witczak-Krempa
Editors

Quantum Theory and Symmetries

Proceedings of the 11th International Symposium, Montréal, Canada

Springer

Editors
M. B. Paranjape
Département de physique
Université de Montréal
Montréal, QC, Canada

Centre de Recherches Mathématiques
Université de Montréal
Montréal, QC, Canada

Zora Thomova
Department of Mathematics and Physics
SUNY Polytechnic Institute
Utica, NY, USA

William Witczak-Krempa
Département de physique
Université de Montréal
Montréal, QC, Canada

Centre de Recherches Mathématiques
Université de Montréal
Montréal, QC, Canada

Richard MacKenzie
Département de physique
Université de Montréal
Montréal, QC, Canada

Pavel Winternitz
Department of Mathematics
and Statistics
Université de Montréal
Montréal, QC, Canada

CRM Series in Mathematical Physics
ISBN 978-3-030-55776-8 ISBN 978-3-030-55777-5 (eBook)
https://doi.org/10.1007/978-3-030-55777-5

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is the proceedings volume of the Symposium Quantum Theory and Symmetries-XI (QTS-XI) that was held in Montréal from July 1st to 5th, 2019. The symposium consisted of nine sessions including a special session in honour of Professor Decio Levi of the University Roma Tre and the sessions: Algebraic Methods, Condensed Matter Physics, Cosmology and Gravitation, Integrability, Non-perturbative QFT, Particle Physics, Quantum Computing and Quantum Information Theory and String Theory/AdS-CFT. Each session was associated with one or two plenary speakers followed by parallel sessions with speakers of invited and contributed talks. For complete information, please visit the website: http://www.crm.umontreal.ca/2019/QTS2019/index_e.php. Several of the plenary speakers and many of the invited and contributing speakers have supplied a contribution to this proceedings volume. It is intended for use by students, researchers and professors in a wide range of fields from mathematical physics to more mainstream physics as evinced by the topics of the various parallel sessions. It should allow the reader to grasp the breadth and scope of the vast field of quantum theory and symmetries.

As these proceedings were being completed, our dear colleague, mentor and friend Pavel Winternitz passed away. We wish to dedicate this volume to him in recognition of his scientific legacy and as an expression of our gratitude.

Montréal, QC, Canada — M. B. Paranjape

Montréal, QC, Canada — Richard MacKenzie

Utica, NY, USA — Zora Thomova

Montréal, QC, Canada — Pavel Winternitz

Montréal, QC, Canada — William Witczak-Krempa

June 2020

Organization

QTS-XI was organized at the Centre de recherches mathématiques, Université de Montréal, Montréal, Québec, Canada from July 5th to 9th, 2019.

Organizing Committee

Conference Chair	M. B. Paranjape (Université de Montréal, Québec, Canada)
Co-Chair	Richard MacKenzie (Université de Montréal, Québec, Canada)
Co-Chair	Yvan Saint-Aubin (Université de Montréal, Québec, Canada)
Co-Chair	Alfred Shapere (University of Kentucky, USA)
Co-Chair	Luc Vinet (Université de Montréal, Québec, Canada)
Co-Chair	William Witczak-Krempa (Université de Montréal, Québec, Canada)

International Advisory Board

Daniel Arovas (University of California San Diego, USA)
Robert Brandenberger (McGill University, Canada)
Jean-Sébastien Caux (University of Amsterdam, NL)
Eric D'Hoker (University of California, Los Angeles, USA)
Philippe Di Francesco (University of Illinois at Urbana-Champaign, USA)
Louise Dolan (University of North Carolina, USA)
Ivette Fuentes Guridi (University of Vienna, Austria)
Shohini Ghose (Wilfrid Laurier University, Canada)
Bum-Hoon Lee (Sogang University, South Korea)
Kimyeong Lee (Korean Institute of Advanced Research, South Korea)
Karyn Le Hur (École Polytechnique, France)
Rob Myers (Perimeter Institute, Canada)

Antti Niemi (Uppsala University, Sweden)
Sumathi Rao (Harish-Chandra Research Institute, India)
Nicolai Reshetikhin (University of California, Berkeley, USA)
Subir Sachdev (Harvard University, USA)
Hubert Saleur (University of Southern California, USA)
Gordon Semenoff (University of British Columbia, Canada)
Ashoke Sen (Harish-Chandra Research Institute, India)
Steve Simon (University of Oxford, UK)
Nandini Trivedi (Ohio State University, USA)
Erick Weinberg (Columbia University, USA)
Frank Wilczek (Massachusetts Institute of Technology, USA)

QTS Conference Board

Cestmir Burdik (Czech Technical University in Prague)
Vladimir K. Dobrev (Bulgarian Academy of Sciences—Chairman)
Heinz-Dietrich Doebner (Technische Universität Clausthal—Honorary Chairman)
Mariano del Olmo (Universidad de Valladolid)
George Pogosyan (Universidad de Guadalajara and Yerevan State University)
Alfred Shapere (University of Kentucky)
Peter Suranyi (University of Cincinnati)
Rohana Wijewardhana (University of Cincinnati)
Kurt Bernardo Wolf (UNAM)

Plenary Speakers

Glenn Barnich (Université Libre de Bruxelles)
Fiona Burnell (University of Minnesota)
Freddy Cachazo (Perimeter Institute for Theoretical Physics)
Alejandra Castro Anich (University of Amsterdam)
Jean-Sébastien Caux (University of Amsterdam)
Chihway Chang (University of Chicago)
Giulio Chiribella (The University of Hong Kong)
Veronika E. Hubeny (University of California, Davis)
Rinat Kedem (University of Illinois at Urbana-Champaign)
Yi Li (The Johns Hopkins University)
Joseph D. Lykken (Fermi National Accelerator Laboratory)
David Poland (Yale University)
Miguel A. Rodríguez (Universidad Complutense de Madrid)
Subir Sachdev (Harvard University)
Daniele Valeri (University of Glasgow)

Claude-Michel Viallet (Centre National de la Recherche Scientifique, Sorbonne Université)
Mark B. Wise (California Institute of Technology)

Session Organizers

Special Session in Honour of Decio Levi Integrability: Continuous and Discrete, Classical and Quantum	Nalini Joshi (The University of Sydney) Zora Thomova (SUNY Polytechnic Institute) Pavel Winternitz (Université de Montréal)
Algebraic Methods	Vincent Bouchard (University of Alberta) Sarah Post (University of Hawai'i)
Condensed Matter Physics	Nandini Trivedi (Ohio State University) William Witczak-Krempa (Université de Montréal)
Cosmology and Gravitation	Robert Brandenberger (McGill University) Shahin M.M. Sheikh-Jabbari (Institute for Research in Fundamental Sciences, Iran)
Integrability	Hubert Saleur (CEA-Saclay/University of Southern California) Robert Weston (Heriot-Watt University)
Non-perturbative QFT	Simon Caron-Huot (McGill University) Andrew Liam Fitzpatrick (Boston University)
Particle Physics	Geneviève Bélanger (CNRS, Université Savoie Mont Blanc) Thomas Grégoire (Carleton University)
Quantum Computing and Quantum Information Theory	Frédéric Dupuis (CNRS, Loria) Christine Muschik (University of Waterloo)
String Theory/AdS-CFT	Keshav Dasgupta (McGill University) Anne Taormina (Durham University)

Sponsors

Centre de recherches mathématiques, Université de Montréal, Montréal, QC, Canada
Institute Transdisciplinaire d'Information Quantique, Polytechnique Montréal, Montréal, QC, Canada
Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada
Asia Pacific Center for Theoretical Physics, Pohang, Republic of Korea
Laboratoire de physique mathématique, Université de Montréal, Montréal, QC, Canada
International Union of Pure and Applied Physics, Singapore
Theoretical Physics Institute, University of Alberta, Edmonton, AB, Canada

Statement of Principles

Free Circulation of Scientists The principle of the Universality of Science is fundamental to scientific progress. This principle embodies the freedom of movement, association, expression and communication for scientists, as well as equitable access to data, information and research materials. In pursuing its objectives with respect to the rights and responsibilities of scientists, the International Union of Pure and Applied Physics (IUPAP) actively upholds this principle and, in so doing, opposes any discrimination on the basis of such factors as ethnic origin, religion, citizenship, language, political stance, gender or age. IUPAP should only sponsor conferences and events at institutions and in countries that uphold this principle. If scientists are excluded from attending IUPAP-sponsored international conferences by a host institution or country on the basis of any of these factors, IUPAP should register its concern at the highest level of that institution or country and should not sponsor any future events in that country until such exclusions have been eliminated.

Harassment at Conferences It is the policy of the International Union of Pure and Applied Physics (IUPAP) that all participants at an IUPAP-supported Conference will enjoy a comfortable experience, and that they will treat each other with respect at all times. The conference organizers will name an advisor who will consult with those who have suffered from harassment and who will suggest ways of redressing their problems, and an advisor who will counsel those accused of harassment.

Contents

Part I
Special Session in Honour of Decio Levi: Integrability: Continuous and Discrete, Classical & Quantum

Spin Chains of Haldane–Shastry Type: A Bird's Eye View

Federico Finkel, Artemio González-López, and Miguel A. Rodríguez

Dedicated to our friend and colleague Decio Levi in his 70th anniversary

Abstract We present a brief report on the main properties of Haldane–Shastry type spin chains and their relation with Calogero–Sutherland spin dynamical models. Recent work on the thermodynamics of these chains is also discussed.

Keywords Integrable systems · Spin chains · Statistical mechanics

1 Introduction

This is a short review of the construction and main properties of integrable spin chains of Haldane–Shastry type stemming from their relation with many-body spin Calogero–Sutherland models, briefly outlining the exact evaluation of their partition function and studying their thermodynamics.

The Haldane–Shastry chain was introduced independently by Haldane [16] and Shastry [24] in the late eighties in connection with the one-dimensional Hubbard model. The explicit computation of its spectrum and its remarkable mathematical properties attracted very quickly the attention of many researchers in condensed matter physics. Its extension to other similar chains was carried out through their relation with spin Calogero–Sutherland models, and many of its integrability properties were also related to those of the latter models. The classification of Calogero–Sutherland type models by Olshanetsky and Perelomov [21], and their relation with the theory of homogeneous spaces and simple Lie algebras and their root systems, was the clue to understanding their properties in a broader context. The freezing trick, as introduced by Polychronakos in [23], makes it possible to

F. Finkel · A. González-López · M. A. Rodríguez (✉)
Depto. de Física Teórica, Universidad Complutense de Madrid, Madrid, Spain
e-mail: ffinkel@ucm.es; artemio@ucm.es; rodrigue@ucm.es

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_1

explain the construction of these chains starting from a spin dynamical model based on those studied by Olshanetsky and Perelomov and the evaluation of their partition function. In this contribution we shall outline the main steps of this construction, and conclude with some recent applications to the study of the thermodynamics of these chains.

2 Calogero–Sutherland Models

The study of integrable systems, in the classical sense of the Liouville–Arnol'd theory, has occupied the activity of many researchers working in different fields of physics and mathematics. However, many-body systems satisfying the required integrability properties are rarely found in physics. Thus the construction of one-dimensional models of many particles by Calogero and Sutherland in the sixties was a breakthrough, and represented a major contribution to this field, as shown by their impressive number of applications of Calogero–Sutherland models in so many areas of physics.[1] We will introduce in this section a short account of these models, thus paving the way for the construction of long-range integrable spin chains based on them.

2.1 The Calogero Model

The Calogero model [5]

$$H = -\sum_i \partial_{x_i}^2 + \omega^2 \sum_i x_i^2 + \sum_{i \neq j} \frac{a(a-1)}{(x_i - x_j)^2} \tag{1}$$

describes a system of N particles on the line with an inverse-square interaction potential. In the previous formulas all sums are understood to run from 1 to N, $\mathbf{x} \equiv (x_1, \ldots, x_N)$ and $a > 1/2$ is the system's coupling constant. Its classical version

$$H = \sum_i p_i^2 + \omega^2 \sum_i x_i^2 + \frac{a(a-1)}{(x_i - x_j)^2}$$

is integrable (in fact, superintegrable [26]), as can be shown by Moser using Lax pair techniques [20]. In the quantum case, the ground state is given by the Jastrow-type expression

[1] Bill Sutherland, Francesco Calogero, and Michel Gaudin were recently awarded the 2019 Dannie Heineman Prize of the American Institute of Physics and the American Physical Society for their seminal contributions to statistical mechanics and many-body physics.

$$\mu(\mathbf{x}) = \mathrm{e}^{-\frac{1}{2}\omega\sum_i x_i^2} \prod_{i<j} |x_i - x_j|^a, \tag{2}$$

in the region $x_1 < x_2 < \cdots < x_N$. The spectrum can be algebraically computed, with the result:

$$E_{\mathbf{n}} = 2\omega \sum_i n_i + E_0, \quad E_0 = [1 + a(N-1)]N\omega,$$

where E_0 is the ground state energy and the multiindex $\mathbf{n} = (n_1, \ldots, n_N)$ satisfies $n_1 \geqslant \cdots \geqslant n_N \geqslant 0$. The ground state can be used as a gauge function, leading to the gauged Hamiltonian

$$H_G = \mu(\mathbf{x})^{-1} H \mu(\mathbf{x}) = -\sum_i \partial_{x_i}^2 + 2\omega \sum_i x_i \partial_{x_i} - 2a \sum_{i<j} \frac{1}{x_i - x_j} \left(\partial_{x_i} - \partial_{x_j}\right).$$

This expression is crucial in the study of the dynamical spin models, where Dunkl operators [8] discussed below also play an important role.

The above simple closed formula for the energy spectrum of the Calogero model allows one to evaluate its partition function in closed form. Indeed,

$$Z(2\omega T) = \sum_{n_1 \geqslant \cdots \geqslant n_N \geqslant 0} \mathrm{e}^{-E_{\mathbf{n}}/(2\omega k_{\mathrm{B}} T)} = q^{E_0/(2\omega)} \sum_{n_1 \geqslant \cdots \geqslant n_N \geqslant 0} q^{\sum_i n_i}, \ q = \mathrm{e}^{-1/(k_{\mathrm{B}} T)},$$

where T is the temperature and k_{B} Boltzmann's constant. If we define the indices k_i as

$$k_i = n_i - n_{i+1}, \ i = 1, \ldots, N-1, \quad k_N = n_N, \quad \sum_i n_i = \sum_j j k_j,$$

a straightforward computation yields

$$Z(2\omega T) = q^{E_0/(2\omega)} \sum_{k_1 \geqslant 0, \ldots, k_N \geqslant 0} q^{\sum_j j k_j} = \prod_{j=1}^{N} \frac{1}{1 - q^j}. \tag{3}$$

2.2 The Sutherland Model

The Sutherland model [25] describes a system of N particles on a circle, the potential being now a trigonometric function with singularities at the collision points:

$$H = -\sum_i \partial_{x_i}^2 + a(a-1)\sum_{i\neq j} \frac{1}{\sin^2(x_i - x_j)}. \tag{4}$$

The ground state is again of Jastrow type,

$$\mu(\mathbf{x}) = \prod_{i<j} |\sin(x_i - x_j)|^a, \quad E_0 = \frac{1}{3}N(N^2-1)a^2,$$

and the gauged Hamiltonian, after the change of variables $z_j = \mathrm{e}^{2\mathrm{i}x_j}$, $j = 1, \ldots, N$, reads:

$$H_G = \mu^{-1} H \mu = 4\sum_i z_i^2 \partial_{z_i}^2 + 4\sum_i z_i \partial_{z_i} - 4a(a-1)\sum_{i\neq j} \frac{z_i z_j}{(z_i - z_j)^2}.$$

As in the case of the Calogero model, the spectrum can be algebraically computed, with the result:

$$E_{\mathbf{n}} = \sum_i (2n_i + (N+1-2i)a)^2,$$

where now $n_i \in \mathbb{R}$ with $n_1 \geqslant \cdots \geqslant n_N$ and (in the center of mass frame) $\sum_i n_i = 0$. Removing the energy of the ground state, the dominant term of $E_{\mathbf{n}}$ when $a \to \infty$ is

$$E_{\mathbf{n}} \underset{a\to\infty}{\sim} 4a\sum_i n_i(N+1-2i) = 4a\sum_j j(N-j)k_j\,,$$

where again $k_i = n_i - n_{i+1}$ for $i \leqslant N-1$ and $k_N = n_N$. It follows that

$$\lim_{a\to\infty} Z(4aT) = \prod_{j=1}^{N-1} \frac{1}{1-q^{j(N-j)}}. \tag{5}$$

This expression is useful for the computation of the partition function of the Haldane–Shastry spin chain [10].

3 Spin Dynamical Models

The one-dimensional systems we have described in the previous sections can be extended to incorporate the spin of the particles (where spin must be understood as internal degrees of freedom, not necessarily su(2) spin) while keeping their integrability properties.

To introduce the spin of the particle, we must consider wavefunctions which are linear combinations of the product $\varphi(\mathbf{x})|\mathbf{s}\rangle$ of a scalar function and a spin state $|\mathbf{s}\rangle$, where

$$|\mathbf{s}\rangle \equiv |s_1, \ldots, s_N\rangle \in \Sigma_m \equiv \bigotimes_{i=1}^{N} \mathbf{C}^m,$$

and $s_i \in \{1, \ldots, m\}$ is the value of the spin of the i-th particle. In order to describe the spin interaction, we introduce spin terms in the Hamiltonian through the spin exchange operators S_{ij} defined by

$$S_{ij}|s_1, \ldots, s_i, \ldots, s_j, \ldots, s_N\rangle = |s_1, \ldots, s_j, \ldots, s_i, \ldots, s_N\rangle\,. \tag{6}$$

For the sake of simplicity, in the rest of the section we shall restrict ourselves to the Calogero spin dynamical model.

3.1 The A_{N-1} Spin Calogero Model

The spin dynamical Hamiltonian corresponding to the extension of the Calogero model can be expressed as

$$H_\epsilon = -\sum_i \partial_{x_i}^2 + a\sum_{i\neq j} \frac{1}{(x_i - x_j)^2}(a - \epsilon S_{ij}) + \omega^2 \sum_i x_i^2, \quad \epsilon = \pm 1. \tag{7}$$

In order to compute its spectrum we introduce the coordinate permutation operators P_{ij}, which act on a scalar function as

$$P_{ij}\varphi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N) = \varphi(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_N).$$

We then introduce the scalar operator

$$H^P = -\sum_i \partial_{x_i}^2 + a\sum_{i\neq j} \frac{1}{(x_i - x_j)^2}(a - P_{ij}) + \omega^2 \sum_i x_i^2,$$

formally obtained by replacing ϵS_{ij} by P_{ij} in the expression (7) for H_ϵ. Using the ground state wavefunction, the operator can be gauge-transformed into

$$H_G^P = \mu^{-1} H^P \mu = -\sum_i \partial_{x_i}^2 + 2\omega \sum_i x_i \partial_{x_i} - 2a \sum_{i<j} \frac{1}{x_i - x_j}(\partial_{x_i} - \partial_{x_j})$$
$$+ a\sum_{i\neq j} \frac{1}{(x_i - x_j)^2}(1 - P_{ij}) + E_0. \tag{8}$$

The operator H_G^P can in turn be expressed in terms of the Dunkl operators

$$J_i = \partial_{x_i} + a \sum_{i \neq j} \frac{1}{x_i - x_j}(1 - P_{ij}) \tag{9}$$

as

$$H_G^P = -\sum_i J_i^2 + 2\omega \sum_i x_i \partial_{x_i} + E_0. \tag{10}$$

The operators (9), introduced by Dunkl [8] in connection with the theory of orthogonal polynomials in several variables, are related to the reflection groups of root systems (in this case, A_{N-1}). These operators form a commuting family, i.e., $[J_i, J_k] = 0$. More importantly for the purposes of this work, they also have the key property of leaving invariant certain polynomial modules [11].

The spectrum of the operator H_G^P can be readily computed from the commutativity of the Dunkl operators J_i, taking advantage of the fact that their action on the invariant modules can be easily triangularized. In this way one can show that the spectrum of the operator H^P is given by the expression

$$E_{\mathbf{n}} = 2\omega \sum_i n_i + E_0,$$

where, by contrast with the scalar Calogero model, the numbers n_i are arbitrary nonnegative integers. If Λ_ϵ is the total symmetrizer (if $\epsilon = 1$) or skew-symmetrizer (if $\epsilon = -1$) in both coordinates and spins, the identity $H_\epsilon \Lambda_\epsilon = H^P \Lambda_\epsilon$ clearly holds. Thus the energies of the dynamical spin Hamiltonian H_ϵ and the operator H^P coincide, although their degeneracies are different due to the spin degrees of freedom.

Following the same steps as in the scalar case, it is not difficult to find the partition function of the Hamiltonian H_ϵ. The intrinsic spin degeneracy of an energy level can be determined as the number of ways of assigning one of the m spin values s_i to each of the components of the multiindex

$$\mathbf{n} = (\underbrace{p_1, \ldots, p_1}_{k_1}, \ldots, \underbrace{p_r, \ldots, p_r}_{k_r}), \qquad \text{with } p_1 > \cdots > p_r \geqslant 0, \quad \sum_{i=1}^r k_i = N,$$

in such a way that in each constant sector $p_i, \ldots, p_i$ we have either a strictly increasing (for $\epsilon = -1$) or nondecreasing (for $\epsilon = 1$) sequence of spin values. Thus the intrinsic spin degeneracy of the energy $E_{\mathbf{n}}$ is given by

$$D_\epsilon(\mathbf{k}; m) \equiv \prod_{i=1}^r d_\epsilon(k_i; m)\,, \qquad \text{with} \quad d_\epsilon(k_i; m) = \binom{m + \delta_{\epsilon 1}(k_i - 1)}{k_i}.$$

Calling $\mathbf{k} = (k_1, \ldots, k_r)$, the partition function of the Hamiltonian Z_ϵ can be expressed as

$$Z_\epsilon(2\omega T) = q^{E_0/(2\omega)} \sum_{\mathbf{k}\in\mathcal{P}_N} \sum_{p_1>\cdots>p_r\geqslant 0} D_\epsilon(\mathbf{k}; m)\, q^{\sum_{i=1}^r k_i p_i} ,$$

where $\mathcal{P}_N$ denotes the set of partitions of the integer N taking order into account. After a straightforward calculation the latter expression yields the closed formula [10]

$$Z_\epsilon(2\omega T) = \frac{q^{E_0/(2\omega)}}{1-q^N} \sum_{\mathbf{k}\in\mathcal{P}_N} D_\epsilon(\mathbf{k}; m) \prod_{i=1}^{r-1} \frac{q^{K_i}}{1-q^{K_i}} , \tag{11}$$

where we have set

$$K_i \equiv \sum_{j=1}^{i} k_i .$$

Similar results are obtained for the Sutherland (trigonometric) model (4) with spin degrees of freedom.

4 Spin Chains of Haldane–Shastry Type

By a spin chain we usually understand a one-dimensional lattice whose sites are occupied by particles with internal degrees of freedom. In this work we shall deal exclusively with spin chains with long-range interactions, involving all the sites. Moreover, the chain sites must be chosen in a very specific way, which is critical for ensuring the symmetry and solvability properties of the models.

For the sake of simplicity, in what follows we shall focus on the rational Calogero model of A_{N-1} type and its associated spin chain introduced below, although the methods applied can be used for any of the models we have previously discussed. Let us then consider the two Hamiltonians we have previously studied, namely the scalar one H in Eq. (1) and the spin dynamical Hamiltonian H_ϵ in Eq. (7). Defining the scalar potential

$$U(\mathbf{x}) = \sum_{i\neq j} \frac{1}{(x_i - x_j)^2} + \sum_i x_i^2 \tag{12}$$

and taking $\omega = a$, we can write the Hamiltonian (7) as

$$H_\epsilon = -\sum_i \partial_{x_i}^2 + a^2 U(\mathbf{x}) - a \sum_{i \neq j} \frac{1}{(x_i - x_j)^2} + 2a h_\epsilon(\mathbf{x}) = H + 2a h_\epsilon(\mathbf{x})) \tag{13}$$

with

$$h_\epsilon(\mathbf{x}) = \epsilon \sum_{i<j} \frac{1 - \epsilon S_{ij}}{(x_i - x_j)^2} .$$

In the limit $a \to \infty$, the wavefunctions of the scalar Hamiltonian $H_{\rm sc}$ (1) are sharply peaked around the minimum $\boldsymbol{\xi}$ of the potential $U(\mathbf{x})$, which can be shown to be unique. Thus in this limit the dynamical and spin degrees of freedom decouple, the latter being governed by the Hamiltonian

$$\mathcal{H}_\epsilon = h_\epsilon(\boldsymbol{\xi}) \equiv \sum_{i<j} \frac{1}{(\xi_i - \xi_j)^2} (1 - \epsilon S_{ij}) . \tag{14}$$

The latter model, which is the rational analogue of the Haldane–Shastry spin chain, is known in the literature as the Polychronakos–Frahm (PF) spin chain [13, 22]. We shall prove in the next sections how this connection between the PF chain and the spin Calogero model can be used to evaluate the partition function of the latter chain in closed form.

4.1 The Chain Sites

In order to complete the construction of the spin chain (14), we still have to compute the minimum of the potential $U(\mathbf{x})$ in Eq. (12) which determines the position of its sites. Since this minimum is clearly also a maximum of the ground state wavefunction (2) of the scalar Calogero model (with $\omega = a$), it is straightforward to derive the following system of algebraic equations satisfied by the coordinates of $\boldsymbol{\xi}$:

$$\sum_{j,\, j \neq i} \frac{1}{\xi_i - \xi_j} - \xi_i = 0, \quad i = 1, \ldots N.$$

As is well known by the results of Stieltjes, and later Calogero and collaborators [1], the solution of this system (which is unique, up to ordering and an overall translation) is the set of zeros of the Hermite polynomial of degree N.

4.2 The Partition Function

The partition function of the spin chain (14) can be computed in a direct way using the partition functions of the models (1) and (7), as first pointed out by Polychronakos [23].

To this end, let $\varphi_i(\mathbf{x})$ be an eigenfunction of the scalar Hamiltonian (1) with energy E_i and $|j\rangle$ an eigenfunction of the spin chain Hamiltonian $\mathcal{H}_\epsilon$ (14) with energy $\mathcal{E}^\epsilon_j$. Since $\varphi_i(\mathbf{x})$ becomes sharply peaked at $\boldsymbol{\xi}$ as $a \to \infty$, in this limit we have

$$\begin{aligned} H_\epsilon(\varphi_i(\mathbf{x})|j\rangle) &= H(\varphi_i(\mathbf{x})|j\rangle) + 2a h_\epsilon(\mathbf{x})(\varphi_i(\mathbf{x})|j\rangle) \simeq E_i\varphi_i(\mathbf{x})|j\rangle + 2a h_\epsilon(\boldsymbol{\xi})(\varphi_i(\mathbf{x})|j\rangle) \\ &= E_i\varphi_i(\mathbf{x})|j\rangle + 2a\varphi_i(\mathbf{x})\mathcal{H}_\epsilon|j\rangle = \left(E_i + 2a\,\mathcal{E}^\epsilon_j\right)\varphi_i(\mathbf{x})|j\rangle\,. \end{aligned} \tag{15}$$

Thus for $a \to \infty$ the wavefunction $\varphi_i(\mathbf{x})|j\rangle$ is an approximate eigenfunction of H_ϵ with energy

$$E^\epsilon_{ij} \simeq E_i + a\,\mathcal{E}^\epsilon_j, \tag{16}$$

and hence the energies of the PF chain (14) can in principle be expressed in terms of those of the scalar and spin Calogero models through the formula

$$\mathcal{E}^\epsilon_j = \lim_{a\to\infty} \frac{E^\epsilon_{ij} - E_i}{2a}\,.$$

Unfortunately, however, we have no rule for determining what is the relation between the indices i, j in the previous formula, i.e. what energies of the scalar and spin Calogero models should be combined to obtain an energy of the PF spin chain. Remarkably, this problem can be bypassed using the partition function. Indeed, from Eq. (16) we can easily deduce the expression

$$\mathcal{Z}_\epsilon(T) = \lim_{a\to\infty} \frac{Z_\epsilon(2aT)}{Z(2aT)}\,, \tag{17}$$

which provides an efficient way for computing the partition function of the PF spin chain (14). Indeed, using Eqs. (5)–(11) for the partition functions of the scalar and spin Calogero models we readily arrive at the closed formula

$$\mathcal{Z}_\epsilon = \sum_{\mathbf{k}\in\mathcal{P}_N} D_\epsilon(\mathbf{k};m)\, q^{\sum_{i=1}^{r-1} K_i} \prod_{i=1}^{N-r}(1-q^{K'_i}),$$

where K'_i is defined by

$$\{K'_1,\ldots,K'_{N-r}\} = \{1,\ldots,N\}\setminus\{K_1,\ldots,K_r = N\}\,.$$

The procedure outlined above can be applied to the Sutherland model (4) and its spin version, which yields the original Haldane–Shastry (HS) spin chain [16, 24]. Likewise, from the hyperbolic Calogero–Sutherland model introduced by Inozemtsev [18] one obtains a hyperbolic counterpart of the HS chain usually known as the Frahm–Inozemtsev (FI) chain [14]. The Hamiltonians of PF, HS, and FI chains can be written in a unified way as

$$\mathcal{H}_\epsilon = \sum_{i<j} J_{ij}(1 - \epsilon S_{ij}), \tag{18}$$

where the couplings J_{ij} are of the form $J_{ij} = Jg(\xi_i - \xi_j)$ with appropriate choices of the interaction potential g and the chain sites ξ_i. More precisely,

$$g(x) = \begin{cases} x^{-2}, & \text{for the PF chain} \\ \frac{1}{2}\sin^{-2} x, & \text{for the HS chain} \\ \frac{1}{2}\sinh^{-2} x, & \text{for the FI chain} \end{cases} \tag{19}$$

and

$$\begin{cases} \xi_i = i\text{-th zero of the Hermite polynomial } H_N, & \text{for the PF chain} \\ \xi_i = \frac{i\pi}{N}, & \text{for the HS chain} \\ \mathrm{e}^{2\xi_i} = i\text{-th zero of the Laguerre polynomial } L_N^{c-1} & \text{for the FI chain} \end{cases} \tag{20}$$

with c a positive parameter. In particular, we see that the chain sites of spin chains of all of these chains coincide with the set of zeros of a family of classical orthogonal polynomials.[2]

The partition function of all three chains of HS type (18) has been computed in closed form [2, 3, 10], and can be written in the unified way

$$\mathcal{Z}_\epsilon = \sum_{\mathbf{k}\in\mathcal{P}_N} d_\epsilon(\mathbf{k}; m)\, q^{\sum_{i=1}^{r-1}\mathcal{E}(K_i)} \prod_{i=1}^{N-r}(1 - q^{\mathcal{E}(K_i')}), \tag{21}$$

where the dispersion function $\mathcal{E}$ is defined by

$$\mathcal{E}(i) = \begin{cases} Ji, & \text{for the PF chain} \\ Ji(N-i), & \text{for the HS chain} \\ Ji(c+i-1), & \text{for the FI chain}. \end{cases} \tag{22}$$

[2] As is well known, the points $\cos(i\pi/N)$ with $i = 1, \ldots, N-1$ are the roots of the Chebyshev polynomial U_{N-1}.

5 Supersymmetric Spin Chains

In the spin models discussed so far the internal degrees of freedom were all fermionic (if $\epsilon = -1$) or bosonic (if $\epsilon = 1$). Moreover, in either case the permutation operator S_{ij} belongs to the enveloping algebra of the defining representation of $\mathfrak{su}(m)$, so these models are naturally regarded as being of su(m) type. We shall introduce in this section a more general class of su($m|n$) supersymmetric spin models, in which the first m internal degrees of freedom are bosonic and the last n fermionic. We shall mainly focus on the models of this type associated with the root system A_{N-1}, for which we shall briefly outline the computation of the spectrum and the exact evaluation of the partition function [4].

5.1 The su($m|n$) *Supersymmetric Exchange Operator*

As we did in the purely fermionic or bosonic cases, we shall start by constructing the spin exchange operators appearing in the Hamiltonian. The possible values of the spin now run from 1 to $m + n$, and the particle will be a boson (fermion) if $s_i \in \{1, \ldots, m\}$ ($s_i \in \{m+1, \ldots, m+n\}$). The exchange operator is then defined as

$$S_{ij}^{(m|n)} |s_1, \ldots, s_i, \ldots, s_j, \ldots, s_N\rangle = \epsilon_{ij}(\mathbf{s}) |s_1, \ldots, s_j, \ldots, s_i, \ldots, s_N\rangle,$$

where

$$\epsilon_{ij}(\mathbf{s}) = \begin{cases} 1, & s_i, s_j \text{ bosons} \\ (-1)^p, & \{s_i, s_j\} = \{\text{fermion, boson}\}, \text{ with } p = \text{number of} \\ & \qquad\qquad \text{fermions in positions } i+1, \ldots, j-1 \\ -1, & s_i, s_j \text{ fermions.} \end{cases}$$

By analogy with the purely bosonic or fermionic case, the Hamiltonian is taken as

$$\mathcal{H}_0^{(m|n)} = \sum_{i<j} J_{ij} (\mathbf{1} - S_{ij}^{(m|n)}), \tag{23}$$

where the coupling constants $J_{ij} = Jg(\xi_i - \xi_j)$ and chain sites ξ_i are still defined by Eqs. (19)–(20). As a matter of fact, it is easily checked that with this definition $\mathcal{H}_0^{(m|0)} = \mathcal{H}_1$ while $\mathcal{H}_0^{(0|m)} = \mathcal{H}_{-1}$.

5.2 Chemical Potential

As usual in the description of thermodynamic systems, we shall add a chemical potential term to the Hamiltonian (23) to control the number of particles of different species. More precisely, we define

$$\mathcal{H}_\mu = -\sum_{\alpha=1}^{m+n-1} \mu_\alpha \mathcal{N}_\alpha,$$

where $\mathcal{N}_\alpha$ is the number operator for the particle of type $\alpha \in \{1, \ldots, n+m\}$ and μ_α is its chemical potential. The complete Hamiltonian of the spin chain is then defined as

$$\mathcal{H}^{(m|n)} = \mathcal{H}_0^{(m|n)} + \mathcal{H}_\mu. \tag{24}$$

The number operators $\mathcal{N}_\alpha$ commute with the exchange operators, and hence with the $\mathcal{H}_0^{(m|n)}$ and the Hamiltonian $\mathcal{H}^{(m|n)}$. It follows that $\mathcal{H}_0^{(m|n)}$ and $\mathcal{H}^{(m|n)}$ can be diagonalized in each subspace $\Sigma(N_1, \ldots, N_{m+n}) \subset \Sigma$ with well-defined numbers N_α of particles of each species, and that in such a subspace the energies of $\mathcal{H}^{(m|n)}$ are obtained adding to each of the energies of $\mathcal{H}_0^{(m|n)}$ the term $\sum_{\alpha=1}^{m+n-1} \mu_\alpha N_\alpha$.

5.3 Partition Function

The construction of the partition function follows the same steps outlined in previous sections for the purely bosonic or fermionic cases. In other words, we first find the (large a limit of the) partition functions of the corresponding scalar and spin dynamical models, and then use Polychronakos's freezing trick to compute the partition function of the spin chain.

Consider, for definiteness, the rational (PF) case, in which the scalar and dynamical spin Hamiltonians are, respectively, given by

$$H_{\rm sc} = -\sum_i \partial_{x_i}^2 + a^2 \sum_i x_i^2 + 2a \sum_{i<j} \frac{a-1}{(x_i - x_j)^2}, \tag{25}$$

$$H_0 = -\sum_i \partial_{x_i}^2 + a^2 \sum_i x_i^2 + 2a \sum_{i<j} \frac{a - S^{(m|n)}}{(x_i - x_j)^2}\,. \tag{26}$$

Defining

$$H = H_0 + \frac{2a}{J}\mathcal{H}_\mu\,,$$

the operators H, $H_{\rm sc}$, and H_0 are related to the spin chain Hamiltonian (24) by the expressions

$$H = H_{\rm sc} + \frac{2a}{J}\mathcal{H}^{(m|n)}\big|_{\xi_i \to x_i}.$$

Reasoning as in the previous sections, we deduce that the partition function of the spin chain (24) can be computed from the freezing trick formula

$$\mathcal{Z}(T) = \lim_{a\to\infty} \frac{Z(2aT/J)}{Z_{\rm sc}(2aT/J)},$$

where Z and $Z_{\rm sc}$, respectively, denote the partition functions of H and $H_{\rm sc}$. In fact, the partition function of the scalar model $H_{\rm sc}$ has already been computed in Sect. 2.2 (cf. Eq. (3)). As to the dynamical spin model, its partition function can be obtained along the same lines as in the purely fermionic or bosonic cases, although the computations are more involved due to the presence of the chemical potential term. The final result is

$$Z\left(\tfrac{2aT}{J}\right) = q^{JE_0/(2a)} \sum_{\mathbf{k}\in\mathcal{P}_N} \Sigma(\mathbf{k})\, q^{J\sum_{i=1}^{r-1} K_i} \prod_{i=1}^{r} \frac{1}{1-q^{JK_i}}, \tag{27}$$

(cf. [12]), where

$$\Sigma(\mathbf{k}) = \prod_{i=1}^{r} \sigma(k_i), \quad \sigma(k_i) = \sum_{j=0}^{k_i} h_{k_i-j}\big(q^{-\mu_1},\dots,q^{-\mu_m}\big) e_j\big(q^{-\mu_{m+1}},\dots,q^{-\mu_{m+n}}\big),$$

$\mu_{m+n} = 0$, and h and e, respectively, denote the complete homogeneous and elementary symmetric polynomials (see, e.g., [19] for their precise definitions and main properties). Taking the quotient of these partition functions we finally obtain the partition function of the su$(m|n)$ supersymmetric PF spin chain with a chemical potential term. Remarkable, the partition function for all three chains of HS type can be expressed in a unified way through the closed formula [12]

$$\mathcal{Z} = \sum_{\mathbf{k}\in\mathcal{P}_N} \Sigma(\mathbf{k})\, q^{\sum_{i=1}^{r-1}\mathcal{E}(K_i)} \prod_{i=1}^{N-r}\left(1-q^{\mathcal{E}(K_i')}\right), \tag{28}$$

where the dispersion function $\mathcal{E}$ is given by Eq. (22).

6 Associated Vertex Models

The closed expression (28) for the partition function of the supersymmetric HS-type chains (24) provides an efficient way of computing their energies and degeneracies, but it is not appropriate for deriving the thermodynamics of these models. In this section we shall relate the spin chains (24) to an inhomogeneous vertex model, finding in this way an alternative expression for the partition function better suited for studying its thermodynamic limit $N \to \infty$.

To this end, consider a (classical) vertex model with $N+1$ vertices and N bonds in which $\sigma_i \in \{1, \ldots, m+n\}$ denotes the state of the i-th bond. We define the energies of these models by the expression

$$E^{(m|n)}(\boldsymbol{\sigma}) = \sum_{i=1}^{N-1} \delta(\sigma_i, \sigma_{i+1}) \mathcal{E}(i), \qquad \boldsymbol{\sigma} \equiv (\sigma_1, \ldots, \sigma_N),$$

where

$$\delta(j,k) = \begin{cases} 1, & j > k, \text{ or } j = k \text{ fermions}, \\ 0, & j < k, \text{ or } j = k \text{ bosons}. \end{cases} \tag{29}$$

The partition function of this vertex model is obtained as the value at $\mathbf{x} = \mathbf{y} = \mathbf{1}$ of the generating function [17]

$$\mathcal{Z}^{\mathrm{V}}(q; \mathbf{x}, \mathbf{y}) = \sum_{\sigma_1, \ldots, \sigma_N} \prod_{\alpha=1}^{m} x_\alpha^{N_\alpha(\boldsymbol{\sigma})} \prod_{\beta=1}^{n} y_\beta^{N_{m+\beta}(\boldsymbol{\sigma})} q^{E^{(m|n)}(\boldsymbol{\sigma})}, \tag{30}$$

where $N_\alpha(\boldsymbol{\sigma})$ denotes the number of bonds of type α in the bond vector $\boldsymbol{\sigma}$.

Remarkably, the generating function (30) can also be expressed in terms of the so-called super-Schur polynomials associated to certain border strips [17, 19]. As a consequence of this relation, it can be shown that the partition function of the supersymmetric spin chain (24) can be expressed in terms of the generating function (30) as

$$\mathcal{Z}(T) = \mathcal{Z}^{V}(q; q^{-\mu_1}, \ldots, q^{-\mu_m} | q^{-\mu_{m+1}}, \ldots, q^{-\mu_{m+n}}) = \sum_{\boldsymbol{\sigma}} q^{E^{(m|n)}(\boldsymbol{\sigma}) - \sum_{\alpha=1}^{m+n-1} \mu_\alpha N_\alpha(\boldsymbol{\sigma})}.$$

It follows that the chain's spectrum can be generated by the formula

$$E(\boldsymbol{\sigma}) = E^{(m|n)}(\boldsymbol{\sigma}) - \sum_{\alpha=1}^{m+n} \mu_\alpha N_\alpha(\boldsymbol{\sigma}) = E^{(m|n)}(\boldsymbol{\sigma}) - \sum_{i=1}^{N} \mu_{\sigma_i}, \tag{31}$$

where $\boldsymbol{\sigma}$ takes all values in $\{1, \ldots, m+n\}^N$. As we shall next discuss, this formula is much better suited for deriving the asymptotic behavior of the partition function in the thermodynamic limit $N \to \infty$.

7 Thermodynamics

We shall only present here a brief overview of the application of the methods outlined in the previous sections to the computation the free energy of a supersymmetric spin chain of Haldane–Shastry type in the thermodynamic limit, referring the interested reader to Refs. [9, 12] for details.

The starting point is to rewrite Eq. (31) for the energy as

$$E(\boldsymbol{\sigma}) = \sum_{i=1}^{N-1} \left(\delta(\sigma_i, \sigma_{i+1}) \mathcal{E}(i) - \frac{1}{2}(\mu_{\sigma_i} + \mu_{\sigma_{i+1}}) \right) - \frac{1}{2}(\mu_{\sigma_1} + \mu_{\sigma_N}), \tag{32}$$

which clearly suggests expressing the partition function in terms of a collection of suitable site-dependent transfer matrices. In fact, for the $N \to \infty$ limit of the partition function per site to be well defined we must first rescale the couplings so that the average energy per particle tends to a finite nonzero limit. More precisely, we set $J = K/N$ for the PF chain and $J = K/N^2$ for the HS and FI chains (with K independent of N), so that setting $x_i = i/N$ (with $i = 0, \ldots, N$) we have

$$\frac{\mathcal{E}(i)}{K} = \varepsilon(x_i)\,, \quad \text{with} \quad \varepsilon(x) = \begin{cases} x, & \text{for the PF chain} \\ x(1-x), & \text{for the HS chain} \\ x(\gamma_n + x), & \text{for the FI chain} \end{cases}$$

After this rescaling, using Eq. (32) we can express the partition function of the spin chain (24) as

$$\mathcal{Z}(T) = \operatorname{tr}(A^{(0)} A^{(1)} \cdots A^{(N-1)}), \tag{33}$$

where the $(m+n) \times (m+n)$ transfer matrices $A^{(i)}$ are defined by $A^{(i)} = A(\varepsilon_i)$ with

$$A^{(i)} = A\big(\varepsilon(x_i)\big)\,, \quad \text{with} \quad A(\varepsilon)_{jk} = q^{\delta(j,k)\varepsilon - \frac{1}{2}(\mu_j + \mu_k)}.$$

The latter formula for $\mathcal{Z}(T)$ can be used to express the Helmholtz free energy per particle

$$f(\mu, T) = -\lim_{N\to\infty} \frac{\log \mathcal{Z}}{N\beta}.$$

in terms of the largest eigenvalue[3] $\lambda(x)$ of the transition matrix $A(\varepsilon(x))$ as

$$f(\boldsymbol{\mu}, T) = -\frac{1}{\beta}\int_0^1 \log\lambda(x)\,\mathrm{d}x\,. \tag{34}$$

We shall outline the derivation of this formula in the simplest nontrivial case $m = n = 1$ in the next section.

7.1 *Example: The* su(1|1) *Case*

Although the free energy per site of the chain (24) can in principle be computed from Eq. (34) for arbitrary values of m and n, this is almost impossible in practice unless m and n are small enough. We shall only present here a brief description of the simplest case $m = n = 1$, in which Eq. (34) yields a simple explicit expression for the free energy per site. In this case we have only one chemical potential $\mu_1 \equiv \mu$ (for the bosonic degree of freedom), and the transfer matrix is simply

$$A(\varepsilon) = \begin{pmatrix} q^{-\mu} & q^{-\mu/2} \\ q^{\varepsilon-\mu/2} & q^{\varepsilon} \end{pmatrix}.$$

The matrices $A^{(i)} = A(\varepsilon(x_i))$ are easily diagonalized:

$$A^{(i)} = P_i D^{(i)} P_i^{-1}, \quad D^{(i)} = \begin{pmatrix} \lambda_i & 0 \\ 0 & 0 \end{pmatrix}, \quad P_i = \begin{pmatrix} q^{-\varepsilon_i-\frac{\mu}{2}} & 1 \\ 1 & -q^{-\frac{\mu}{2}} \end{pmatrix},$$

with $\varepsilon_i \equiv \varepsilon(x_i)$ and $\lambda_i = q^{\varepsilon_i} + q^{-\mu}$. We can thus write

$$A^{(0)}A^{(1)}A^{(2)}\cdots A^{(N-1)} = A^{(0)}P_1 D^{(1)} P_1^{-1} P_2 D^{(2)} P_2^{-1}\cdots P_{N-1} D^{(N-1)} P_{N-1}^{-1}\,.$$

Let us take, as an example, the PF spin chain. In this case

$$\lim_{N\to\infty}(\varepsilon_{i+1} - \varepsilon_i) = K \lim_{N\to\infty}\frac{1}{N}(i+1-i) = 0, \quad P_i^{-1}P_{i+1} \underset{N\to\infty}{\to} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus when $N \to \infty$ we have

$$A^{(0)}A^{(1)}A^{(2)}\cdots A^{(N-1)} \simeq A^{(0)}P_1 \begin{pmatrix} \prod_{i=1}^{N-1}\lambda_i & 0 \\ 0 & 0 \end{pmatrix} P_{N-1}^{-1},$$

[3] Since all the matrix elements of the matrix $A(\varepsilon)$ are positive, it follows from the Perron–Frobenius theorem that its largest eigenvalue in module is positive and simple.

and the partition function is given by

$$\mathcal{Z}(T) = \mathrm{tr}(A^{(0)}A^{(1)}A^{(2)}\cdots A^{(N-1)}) \simeq \mathrm{tr}\left[P_{N-1}^{-1}A^{(0)}P_1\begin{pmatrix}\prod_{i=1}^{N-1}\lambda_i & 0\\ 0 & 0\end{pmatrix}\right] \simeq U\prod_{i=1}^{N-1}\lambda_i,$$

where $U > 0$ does not depend on N. It follows that

$$-\frac{1}{N\beta}\log\mathcal{Z}(T) \underset{N\gg 1}{\simeq} -\frac{1}{N\beta}\sum_{i=1}^{N-1}\log\lambda_i = -\frac{1}{N\beta}\sum_{i=1}^{N-1}\log\left(\mathrm{e}^{-\beta\mu}+\mathrm{e}^{-\beta\varepsilon(x_i)}\right),$$

and letting $N \to \infty$ we obtain the following explicit closed formula for the free energy per site of the su(1|1) PF chain in the thermodynamic limit:

$$f(\mu,T) = -\frac{1}{\beta}\int_0^1 \log\left(\mathrm{e}^{-\beta\mu}+\mathrm{e}^{-\beta\epsilon(x)}\right)\mathrm{d}x.$$

8 Conclusions

Spin chains, and in particular the class of long-range solvable models discussed in this contribution, are a powerful theoretical laboratory for realizing in a simple way the fundamental properties of many physical systems, particularly in condensed matter physics. In this short review we have presented a discussion of their relation with integrable many-body (spin) dynamical models, which is at the root of their remarkable symmetry properties and their exact solvability. We have also briefly outlined the recently developed method for deriving the thermodynamics of these chains based on their connection with certain inhomogeneous classical vertex model, illustrating it in some detail for the su(1|1) Polychronakos–Frahm chain. Many interesting new developments which have emerged over the last years have of necessity been omitted in this short overview. To name only a few recent ones, we shall mention the remarkable entanglement and criticality properties of HS-type chains, which stem from their close connection with two-dimensional conformal field theories (see, e.g., [6, 12, 15]), or their relation with matrix product states in quantum field theory [7].

References

1. S. Ahmed, M. Bruschi, F. Calogero, M.A. Olshanetsky, A.M. Perelomov, Il Nuovo Cimento B **49**, 173–198 (1979)
2. J.C. Barba, F. Finkel, A. González-López, M.A. Rodríguez, Europhys. Lett. **83**(6), 27005 (2008)

3. J.C. Barba, F. Finkel, A. González-López, M.A. Rodríguez, Nucl. Phys. B **839**, 499–525 (2010)
4. B. Basu-Mallick, N. Bondyopadhaya, K. Hikami, D. Sen, Nucl. Phys. B **782**, 276–295 (2007)
5. F. Calogero, J. Math. Phys. **10**, 2191–2197 (1969)
6. J.A. Carrasco, F. Finkel, A. González-López, M.A. Rodríguez, P. Tempesta, Phys. Rev. E **93**, 062103 (2016)
7. J. Cirac, G. Sierra, Phys. Rev. B **81**(4), 104431 (2010)
8. C.F. Dunkl, Trans. Am. Math. Soc. **311**, 167–183 (1989)
9. A. Enciso, F. Finkel, A. González-López, Ann. Phys. **327**, 2627–2665 (2012)
10. F. Finkel, A. González-López, Phys. Rev. B **72**, 174411 (2005)
11. F. Finkel, D. Gómez-Ullate, A. González-López, M.A. Rodríguez, R. Zhdanov, Commun. Math. Phys. **221**, 477–497 (2001)
12. F. Finkel, A. González-López, I. León, M.A. Rodríguez, J. Stat. Mech., 043101 (2018)
13. H. Frahm, J. Phys. A **26**, L473–L479 (1993)
14. H. Frahm, V.I. Inozemtsev, J. Phys. A **27**, L801–L807 (1994)
15. D. Giuliano, A. Sindona, G. Falcone, F. Plastina, L. Amico, New J. Phys. **12**(15), 025022 (2010)
16. F.H.D. Haldane, Phys. Rev. Lett. **60**, 635–638 (1988)
17. K. Hikami, B. Basu-Mallick, Nucl Phys. B **566** 511–528 (2000)
18. V.I. Inozemtsev, Phys. Scr. **53**, 516 (1996)
19. I.G. MacDonald, *Symmetric Functions and Hall Polynomials* (Oxford Press, 1995)
20. J. Moser, Adv. Math. **16**, 197–220 (1975)
21. M.A. Olshanetsky, A.M. Perelomov, Phys. Rep. **71**, 313–400 (1981); **94**, 313–404 (1983)
22. A.P. Polychronakos, Phys. Rev. Lett. **70**, 2329–2331 (1993)
23. A.P. Polychronakos, Nucl. Phys. B **419**, 553–566 (1994)
24. B. Shastry, Phys. Rev. Lett. **60**, 639–642 (1988)
25. B. Sutherland, Phys. Rev. A **4**, 2019–2021 (1971)
26. S. Wojciechowski, Phys. Lett. A **95**, 279–281 (1983)

Features of Discrete Integrability

Claude M. Viallet

Abstract We describe some standard features of integrability for a class of integrable partial difference equations: the quad equations. These features are the existence of Lax pairs, higher dimensional consistency, singularity properties, existence of symmetries, and low complexity (vanishing algebraic entropy). All these features have pros and cons, and we give a glimpse of them.

Keywords Discrete integrability

1 Introduction

For ages we have been inclined to think of evolution equations as differential equations, the discrete versions coming at a later stage, in particular when one is constructing a numerical scheme for their resolution. What gained in the recent years is the consideration of discrete equations per se. There are a number of reasons for this change. One is the considerable increase of the computational power of our machines, especially for the formal calculus. Another one is the advent of discrete equations in diverse branches of theoretical physics and mathematics, from 2D gravity [1–3] to statistical models on the lattice [4–7], not forgetting the fundamental contribution of [8], which gave its letters of nobility to the study of self-maps of spaces of finite dimension.

A special interest was taken in the integrable cases, their rich structure leading to new developments. One basic question arising immediately: given a discrete system, in the form of a recurrence relation, a discrete time evolution, or a lattice equation, which are the discrete forms of ordinary and partial differential equations, how do we characterise its integrability?

C. M. Viallet (✉)
Centre National de la Recherche Scientifique, Sorbonne Université, Paris, France
e-mail: claude.viallet@upmc.fr

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_2

Furthermore, if a discrete equation is obtained by discretisation of a continuous one, we would like to preserve as much as possible of the structure of the original equation, and integrability is a crucial one, since it conditions the fundamental properties of the solutions.

To get more information on the subject, explore the proceedings of the "SIDE" conferences, http://side-conferences.net/. It is worth recalling that the first meeting of this series was organised by the Centre de Recherches Mathématiques in 1994 under the governance of Luc Vinet. The meeting was so successful that it gave rise to a series under the acronym of SIDE, organised in Europe (x5), Japan (x2), Australia, China, India, and back to La Belle Province twice (2008 and 2016).

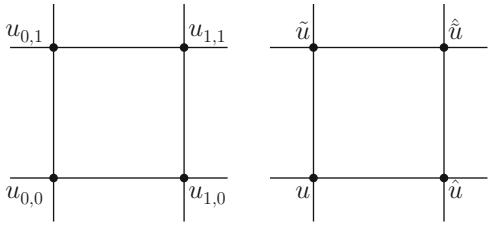
We will avoid giving a precise definition of discrete integrability (see the monograph [9]) but rather describe some of its features, in the specific case of discrete partial difference equations (quad equations) on a two-dimensional square lattice. These are Existence of a Lax pair, Higher dimensional consistency (consistency around the cube, in short CAC), Symmetries, Singularities, Low complexity (vanishing algebraic entropy). We will also comment on the respective merits and limitations of these features.

We will examine one specific example of quad equation.

Warning: We will impose a restriction on the class of evolutions we consider: there will always be a forward and backward evolution, both given by *rational transformations*. One keyword in all parts of our analysis is then *birationality*.

2 What Is a Quad Equation?

It is a discrete equation on a 2-dimensional square lattice, that is to say a discrete version of a partial differential equation in 1+1 dimension. The unknown function u (sometimes called dependent variable) is located at the vertices of the lattice (Fig. 1). The vertices of the lattice are labelled by their integer coordinates (n, m) (independent variables). Different notations are commonly used to represent the values of u at the vertices. We show here two standard ones:



On the left side, the indices have been shifted to the origin, and on the right side, the indices do not appear but $\hat{}$ means a shift by 1 of the first index, i.e. $\hat{u}_{n,m} = u_{n+1,m}$ and $\tilde{}$ means a shift of the second index, i.e. $\tilde{u}_{n,m} = u_{n,m+1}$.

The model is defined by the relation between the corners of the basic square cell, and a solution is given when all the $u_{n,m}$ are known.

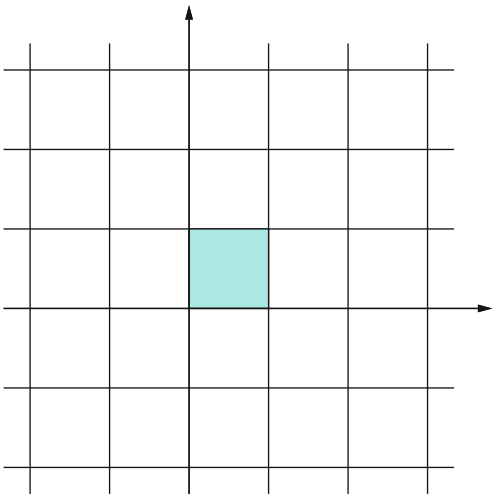


Fig. 1 2D lattice

The conditions we consider are of the form:

$$\begin{aligned} Q = {} & p_1 \cdot u_{0,0}\, u_{1,0}\, u_{0,1}\, u_{1,1} + p_2 \cdot u_{0,0}\, u_{1,0}\, u_{0,1} + p_3 \cdot u_{0,0}\, u_{1,0} u_{1,1} + p_4 \cdot u_{1,0}\, u_{0,1}\, u_{1,1} \\ & + p_5 \cdot u_{0,0}\, u_{0,1}\, u_{1,1} + p_6 \cdot u_{0,0}\, u_{0,1} + p_7 \cdot u_{1,0}\, u_{0,1} + p_8 \cdot u_{0,1}\, u_{1,1} \\ & + p_9 \cdot u_{0,0}\, u_{1,0} + p_{10} \cdot u_{0,0}\, u_{1,1} + p_{11} \cdot u_{1,0}\, u_{1,1} + p_{12} \cdot u_{0,1} \\ & + p_{13} \cdot u_{0,0} + p_{14} \cdot u_{1,0} + p_{15} \cdot u_{1,1} + p_{16} = 0 \end{aligned} \tag{1}$$

Notice that the multilinear nature of the relation implies that, for all cells, any of the four corner values can be rationally expressed in terms of the three others.

3 A Few Standard Features of Integrable Quad Equations

3.1 Discrete Lax Pairs

Lax pairs are a characteristic feature of integrable differential equations, and they have been extremely important in the development of the subject. For $1+1$ dimensional partial differential equations, they become rather zero curvature equations, and these have a straightforward discrete form. A discrete Lax pair is a pair of matrices $L(u)$, $M(u)$ such that

$$\hat{L} \cdot M \simeq \tilde{M} \cdot L, \tag{2}$$

where $\hat{L} = L(\hat{u})$ (shift along the horizontal direction), $\tilde{M} = M(\tilde{u})$ (shift along the vertical direction), and $\simeq$ means proportionality, is equivalent to the defining

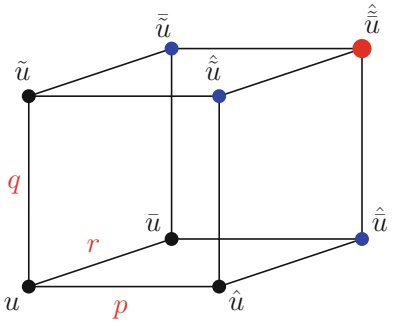
equation (1). Discrete quad equations have soliton type solutions, and Lax pairs can be used to produce explicit solutions [10].

3.2 Consistency Around the Cube (CAC)

Consider the archetypal case of discrete potential KdV:

$$(\hat{u} - \tilde{u})(u - \hat{\tilde{u}}) = p^2 - q^2 \tag{3}$$

It is possible to embed the two-dimensional cell into a three-dimensional one:



where one imposes a similar relation to all faces (the same for opposite faces), and $^{-}$ means shift in the third dimension.

Consistency around the cube means that *the value of* $\hat{\bar{\tilde{u}}}$ *is independent of the way it is calculated*, as there are three ways to evaluate it from the initial condition $u, \hat{u}, \bar{u}, \tilde{u}$. The major, and remarkable major output of this property is to ensure the existence of a Lax pair [11, 12], which is accepted as a proof of integrability. The interest of the higher dimensional consistency approach is that it also provides us with a classification of a set of integrable quad equations [13, 14], referred to as ABS list. There exists a simple rationally parametrised interpolating form [15] whose integrability was discovered by algebraic entropy argument (see Sect. 3.6), and confirmed by symmetry arguments [16] (see Sect. 3.4).

3.3 Lax Pair from the Consistency Around the Cube

The left face and the bottom face give, respectively,

$$\bar{\tilde{u}} = \frac{q^2 - r^2 + \tilde{u}u - \bar{u}\,u}{\tilde{u} - \bar{u}}, \qquad \hat{\bar{u}} = \frac{p^2 - r^2 - \bar{u}\,u + u\hat{u}}{\hat{u} - \bar{u}} \tag{4}$$

Projectivisation: writing $\bar{u} = F/G$, $\tilde{\bar{u}} = \tilde{F}/\tilde{G}$, $\hat{\bar{u}} = \hat{F}/\hat{G}$ and

$$\Phi = \begin{pmatrix} F \\ G \end{pmatrix} \quad L = \begin{pmatrix} u\ q^2 - r^2 + u\,\tilde{u} \\ 1 \qquad\quad -\tilde{u} \end{pmatrix} \quad M = \begin{pmatrix} u\ r^2 - p^2 + u\,\hat{u} \\ 1 \qquad\quad -\hat{u} \end{pmatrix}$$

$$\tilde{\Phi} = L \cdot \Phi, \qquad \hat{\Phi} = M \cdot \Phi$$

The consistency around the cube yields

$$\hat{L} \cdot M \simeq \tilde{M} \cdot L, \tag{5}$$

where $\simeq$ means proportionality. Equation (5) is a discrete zero curvature condition.

By above argument, the "consistency around the cube" condition is considered as a major integrability condition for quad equations, We will see in Sect. 4 that the two notions are not equivalent. See also concluding remark.

3.4 Symmetries

The existence of continuous symmetries is a characteristic feature of integrability. It is of course related to the existence of conserved quantities, but is more easily tractable. There is an important literature on the subject, originally for the differential case [17, 18], and now for the discrete case [19, 19–30].

What is a continuous symmetry of the quad equation $u_{n+1,m+1} = F(u_{n,m+1}, u_{n,m}, u_{n+1,m})$? It is a vector field over the space of solutions.

We restrict ourselves to the specific form ($k > l$, $p > q \in \mathbb{Z}$)

$$\partial u_{n,m} = G\left(u_{n+k,m}, u_{n+k-1,m}, \ldots, u_{n+l,m}, u_{n,m+p}, u_{n,m+p-1}, \ldots, u_{n,m+q}\right) \tag{6}$$

The figure shows the reach of the symmetry: it is a rectangle of size $(k-l+1)\times(p-q+1)$. Since the symmetry acts on the space of solutions, it may be described in terms of the values on the lines $(n+k), m \ldots (n+l), m$ and $(n, m+p) \ldots (n, m+q)$. This is a local symmetry with a finite extension (Fig. 2).

The symmetry condition being a constraint between F and G, the symmetry approach is then to look for models (determined by F) for which there are symmetries (given by G). The method is *constructive*, with the input of Ansätze for F and G, the idea being to ask for the existence of more than one symmetry.

Remarkably the symmetries split into two independent pieces, vertical and horizontal (meaning $p = q = 0$ and $k = l = 0$, respectively).

This means that Eqs. (6) become ordinary differential-difference equations, which turn out to be integrable themselves [16, 26, 27, 31, 32]!

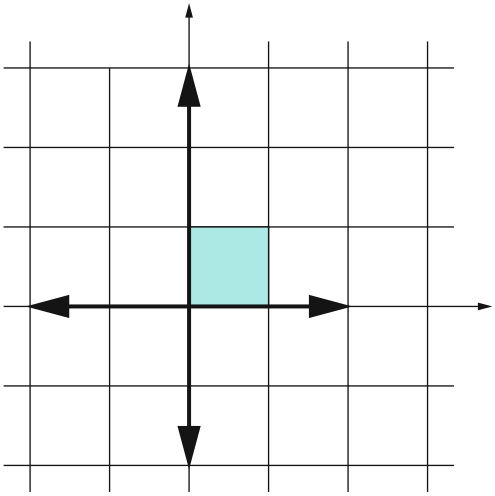


Fig. 2 A symmetry pattern

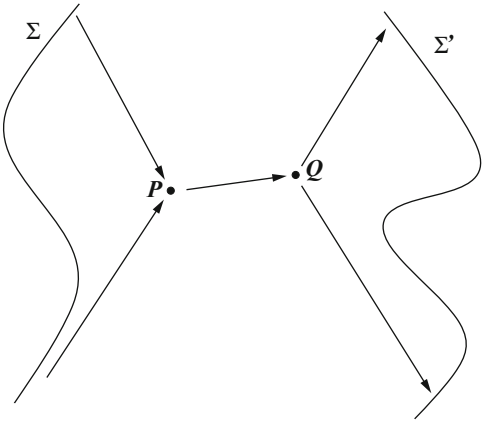


Fig. 3 A singularity pattern in two dimensions

3.5 Singularity Analysis

There is a deep link between integrability and singularity structure. This structure is best illustrated in the case of maps, that is to say ordinary difference equations.

Figure 3 shows what a birational map in two dimensions can do. The curve Σ is sent to a point P by the evolution. This means that the backward evolution is not defined at P. The point P is then sent to the point Q, which is singular as the backward evolution sends the whole curve Σ' to Q.

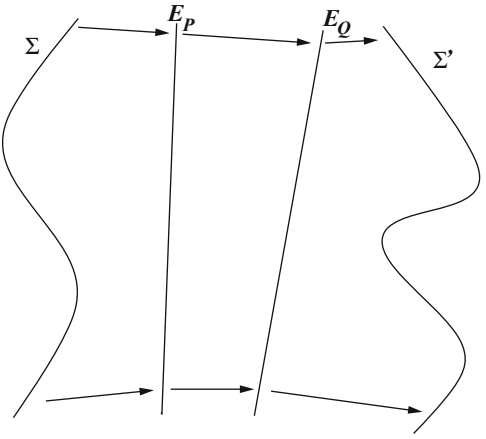


Fig. 4 Desingularisation pattern in two dimensions

The best description of the geometry is obtained using simple tools of algebraic geometry [33]. It is possible to remove the singularities by blowing up the point P and Q. This amounts to adding curves to the space, replacing the points P and Q by the lines E_P and E_Q, respectively. After the blow-ups, the map sends the curve Σ onto the line E_P then to E_Q and finally to the curve Σ' birationally. The setting described by Figs. 3, 4 is similar to the one we encounter for the QRT maps [34, 35]. This was the key to an important classification of the non-autonomous generalisation of these maps yielding the discrete Painlevé equations [36]. See, for example, the monograph [37].

Remark Two important facts (see, for example, [37, 38]) should be noticed:

- The curves Σ and Σ' are algebraic
- The singularity of the forward (resp. backward) map at Q (resp. P) shows up naturally when we use homogeneous (projective) coordinates, in relation with the equation of Σ (resp. Σ'). If one calculates the third iterate of the forward map, the polynomial expression of the coordinates of the image all vanish. This means that the equation of Σ factors out from these coordinates. Once factored out we get a perfectly well-defined image.

There is a whole class of maps in two dimensions (actually order 2 recurrences) for which the singularities may be removed by blowing up a finite number of points, and this has led to very important results on integrable maps, autonomous ones as well as non-autonomous ones [34–37]. This was even considered to characterise integrability [39], but was eventually shown not to be the case [40].

A similar singularity analysis can be performed for the quad equations.

3.6 Complexity Analysis: Algebraic Entropy

Explicit calculations of the iterates of maps in the integrable case go much faster than in the non-integrable case. The reason is that the degree of the iterates grows slowly (polynomially) in the integrable case and rapidly (exponentially) in the non-integrable case. This has led to the definition of the algebraic entropy from the sequence of degrees $\{d_n\}$ of the iterates.

$$\epsilon = \lim_{n\to\infty} \frac{1}{n} Log(d_n) \tag{7}$$

- This limit always exists by the subadditivity of $Log(d_n)$, and is canonical, being invariant by birational changes of coordinates.
- Vanishing of ϵ is the hallmark of integrability.
- The entropy has remarkable arithmetic properties (conjectured to be the Log of an algebraic integer).

Claim: The entropy is consubstantial with the singularity structure. Indeed going back to Figs. 3, 4 it is easy to convince oneself that—in the case shown—there will be a drop of the degree for the third iterate of the map. The equation of the hypersurface Σ will factor out from the rational expressions of the iterate [38, 41].

The entropy first defined for maps acting on finite dimensional spaces (ordinary difference equations), has been further generalised to the infinite dimensional case, allowing to consider semi-discrete equations [32] involving maps over functional spaces and quad equations as well [42, 43].

3.7 What About Quad Equations?

The point is to define an evolution. Figure 5 shows how this can be done from a staircase initial condition. The initial condition at time 0 of line (0) determines the values on line (1) at time 1, line (2) at time 2, and so on.

The evolution after k steps is expressible as a rational fraction in terms of $2k + 1$ initial values. Evaluating the degrees of these rational fractions gives a sequence of degrees $\{d_n\}$, providing in turn the value of the entropy.

We will not detail the calculation method here, but the *asymptotic property which the entropy measures can most of the time be extracted from a finite piece of the sequence of degrees*. This is a manifestation of the fact that a local property governs the global behaviour.

Claim: The vanishing of the entropy is a good criterion of integrability for quad equations as it is for maps.

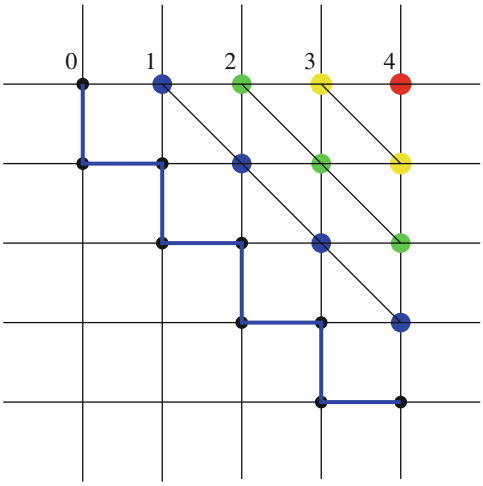


Fig. 5 Evolution on the square lattice

4 A Case Study

We will concentrate on one specific quad equation, taken from [44]. *It does not belong to the ABS list, as it is not symmetric. It does not verify the consistency around the cube condition.*

$$E = \left(u_{1,0}+1\right)\left(u_{0,0}-1\right) - \left(u_{1,1}-1\right)\left(u_{0,1}+1\right) = 0 \tag{8}$$

The variable $u_{n,m}$ at site (n, m) is denoted $u_{0,0}$ with $u_{k,l}$ standing for $u_{n+k,m+l}$ as shown in Sect. 2.

4.1 Symmetries

Equation (E) possesses continuous symmetries. These symmetries split into vertical and horizontal symmetries (H and V), as is usual for this type of quad equations.

Three points symmetries:

$$H_3: \ \partial_{t_1} u_{0,0} = (u_{0,0}^2 - 1)(u_{1,0} - u_{-1,0}) \tag{9}$$

$$V_3: \ \partial_{\tau_1} u_{0,0} = (u_{0,0}^2 - 1)\left(\frac{1}{u_{0,1}+u_{0,0}} - \frac{1}{u_{0,0}+u_{0,-1}}\right) \tag{10}$$

Five points symmetries:

$$H_5 : \partial_{t_2} u_{0,0} = (u_{0,0}^2-1)((u_{1,0}^2-1)(u_{2,0}+u_{0,0})-(u_{-1,0}^2-1)(u_{0,0}+u_{-2,0})), \quad (11)$$

$$V_5 : \partial_{\tau_2} u_{0,0} = \frac{u_{0,0}^2-1}{(u_{0,1}+u_{0,0})^2}\left(\frac{u_{0,1}^2-1}{u_{0,2}+u_{0,1}}+\frac{u_{0,0}^2-1}{u_{0,0}+u_{0,-1}}\right) - \frac{u_{0,0}^2-1}{(u_{0,0}+u_{0,-1})^2}\left(\frac{u_{0,0}^2-1}{u_{0,1}+u_{0,0}}+\frac{u_{0,-1}^2-1}{u_{0,-1}+u_{0,-2}}\right). \quad (12)$$

Claims:

- The existence of these symmetries is a strong constraint on the equations, and it may serve as an integrability criterion.
- All these equations (E and the successive symmetries H_i, V_j) are integrable, as can be checked by direct calculation.
- A confirmation of the integrability comes from the calculation of the algebraic entropy, which vanishes, as we will see later.

Notice that *the form of the symmetries is local*. Of course the symmetries are verified modulo the ideal generated by the local quad relations ("on shell").

4.2 A Lax Pair

Although it is not 3D consistent, equation (E) has a Lax pair.

$$L_{i,j} = \frac{1}{u_{i,j+1}-1}\begin{bmatrix} \lambda-\lambda^{-1} & 2\,(u_{i,j+1}^2-1) \\ -2 & \left(\lambda-\lambda^{-1}\right)(u_{i,j+1}+1) \end{bmatrix} \quad (13)$$

$$M_{i,j} = \begin{bmatrix} \lambda-\lambda^{-1} & -\left(u_{i+1,j}+1\right)\left(u_{i,j}-1\right) \\ 1 & 0 \end{bmatrix} \quad (14)$$

The equation is obtained by the discrete zero curvature condition

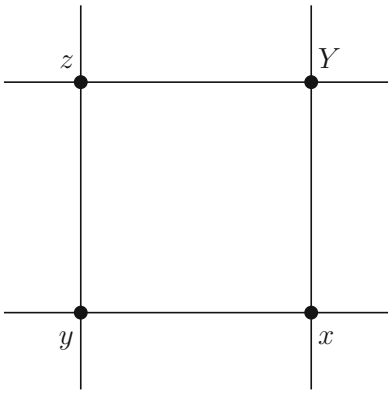
$$L_{i+1,j}\cdot M_{i,j} \simeq M_{i,j+1}\cdot L_{i,} \quad (15)$$

which we already wrote

$$\hat{L}\cdot M \simeq \tilde{M}\cdot L \quad (16)$$

4.3 What About Singularities?

Singularities are known to play a fundamental rôle in the game. For quad equations, singularities appear already at the level of the basic cell of the lattice. Suppose we look at the elementary cell



The defining relation E gives a projective linear map $\varphi_{xz} : y \longrightarrow Y$, whose inverse φ^{-1} is projective linear. The composed map $\varphi \cdot \varphi^{-1}$ comes with an overall factor

$$H(x, z) = (1 + x)\,(1 + z) \tag{17}$$

which is the locus of the singularities (remember the picture in two dimensions) *This quantity contains the information about the singularities, and it is the key of the factorisations (and simplifications) appearing in the evolution*. Labelling the cells by their lower left corner, denote $H_{n,m}$ the value of H calculated on the (n, m) cell.

One may define

$$\Omega_{n,m} = gcd(H_{n,m}, H_{n+2,m+1}) \tag{18}$$

then, remarkably

$$H_{n,m} = \Omega_{n,m}.\Omega_{n-2,m-1} \tag{19}$$

The values of $u_{n,m}$ over the plane are polynomials (if we work in projective coordinates) in the initial conditions. The *infinitesimal* relations defining the map imply *local* relations! This is a constant feature of the birational evolutions we consider (Fig. 6).

We also see here the effect of a specific feature of the model: it is asymmetric (contrary to the members of the ABS list).

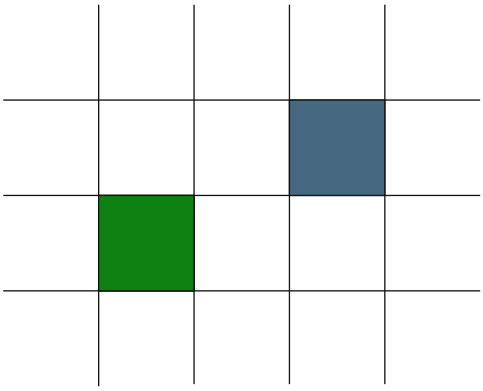


Fig. 6 Localisation of the factors Ω

4.4 What About Entropy?

The factors Ω contain all the information on

- How the successive iterates hit singularities (remember the picture in the plane, but now we work in an infinite dimensional space of initial conditions)
- the sequence of degrees of the iterates and leading to the vanishing of the entropy.

The sequence of degrees for this specific model is obtained by direct calculation

$$\{d_n\} = 1, 2, 4, 7, 11, 16, 22, 29, 37, 46, 56, 67, 79, 92, 106, 121, \ldots \tag{20}$$

It can be fitted by the generating function

$$g(s) = \sum_{k=0}^{\infty} d_k\, s^k = \frac{1 - s + s^2}{(1-s)^3} \tag{21}$$

that is to say

$$d_n = 1 + \frac{n(n+1)}{2} \tag{22}$$

quadratic growth, vanishing entropy.

Remark the function g *can be guessed from the first* 5 *iterations, and the further terms of the sequence give a check*. Moreover the result can be proved along the lines of [41].

5 Comments and Perspectives

We have given a glimpse of the subject, leaving aside some of the recent developments like, for example, the Lagrangian multiform approach [9, 45, 46], which is a promising approach.

Our point was rather to give an idea of some of the features of integrability, with their strength and their limitations, which we can summarise as follows:

- *Lax pairs.* pro: Lax pairs are a powerful instrument to produce solutions, as they were in the continuous case; con: They are not always easy to find, and they may have different forms (the size of the linear system is not known in advance, and there may be different pairs of different sizes). In addition one should beware of fakes [47, 48].
- *3D consistency.* pro: provides a Lax pair, and allows a classification; con: not equivalent to integrability.
- *Singularity confinement.* pro: constructive and allowing classification at least for the order two equations; con: not necessary nor sufficient for integrability.
- *Symmetries.* pro: constructive and necessary; con: not always sufficient as one needs a sufficiently large number of them.
- *Algebraic entropy.* pro: it is canonical (invariant by birational changes of coordinates), and the vanishing of the entropy may serve as a characterisation of integrability, as the sign of catastrophic drop of the complexity, con: destructive rather than constructive, since it gives a yes/no answer to the question "is this model integrable?"

We have alluded to the relation of the singularity structure with the measure of complexity of the evolutions. What is still missing is a better understanding of its relations with Lax pairs and Symmetries.

One hint may come from the following fact: when analysing the singularities, the sequence of degrees of the iterates, as well as the symmetries, the most important relations bear on local properties with a finite extension. This means that, even in the discrete world, we may distinguish three scales—as we have in the continuous world—infinitesimal, local, global. Infinitesimal would be the defining relation (one cell), local would be any relation extending over more than one cell, and of course global would be extending to infinity. What we see is that we can reach a conclusion on a global property like integrability from local ones. The algebraic nature of the models we consider is probably at the origin of this phenomenon. Since this algebraic nature also conditions the singularity structure, we should look further into relations of the singularities with symmetries and even Lax pairs.

We have work to do.

Acknowledgments The author would like to thank the organisers of the 11th symposium "Théorie Quantique et Symétries" for the invitation at the Centre de Recherches Mathématiques, in particular at the occasion of the "Special Session in honour of Decio Levi: Integrability: continuous and discrete, classical & quantum."

References

1. E. Brezin, V. Kazakov, Exactly solvable field theories of closed strings. Phys. Lett. **B 236**(2), 144–150 (1990)
2. D. Gross, A. Migdal, Non perturbative two-dimensional quantum gravity. Phys. Rev. Lett. **64**(2), 127–130 (1990)
3. A.S. Fokas, A.R. Its, A.V. Kitaev, Discrete Painlevé equations and their appearance in quantum gravity. Commun. Math. Phys **142**, 313–344 (1991)
4. R.J. Baxter, The inversion relation method for some two-dimensional exactly solved models in lattice statistics. J. Stat. Phys. **28**, 1–41 (1982)
5. M.P. Bellon, J.-M. Maillard, C.-M. Viallet, Infinite discrete symmetry group for the Yang-Baxter equations: spin models. Phys. Lett. A **A 157**, 343–353 (1991)
6. M.P. Bellon, J.-M. Maillard, C.-M. Viallet, Infinite discrete symmetry group for the Yang-Baxter equations: vertex models. Phys. Lett. A **B 260**, 87–100 (1991)
7. V.G. Drinfeld, On some unsolved problems in quantum group theory, in "*Quantum Groups*", vol. 1510 of *Lecture Notes in Math* (Springer, 1992), pp. 1–8
8. M. Hénon, A two-dimensional mapping with a strange attractor. Commun. Math. Phys. **50**, 69–77 (1976)
9. J. Hietarinta, N. Joshi, F.W. Nijhoff, *Discrete Systems and Integrability*. Cambridge Texts in Applied Mathematics (Cambridge University Press, 2016)
10. S. Butler, N. Joshi, An inverse scattering transform for the lattice potential KdV equation. Inverse Problems **26**, 115012 (2010)
11. F. Nijhoff, Lax pair for the Adler (lattice Krichever-Novikov) system. Phys. Lett. **A 297**, 49–58 (2002). arXiv:nlin.SI/0110027
12. A.I. Bobenko, Yu.B. Suris, Integrable systems on quad-graphs. Int. Math. Res. Notices **11**, 573–611 (2002). arXiv:nlin/0110004
13. V.E. Adler, A.I. Bobenko, Yu.B. Suris, Classification of integrable equations on quad-graphs. The consistency approach. Commun. Math. Phys. **233**(3), 513–543 (2003). arXiv:nlin.SI/0202024
14. V.E. Adler, A.I. Bobenko, Yu.B. Suris, Discrete nonlinear hyperbolic equations. Classification of integrable cases. Funct. Anal. Appl. **43**, 3–17 (2009). arXiv:0705.1663
15. C.-M. Viallet, Integrable lattice maps: Q_V, a rational version of Q_4. Glasgow Math. J. **51 A**, 157–163 (2009). arXiv:0802.0294
16. A.V. Mikhailov, J.P. Wang, A new recursion operator for Adler's equation in the Viallet form. Physics Letters **A 375**, 3960–3963 (2011). arXiv:1105.1269
17. P.J. Olver, Evolution equations possessing an infinitely many symmetries. J. Math. Phys. **18**(6), 1212–1215 (1977)
18. M.J. Ablowitz, D.J. Kaup, A.C. Newell, H. Segur, Inverse scattering transform-Fourier analysis for nonlinear problems. Stud. Appl. Math. **53**(4), 249–315 (1974)
19. S. Maeda, The similarity method for difference equations. IMA J. Appl. Math. **38**, 129 (1987)
20. D. Levi, P. Winternitz, Continuous symmetries of difference equations. J. Phys. A Math. Gen. **39**, R1–R63 (2006)
21. D. Levi, R.I. Yamilov, Conditions for the existence of higher symmetries of evolutionary equations on the lattice. J. Math. Phys. **38**, 6648–6674 (1997)
22. D. Levi, R.I. Yamilov, The generalized symmetry method for discrete equations. J. Phys. A Math. Theor. **42**, 454012 (2009)
23. D. Levi, R.I. Yamilov, Generalized symmetry integrability test for discrete equations on the square lattice. J. Phys. A Math. Theor. **44**, 145207 (2011)
24. G.R.W. Quispel, H.W. Capel, R. Sahadevan, Continuous symmetries of differential-difference equations: the Kac-van Moerbeke equation and the Painlevé reduction. Phys. Lett. **A**(170), 379–383 (1992)
25. D. Levi, L. Vinet, P. Winternitz, Lie group formalism for difference equations. J. Phys. A Math. Gen. **30**(2), 633–649 (1997)

26. A.V. Mikhailov, J.P. Wang, P. Xenitidis, Recursion operators, conservation laws, and integrability conditions for difference equations. Teore. Mat. Fiz. **1**, 23–49 (2011). Transl. Theore. Math. Phys. **167**, 421–443 (2011). arXiv:1004.5346
27. A.V. Mikhailov, J.P. Wang, P. Xenitidis, Cosymmetries and Nijenhuis recursion operators for difference equation. Nonlinearity **24**, 2079–2097 (2011). arXiv:1009.2403
28. R.I. Yamilov, Symmetries as integrability criteria for differential difference equations. J. Phys. A Math. Gen. **39**, R541–R623 (2006)
29. P. Hydon. *Difference Equations by Differential Equation Methods*. Number 27 in Cambridge Monographs on Applied and Computational Mathematics (Cambridge University Press, 2014)
30. O. Rasin, P. Hydon, Symmetries of integrable difference equations on the quad graph. Stud. Appl. Math. **119**, 253–269 (2007)
31. P.D. Xenitidis, V.G. Papageorgiou, Symmetries and integrability of discrete equations defined on a black–white lattice. J. Phys. A Math. Theor. **42**, 454025 (2009)
32. D.K. Demskoy, C.-M. Viallet, Algebraic entropy for semi-discrete equations. J. Phys. A Math. Theor. **45**, 352001 (2012). arXiv:1206.1214
33. I.R. Shafarevich, *Basic Algebraic Geometry*. Number 231 in Grundlehren der mathematischen Wissenschaften (Springer, Berlin, 1977)
34. G.R.W. Quispel, J.A.G. Roberts, C.J. Thompson, Integrable mappings and soliton equations. Phys. Lett. **A 126**, 419 (1988)
35. G.R.W. Quispel, J.A.G. Roberts, C.J. Thompson, Integrable mappings and soliton equations II. Physica **D34**, 183–192 (1989)
36. H. Sakai, Rational surfaces associated with affine root systems and geometry of the Painlevé equations. Commun. Math. Phys. **220**(1), 165–229 (2001)
37. J.J. Duistermaat, *Discrete Integrable Systems: QRT Maps and Elliptic Surfaces*. Springer Monographs in Mathematics (Springer, New York, 2010)
38. M.P. Bellon, C.-M. Viallet, Algebraic entropy. Commun. Math. Phys. **204**, 425–437 (1999). arXiv:chao-dyn/9805006
39. B. Grammaticos, A. Ramani, V. Papageorgiou, Do integrable mappings have the Painlevé property? Phys. Rev. Lett. **67**, 1825–1827 (1991)
40. J. Hietarinta, C.-M. Viallet, Singularity confinement and chaos in discrete systems. Phys. Rev. Lett. **81**(2), 325–328 (1998). arXiv:solv-int/9711014
41. C.M. Viallet, On the algebraic structure of rational discrete dynamical systems. J. Phys. A Math. Theor. **48**, 16FT01 (2015)
42. C.-M. Viallet, *Algebraic Entropy for Lattice Equations*. arXiv:math-ph/0609043
43. S. Tremblay, B. Grammaticos, A. Ramani, Integrable lattice equations and their growth properties. Phys. Lett. **A 278**, 319–324 (2001)
44. D. Levi, R. Yamilov, On a nonlinear integrable difference equation on the square 3D-inconsistent. Ufa Math. J. **1**, 03 (2009). arXiv:0902.2126
45. S. Lobb, F. Nijhoff, Lagrangian multiforms and multidimensional consistency. J. Phys. A Math. Theor. **42**(45), 454013 (2009)
46. Y. Suris, Variational symmetries and pluri-Lagrangian systems, in *Dynamical Systems, Number Theory and Applications. A Festschrift in Honor of Armin Leutbecher's 80th Birthday*, ed. by Th. Hagen, F. Rupp, J. Scheurle. (World Scientific, 2016), pp. 255–266 arXiv:1307.2639
47. J. Hietarinta, C.M. Viallet, Weak Lax pairs for lattice equations. Nonlinearity **25**, 1955–1966 (2012). arXiv:1105.3329
48. S. Butler, M. Hay. *Simple Identification of Fake Lax Pairs* (2013). arXiv:1311.2406

Darboux-Bäcklund Transformations for Spin-Valued Linear Problems

Jan L. Cieśliński

This paper is dedicated to Prof. Decio Levi on the occasion of his 70th birthday.

Abstract Motivated by geometry of submanifolds we develop an algebraic construction of Darboux transformations using Clifford numbers and Spin groups. Eigenvalues parameterizing solitons, usually computed as zeros of determinants, are identified as zeros of the spinor norm. Reduction groups (loop groups) for Spin-valued linear problems are identified with involutions in Clifford algebras.

Keywords Darboux matrix · Spin groups · Clifford algebras · Soliton surfaces

1 Introduction

This paper is a continuation of my long research on Darboux transformations, which started more than 30 years ago and the interaction with Decio Levi has been crucial in this context. My PhD thesis was focused on the following problems, all of which stemmed from the cooperation between Antoni Sym (my supervisor) and Decio Levi:

- Group interpretation of the spectral parameter [1–6].
- Darboux transformations (non-isospectral and non-canonical cases) [1, 7–9].
- Symmetric formulas for multi-soliton surfaces [10–12].

My several visits to Rome were very stimulating for this research. Recently, I returned to these topics, see [13, 14] and the formula (24) in the present paper, with many hopes for further progress in the near future. In the meantime I devoted

J. L. Cieśliński (✉)
University of Bialystok, Faculty of Physics, Białystok, Poland
e-mail: j.cieslinski@uwb.edu.pl

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_3

a lot of time to structure preserving numerical methods (see, e.g., [15, 16]), which is also close to Decio's interests [17], but this subject evolved rather independently.

Spin-valued spectral problems seem to be quite natural in the geometric context because Spin groups and Clifford algebras are very natural structures to deal with orthogonal transformations, see the next section. Actually, even the standard $su(2)$-valued Lax pairs for some well-known soliton equations (including sine-Gordon, nonlinear Schrödinger and modKdV equations) can be rewritten in terms of the group Spin(3) (which is isomorphic to $SU(2)$, see, e.g., [18]). In particular, the sine-Gordon equation $\phi_{,xy} = \sin\phi$ arises as compatibility conditions for the Lax pair

$$\begin{aligned} &\Psi_{,x} = U\Psi \ , \quad U = \begin{pmatrix} i\lambda & -\frac{1}{2}\phi_{,x} \\ \frac{1}{2}\phi_{,x} & -i\lambda \end{pmatrix} = \lambda\,\sigma_1\sigma_2 + \frac{1}{2}\phi_{,x}\,\sigma_1\sigma_3 \ , \\ &\Psi_{,t} = U\Psi \ , \quad V = \frac{1}{4i\lambda}\begin{pmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{pmatrix} = \frac{\sigma_2\sigma_1\cos\phi + \sigma_3\sigma_2\sin\phi}{4\lambda} \ , \end{aligned} \tag{1}$$

where σ_k denote Pauli matrices (and we replaced $i\sigma_3$ by $\sigma_1\sigma_2$, etc.). In order to see the Lie algebra of a Spin group in the above formulas, one needs some basic information on Clifford algebras and Spin groups, which will be provided in the next section.

In the framework of Sym's soliton surfaces approach $su(2)$-valued linear problems correspond to surfaces immersed in the Euclidean 3-space and other semi-simple Lie algebras can be associated with surfaces in multi-dimensional (pseudo)-Euclidean spaces [19]. Another natural possibility, followed in this paper, is to extend this approach from $SU(2) \cong \mathrm{Spin}(3)$ on any Spin groups [20].

2 Clifford Algebras, Spin Groups, and Isometries of $\mathbb{R}^{p,q}$

Given linear space V and bilinear form $\langle\,\cdot\,|\,\cdot\,\rangle$ (or, equivalently, the quadratic form $Q(v) := \langle v\,|\,v\rangle$) we define the Clifford product by the following condition:

$$vw + wv = 2\langle v\,|\,w\rangle\mathbf{1} \ , \qquad (v, w \in V) \ , \tag{2}$$

where $\mathbf{1}$ is the unity (multiplicative identity) of the algebra. In other words, parallel vectors commute and orthogonal vectors anti-commute. Hence, in particular,

$$v^{-1} = \frac{v}{\langle v\,|\,v\rangle} \ , \tag{3}$$

i.e., vectors with $Q(v) \neq 0$ are invertible and the result is geometrically interpreted as inversion with respect to the unit sphere.

Clifford algebra (denoted by $\mathcal{C}(V, Q)$ or $\mathcal{C}(p, q)$, where (p, q) is the signature of Q) is generated by V using linear operations and the Clifford product, see, e.g., [18, 21].

Let $\mathbf{e}_1, \ldots, \mathbf{e}_N$ be an orthonormal basis in V. In other words, see (2), we have

$$\mathbf{e}_j^2 = \pm 1 \,, \qquad \mathbf{e}_j \mathbf{e}_k = -\mathbf{e}_k \mathbf{e}_j \,. \tag{4}$$

Then, $\dim \mathcal{C}(V, Q) = 2^N$ and the Clifford algebra is spanned by $\mathbf{1}$ (scalars), $\mathbf{e}_k$ (vectors), $\mathbf{e}_j \mathbf{e}_k$ $(j < k)$ (bivectors) , multivectors, and $\mathbf{e}_1 \mathbf{e}_2 \ldots \mathbf{e}_N$ (pseudoscalars).

Well-known examples: Pauli matrices ($\mathbf{e}_k = \sigma_k$, $N = 3$, $p = 3$, $q = 0$) and Dirac matrices ($\mathbf{e}_k = \gamma_k$, $N = 4$, $p = 1$, $q = 3$).

The vector space V, generating the Clifford algebra $\mathcal{C}(p, q)$ $(p + q = N)$, can be identified with the pseudo-Euclidean space spanned by $\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_N$. Reflection with respect to the hyperplane orthogonal to an invertible $n \in V$ can be represented as

$$v' = -nvn^{-1} \,, \qquad \text{hence} \quad v' = -(2\langle n \mid v\rangle - vn)n^{-1} = v - \frac{2\langle n \mid v\rangle}{\langle n \mid n\rangle} n \,. \tag{5}$$

Indeed, if $v = v_n + v_t$ (where v_n is normal and v_t is parallel to the hyperplane), then $v' = v_t - v_n$. Thus: $-n(v_t + v_n)n^{-1} = v_t n n^{-1} - v_n n n^{-1} = v'$.

By the Cartan–Dieudonné theorem any isometry of V can be represented as a composition of at most N, say k, reflections. Therefore, using (5), we have

$$v' = (-1)^k n_k n_{k-1} \ldots n_1 \, v \, n_1^{-1} n_2^{-1} \ldots n_k^{-1} \,. \tag{6}$$

Hence, it is natural to consider multiplicative groups in Clifford algebras:

- **Lipschitz** group $\Gamma(V, Q)$: products of invertible vectors,
- **Pin** group $\mathrm{Pin}(V, Q)$: products of unit vectors,
- **Spin** group $\mathrm{Spin}(V, Q)$: products of even number of unit vectors

(often we omit Q, writing $\mathrm{Spin}(V)$, etc.). Thus Eq. (6) can be rewritten as

$$\mathcal{C}(p, q) \supset V \ni v \quad \mapsto \quad v' = \Psi \, v \, \Psi^{-1} \in V, \tag{7}$$

where $\Psi \in \mathrm{Pin}(p, q)$, which is equivalent to an orthogonal transformation in V:

$$\mathbb{R}^{p,q} \supset V \ni v \quad \mapsto \quad v' = R \, v \in V \tag{8}$$

where $R \in O(p, q)$. Obviously, $-\Psi$ and Ψ yield the same isometry R. Therefore, we have double coverings, like $\mathrm{Pin}(N) \to O(N)$ or $\mathrm{Spin}(N) \to SO(N)$.

The Lie algebra $\mathrm{spin}(p, q)$ of the Spin group $\mathrm{Spin}(p, q)$ is spanned by bivectors $\mathbf{e}_j \mathbf{e}_k$ $(1 \le j < k \le N)$. Note that $[\mathbf{e}_m \mathbf{e}_k, \mathbf{e}_j \mathbf{e}_k] = 2\langle \mathbf{e}_k \mid \mathbf{e}_k\rangle \, \mathbf{e}_j \mathbf{e}_m$ $(m \ne j \ne k \ne m)$.

One can define following automorphisms and anti-automorphisms in any Clifford algebra (using also linearity, in the first two cases, or anti-linearity, in the third case):

- grade involution $\boldsymbol{\alpha}$: $\quad \boldsymbol{\alpha}(XY) = \boldsymbol{\alpha}(X)\boldsymbol{\alpha}(Y), \quad \boldsymbol{\alpha}(v) = -v.$
- reversion $\boldsymbol{\beta}$: $\quad \boldsymbol{\beta}(XY) = \boldsymbol{\beta}(Y)\boldsymbol{\beta}(X)\,, \quad \boldsymbol{\beta}(v) = v.$
- complex conjugation: $\quad \overline{XY} = \bar{X}\bar{Y}, \quad \bar{v} = v.$
- Clifford conjugation (complexified): $\quad X^* := \boldsymbol{\alpha}(\boldsymbol{\beta}(\bar{X})).$

The reversion can be used to compute the so-called spinor norm:

$$N(X) = X\beta(X)\,, \tag{9}$$

which is a real number for $X \in \Gamma(V, Q)$.

3 Spin-Valued Linear Problems

We consider linear problems of the form:

$$\Psi_{,\mu} = U_\mu \Psi\,, \qquad U_\mu = \sum_{j<k} u_{jk}\mathbf{e}_j\mathbf{e}_k\,, \tag{10}$$

where $\mathbf{e}_j$ are generators of a Clifford algebra, and u_{jk} depend on x^μ, λ and are real for $\lambda \in \mathbb{R}$. However, in the following, we suppress the dependence on x^μ and, sometimes, on λ). Then, Ψ is Spin-valued (provided that the initial condition is Spin-valued).

Our original motivation came from studying isothermic surfaces (which, by definition, admit conformal parameterization of curvature lines). Using isomorphisms

$$\mathrm{so}(4,1) \;\cong\; \mathrm{sp}(1,1) \;\cong\; \mathrm{spin}(4,1) \tag{11}$$

we transformed $SO(4,1)$-valued Lax pair into the form (10). Clifford algebras are, in general, a useful tool in dealing with isothermic surfaces [22].

We point out that Lie algebra $su(2)$ is spanned by $\sigma_k\sigma_j$, hence $SU(2) \cong \mathrm{Spin}(3)$ and all $SU(2)$-valued linear problems, including (1), belong to the class (10).

The case of isothermic surfaces suggested further restrictions on the form of the linear problem. We consider two vector spaces, V and W, equipped with quadratic forms and orthogonal to each other, such that

- $\dim V = r$ and $\mathbf{e}_1, \ldots, \mathbf{e}_r$ is an orthonormal basis in V,
- $\dim W = q$ and $\mathbf{e}_{r+1}, \ldots, \mathbf{e}_{r+q}$ is an orthonormal basis in W.

In this paper we focus on the following class of linear problems [23]:

$$\Psi_{,\mu} = U_\mu \Psi \ , \qquad U_\mu = \tfrac{1}{2}\mathbf{e}_\mu(\lambda \mathbf{a}_\mu + \mathbf{b}_\mu) \ , \qquad (\mu = 1, \ldots, m), \tag{12}$$

where $m \leqslant r$ (which implies $\mathbf{e}_\mu \in V$ for $\mu = 1, \ldots, m$), and

$$\mathbf{a}_\mu = \mathbf{a}_\mu(x^1, \ldots, x^m) \in W, \quad \mathbf{b}_\mu = \mathbf{b}_\mu(x^1, \ldots, x^m) \in V, \quad \langle \mathbf{b}_\nu \mid \mathbf{e}_\nu \rangle = 0. \tag{13}$$

Therefore, $U_\mu = U_\mu(x^1, \ldots, x^m, \lambda) \in \mathrm{spin}(V \oplus W) \cong \mathrm{so}(V \oplus W)$, and, as a consequence, $\Psi = \Psi(x^1, \ldots, x^m, \lambda) \in \mathrm{Spin}(V \oplus W)$ (provided that Ψ belongs to the Spin group at an initial point).

3.1 Geometric Interpretation: Soliton Surfaces Approach

The so-called Sym's (or Sym-Tafel's) formula:

$$F := \Psi^{-1}\Psi_{,\lambda} \tag{14}$$

provides a geometric interpretation for integrable systems associated with a given linear problem. Note that if $\Psi \in G$ (where G is a Lie group), then F takes values in the Lie algebra of G. One can easily verify:

$$F_{,\mu} = \Psi^{-1} U_{\mu,\lambda}\, \Psi \ , \qquad F_{,\mu\nu} = \Psi^{-1} \left(U_{\mu,\lambda\nu} + [U_{\mu,\lambda} \, , U_\nu] \right) \Psi \ . \tag{15}$$

Hence, fundamental forms of F (including $g_{\mu\nu} \equiv \langle F_{,\mu} \mid F_{,\mu} \rangle$) are expressed in terms of U_μ (explicit form of Ψ is not needed).

In the case of the linear problems (10) it is sufficient to consider the Sym-Tafel formula evaluated at $\lambda = 0$. Then, F is a submanifold in $V \wedge W$. In order to obtain interesting immersions in lower dimensional spaces we can use appropriately chosen projection P, i.e., we consider $\boldsymbol{r} = P(F)$. In particular, we get

- **Isothermic surfaces** in $\mathbb{R}^n$. $\dim V = n, \dim W = 2$,
  $W \cong \mathbb{R}^{1,1}$, $\ker P$ is any isotropic (null) vector in W.
- **Orthogonal nets** in $\mathbb{R}^n$ such that $\sum_{k=1}^n h_k^2 = \text{const}$.
  $\dim V = \dim W = n$, P is a projection on any $\mathbf{e}_k \in W$.
- **Guichard nets** in R^3 $(h_1^2 + h_2^2 = h_3^2)$. $\dim V = n, \dim W = 3$,
  $W \cong \mathbb{R}^{2,1}$, $\ker P$ is a light front (tangent space to the light cone) in W.

4 Darboux-Bäcklund Transformations

Darboux transformation is a gauge-like transformation using the "Darboux matrix" D:

$$\tilde{\Psi} = D\Psi, \qquad \tilde{\Psi}_{,\mu} = \tilde{U}_\mu \tilde{\Psi}, \qquad \tilde{U}_\mu = D_{,\mu}\, D^{-1} + D U_\mu D^{-1}\,, \tag{16}$$

provided that $\tilde{U}_\mu$ has the same dependence on dependent variables as U_μ.

Instead of caring about dependent variables one can try to describe the structure of the Lax pair. The considered nonlinear system follows uniquely from compatibility conditions [8, 9, 13]. Then the Darboux transformation has to preserve this structure.

The structure is characterized primarily by the dependence on λ (e.g., divisor of poles) [24, 25], reduction group (loop group) [26], and other invariants of Darboux transformations, like linear and multilinear constraints on coefficients of the Laurent expansion around poles [13].

Different methods of constructing the Darboux matrix need different form of λ-dependence of D (these forms are equivalent up to a λ-dependent scalar factor). In particular, one can assume D as polynomial in λ (*eigenvalues*, corresponding to solitons, are zeros of $\det D$) [27, 28], sum of simple fractions (*eigenvalues*: poles of D and D^{-1}) [24, 26], or a "realization" ($D = N + F(\lambda - A)^{-1}G$) [29, 30].

The motivation for the case of Spin groups came from yet another approach [31]. Multiplying (16) by $D^2(\lambda)$ we get

$$D_{,\mu}\, D + D U_\mu D = \tilde{U}_\mu D^2\,. \tag{17}$$

It is crucial point that the right-hand side vanishes for λ_+ and λ_- such that $D^2(\lambda_\pm) = 0$. Then, we obtain a solution of the remaining equation: $D(\lambda_\pm) = \varphi_\pm \Psi(\lambda_\pm) d_\pm \Psi(\lambda_\pm)^{-1}$, where $d_\pm = \text{const}$, $(d_\pm)^2 = 0$ and $\varphi_\pm$ are two scalar functions. Finally, $D(\lambda)$ is given as a linear combination of $D(\lambda_+)$ and $D(\lambda_-)$ with coefficients linear in λ [31], which yields one-soliton Darboux matrix.

This approach was extended on the multi-soliton case for 2×2 matrix problems [32], and analogous generalization for Spin-valued linear problems is introduced below.

4.1 The Darboux-Bäcklund Transformation in the Case of Spin Groups

In this section we focus on the linear problem (12) and assume $\Psi \in \mathrm{Spin}(V \oplus W)$. Using commutation relations (4) one can easily verify that

$$\boldsymbol{\beta}(U_\mu) = -U_\mu\,, \qquad \mathbf{e}\, U_\mu(\lambda) = U_\mu(-\lambda)\mathbf{e}\,, \qquad \overline{U_{,\mu}(\lambda)} = U_\mu(\bar{\lambda})\,, \tag{18}$$

where $\mathbf{e} = \mathbf{e}_{r+1}\mathbf{e}_{r+2}\ldots\mathbf{e}_{r+q}$. In a way analogous to the matrix case one can obtain corresponding formulas for Ψ:

$$N(\Psi) \equiv \Psi\boldsymbol{\beta}(\Psi) = \text{const}\,, \qquad \mathbf{e}\,\Psi(\lambda) = \Psi(-\lambda)\mathbf{e}\,, \qquad \overline{\Psi(\lambda)} = \Psi(\bar{\lambda})\,. \tag{19}$$

We derive the first formula, which is most difficult:

$$(N(\Psi))_{,\mu} = \Psi_{,\mu}\,\boldsymbol{\beta}(\Psi) + \Psi\boldsymbol{\beta}(\Psi_{,\mu}) = (U_\mu + \boldsymbol{\beta}(U_\mu))N(\Psi) = 0\,, \tag{20}$$

where one has to remember that $N(\Psi)$ is a scalar, so it commutes with any elements.

The constraints (19) are obviously satisfied when we put $D(\lambda)$ in place of $\Psi(\lambda)$. We point out that then constant in the first equation of (19) will depend on λ and zeros of $N(D(\lambda))$ are *eigenvalues* used in the construction of soliton solutions.

The simplest case is the Darboux transformation which is a Clifford vector and is linear in λ. Then, the result can be obtained using (17), see [31]:

$$D = \frac{\lambda n + \kappa p}{\sqrt{p^2}} = \lambda\hat{n} + \kappa\,\hat{p}\,, \tag{21}$$

where hat denotes unit vectors and

$$p + in := \Psi(i\kappa)(p_0 + in_0)\Psi^{-1}(i\kappa)\,, \tag{22}$$

where $p_0 \in V$ and $n_0 \in W$ are constant Clifford vectors such that $p_0^2 = n_0^2$ and $\kappa \in \mathbb{R}$. Transformations for soliton submanifolds (14) (evaluated at $\lambda = 0$) read

$$\tilde{F} = F + \frac{1}{\kappa}\hat{p}^{-1}\hat{n}\,, \qquad \tilde{\mathbf{r}} = \mathbf{r} + \frac{1}{\kappa}\hat{p}^{-1}P(\hat{n})\,. \tag{23}$$

The Darboux "matrix" (21), being a Clifford vector, produces $\tilde{\Psi}$ which does not belong to the Spin group. However, due to the invariance of the linear problem (12) with respect to the transformation $\Psi \to \Psi w$, we can take $w \in V \oplus W$ and then $D\Psi w \in \mathrm{Spin}(V \oplus W)$. Then formulas (23) have to be changed as well (geometrically this is just a reflection).

This approach has many practical advantages. Calculations are much shorter using Clifford numbers than matrix representations. It is enough to compare the length and content of two papers, [31] and [33], which present in fact the same final result.

Iterating twice the transformation (21) (with parameters κ_1 and κ_2, respectively) and performing some algebraic calculations, we succeeded to obtain the following symmetric form of the two-soliton Darboux transformation:

$$D^{[2]}(\lambda) = \frac{S(\lambda) - (\kappa_1^2 - \kappa_2^2)D_{[0]1} \wedge D_{[0]2}}{K}\,, \tag{24}$$

where $\wedge$ denotes the exterior (or wedge) product, $D_{[0]j}(\lambda) = \lambda\hat{n}_j + \kappa_j\hat{p}_j$

$$\begin{aligned} S(\lambda) &= \kappa_1\kappa_2(2\lambda^2 + \kappa_1^2 + \kappa_2^2)\sigma - (2\kappa_1^2\kappa_2^2 + \lambda^2(\kappa_1^2 + \kappa_2^2))\nu\,, \\ K^2 &:= 4\kappa_1^2\kappa_2^2(\sigma^2 + \nu^2) - 4\kappa_1\kappa_2(\kappa_1^2 + \kappa_2^2)\sigma\nu + (\kappa_1^2 - \kappa_2^2)^2 \end{aligned} \tag{25}$$

and, finally, $\sigma := \langle\hat{p}_1 \mid \hat{p}_2\rangle$ and $\nu := \langle\hat{n}_1 \mid \hat{n}_2\rangle$.

The full multi-soliton case is still under construction but basic results are already obtained. In the general case, when $D(\lambda) \in \mathrm{Pin}(V \oplus W)$ is a polynomial in λ, we multiply (16) by the spinor norm $N(D)$, defined by (9), and obtain:

$$D_{,\mu}\, \boldsymbol{\beta}(D) + DU_\mu \boldsymbol{\beta}(D) = \tilde{U}_\mu N(D) \tag{26}$$

Note that the special case $D \in V \oplus W$ reduces to (17), because then $\boldsymbol{\beta}(D) = D$ and $N(D) = D^2$. In the general case the right-hand side of (26) vanishes if $N(D(\lambda)) = 0$. This is a polynomial in λ. Its roots, denoted by λ_k, are constant by virtue of (20). Similarly as in the case (17) we define

$$C_k := \Psi^{-1}(\lambda_k) D(\lambda_k) \Psi(\lambda_k)\ . \tag{27}$$

Substituting it to (26) we obtain

$$C_k \boldsymbol{\beta}(C_k) = 0\ , \qquad C_{k,\mu}\, \boldsymbol{\beta}(C_k) = 0\ . \tag{28}$$

Following [34], we can solve this system:

$$C_k = \hat{C}_k d_k\ , \quad d_k = \mathrm{const} \in V\ , \quad d_k^2 = 0\ , \quad \hat{C}_k \in \mathrm{Pin}(V \oplus W)\ . \tag{29}$$

Therefore, $D(\lambda_k) = \hat{D}_k\, \Psi(\lambda_k) d_k \Psi^{-1}(\lambda_k)$, where $\hat{D}_k \in \mathrm{Pin}(V \oplus W)$. It is natural to conjecture that $D(\lambda)$ can be uniquely (up to a normalization) defined by eigenvalues λ_k and null vectors d_k and can be expressed in an algebraic way by elements $\Psi(\lambda_k) d_k \Psi^{-1}(\lambda_k)$. Reduction groups will be expressed in terms of involutions in the Clifford algebra.

5 Conclusions and Open Problems

Extending results of [31] we presented a method of construction of the Darboux matrix for Spin-valued linear problems. The construction is fully done only in the one-soliton case. The iteration of one-soliton transformations is not difficult, and two-soliton Darboux matrix is shown. However, the multi-soliton case needs further elaboration. Then, our approach should work also for multi-dimensional Lobachevsky spaces [20, 35] and, possibly, for all problems described in [35, 36].

References

1. D. Levi, A. Sym, Integrable systems describing surfaces of non-constant curvature. Phys. Lett. A **149**, 381–387 (1990)
2. D. Levi, A. Sym, G.-Z. Tu, A working Algorithm to Isolate Integrable Surfaces in E^3. Dip. Fis. INFN N.761, 10/10/1990. Preprint (1990)

3. J. Cieśliński, Lie symmetries as a tool to isolate integrable geometries, in *Nonlinear Evolution Equations and Dynamical System*, ed. by M. Boiti, L. Martina, F. Pempinelli (World Scientific, Singapore 1992), pp. 260–268
4. J. Cieśliński, Non-local symmetries and a working algorithm to isolate integrable geometries. J. Phys. A Math. Gen. **26**, L267–L271 (1993)
5. J. Cieśliński, Group interpretation of the spectral parameter in the case of nonhomogeneous, nonlinear Schrödinger system. J. Math. Phys. **34**, 2372–2384 (1993)
6. J.L. Cieśliński, P. Goldstein, A. Sym, On integrability of the inhomogeneous Heisenberg ferromagnet model: Examination of a new test. J. Phys. A Math. Gen. **27**, 1645–1664 (1994)
7. D. Levi, Hierarchies of integrable equations obtained as non-isospectral (in x and t) deformations of the Schrödinger spectral problem. Phys. Lett. A **119**, 453–456 (1987)
8. J. Cieśliński, Algebraic representation of the linear problem as a method to construct the Darboux-Bäcklund transformation. Chaos Solitons Fractals **5**, 2303–2313 (1995)
9. J. Cieśliński, An algebraic method to construct the Darboux matrix. J. Math. Phys. **36**, 5670–5706 (1995)
10. J. Cieśliński, D. Levi, A. Sym, Solitons on a Relativistic String. DF-INFN N.496, 13/1/86. Preprint (1986)
11. J. Cieśliński, An effective method to compute N-fold Darboux matrix and N-soliton surfaces. J. Math. Phys. **32**, 2395–2399 (1991)
12. J. Cieśliński, Two solitons on a thin vortex filament. Phys. Lett. A **171**, 323–326 (1992)
13. J.L. Cieśliński, Algebraic construction of the Darboux matrix revisited. J. Phys. A Math. Theor. **42**, 404003 (2009)
14. J.L. Cieśliński, A. Kobus, Group interpretation of the spectral parameter. The case of isothermic surfaces. J. Geom. Phys. **113**, 28–37 (2017)
15. J.L. Cieśliński, An orbit-preserving discretization of the classical Kepler problem. Phys. Lett. A **370**, 8–12 (2007)
16. J.L. Cieśliński, B. Ratkiewicz, Energy-preserving numerical schemes of high accuracy for one-dimensional Hamiltonian systems. J. Phys. A Math. Theor. **44**, 155206 (2011)
17. D. Levi, L. Martina, P. Winternitz, Structure preserving discretizations of the Liouville equation and their numerical tests. SIGMA **11**, 080 (2015)
18. J. Vaz, R. da Rocha, *An Introduction to Clifford Algebras and Spinors* (Oxford University Press, 2017)
19. A. Sym, Soliton surfaces and their application (Soliton geometry from spectral problems), in *Geometric Aspects of the Einstein Equations and Integrable Systems*, ed. by R. Martini. Lecture Notes in Physics, vol. 239 (Springer, Berlin–Heidelberg, 1985), pp. 154–231
20. J.L. Cieśliński, Geometry of submanifolds derived from spin-valued spectral problems. Theor. Math. Phys. **137**, 1396–1405 (2003)
21. P. Lounesto, *Clifford Algebras and Spinors*, 2nd edn. (Cambridge University Press, Cambridge, 2001)
22. F.E. Burstall, Isothermic surfaces: conformal geometry, Clifford algebras and integrable systems, in *Integrable systems, Geometry and Topology*, ed. by C.-L. Terng. AMS/IP Studies in Advanced Math., vol. 36, pp. 1–82 (2006)
23. J.L. Cieśliński, A class of spectral problems in Clifford algebras. Phys. Lett. A **267**, 251–255 (2000)
24. S.P. Novikov, S.V. Manakov, L.P. Pitaievsky, V.E. Zakharov, *Theory of Solitons* (Springer US, New York, 1984)
25. C.H. Gu, Bäcklund transformations and Darboux transformations, in *Soliton Theory and Its Applications*, ed. by C.H. Gu. (Springer, Berlin–Heidelberg, 1995), pp. 122–151
26. A.V. Mikhailov, The reduction problem and the inverse scattering method. Physica D **3**, 73–117 (1981)
27. G. Neugebauer, R. Meinel, General N-soliton solution of the AKNS class on arbitrary background. Phys. Lett. A **100**, 467–470 (1984)
28. C. Rogers, W.K. Schief, *Bäcklund and Darboux Transformations: Geometry and Modern Applications in Soliton Theory* (Cambridge University Press, Cambridge, 2002)

29. A.L. Sakhnovich, Dressing procedure for solutions of nonlinear equations and the method of operator identities. Inverse Problems **10**, 699–710 (1994)
30. A.L. Sakhnovich, L.A. Sakhnovich, I.Ya. Roitberg, *Inverse Problems and Nonlinear Evolution Equations* (De Gruyter, 2013)
31. W. Biernacki, J.L. Cieśliński, A compact form of the Darboux-Bäcklund transformation for some spectral problems in Clifford algebras. Phys. Lett. A **288**, 167–172 (2001)
32. J.L. Cieśliński, W. Biernacki, A new approach to the Darboux-Bäcklund transformations versus the standard dressing method. J. Phys. A Math. Gen. **38**, 9491 (2005)
33. J. Cieśliński, The Darboux-Bianchi transformation for isothermic surfaces. Classical results versus the soliton approach. Differ. Geom. Appl. **7**, 1–28 (1997)
34. J.L. Cieśliński, Divisors of zero in the Lipschitz semigroup. Adv. Appl. Clifford Algebras **17**, 153–157 (2007)
35. K. Tenenblat, Transformations of Manifolds and Applications to Differential Equations (Addison Wesley Longman, 1998)
36. M.J. Ablowitz, R. Beals, K. Tenenblat, On the solution of the generalized wave and generalized sine-Gordon equations. Stud. Appl. Math. **74**, 177–203 (1986)

Painlevé IV Transcendents Generated from the Complex Oscillator

David J. Fernández

Abstract Supersymmetry transformations are used to generate exactly solvable potentials departing from the complex oscillator. The corresponding Hamiltonians are shown to be ruled by polynomial Heisenberg algebras. A process for reducing the degree of these algebras to 2 is used to connect such systems with the Painlevé IV equation, thus leading to a simple algorithm for generating Painlevé IV transcendents.

Keywords Painlevé transcendents · Complex oscillator · Supersymmetric quantum mechanics

1 Introduction

The recent scientific advances make it important to study the links that could exist between supersymmetric quantum mechanics (SUSY QM) and nonlinear differential equations [1]. Indeed, there is a well known connection between SUSY partners of the free particle and solutions of the KdV equation [2–4]. Similarly, it has been shown that there is a link between systems ruled by second-degree polynomial Heisenberg algebras and Painlevé IV (PIV) equation [5–12]. This connection helped to design further an algorithm for generating solutions to the PIV equation, called Painlevé IV transcendents in the literature [13, 14]. The simplest systems that can be used to supply explicit expressions for PIV transcendents are the harmonic oscillator and its SUSY partners [15, 16]. It would be important to know if the so-called complex oscillator [17], which arises from making complex the oscillator frequency and includes the harmonic oscillator as a limit, does the same. This subject is going to be explored in this article. In order to do this, in Sect. 2 we will make a quick

D. J. Fernández (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: david@fis.cinvestav.mx

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
http://www.fis.cinvestav.mx/~david

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_4

survey of supersymmetric quantum mechanics (SUSY QM) [1]. Then, in Sect. 3, we will sketch the polynomial Heisenberg algebras (PHA), paying special attention to the second-degree ones. In Sects. 4 and 5 we shall address the complex oscillator and its SUSY partners, respectively. In Sect. 6 we will derive Painlevé IV transcendents from these two examples. Our conclusions shall be presented in Sect. 7.

2 Supersymmetric Quantum Mechanics

The supersymmetry algebra with two generators introduced by Witten in 1981

$$[H_{\rm ss}, Q_i] = 0, \quad \{Q_i, Q_j\} = \delta_{ij} H_{\rm ss}, \quad i, j = 1, 2,$$

when realized as follows

$$Q_1 = \frac{Q^+ + Q}{\sqrt{2}}, \quad Q_2 = \frac{Q^+ - Q}{i\sqrt{2}}, \quad Q = \begin{pmatrix} 0 & 0 \\ B & 0 \end{pmatrix},$$
$$Q^+ = \begin{pmatrix} 0 & B^+ \\ 0 & 0 \end{pmatrix}, \quad H_{\rm ss} = \{Q, Q^+\} = \begin{pmatrix} B^+B & 0 \\ 0 & BB^+ \end{pmatrix},$$

is called supersymmetric quantum mechanics, where $H_{\rm ss}$ is the supersymmetric Hamiltonian and Q_1, Q_2 are the supercharges. There exist two Schrödinger Hamiltonians H, $\widetilde{H}$ and a kth order differential operator B^+ intertwining them:

$$\widetilde{H}B^+ = B^+H, \quad H = -\frac{1}{2}\frac{d^2}{dx^2} + V(x), \quad \widetilde{H} = -\frac{1}{2}\frac{d^2}{dx^2} + \widetilde{V}(x).$$

The two different products of B and B^+ turn out to be given by

$$B^+B = (\widetilde{H} - \epsilon_1)\cdots(\widetilde{H} - \epsilon_k), \qquad BB^+ = (H - \epsilon_1)\cdots(H - \epsilon_k),$$

which implies that

$$H_{\rm ss} = (H_{\rm p} - \epsilon_1)\cdots(H_{\rm p} - \epsilon_k), \qquad H_{\rm p} = \begin{pmatrix} \widetilde{H} & 0 \\ 0 & H \end{pmatrix}.$$

If H is a given initial Hamiltonian from which we wish to construct $\widetilde{H}$, then k *seed solutions* u_i, $i = 1, \ldots, k$ are required, such that

$$Hu_i = \epsilon_i u_i.$$

Thus, the new potential is given by

$$\widetilde{V}(x) = V(x) - [\log W(u_1, \ldots, u_k)]'',$$

with $W(u_1, \ldots, u_k)$ being the Wronskian of the k seed solutions, while the eigenfunctions (perhaps just formal) of $\widetilde{H}$ associated with E_n and ϵ_i become, respectively,

$$\widetilde{\psi}_n \propto B^+\psi_n \propto \frac{W(u_1, \ldots, u_k, \psi_n)}{W(u_1, \ldots, u_k)},$$
$$\widetilde{\psi}_{\epsilon_i} \propto \frac{W(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_k)}{W(u_1, \ldots, u_k)},$$

where we have assumed that $H\psi_n = E_n\psi_n$.

3 Polynomial Heisenberg Algebras

The polynomial Heisenberg algebras (PHA) are deformations of the Heisenberg-Weyl algebra with three generators H, $\mathcal{L}^+$, and $\mathcal{L}^-$ such that [9]

$$[\mathsf{H}, \mathcal{L}^\pm] = \pm\mathcal{L}^\pm,$$
$$[\mathcal{L}^-, \mathcal{L}^+] \equiv Q_{m+1}(\mathsf{H}+1) - Q_{m+1}(\mathsf{H}) = P_m(\mathsf{H}),$$
$$Q_{m+1}(\mathsf{H}) \equiv \mathcal{L}^+\mathcal{L}^- = \prod_{i=1}^{m+1}(\mathsf{H} - \mathcal{E}_i)\,.$$

The energy spectra of systems ruled by PHA depend on how many extremal states in the kernel of $\mathcal{L}^-$ become also physical eigenstates of H. If s of those extremal states satisfy

$$\mathcal{L}^-\psi_{\mathcal{E}_i} = 0, \quad \mathsf{H}\psi_{\mathcal{E}_i} = \mathcal{E}_i\psi_{\mathcal{E}_i}, \quad i = 1, \ldots, s,$$

as well as the defining boundary conditions, then from the iterated action of $\mathcal{L}^+$ onto each one of them we can construct s infinite energy ladders for H. It could happen that one of those infinite ladders (let us say the jth one) truncates after the nth step, i.e., $(\mathcal{L}^+)^{n-1}\psi_{\mathcal{E}_j} \neq 0$, $(\mathcal{L}^+)^n\psi_{\mathcal{E}_j} = 0$. In such a case it must happen that $\mathcal{E}_\ell = \mathcal{E}_j + n$ for some $\ell \in \{s+1, \ldots, k\}$.

An important differential realization of the PHA arises if H is a 1-dim Schrödinger Hamiltonian

$$\mathsf{H} = -\frac{1}{2}\frac{d^2}{dx^2} + \mathsf{V}(x),$$

while $\mathcal{L}^\pm$ are $(m+1)$th order differential ladder operators. In particular, the case with $m = 2$ is worth of further study.

3.1 Second-Degree PHA

By taking $m = 2$ we arrive at the second-degree PHA, for which

$$Q_3(\mathsf{H}) = (\mathsf{H} - \mathcal{E}_1)(\mathsf{H} - \mathcal{E}_2)(\mathsf{H} - \mathcal{E}_3), \quad P_2(\mathsf{H}) = 3\mathsf{H}^2 + (3 - 2\mathcal{S})\mathsf{H} + \mathcal{P} + 1 - \mathcal{S},$$

where $\mathcal{S} = \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3$, $\mathcal{P} = \mathcal{E}_1\mathcal{E}_2 + \mathcal{E}_1\mathcal{E}_3 + \mathcal{E}_2\mathcal{E}_3$, and $\mathcal{L}^{\pm}$ are third-order differential operators. Systems ruled by second-degree PHA could have up to 3 infinite energy ladders starting from $\mathcal{E}_1,\ \mathcal{E}_2, \mathcal{E}_3$.

It is important to look for the most general Schrödinger Hamiltonians ruled by second-degree PHA. In order to find them, let us take $\mathcal{L}^{\pm}$ as

$$\mathcal{L}^+ = L_1^+ L_2^+, \quad L_1^+ = \frac{1}{\sqrt{2}}\left(-\frac{d}{dx} + f(x)\right), \quad L_2^+ = \frac{1}{2}\left(\frac{d^2}{dx^2} + g(x)\frac{d}{dx} + h(x)\right),$$

$$\mathsf{H}L_1^+ = L_1^+(\mathsf{H}_a + 1), \quad \mathsf{H}_a L_2^+ = L_2^+\mathsf{H} \quad \Rightarrow \quad [\mathsf{H}, \mathcal{L}^+] = \mathcal{L}^+.$$

A straightforward calculation leads to

$$f = x + g(x), \qquad h = -x^2 + \frac{g'}{2} - \frac{g^2}{2} - 2xg + a,$$

$$\mathsf{V} = \frac{x^2}{2} - \frac{g'}{2} + \frac{g^2}{2} + xg + \mathcal{E}_3 - \frac{1}{2},$$

where the key function g satisfies the Painlevé IV equation:

$$g'' = \frac{g'^2}{2g} + \frac{3}{2}g^3 + 4xg^2 + 2\left(x^2 - a\right)g + \frac{b}{g},$$

with $a = \mathcal{E}_1 + \mathcal{E}_2 - 2\mathcal{E}_3 - 1,\ b = -2\Delta^2,\ \Delta = \mathcal{E}_1 - \mathcal{E}_2$. The three extremal states can be expressed in terms of g as follows

$$\psi_{\mathcal{E}_1} \propto \left(\tfrac{g'}{2g} - \tfrac{g}{2} - \tfrac{\Delta}{g} - x\right)\exp\left[\int\left(\tfrac{g'}{2g} + \tfrac{g}{2} - \tfrac{\Delta}{g}\right)\mathrm{d}x\right], \tag{1}$$

$$\psi_{\mathcal{E}_2} \propto \left(\tfrac{g'}{2g} - \tfrac{g}{2} + \tfrac{\Delta}{g} - x\right)\exp\left[\int\left(\tfrac{g'}{2g} + \tfrac{g}{2} + \tfrac{\Delta}{g}\right)\mathrm{d}x\right], \tag{2}$$

$$\psi_{\mathcal{E}_3} \propto \exp\left(-\tfrac{x^2}{2} - \int g\,\mathrm{d}x\right). \tag{3}$$

We conclude that the most general Hamiltonians ruled by second-degree PHA have potentials expressed in terms of Painlevé IV transcendents. Conversely, Eq. (3) leads to

$$g(x) = -x - \{\ln[\psi_{\mathcal{E}_3}(x)]\}',$$

thus Painlevé IV transcendents can be found by simply supplying the extremal states of Hamiltonians ruled by second-degree PHA, having third-order differential ladder operators.

4 Complex Oscillator

The complex oscillator potential is given by [17]

$$V(x) = \frac{1}{2}\omega^2 x^2, \quad \omega = e^{i\theta}, \quad -\frac{\pi}{2} \leq \theta < \frac{3\pi}{2}.$$

The general solution to the associated Schrödinger equation reads

$$\begin{aligned} u(x,\varepsilon) &= e^{-\frac{\omega x^2}{2}}\left[{}_1F_1\left(\frac{1}{4} - \frac{\varepsilon}{2\omega}, \frac{1}{2}; \omega x^2\right) + \lambda\, x\, {}_1F_1\left(\frac{3}{4} - \frac{\varepsilon}{2\omega}, \frac{3}{2}; \omega x^2\right)\right] \\ &= e^{\frac{\omega x^2}{2}}\left[{}_1F_1\left(\frac{1}{4} + \frac{\varepsilon}{2\omega}, \frac{1}{2}; -\omega x^2\right) + \lambda\, x\, {}_1F_1\left(\frac{3}{4} + \frac{\varepsilon}{2\omega}, \frac{3}{2}; -\omega x^2\right)\right], \end{aligned}$$

where $\lambda = 2\,\nu\,\Gamma\left(\frac{3}{4} - \frac{\varepsilon}{2\omega}\right)/\Gamma\left(\frac{1}{4} - \frac{\varepsilon}{2\omega}\right)$. The square-integrable solutions of such non-Hermitian Hamiltonian are given by

$$\phi_n(x) = C_n \mathrm{H}_n(\sqrt{\omega}x)e^{-\frac{1}{2}\omega x^2}, \quad E_n(\theta) = (n + \tfrac{1}{2})e^{i\theta},$$

for $\sqrt{\omega} = e^{i\frac{\theta}{2}}$, $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ and

$$\phi_n(x) = D_n \mathrm{H}_n(\sqrt{-\omega}x)e^{\frac{1}{2}\omega x^2}, \quad E_n(\theta) = (n + \tfrac{1}{2})e^{i(\theta-\pi)},$$

for $\sqrt{-\omega} = e^{i(\frac{\theta-\pi}{2})}$, $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$, where C_n, D_n are normalization factors and $\mathrm{H}_n(z)$ are the Hermite polynomials of complex argument z. Note that for $\theta = \pm\frac{\pi}{2}$ there are no square-integrable solutions for the stationary Schrödinger equation since the complex oscillator potential reduces then to the repulsive oscillator. As the eigenvalues lie in the first or in the fourth quadrant in the complex E-plane we can take $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$ without loss of generality. Moreover, since $E_n(-\theta) = [E_n(\theta)]^*$ we can further restrict to $0 \leq \theta < \frac{\pi}{2}$.

Similarly, as for the standard oscillator, the analogues of the annihilation and creation operators

$$a_\omega^\pm = \frac{1}{\sqrt{2}}\left(\mp\frac{\mathrm{d}}{\mathrm{dx}} + \omega x\right),$$

fulfill the following relations:

$$[H, a_\omega^\pm] = \pm\omega a_\omega^\pm, \quad \{a_\omega^-, a_\omega^+\} = 2H, \quad [a_\omega^-, a_\omega^+] = \omega.$$

From them it is possible to determine algebraically the eigenfunctions of H [17]:

$$\phi_n(x) = \frac{\widetilde{C}_n}{\sqrt{n!}} \left(a_\omega^+\right)^n \phi_0(x), \qquad \phi_0(x) = \left(\frac{\cos\theta}{\pi}\right)^{\frac{1}{4}} e^{-\frac{\omega x^2}{2}}.$$

5 SUSY Partners of the Complex Oscillator

In order to perform a kth order SUSY transformation, k seed solutions $u_1, \ldots, u_k$ associated with complex factorization energies $\varepsilon_1, \ldots, \varepsilon_k$ are to be taken, so that k new levels for $\widetilde{H}$ will be created [1]:

$$\mathsf{Sp}(\widetilde{H}) = \left\{\varepsilon_j, E_n;\ j = 1, \cdots, k;\ n = 0, 1, 2, \cdots\right\}.$$

The new potential reads

$$\widetilde{V} = \tfrac{1}{2}\omega^2 x^2 - [\log W(u_1, \ldots, u_k)]''.$$

Since the oscillation theorem is no longer valid, the factorization energies $\varepsilon_1, \ldots, \varepsilon_k$ can be chosen essentially at any position on the complex E-plane.

The natural ladder operators for $\widetilde{H}$ are given by

$$\mathcal{L}^\pm = B^+ a_\omega^\pm B^-.$$

They fulfill a PHA of degree $2k$, since:

$$[\widetilde{H}, \mathcal{L}^\pm] = \pm\omega\mathcal{L}^\pm, \qquad [\mathcal{L}^-, \mathcal{L}^+] = P_{2k}(\widetilde{H}),$$
$$Q_{2k+1}(\widetilde{H}) = \left(\widetilde{H} - \frac{\omega}{2}\right) \prod_{i=1}^{k} \left(\widetilde{H} - \varepsilon_i - \omega\right)\left(\widetilde{H} - \varepsilon_i\right).$$

As an example, for $k = 1$ the simplest non-singular transformations can be implemented by using the bound state seed solutions with $n = 2j$:

$$u_1(x) = \mathrm{H}_{2j}(\sqrt{\omega}x) e^{-\frac{1}{2}\omega x^2}, \quad j = 0, 1, 2, \cdots \tag{4}$$

The first-order SUSY partner potential now takes the form

$$\widetilde{V}(x) = \tfrac{1}{2}\omega^2 x^2 + \omega - 8j\omega\left[(2j-1)\frac{\mathrm{H}_{2j-2}(\sqrt{\omega}x)}{\mathrm{H}_{2j}(\sqrt{\omega}x)} - 2j\left(\frac{\mathrm{H}_{2j-1}(\sqrt{\omega}x)}{\mathrm{H}_{2j}(\sqrt{\omega}x)}\right)^2\right].$$

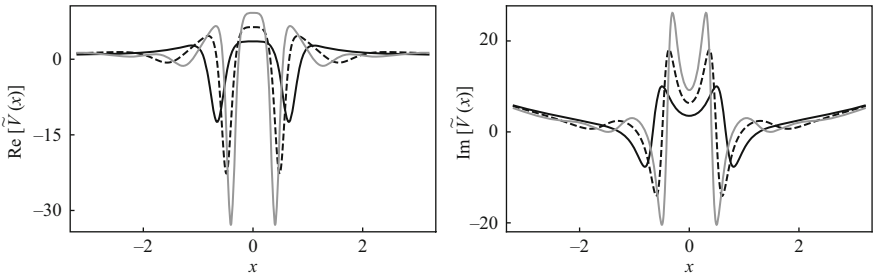


Fig. 1 Real (left) and imaginary (right) parts of $\widetilde{V}(x)$ for $k = 1$, $j = 1$ (black continuous line), $j = 2$ (dashed line), and $j = 3$ (gray continuous line), with $\theta = \frac{\pi}{4}$

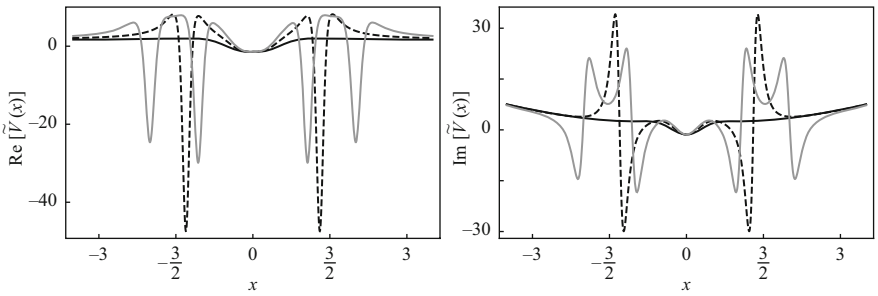


Fig. 2 Real (left) and imaginary (right) parts of $\widetilde{V}(x)$ for $k = 2$, $j = 1$ (black continuous line), $j = 2$ (dashed line), and $j = 3$ (gray continuous line), with $\theta = \frac{\pi}{4}$

Plots for some of these potentials can be seen in Fig. 1.

On the other hand for $k = 2$, with $u_1(x)$ as given in Eq. (4) and $u_2 = a_\omega^- u_1$, it is obtained

$$\widetilde{V}(x) = \tfrac{1}{2}\omega^2 x^2 - \frac{W''}{W} + \left(\frac{W'}{W}\right)^2,$$

$$W \equiv W(u_1, u_2) \propto e^{-\omega x^2}\left\{(2j-1)\mathrm{H}_{2j}(\sqrt{\omega}x)\mathrm{H}_{2j-2}(\sqrt{\omega}x) - 2j[\mathrm{H}_{2j-1}(\sqrt{\omega}x)]^2\right\}.$$

Plots of these potentials for $j = 1, 2, 3$ can be seen in Fig. 2.

6 PIV Transcendents

Let us remind that for $k = 1$ the natural ladder operators $\mathcal{L}^{\pm}$ are of order 3, thus the first-order SUSY partners of the complex oscillator are directly linked to the PIV equation. On the other hand, for $k > 1$ this order is necessarily greater than 3, but it can be reduced precisely to 3 by connecting all the seed solutions in the way

$u_j(x) = (a_\omega^-)^{j-1}u_1(x)$, $\varepsilon_j = \varepsilon_1 - (j-1)\omega$, $j = 1, \cdots, k$. As a consequence, $\mathcal{L}^+ = \mathcal{P}_{k-1}(\widetilde{H})\, l^+$, where $l^\pm$ are third-order differential ladder operators fulfilling

$$[\widetilde{H}, l^\pm] = \pm\omega\, l^\pm, \qquad l^+l^- = \left(\widetilde{H} - \frac{\omega}{2}\right)\left(\widetilde{H} - \varepsilon_1 - \omega\right)\left(\widetilde{H} - \varepsilon_k\right).$$

The roots of the polynomial l^+l^- suggest the following 3 extremal states

$$\begin{aligned}
&\psi_{\mathcal{E}_1} \propto B_k^+ \exp\left(-\frac{\omega x^2}{2}\right), && \mathcal{E}_1 = \frac{\omega}{2},\\
&\psi_{\mathcal{E}_2} \propto B_k^+ a_\omega^+ u_1, && \mathcal{E}_2 = \varepsilon_1 + \omega,\\
&\psi_{\mathcal{E}_3} \propto \frac{W(u_1, \ldots, u_{k-1})}{W(u_1, \ldots, u_k)}, && \mathcal{E}_3 = \varepsilon_k = \varepsilon_1 - (k-1)\omega.
\end{aligned}$$

In order to generate the PIV transcendents, we have to scale the Hamiltonian $\widetilde{H}$, as well as the involved factorization energies, and introduce the variable $z = \sqrt{\omega}x$, as follows:

$$\begin{aligned}
\mathsf{H} = \frac{\widetilde{H}}{\omega} &= -\tfrac{1}{2\omega}\tfrac{\mathrm{d}^2}{\mathrm{d}x^2} + \tfrac{1}{2}\omega x^2 - \tfrac{1}{\omega}\tfrac{\mathrm{d}^2}{\mathrm{d}x^2}\left[\ln W\left(u_1, \cdots, u_k\right)\right]\\
&= -\tfrac{1}{2}\tfrac{\mathrm{d}^2}{\mathrm{d}z^2} + \tfrac{z^2}{2} - \tfrac{\mathrm{d}^2}{\mathrm{d}z^2}\left[\ln W\left(u_1, \cdots, u_k\right)\right].
\end{aligned}$$

The corresponding PIV transcendents are simply calculated through

$$g(z) = -z - \tfrac{\mathrm{d}}{\mathrm{d}z}\ln\left[\psi_{\mathcal{E}_3}\left(\tfrac{z}{\sqrt{\omega}}\right)\right].$$

Some results are shown in Tables 1 and 2 for $k = 1$ and $k = 2$, respectively.

Table 1 PIV transcendents for $k = 1$, $\varepsilon_1 = \frac{5\omega}{2}$, $u_1(x) = \phi_2(x)$

$\psi_{\mathcal{E}_3}$	$\frac{1}{u_1}$	$B^+\phi_0$	$B^+a_\omega^+u_1$
$\frac{\mathcal{E}_3}{\omega}$	$\frac{5}{2}$	$\frac{1}{2}$	$\frac{7}{2}$
$g(z)$	$\frac{6z-4z^3}{2z^2-1}$	$-\frac{2z^2+1}{z-2z^3}$	$\frac{4z\left(-4z^4+4z^2+3\right)}{8z^6-4z^4+6z^2-3}$
a	-2	4	-5
b	-18	-2	-8

Table 2 PIV transcendents for $k = 2$, $\epsilon_1 = \frac{5\omega}{2}$, $u_1(x) = \phi_2(x)$, $u_2 = a_\omega^- u_1$

$\psi_{\mathcal{E}_3}$	$\frac{u_1}{W(u_1,u_2)}$	$B^+\phi_0$	$B^+a_\omega^+u_1$
$\frac{\mathcal{E}_3}{\omega}$	$\frac{3}{2}$	$\frac{1}{2}$	$\frac{7}{2}$
$g(z)$	$\frac{8z^5+6z}{1-4z^4}$	$\frac{4z}{2z^2+1}$	$-\frac{4z^4+3}{4z^5+8z^3+3z}$
a	0	3	-6
b	-18	-8	-2

7 Conclusions

In this article we have analyzed the link that exists between systems ruled by second-degree PHA (which have third-order ladder operators) and the PIV equation. Starting from the SUSY partners of the complex oscillator, an algorithm for generating PIV transcendents has been implemented (see also [17]). The non-Hermitian Hamiltonians giving place to these transcendents have two infinite energy ladders: an infinite one starting from $\frac{\omega}{2}$ and going outside the origin along θ-direction, plus a finite ladder beginning from an arbitrary complex number ε_k and pointing in the same direction. A deeper study about the classification of the PIV transcendents generated from the SUSY partners of the complex oscillator seems to be required.

References

1. D.J. Fernández, in *Integrability, Supersymmetry and Coherent States*, ed. by S. Kuru, et al. CRM Series in Mathematical Physics, (Springer, Cham, 2019), pp. 37–68
2. V.B. Matveev, M.A. Salle, *Darboux Transformations and Solitons* (Springer, Berlin, 1991)
3. B.K. Bagchi, *Supersymmetry in Quantum and Classical Mechanics* (Chapman & Hall, Boca Raton, 2001)
4. C.V. Sukumar, in *AIP Conf. Proc.*, vol. 744, 166 (2004)
5. A.P. Veselov, A.B. Shabat, Funct. Anal. Appl. **27**, 81 (1993)
6. S.Y. Dubov, V.M. Eleonsky, N.E. Kulagin, Chaos **4**, 47 (1994)
7. V.E. Adler, Physica D **73**, 335 (1994)
8. A. Andrianov, F. Cannata, M. Ioffe, D. Nishnianidze, Phys. Lett. A **266**, 341 (2000)
9. J.M. Carballo, D.J. Fernández, J. Negro, L.M. Nieto, J. Phys. A **37**, 10349 (2004)
10. J. Mateo, J. Negro, J. Phys. A Math. Theor. **41**, 045204 (2008)
11. D. Bermudez, D.J. Fernández, SIGMA **7**, 025 (2011)
12. C. Rogers, A.P. Bassom, P.A. Clarkson, J. Math. Ann. App. **462**, 1225 (2018)
13. D. Levi, P. Winternitz, *Painlevé Transcendents, NATO ASI B278* (Plenum Press, New York, 1992)
14. A.P. Bassom, P.A. Clarkson, A.C. Hicks, Stud. Appl. Math. **95**, 1 (1995)
15. D.J. Fernández, V. Hussin, J. Phys. A **32**, 3603 (1999)
16. I. Marquette, J. Math. Phys. **50**, 095202 (2009)
17. D.J. Fernández, J.C. González, Ann. Phys. **359**, 213 (2015)

The Veronese Sequence of Analytic Solutions of the $\mathbb{C}P^{2s}$ Sigma Model Equations Described via Krawtchouk Polynomials

Nicolas Crampé and Alfred Michel Grundland

In honour of Decio Levi (University of Roma Tre)

Abstract The objective of this paper is to establish a new relationship between the Veronese sequence of analytic solutions of the Euclidean $\mathbb{C}P^{2s}$ sigma model in two dimensions and the orthogonal Krawtchouk polynomials. We show that such solutions of the $\mathbb{C}P^{2s}$ model, defined on the Riemann sphere and having a finite action, can be explicitly parametrized in terms of these polynomials. We apply the obtained results to the analysis of surfaces associated with $\mathbb{C}P^{2s}$ sigma models, defined using the generalized Weierstrass formula for immersion. We show that these surfaces are spheres immersed in the $\mathfrak{su}(2s+1)$ Lie algebra, and express several other geometrical characteristics in terms of the Krawtchouk polynomials. Finally, a new connection between the $\mathfrak{su}(2)$ spin-s representation and the $\mathbb{C}P^{2s}$ model is explored in detail. It is shown that for any given holomorphic vector function in $\mathbb{C}^{2s+1}$ written as a Veronese sequence, it is possible to derive a sequence of analytic solutions of the $\mathbb{C}P^{2s}$ model through algebraic recurrence relations which turn out to be simpler than the analytic relations known in the literature.

Keywords Sigma models · Projector formalism · Integrable systems · Soliton surfaces · Weierstrass formula for immersion · Spin matrices

Mathematical Subject Classification 81T45, 53C43, 35Q51

N. Crampé
Institut Denis Poisson, Université de Tours – Université d'Orléans, Orléans, France

A. M. Grundland (✉)
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC H3C 3J7, Canada
Université du Québec à Trois-Rivières, Trois-Rivières, QC G9A 5H7, Canada
e-mail: grundlan@crm.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_5

1 The $\mathbb{C}P^{2s}$ Sigma Model

The dynamical fields in the $\mathbb{C}P^{2s}$ sigma models are maps from the Riemann sphere $\mathbb{S}^2$ to the complex projective space $\mathbb{C}P^{2s} \simeq \mathbb{S}^{4s(s+1)}/U(1)$

$$\mathbb{S}^2 \ni \xi_\pm = \xi^1 \pm i\xi^2 \mapsto z = (z_0, z_1, \ldots, z_{2s}) \in \mathbb{C}^{2s+1} \setminus \{\emptyset\},$$

(where the value of the index s is either an integer or half-integer) which are stationary points of the action functional [1]

$$\mathcal{A} = \frac{1}{4} \iint_{\mathbb{S}^2} (D_\mu z)^\dagger \cdot (D_\mu z) d\xi_+ d\xi_-, \tag{1}$$

and hence are solutions of the Euler–Lagrange (EL) equations

$$D_\mu D_\mu z + (D_\mu z)^\dagger \cdot (D_\mu z) z = 0, \tag{2}$$

subjected to $z^\dagger z = 1$, where D_μ are the covariant derivatives defined by

$$D_\mu z = \partial_\mu z - (z^\dagger \partial_\mu z) z, \qquad \partial_\mu = \frac{\partial}{\partial \xi^\mu}, \qquad \mu = 1, 2.$$

We require that the action (1) over the whole Riemann sphere $\mathbb{S}^2$ be finite.

2 Projective Formalism

Equivalently, representing the z's by their homogeneous representatives, i.e. maps into $\mathbb{C}^{2s+1} \setminus \{\emptyset\}$

$$z = \frac{f}{(f^\dagger \cdot f)^{1/2}},$$

we may use (fields of) rank-1 Hermitian projectors

$$P = \frac{f \otimes f^\dagger}{f^\dagger \cdot f}, \qquad P^2 = P, \qquad P^\dagger = P. \tag{3}$$

This places the EL equations in the form of the conservation law (CL)

$$\partial[\bar{\partial} P, P] + \bar{\partial}[\partial P, P] = 0, \tag{4}$$

where the symbols ∂ and $\bar{\partial}$ stand for the complex derivatives with respect to ξ_+ and ξ_- given by

$$\partial = \frac{1}{2}\left(\frac{\partial}{\partial\xi^1} - i\frac{\partial}{\partial\xi^2}\right), \qquad \bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial\xi^1} + i\frac{\partial}{\partial\xi^2}\right).$$

Under the above assumptions every solution can be obtained from a holomorphic (respectively, antiholomorphic) solution $f : \mathbb{S}^2 \to \mathbb{C}^{2s+1} \setminus \{\emptyset\}$, $\bar{\partial} f = 0$, by successive applications of the raising or lowering operator [1],

$$\begin{gathered} f_{k+1} = P_+(f_k) := (\mathbb{I}_{2s+1} - P_k)\partial f_k, \qquad f_{k-1} = P_-(f_k) := (\mathbb{I}_{2s+1} - P_k)\bar{\partial} f_k, \\ P_\pm^0 = \mathbb{I}_{2s+1}, \qquad P_\pm^{2s+1} f_k = 0, \qquad k = 0, 1, \ldots, 2s, \end{gathered} \tag{5}$$

where $P_+(f_k)$ is a creation operator and $P_-(f_k)$ is an annihilation operator. Thus the sequence of solutions in the $\mathbb{C}P^{2s}$ model consists of $2s+1$ vectors f_k or $2s+1$ rank-1 Hermitian projectors P_k. The action integral (1) in terms of the projectors P_k has a more compact form

$$\mathcal{A}(P_k) = \iint_{\mathbb{S}^2} tr\left(\partial P_k \cdot \bar{\partial} P_k\right) d\xi_+ d\xi_-. \tag{6}$$

In terms of the nonconstant projectors P_k, the recurrence relations (5) become [2–4]

$$P_{k\pm1} = \Pi_\pm(P_k) := \frac{(\partial_\pm P_k)P_k(\partial_\mp P_k)}{tr[(\partial_\pm P_k)P_k(\partial_\mp P_k)]}, \tag{7}$$

for $tr[(\partial_\pm P_k)P_k(\partial_\mp P_k)] \neq 0$ and are equal to zero when $tr[(\partial_\pm P_k)P_k(\partial_\mp P_k)] = 0$, where ∂_+ and ∂_- stand for ∂ and $\bar{\partial}$, respectively. Here P_k stands for one of the projectors $\{P_0, P_1, \ldots, P_{2s}\}$. This set satisfies the orthogonality and completeness relations

$$P_j P_k = \delta_{jk} P_j, \quad 0 \leq k, j \leq 2s, \qquad \sum_{j=0}^{2s} P_j = \mathbb{I}_{2s+1}. \tag{8}$$

3 Solutions of the $\mathbb{C}P^{2s}$ Sigma Model

A particular holomorphic solution of the $\mathbb{C}P^{2s}$ model equations (4) expressed in terms of the f's

$$\left(\mathbb{I}_{2s+1} - \frac{f_k \otimes f_k^\dagger}{f_k^\dagger \cdot f_k}\right)\left[\partial\bar{\partial} f_k - \frac{1}{f_k^\dagger \cdot f_k}\left(\left(f_k^\dagger \cdot \bar{\partial} f_k\right)\partial f_k + \left(f_k^\dagger \cdot \partial f_k\right)\bar{\partial} f_k\right)\right] = 0, \tag{9}$$

for $0 \leq k \leq 2s$, can be written as the Veronese sequence [5]

$$f_0 = \left(1, \binom{2s}{1}^{1/2} \xi_+, \ldots, \binom{2s}{r}^{1/2} \xi_+^r, \ldots, \xi_+^{2s}\right) \in \mathbb{C}^{2s+1} \backslash \{\emptyset\}, \qquad \text{for } k = 0. \tag{10}$$

The Veronese sequence of analytic solutions of (9) can be obtained by acting with the creation operators (5). Thus for $k > 2$ this procedure allows us to construct three classes of solutions: holomorphic f_0, antiholomorphic f_{2s} and mixed solutions f_k, $1 \leq k \leq 2s - 1$.

Under the above assumptions we show that any Veronese sequence of solutions P_k of the EL equations (9) can be expressed explicitly in terms of the Krawtchouk orthogonal polynomials.

Theorem 1 (The Main Result [6]) *Let the $\mathbb{C}P^{2s}$ model be defined on the Riemann sphere $\mathbb{S}^2$ and have a finite action functional. Then the Veronese sequence of analytic solutions f_k of the $\mathbb{C}P^{2s}$ model (9) takes the form*

$$(f_k)_j = \frac{(2s)!}{(2s-k)!} \left(\frac{-\xi_-}{1+\xi_+\xi_-}\right)^k \sqrt{\binom{2s}{j}} \xi_+^j K_j(k; p, 2s), \qquad 0 \leq k, j \leq 2s \tag{11}$$

$$0 < p = \frac{\xi_+\xi_-}{1+\xi_+\xi_-} < 1,$$

where $(f_k)_j$ is the jth component of the vector $f_k \in \mathbb{C}^{2s+1} \backslash \{\emptyset\}$ and $K_j(k; p, 2s)$ are Krawtchouk orthogonal polynomials defined in terms of the hypergeometric function

$$K_j(k) = K_j(k; p, 2s) = {}_2F_1(-j, -k; -2s; 1/p), \quad 0 \leq k \leq 2s. \tag{12}$$

Here j,k and $2s$ are parameters, while p is an argument in (12). We use the convention

$$K_j(0; p, 2s) = 1, \quad \text{for } k = 0. \tag{13}$$

The vectors f_k can be used to construct the rank-1 Hermitian matrix projector P_k with an entry in the ith row and jth column given by

$$(P_k)_{ij} = \binom{2s}{k} \frac{(\xi_+\xi_-)^k}{(1+\xi_+\xi_-)^{2s}} \xi_+^i \xi_-^j \sqrt{\binom{2s}{i}\binom{2s}{j}} K_i(k) K_j(k), \tag{14}$$

where, in what follows, we use the following abbreviated notation

$$K_j(k) := K_j(k; p, 2s), \qquad K_j(k \pm 1) := K_j(k \pm 1; p, 2s). \tag{15}$$

The EL equations (9) with the idempotency condition $P_k^2 = P_k$ admit a larger class of solutions [4] than the rank-1 Hermitian projector P_k.

Proposition 1 (Higher-Rank Projectors) *Let the linear combinations of the rank-1 Hermitian projectors P_l be*

$$P = \sum_{l=0}^{2s} \lambda_l P_l, \qquad \lambda_l = 0 \quad or \quad \lambda_l = 1 \quad for \quad all \quad l \in \{0, 1, \ldots, 2s\}, \tag{16}$$

for which P_l satisfy the EL equations (9). The higher-rank projector P can be expressed in terms of the Krawtchouk polynomials

$$(P)_{ij} = \sum_{l=0}^{2s} \lambda_l \binom{2s}{l} \frac{(\xi_+\xi_-)^l}{(1+\xi_+\xi_-)^{2s}} \xi_-^i \xi_+^j \sqrt{\binom{2s}{i}\binom{2s}{j}} K_i(l) K_j(l) \tag{17}$$

which satisfy both the EL equations (9) and the idempotency condition $P^2 = P$. In this case the projector P maps the $\mathbb{C}^{2s+1}$ space onto $\mathbb{C}^k$, where $k = \sum_{l=0}^{2s} \lambda_l$.

Proof The proof is straightforward if we use (16) and the rank-1 Hermitian projector P_k in terms of the Krawtchouk polynomials (14).

4 The $\mathfrak{su}(2)$ Spin-s Representation

A direct connection was established between the $\mathbb{C}P^{2s}$ model and the spin-s $\mathfrak{su}(2)$ representation [3, 7]. The spin matrix S^z is defined as a linear combination of the (2s+1) rank-1 Hermitian projectors P_k, i.e.

$$S^z(\xi_+, \xi_-) = \sum_{k=0}^{2s} (k-s) P_k, \qquad (S^z)^\dagger = S^z, \tag{18}$$

where the eigenvalues of the generator S^z are $\{-s, -s+1, \ldots, s-1, s\}$. They are either integer (for odd $2s+1$) or half-integer (for even $2s+1$) values. From Eq. (18) we obtain that the spin matrix S^z is given by the tridiagonal matrix with an entry in the ith row and jth column [6]

$$(S^z)_{ij} = \delta_{ij} \left(\frac{1-\xi_+\xi_-}{1+\xi_+\xi_-} \right) (i-s) - \delta_{i-1,j} \left(\frac{\xi_+}{1+\xi_+\xi_-} \right) \sqrt{i(2s+1-i)}$$

$$- \delta_{i,j-1} \left(\frac{\xi_-}{1+\xi_+\xi_-} \right) \sqrt{j(2s-j+1)}, \qquad 0 \le i, j \le 2s. \tag{19}$$

The generators S^z and $S^{\pm}$ of the $\mathfrak{su}(2)$ Lie algebra satisfy the commutation relations

$$[S^z, S^{\pm}] = \pm S^{\pm}, \qquad [S^+, S^-] = 2S^z, \tag{20}$$

and they are identified with the following $(2s+1) \times (2s+1)$ matrices [8]

$$(\sigma^z)_{ij} = (s-i)\delta_{ij}, \tag{21}$$

$$(\sigma^+)_{ij} = \sqrt{(2s-j+1)j}\,\delta_{i,j-1}, \qquad 0 \le i, j \le 2s \tag{22}$$

$$(\sigma^-)_{ij} = \sqrt{(2s-i+1)i}\,\delta_{i-1,j}. \tag{23}$$

Hence the matrices S^z and $S^{\pm}$ can be decomposed as a linear combination of the matrices σ^z and $\sigma^{\pm}$, namely

$$\begin{pmatrix} S^z \\ S^+ \\ S^- \end{pmatrix} = \frac{1}{1+\xi_+\xi_-} \begin{pmatrix} \xi_+\xi_- - 1 & -\xi_- & -\xi_+ \\ 2\xi_- & \xi_-^2 & -1 \\ 2\xi_+ & -1 & \xi_+^2 \end{pmatrix} \begin{pmatrix} \sigma^z \\ \sigma^+ \\ \sigma^- \end{pmatrix}, \tag{24}$$

where $(S^+)^\dagger = S^-$ and $(S^-)^\dagger = S^+$. The eigenvalue problem for the spin matrix S^z is given by

$$S^z f_k = (k-s) f_k, \qquad S^z(S^{\pm} f_k) = (k \pm 1 - s)(S^{\pm} f_k), \qquad \text{for} \quad 0 \le k \le 2s.$$

Under these circumstances the following holds

Proposition 2 (Recurrence Relations Associated with the $\mathbb{C}P^{2s}$ Models) *For the Veronese sequence of analytic solutions f_k of the $\mathbb{C}P^{2s}$ model (9), the algebraic recurrence relations for the vectors $S^z f_k$ and $S^{\pm} f_k$ are given by [6]*

$$S^+ f_k = \begin{cases} -(1+\xi_+\xi_-) f_{k+1} & \text{for} \quad 0 \le k \le 2s-1, \\ 0 \quad \text{for} \quad k = 2s, \end{cases} \tag{25}$$

$$S^- f_k = \frac{1}{1+\xi_+\xi_-} k(k-1-2s) f_{k-1} \quad \text{for} \quad 0 \le k \le 2s. \tag{26}$$

In terms of the projectors P_k, the recurrence relations (7) take the algebraic form

$$P_{k+1} = \Pi_+(P_k) := \frac{S^+ P_k S^-}{tr(S^+ P_k S^-)}, \qquad P_{k-1} = \Pi_-(P_k) := \frac{S^- P_k S^+}{tr(S^- P_k S^+)}, \tag{27}$$

where $tr(S^+ P_k S^-) \neq 0$.

Proof The proof of the formulae (27) follows immediately from (3) and the relations (25) and (26), i.e.

$$\frac{S^+P_kS^-}{tr(S^+P_kS^-)} = \frac{S^+f_k \otimes f_k^\dagger S^-}{tr(S^+f_k \otimes f_k^\dagger S^-)} \sim \frac{f_{k+1} \otimes f_{k+1}^\dagger}{f_{k+1}^\dagger \cdot f_{k+1}} = P_{k+1}, \tag{28}$$

since $(S^+f_k)^\dagger = f_k^\dagger S^-$. Similarly, it is easy to show that the following relation holds

$$\frac{S^-P_kS^+}{tr(S^-P_kS^+)} = P_{k-1}. \tag{29}$$

Note that the relations (25) and (26) allow us to recursively construct the Veronese sequence of analytic solutions f_k from the holomorphic solution f_0 in a simpler way than the ones obtained from the analytic recurrence relation (5). Therefore, the matrices $S^\pm$ are the creation and annihilation operators for the vectors f_k and the projectors P_k. The result given in the above proposition can be interpreted as the matrix elements of the $SU(2)$ irreducible representations, known as the Wigner D function. It is known [9, 10] that these matrix elements can be expressed in terms of the Krawtchouk polynomials.

5 Geometrical Aspects of Surfaces

The generalized Weierstrass formula for the immersion of 2D-surfaces associated with the $\mathbb{C}P^{2s}$ model (9) is given by [11]

$$X_k(\xi_+, \xi_-) = -i\left(P_k + 2\sum_{j=0}^{k-1} P_j\right) + i\left(\frac{1+2k}{1+2s}\right)\mathbb{I}_{2s+1} \in \mathfrak{su}(2s+1) \tag{30}$$

and the raising and lowering operators for X_k are [2]

$$X_{k\pm1} = \Pi_\pm(X_k) = \frac{(\partial_\pm X_k)X_k(\partial_\mp X_k)}{tr((\partial_\pm X_k)X_k(\partial_\mp X_k))},$$

where ∂_+ and ∂_- stand for ∂ and $\bar{\partial}$, respectively. It follows from (27) that the creation or annihilation operator for the immersion functions X_k can be defined algebraically by

$$X_{k+1} = X_k - i\left(\frac{S^+ P_k S^-}{tr(S^+ P_k S^-)} + P_k - \frac{2}{2s+1}\mathbb{I}\right),$$
$$X_{k-1} = X_k + i\left(\frac{S^- P_k S^+}{tr(S^- P_k S^+)} + P_k - \frac{2}{2s+1}\mathbb{I}\right).$$

For the sake of uniformity, the inner product is defined by

$$(A, B) = -\frac{1}{2} tr(A \cdot B), \qquad A, B \in \mathfrak{su}(2s+1).$$

The first and second fundamental forms are

$$I_k = tr(\partial P_k \cdot \bar{\partial} P_k) d\xi_+ d\xi_- = \frac{2(2sk + s - k^2)}{(1+\xi_+\xi_-)^2} d\xi_+ d\xi_-,$$
$$II_k = -tr(\partial P_k \cdot \bar{\partial} P_k)\partial\left(\frac{[\partial P_k, P_k]}{tr(\partial P_k \cdot \bar{\partial} P_k)}\right) d\xi_+^2 + 2i[\bar{\partial} P_k, \partial P_k] d\xi_+ d\xi_- \qquad (31)$$
$$- tr(\partial P_k \cdot \bar{\partial} P_k)\bar{\partial}\left(\frac{[\bar{\partial} P_k, P_k]}{tr(\partial P_k \cdot \bar{\partial} P_k)}\right) d\xi_-^2.$$

Proposition 3 (Non-intersecting Spheres) *For any value of the Veronese sequence of analytic solutions f_k of the $\mathbb{C}P^{2s}$ model (9), all the 2D-surfaces X_k are non-intersecting spheres with the radius*

$$R_k = (X_k, X_k)^{1/2} = \left(-\frac{1}{2} tr(X_k)^2\right)^{1/2} = \left|\frac{-2k^2 + 2k(2s-1) + s - 1}{1+2s}\right|^{1/2}, \qquad (32)$$

immersed in the Lie algebra $\mathfrak{su}(2s+1) \simeq \mathbb{R}^{4s(s+1)}$.

Proof Let us assume that $l > k$ are two different indices of the induced surfaces. Subtracting (30) from the analogous expression for X_l, we get

$$P_l - P_k + 2\sum_{j=k}^{l-1} P_j - \frac{2(l-k)}{2s+1}\mathbb{I}_{2s+1} = 0. \qquad (33)$$

Multiplying Eq. (33) by P_k, P_l or P_{l-1} and solving the obtained system of equations, we obtain that the 2D-surfaces X_k and X_l do not intersect if $k \neq l$ with the exceptions of X_0 and X_1 in the $\mathbb{C}P^1$ model since X_0 and X_1 coincide [7]. The fundamental forms (31) imply that the Gaussian curvatures of the 2D-surfaces have constant positive values

$$K_k = \frac{2}{2sk + s - k^2}. \qquad (34)$$

The Kähler angles are given by

$$\tan\left(\frac{1}{2}\theta_k(m)\right) = \left|\frac{df_k(m)(\partial/\partial\xi_-)}{df_k(m)(\partial/\partial\xi_+)}\right|, \qquad m \in \mathbb{S}^2$$

and have constant positive values

$$\cos\theta_k = \frac{s-k}{2sk+s-k^2}.$$

The Euler–Poincaré characters of the 2D-surfaces X_k are the integer $\Delta_k = 2$ for all k such that $0 \leq k \leq 2s$. This means that all 2D-surfaces associated with the $\mathbb{C}P^{2s}$ model are non-intersecting spheres with radius R_k given by (32).

The technique for obtaining surfaces via projective structures and their links with orthogonal polynomials, elaborated from the $\mathbb{C}P^{2s}$ models, can be extended to different types of Grassmannian manifolds. An analysis of these manifolds can provide us with much more diverse types of surfaces.

Acknowledgments This research was supported by the NSERC operating grant of one of the authors (A.M.G.). N.C. is indebted to the Centre de Recherches Mathématiques (CRM), Université de Montréal for the opportunity to hold a CRM-Simons professorship.

References

1. W.J. Zakrzewski, *Low Dimensional Sigma Models* (Hilger, Bristol, 1989)
2. P.P. Goldstein, A.M. Grundland, Invariant recurrence relations for $\mathbb{C}P^{N-1}$ models. J. Phys. A Math. Theor. **43**(265206), 1–18 (2010)
3. P.P. Goldstein, A.M. Grundland, S. Post, Soliton surfaces associated with sigma models: differential and algebraic aspects. J. Phys. A Math. Theor. **45**(395208), 1–19 (2012)
4. P.P. Goldstein, A.M. Grundland, On a stack of surfaces obtained from the $\mathbb{C}P^{N-1}$ sigma model. J. Phys. A Math. Theor. **51**(095201), 1–13 (2018)
5. J. Bolton, G.R. Jensen, M. Rigoli, L.M. Woodward, On conformal minimal immersion of S^2 into $\mathbb{C}P^N$. Math. Ann. **279**, 599–620 (1988)
6. N. Crampé, A.M. Grundland, $\mathbb{C}P^{2S}$ sigma models described through hypergeometric orthogonal polynomials. Ann. Henri Poincaré **20**(10), 1–23 (2019)
7. P.P. Goldstein, A.M. Grundland, Analysis of $\mathbb{C}P^{N-1}$ sigma models via soliton surfaces, in *A Volume in Honour of V. Hussin*, ed. by S. Kuru, et al. CRM series in Math. Phys. (Springer, New York, 2019), pp. 341–358
8. E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1998)
9. T. Koornwinder, Krawtchouk polynomials, a unification of two different group theoretic interpretation. SIAM J. Math. Anal. **13**, 10111023 (1982)
10. Y.I. Granovski, A. Zhedanov, Orthogonal polynomials in the Lie algebras. Sov. Phys. J. **29**, 387393 (1986)
11. A.M. Grundland, I. Yurdusen, On analytic descriptions of two-dimensional surfaces associated with the $\mathbb{C}P^{N-1}$ model. J. Phys. A Math. Theor. **42**(172001), 1–5 (2009)

A Novel Integrable Fourth-Order Difference Equation Admitting Three Invariants

Giorgio Gubbiotti

Abstract In this short note we present a novel integrable fourth-order difference equation. This equation is obtained as a stationary reduction from a known integrable differential-difference equation. The novelty of the equation is inferred from the number and shape of its invariants.

Keywords Difference equations · Integrability · Algebraic entropy

1 Introduction

The interest in discrete systems, that is, of systems whose independent degrees of freedom take values in a discrete set, grew enormously during the past decades, for reasons which span from very philosophical [23] to purely practical ones [29]. Nowadays, discrete systems are studied from different points of view and perspective, see [12, 24].

In this short note we will present a new *integrable fourth-order difference equation*. An Nth-order difference equation is a functional equation for an unknown sequence $\{x_n\}_{n\in\mathbb{Z}}$ where the x_{n+N} element is expressible in terms of the previous x_{n+i}, $i = 0, 1, \ldots, N-1$. Such kind of functional equations are also called *recurrence relations*. Without entering in the details, which will be given in Sect. 2, we say that an Nth-order difference equation is integrable when its dynamics is *sufficiently regular* and *predictable*.

The integrability of second-order difference equations is a well understood topic, as it is known that most of the integrable second-order difference equations belong to the QRT class [30, 31], even though there are some notable exceptions [11, 33, 44].

In higher dimension an analogous general framework does not exist, whereas some approaches similar to the one of QRT [30, 31] have been pursued in the

G. Gubbiotti (✉)
School of Mathematics and Statistics, The University of Sydney, Sydney, NSW, Australia
e-mail: giorgio.gubbiotti@sydney.edu.au

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_6

literature [7, 21, 22]. These approaches relied on searching for difference equations admitting some extra structures allowing to claim integrability, i.e. invariants of a fixed form and/or symplectic structures. Higher-order difference equations have been produced in the literature with different methods, like periodic reductions of partial difference equations [9, 35–38, 40].

In an upcoming paper [20] we propose a new approach to generate integrable higher-order difference equations through stationary solutions of integrable differential-difference equations. A differential-difference equation is a function equation for an unknown sequence $\{x_n(t)\}_{n\in\mathbb{Z}}$ of functions a continuous variable, depending on both shifts and derivatives. A well-known class of differential-difference equations are the *Volterra-like equations*:

$$\frac{\partial x_n}{\partial t} = f\left(x_{n+k}, \ldots, x_n, \ldots, x_{n-k'}\right), \quad k > k'. \tag{1}$$

The stationary reduction of a differential-difference equation is obtained by letting $\partial x_n/\partial t \equiv 0$ and it is clearly a difference equation, since we suppressed the dependence on the continuous variable. To be more precise, in [20] we will present the integrability properties of the stationary solutions for two classes fourth-order Volterra-like equations, recently classified in [15, 16]. In this short note we present an interesting example out of this general picture.

The plan of the paper is following: in Sect. 2 we introduce the formal definitions of integrability we will be using throughout this note. In Sect. 3 we will present our new example, and show its integrability in the sense of Sect. 2. Finally, in Sect. 4 we give some final comments and an outlook towards the general results.

2 Integrability of Difference Equations

Integrability both for continuous and discrete systems can be defined in several different ways. In this note we will limit ourselves to two alternative definition, out of all the possible ones.

Consider an autonomous Nth-order difference equation:

$$x_{n+N} = Q\left(x_{n+N-1}, x_{n+N-2}, \ldots, x_n\right). \tag{2}$$

A function

$$I = I\left(x_{n+N-1}, x_{n+N-2}, \ldots, x_n\right) \tag{3}$$

is called an *invariant* if:

$$I\left(x_{n+N}, x_{n+N-1}, \ldots, x_{n+1}\right) = I\left(x_{n+N-1}, x_{n+N-2}, \ldots, x_n\right) \tag{4}$$

on the solutions of equation (2).

If there exist $N-1$ functionally independent invariants I_l, $l=1,\ldots,N-1$, then it is possible to reduce the difference equation (2) to first-order one by solving the relations:

$$I_j = \kappa_j, \tag{5}$$

where κ_j are the value of the invariants on a set of initial data. In such case we say that the difference equation (2) is *integrable*.

This definition of integrability is very general. If some additional structures, like Poisson or symplectic structures, are present, then the number of invariants needed for integrability can be significantly lowered: this is the content of the discrete Liouville–Poisson theorem [6, 28, 41].

In general to search for invariants is difficult procedure, see remark 1. For this reason, several *integrability indicators*, that is, *necessary conditions* for integrability, have been introduced. A well-known integrability indicator, which is also an equivalent definition of integrability, is the *algebraic entropy* [5, 13, 42]. Algebraic entropy is defined for bi-rational maps of the complex projective space $\mathbb{CP}^N$ to itself. *Rational* difference equations (2) which can uniquely solve with respect to x_n, that is, which are *fractionally linear in* x_n, are equivalent to such maps. To see this first notice that computing the orbit of such an equation is equivalent to iterate the complex map $\Phi : \mathbb{C}^n \to \mathbb{C}^n$ defined as follows:

$$\Phi\left(x_{N-1},\ldots,x_0\right) = \left(Q\left(x_{n+N-1},\ldots,x_n\right), x_{N-1},\ldots,x_1\right). \tag{6}$$

The condition of unique solvability with respect to x_0 of (2) ensures us that the map Φ has a rational inverse Ψ. Then, introducing the homogeneous coordinates $[X_{N-1} : \ldots, X_0 : T]$ by

$$\left(x_{N-1},\ldots,x_0\right) = \left(\frac{X_{N-1}}{T},\ldots,\frac{X_0}{T}\right), \tag{7}$$

we have that the map (6) can be lifted to a rational map $\varphi : \mathbb{CP}^N \to \mathbb{CP}^N$. Lifting the inverse map Ψ to $\psi : \mathbb{CP}^N \to \mathbb{CP}^N$ we conclude that the map we obtain is actually a bi-rational map of the complex projective space $\mathbb{CP}^N$ to itself.

Given a bi-rational map, we can take as measure of its *complexity*, in the sense of Arnol'd [4], the growth of the number of intersections of the successive images of a straight line with a generic hyperplane in complex projective space [42]. This actually corresponds to compute the degrees of its iterates with respect to a generic initial condition:

$$d_k = \deg \varphi^k, \quad k \in \mathbb{N}. \tag{8}$$

Then we consider the following limit:

$$\varepsilon = \lim_{k\to\infty} \frac{1}{k} \log d_k, \tag{9}$$

called the *algebraic entropy*. If the growth of the map φ is sub-exponential, then the algebraic entropy (9) vanishes and we say the map φ is *integrable in the sense of the algebraic entropy* [5, 13, 42]. As a particular case, when the growth of a map is linear the map is believed to be *linearizable* [25].

Algebraic entropy is an *invariant* of bi-rational maps, meaning that its value is unchanged up to bi-rational equivalence. Moreover, its value is determined by the *singularity structure* of a map [5, 34, 43].

To compute the algebraic entropy from (9) we need to know the *asymptotic behaviour* of the sequence d_n. For the majority of applications such behaviour can be inferred by using a generating function [27], that is, a function $g = g(z)$ such that:

$$g(z) = \sum_{n=0}^{\infty} d_k z^k. \tag{10}$$

A generating function is a predictive tool which can be used to test the successive members of a finite sequence. It follows that the algebraic entropy is given by the logarithm of the smallest pole of the generating function, see [17, 18].

Remark 1 Finding invariants is a hard task. Here we recall briefly a method for finding invariants of bi-rational maps presented first in [13] and recently reprised in [8]. If the ratio P/Q is an invariant of a map φ, then the pullback of φ on P/Q is invariant: $\varphi^*(P/Q) = P/Q$. This implies

$$\varphi^*(P) = aP \quad \text{and} \quad \varphi^*(Q) = aQ \tag{11}$$

for some polynomial factor a. Using the fact that $\psi \circ \varphi = \kappa\,\mathrm{Id}$ where κ is a polynomial one gets that a must contain some of the factors dividing κ. Hence one can search for invariants imposing the form of P, then searching for the appropriate factors. We get an invariant when we obtain more than one solution for the same a. By taking ratios of the solutions we obtain the invariants.

The problem with this method is that it is not bounded as we do not know a priori the degree of P. However, in practice this method is quite useful for the explicit computation of the invariants, since the conditions in (11) are *linear*, even though their number can become huge as $\deg(P)$ grows.

3 A Novel Example

Consider the following differential-difference equation:

$$\frac{\partial x_n}{\partial t} = (x_n + 1)\left[\frac{x_{n+2}x_n(x_{n+1}+1)^2}{x_{n+1}} - \frac{x_{n-2}x_n(x_{n-1}+1)^2}{x_{n-1}} + (1+2x_n)(x_{n+1}-x_{n-1})\right]. \tag{12}$$

Equation (12) has been found in [2] and it is related to the discrete Sawada–Kotera equation found in [1, 39]. Equation (12) emerged again in [15] where the authors classified the fourth-order Volterra-like equations (1) linear in $x_{n\pm2}$. Imposing $\partial x_n/\partial t \equiv 0$ in (12) we obtain its stationary reduction:

$$\frac{x_{n+2}x_n(x_{n+1}+1)^2}{x_{n+1}} - \frac{x_{n-2}x_n(x_{n-1}+1)^2}{x_{n-1}} = -(1+2x_n)(x_{n+1}-x_{n-1}). \tag{13}$$

Equation (13) is not *resummable*, in the sense of [3], to a second-order difference equation, nor it is *deflatable*, in the sense of [26], to a third-order difference equation.

In [19] it was heuristically shown that Eq. (12) has quadratic growth [10]. Since (13) is a reduction of (12) we have that it can have *at most* quadratic growth. Computing the growth of degrees of equation (13) we obtain

$$\begin{aligned} &1, 5, 15, 35, 67, 113, 167, 229, 297, 375, 463, 561, 667, \\ &785, 911, 1047, 1193, 1349, 1511, 1681, 1859, 2051, \\ &2255, 2469, 2689, 2917, 3151, 3395, 3651, 3921, 4199\ldots \end{aligned} \tag{14}$$

whose generating function is

$$g(z) = \frac{\left[\begin{array}{r} 5z^{13} + 5z^{12} + 15z^{11} + 17z^{10} + 29z^9 + 25z^8 \\ +32z^7 + 26z^6 + 27z^5 + 19z^4 + 13z^3 + 7z^2 + 3z + 1 \end{array}\right]}{(1-z)^3(z+1)(z^2+1)(z^4+1)(z^2-z+1)(z^2+z+1)}. \tag{15}$$

Since all the roots of the denominator of (15) lie on the unit circle we have, as expected, that the algebraic entropy of equation (13) vanishes. Moreover, due to the presence of the factor $(1-z)^3$ we obtain that its growth is asymptotically quadratic [14]. That is, Eq. (13) is integrable in the sense of algebraic entropy and it is not expected to be linearizable, since its growth is quadratic.

Applying the method presented in Remark 1 we find that the map (13) admits the following functionally independent invariants:

$$
\begin{aligned}
I_1 &= \frac{x_n x_{n-2}}{x_{n-1}} + \frac{x_{n-1}x_{n+1}}{x_n} + x_n + x_{n-1} + 2x_n x_{n-2} \\
&+ 2x_{n-1}x_{n+1} + 2x_n x_{n-1} + x_n x_{n-1} x_{n-2} + x_n x_{n-1} x_{n+1}
\end{aligned} \tag{16a}
$$

$$
\begin{aligned}
I_2 &= \left(\frac{x_{n-2}}{x_{n-1}}\right)^2 + \left(\frac{x_{n+1}}{x_n}\right)^2 + x_{n-1}^2 x_{n-2}^2 + x_n^2 x_{n+1}^2 + x_n^2 x_{n-1}^2 \\
&- 2\Bigg[x_{n-2}x_{n+1}\left(\frac{x_n}{x_{n-1}} + \frac{1}{x_n x_{n-1}} + \frac{x_{n-1}}{x_n} + x_n x_{n-1}\right) \\
&\qquad + x_{n-1}x_n\left(x_{n-1}x_{n-2} + x_n x_{n+1}\right)\Bigg] \\
&+ 4\left[x_{n-1}x_{n-2}^2 + x_n x_{n+1}^2 + (x_n + x_{n-1})(1 - x_{n-2}x_{n+1})\right] \\
&+ 4\left[\frac{x_{n+1}(x_{n-1} + x_{n+1} - x_{n-2})}{x_n} + \frac{x_{n-2}(x_n + x_{n-2} - x_{n+1})}{x_{n-1}}\right] \\
&+ 6\left(x_{n+1}^2 + x_{n-1}x_{n+1} + x_n x_{n-2} + x_{n-2}^2\right) + 8\left(x_n x_{n-1} - x_{n-2}x_{n+1}\right)
\end{aligned} \tag{16b}
$$

$$
I_3 = \frac{N_3}{D_{3,1} D_{3,2}}, \tag{16c}
$$

where

$$
\begin{aligned}
N_3 &= x_n^3 x_{n-2}^2 x_{n-1}^3 + x_n^3 x_{n-2} x_{n-1}^3 x_{n+1} + x_n^3 x_{n-1}^3 x_{n+1}^2 \\
&+ 3x_n^3 x_{n-2}^2 x_{n-1}^2 + 3x_n^3 x_{n-2} x_{n-1}^3 + 3x_n^3 x_{n-1}^3 x_{n+1} \\
&+ 3x_n^3 x_{n-2}^2 x_{n-1} + 4x_n^3 x_{n-2} x_{n-1}^2 - x_n^3 x_{n-2} x_{n-1} x_{n+1} \\
&+ 2x_n^3 x_{n-1}^3 + x_n^3 x_{n-1}^2 x_{n+1} + x_n^2 x_{n-2} x_{n-1}^3 - x_n^2 x_{n-2} x_{n-1}^2 x_{n+1} \\
&+ 4x_n^2 x_{n-1}^3 x_{n+1} - x_n x_{n-2} x_{n-1}^3 x_{n+1} + 3x_n x_{n-1}^3 x_{n+1}^2 + x_n^3 x_{n-2}^2 \\
&+ x_n^3 x_{n-2} x_{n-1} + x_n^3 x_{n-1}^2 + x_n{}^2 x_{n-2} x_{n-1}^2 - x_n^2 x_{n-2} x_{n-1} x_{n+1} \\
&+ x_n^2 x_{n-1}^2 x_{n+1} - x_n x_{n-2} x_{n-1}{}^2 x_{n+1} + x_n x_{n-1}^3 x_{n+1} + x_{n-1}^3 x_{n+1}^2 \\
&+ x_n^2 x_{n-1}^2 + 3x_n^2 x_{n-1}^3 x_{n+1}^2 + x_n^2 x_{n-1}^3,
\end{aligned} \tag{17a}
$$

$$
\begin{aligned}
D_{3,1} &= x_n x_{n-1}^2 x_{n-2} + 2x_n x_{n-1} x_{n-2} + x_n x_{n-1}^2 \\
&- x_n x_{n-1} x_{n+1} + x_n x_{n-2} - x_{n-1} x_{n+1},
\end{aligned} \tag{17b}
$$

$$
\begin{aligned}
D_{3,2} &= x_n x_{n-1} x_{n-2} - x_n^2 x_{n-1} x_{n+1} - x_n^2 x_{n-1} \\
&- 2x_n x_{n-1} x_{n+1} + x_n x_{n-2} - x_{n-1} x_{n+1}.
\end{aligned} \tag{17c}
$$

This proves that Eq. (13) is an integrable equation according to both definitions stated in Sect. 2. Moreover, the number and shape of the invariants proves that this equation is outside the known classifications given in [7, 22]. For instance, the invariants are rational in the affine coordinate, differently from [22], where the invariants are polynomials in affine coordinates. Furthermore, these invariants are not ratios of biquadratic polynomials, therefore are outside the class considered in [7].

From the invariants (16) it is possible to construct a *dual map* [32]. Differently from the dual maps appearing in [7, 26, 32] this dual map are not integrable according to the algebraic entropy test. We underline that dual maps with such features already appeared in [22].

Remark 2 The invariant I_1 is linear in x_{n+1} and x_{n-2}. This implies that $T_n I_1 - I_1$ is actually equivalent to Eq. (13). That is, Eq. (13) is resummable to an *autonomous third-order difference equation*.

4 Final Remarks

In this short note we showed that the stationary reduction of equation (12), namely Eq. (13), is integrable in the sense of algebraic entropy and in the sense of the existence of invariants. The first properties follow from our previous work [19], yet we showed that the quadratic growth is preserved.

In our upcoming paper [20] we will consider all the stationary reductions of the fourth-order Volterra-like differential-difference equations classified in [15, 16]. The application of the algebraic entropy test will give rise to a vast "zoology" of possibilities, consisting in periodic equations, explicitly linear equations, linearizable equations (linear growth), and integrable equations (quadratic growth). We will explain these growth properties using the following notions:

1. correspondence to *idempotent* maps,
2. explicit *linearization*,
3. *resummation*, in the sense of [3], to integrable second-order non-autonomous difference equations,
4. *deflation*, in the sense of [26], to integrable third-order difference equations,
5. existence of *three invariants*.

Genuinely new integrable fourth-order difference equations belong to the last class. With these new examples we aim to broaden our knowledge of higher-order integrable difference equations and give the foundation for a new algorithmic search method based on hierarchies of differential-difference equations.

Acknowledgments We thank Dr. D. T. Tran for the helpful discussions during the preparation of this paper.

GG is supported through Prof. N. Joshi's Australian Laureate Fellowship #FL120100094.

References

1. V.E. Adler, On a discrete analog of the Tzitzeica equation. Preprint (2011). arXiv:1103.5139
2. V.E. Adler, Integrable Möbius invariant evolutionary lattices of second order. Preprint (2016). arXiv:1605.00018
3. V.E. Adler, Integrability test for evolutionary lattice equations of higher order. J. Symb. Comput. **74**, 125–139 (2016)
4. V.I. Arnol'd, Dynamics of complexity of intersections. Bol. Soc. Brasil. Mat. (N.S.) **21**, 1–10 (1990)
5. M. Bellon, C.M. Viallet, Algebraic entropy. Commun. Math. Phys. **204**, 425–437 (1999)
6. M. Bruschi, O. Ragnisco, P.M. Santini, G.Z. Tu, Integrable symplectic maps. Physica D **49**(3), 273–294 (1991)
7. H.W. Capel, R. Sahadevan, A new family of four-dimensional symplectic and integrable mappings. Physica A **289**, 80–106 (2001)
8. E. Celledoni, C. Evripidou, D.I. McLaren, B. Owren, G.R.W. Quispel, B.K. Tapley, P.H. van der Kamp, Using discrete Darboux polynomials to detect and determine preserved measures and integrals of rational maps. J. Phys. A Math. Theor. **52**, 31LT01 (11pp) (2019)
9. D.K. Demskoi, D.T. Tran, P.H. van der Kamp, G.R.W. Quispel, A novel nth order difference equation that may be integrable. J. Phys. A **45**, 135,202, (10pp) (2012)
10. D.K. Demskoy, C.M. Viallet, Algebraic entropy for semi-discrete equations. J. Phys. A Math. Theor. **45**, 352,001 (10 pp) (2012)
11. J. Duistermaat, *Discrete Integrable Systems: QRT Maps and Elliptic Surfaces*. Springer Monographs in Mathematics (Springer, New York, 2011)
12. S. Elaydi, *An Introduction to Difference Equations*, 3rd edn. (Springer, 2005)
13. G. Falqui, C.M. Viallet, Singularity, complexity, and quasi-integrability of rational mappings. Commun. Math. Phys. **154**, 111–125 (1993)
14. P. Flajolet, A. Odlyzko, Singularity analysis of generating functions. SIAM J. Discr. Math. **3**(2), 216–240 (1990)
15. R.N. Garifullin, R.I. Yamilov, D. Levi, Classification of five-point differential-difference equations. J. Phys. A Math. Theor. **50**, 125,201 (27pp) (2017)
16. R.N. Garifullin, R.I. Yamilov, D. Levi, Classification of five-point differential-difference equations II. J. Phys. A Math. Theor. **51**, 065,204 (16pp) (2018)
17. B. Grammaticos, R.G. Halburd, A. Ramani, C.M. Viallet, How to detect the integrability of discrete systems. J. Phys. A Math. Theor. **42**, 454,002 (41 pp) (2009). Newton Institute Preprint NI09060-DIS
18. G. Gubbiotti, Integrability of difference equations through algebraic entropy and generalized symmetries, in *Symmetries and Integrability of Difference Equations: Lecture Notes of the Abecederian School of SIDE 12*, Montreal 2016, ed. by D. Levi, R. Verge-Rebelo, P. Winternitz. CRM Series in Mathematical Physics, chap. 3 (Springer International Publishing, Berlin, 2017), pp. 75–152
19. G. Gubbiotti, Algebraic entropy of a class of five-point differential-difference equations. Symmetry **11**, 432 (24pp) (2019)
20. G. Gubbiotti, Stationary reductions of five-point differential-difference equations and their integrability properties (2020). In preparation
21. G. Gubbiotti, N. Joshi, D.T. Tran, C.M. Viallet, Complexity and integrability in 4D bi-rational maps with two invariants (2020). Accepted for publication in Springer's PROMS series: "Asymptotic, Algebraic and Geometric Aspects of Integrable Systems"
22. G. Gubbiotti, N. Joshi, D.T. Tran, C.M. Viallet, Bi-rational maps in four dimensions with two invariants. J. Phys. A: Math. Theor **53**, 115201 (24pp) (2020)
23. A. Hagar, *Discrete or Continuous? The Quest for Fundamental Length in Modern Physics* (Cambridge University Press, Cambridge, 2014)
24. J. Hietarinta, N. Joshi, F. Nijhoff, *Discrete Systems and Integrability*. Cambridge Texts in Applied Mathematics (Cambridge University Press, 2016)

25. J. Hietarinta, C.M. Viallet, Searching for integrable lattice maps using factorization. J. Phys. A Math. Theor. **40**, 12,629–12,643 (2007)
26. N. Joshi, C.M. Viallet, Rational maps with invariant surfaces. J. Int. Syst. **3**, xyy017 (14pp) (2018)
27. S.K. Lando, *Lectures on Generating Functions* (American Mathematical Society, 2003)
28. S. Maeda, Completely integrable symplectic mapping. Proc. Jpn. Acad. A Math. Sci. **63**, 198–200 (1987)
29. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes*, 3rd edn. (Cambridge University Press, Cambridge, 2007). The Art of Scientific Computing
30. G.R.W. Quispel, J.A.G. Roberts, C.J. Thompson, Integrable mappings and soliton equations. Phys. Lett. A **126**, 419 (1988)
31. G.R.W. Quispel, J.A.G. Roberts, C.J. Thompson, Integrable mappings and soliton equations II. Physica D **34**(1), 183–192 (1989)
32. G.W.R. Quispel, H.R. Capel, J.A.G. Roberts, Duality for discrete integrable systems. J. Phys. A Math. Gen. **38**(18), 3965 (2005)
33. J.A.G. Roberts, D. Jogia, Birational maps that send biquadratic curves to biquadratic curves. J. Phys. A Math. Theor. **48**, 08FT02 (2015)
34. T. Takenawa, Algebraic entropy and the space of initial values for discrete dynamical systems. J. Phys. A **34**, 10,533 (2001)
35. D.T. Tran, P.H. van der Kamp, G.R.W. Quispel, Closed-form expressions for integrals of traveling wave reductions of integrable lattice equations. J. Phys. A **42**, 225,201, (20pp) (2009)
36. D.T. Tran, P.H. van der Kamp, G.R.W. Quispel, Sufficient number of integrals for the pth-order Lyness equation. J. Phys. A **43**, 302,001, (11pp) (2010)
37. D.T. Tran, P.H. van der Kamp, G.R.W. Quispel, Involutivity of integrals of sine-Gordon, modified KdV and potential KdV maps. J. Phys. A **44**, 295,206, (13pp) (2011)
38. D.T. Tran, P.H. van der Kamp, G.R.W. Quispel, Poisson brackets of mappings obtained as $(q, -p)$ reductions of lattice equations. Regul. Chaotic Dyn. **21**(6), 682–696 (2016)
39. S. Tsujimoto, R. Hirota, Pfaffian representation of solutions to the discrete BKP hierarchy in bilinear form. J. Phys. Soc. Jpn. **65**, 2797–2806 (1996)
40. P. van der Kamp, G.W.R. Quispel, The staircase method: integrals for periodic reductions of integrable lattice equations. J. Phys. A Math. Theor. **43**, 465,207 (2010)
41. A.P. Veselov, Integrable maps. Russ. Math. Surv. **46**, 1–51 (1991)
42. A.P. Veselov, Growth and integrability in the dynamics of mappings. Commun. Math. Phys. **145**, 181–193 (1992)
43. C.M. Viallet, On the algebraic structure of rational discrete dynamical systems. J. Phys. A **48**(16), 16FT01 (2015)
44. C.M. Viallet, B. Grammaticos, A. Ramani, On the integrability of correspondences associated to integral curves. Phys. Lett. A **322**, 186–93 (2004)

Weighted Hurwitz Numbers, τ-Functions, and Matrix Integrals

J. Harnad

Dedicated to Prof. Decio Levi on the occasion of his 70th birthday

Abstract The basis elements spanning the Sato Grassmannian element corresponding to the KP τ-function that serves as generating function for rationally weighted Hurwitz numbers are shown to be Meijer G-functions. Using their Mellin-Barnes integral representation the τ-function, evaluated at the trace invariants of an externally coupled matrix, is expressed as a matrix integral. Using the Mellin-Barnes integral transform of an infinite product of Γ functions, a similar matrix integral representation is given for the KP τ-function that serves as generating function for quantum weighted Hurwitz numbers.

Keywords Hurwitz numbers · τ-functions

1 Hurwitz Numbers: Classical and Weighted

The fact that KP and $2D$-Toda τ-functions of hypergeometric type serve as generating functions for *weighted Hurwitz numbers* was shown in [3–6], generalizing the case of *simple* (single and double) Hurwitz numbers [8, 9]. Sections 1.1 and 1.2 below, and Sect. 2 give a brief review of this theory, together with two illustrative examples: rational and quantum weighted Hurwitz numbers. In Sect. 3, it is shown how evaluation of such τ-functions at the trace invariants of a finite matrix may be expressed either as a Wronskian determinant or as a matrix integral. The content of

J. Harnad (✉)
Department of Mathematics and Statistics, Concordia University, Montréal, QC, Canada

Centre de recherches mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: harnad@crm.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_7

Sects. 3.2–3.4 are largely drawn from [2, 7], in which further details and proofs of the main results may be found.

1.1 Geometric Meaning of Classical Hurwitz Numbers

The Hurwitz number $H(\mu^{(1)}, \dots, \mu^{(k)})$ is the number of inequivalent branched N-sheeted covers $\Gamma \to \mathbf{P}^1$ of the Riemann sphere, with k branch points $(Q_1, \dots, Q_k)$, whose *ramification profiles* are given by k *partitions* $(\mu^{(1)}, \dots, \mu^{(k)})$ of N, normalized by dividing by the order $|\operatorname{aut}(\Gamma)|$ of its automorphism group. The *Euler characteristic* χ and *genus* g of the covering curve are given by the *Riemann–Hurwitz formula*:

$$\chi = 2 - 2g = 2N - d, \quad d := \sum_{i=1}^{l} \ell^*(\mu^{(i)}), \tag{1}$$

where $\ell^*(\mu) := |\mu| - \ell(\mu) = N - \ell(\mu)$ is the *colength* of the partition.

The *Frobenius–Schur* formula gives $H(\mu^{(1)}, \dots \mu^{(k)})$ in terms of S_N characters:

$$H(\mu^{(1)}, \dots \mu^{(k)}) = \sum_{\lambda, |\lambda| = N} h^{k-2}(\lambda) \prod_{j=1}^{k} \frac{\chi_\lambda(\mu^{(i)})}{z_{\mu^{(j)}}}, \quad |\mu^{(i)}| = N, \tag{2}$$

where $h(\lambda) = \left(\det \frac{1}{(\lambda_i - i + j)!}\right)^{-1}$ is the product of the hook lengths of the partition $\lambda = (\lambda_1 \geqslant \cdots \geqslant \lambda_{\ell(\lambda} > 0)$, $\chi_\lambda(\mu^{(j)})$ is the *irreducible character* of representation λ evaluated on the conjugacy class $\mu^{(j)}$, and

$$z_{\mu^{(j)}} := \prod_i i^{m_i(\mu^{(j)})} (m_i(\mu^{(j)}))! \tag{3}$$

is the order of the stabilizer of any element of $\operatorname{cyc}(\mu^{(j)})$ (and $m_i(\mu^{(j)}) = \#$ parts of partition $\mu^{(j)}$ equal to i)

1.2 Weighted Hurwitz Numbers [3–6]

Define the *weight generating function* $G(z)$, or its dual $\widetilde{G}(z)$, as an infinite (or finite) product or sum (formal or convergent).

$$G(z) = \prod_{i=1}^{\infty}(1+zc_i) = 1+\sum_{j=1}^{\infty} g_j z^j$$

$$\widetilde{G}(z) = \prod_{i=1}^{\infty}(1-zc_i)^{-1} = 1+\sum_{j=1}^{\infty} \widetilde{g}_j z^j. \tag{4}$$

The weight for a branched covering with ramification profiles $(\mu^{(1)},\dots,\mu^{(k)})$ is defined to be:

$$\mathcal{W}_G(\mu^{(1)},\dots,\mu^{(k)}) := \frac{1}{k!}\sum_{\sigma\in S_k}\sum_{1\leqslant i_1<\cdots<i_k} c_{i_{\sigma(1)}}^{\ell^*(\mu^{(1)})}\cdots c_{i_{\sigma(k)}}^{\ell^*(\mu^{(k)})}$$

$$\widetilde{\mathcal{W}}_{\widetilde{G}}(\mu^{(1)},\dots,\mu^{(k)}) := \frac{(-1)^{\sum_{i=1}^k \ell^*(\mu^{(i)})+k}}{k!}\sum_{\sigma\in S_k}\sum_{1\leqslant i_1\leqslant\cdots\leqslant i_k} c_{i_{\sigma(1)}}^{\ell^*(\mu^{(1)})}\cdots c_{i_{\sigma(k)}}^{\ell^*(\mu^{(k)})}. \tag{5}$$

Weighted double Hurwitz numbers $H_G^d(\mu,\nu)$, $H_{\widetilde{G}}^d(\mu,\nu)$ for n-sheeted branched coverings of the Riemann sphere having a pair of unweighted branch points (Q_0, Q_∞), with ramification profiles of type (μ,ν), and k additional weighted branch points $(Q_1,\dots,Q_k)$ with ramification profiles $(\mu^{(1)},\dots,\mu^{(k)})$ are defined as:

$$H_G^d(\mu,\nu) := \sum_{k=1}^{d}\sum_{\substack{\mu^{(1)},\dots\mu^{(k)}\\ \sum_{i=1}^k \ell^*(\mu^{(i)})=d}}{}' \mathcal{W}_G(\mu^{(1)},\dots,\mu^{(k)})H(\mu^{(1)},\dots,\mu^{(k)},\mu,\nu),$$

$$H_{\widetilde{G}}^d(\mu,\nu) := \sum_{k=1}^{d}\sum_{\substack{\mu^{(1)},\dots\mu^{(k)}\\ \sum_{i=1}^k \ell^*(\mu^{(i)})=d}}{}' \widetilde{\mathcal{W}}_{\widetilde{G}}(\mu^{(1)},\dots,\mu^{(k)})H(\mu^{(1)},\dots,\mu^{(k)},\mu,\nu),$$

where $\sum'$ denotes the sum over all partitions other than the cycle type of the identity element $(1)^n$. If Q_∞ is not a branch point; i.e. $\nu=(1)^n$, we have a weighted *single* Hurwitz number

$$H_G^d(\mu) := H_G^d(\mu,(1)^n). \tag{6}$$

Two cases of particular interest are: *rational weight generating functions:*

$$G_{\mathbf{c},\mathbf{d}}(z) := \frac{\prod_{l=1}^{L}(1+c_l z)}{\prod_{m=1}^{M}(1-d_m z)} \tag{7}$$

and *quantum weight generating function (quantum exponential):*

$$\widetilde{G}(z) = H_q(z) := \prod_{i=0}^{\infty}(1-q^i z)^{-1} = \sum_{n=0}^{\infty}\frac{z^n}{(q;q)_n}, \tag{8}$$

where

$$(q;q)_n := (1-q)(1-q^2)\cdots(1-q^n) \tag{9}$$

for some parameter q, with $|q| < 1$.

The corresponding *rationally weighted* (single) Hurwitz numbers are

$$H^d_{G_{\mathbf{c},\mathbf{d}}}(\mu,\nu) := \sum_{\substack{1\leqslant k,l\\ k+l\leqslant d}} \sum_{\substack{\mu^{(1)},\dots\mu^{(k)},\nu^{(1)},\dots\nu^{(l)},\\ \sum_{i=1}^k \ell^*(\mu^{(i)})+\sum_{j=1}^l \ell^*(\nu^{(j)})=d\\ |\mu^{(i)}|=|\nu^{(j)}|=N}} \mathcal{W}_{G_{\mathbf{c},\mathbf{d}}}(\mu^{(1)},\dots,\mu^{(k)};\nu^{(1)},\dots,\nu^{(l)})$$
$$\times H(\mu^{(1)},\dots,\mu^{(k)},\nu^{(1)},\dots,\nu^{(l)},\mu),$$

where the rational weight factor is

$$\mathcal{W}_{G_{\mathbf{c},\mathbf{d}}}(\mu^{(1)},\dots,\mu^{(k)};\nu^{(1)},\dots,\nu^{(l)})$$
$$:= \frac{(-1)^{\sum_{j=1}^l \ell^*(\nu^{(j)})-l}}{k!l!}\sum_{\substack{\sigma\in S_k\\ \sigma'\in S_l}}\sum_{\substack{1\leq a_1<\cdots<a_k\leq M\\ 1\leq b_1\cdots\leq b_k\leq L}} c_{a_{\sigma(1)}}^{\ell^*(\mu^{(1)})}\cdots c_{a_{\sigma(k)}}^{\ell^*(\mu^{(k)})} d_{b_{\sigma'(1)}}^{\ell^*(\nu^{(1)})}\cdots d_{b_{\sigma'(l)}}^{\ell^*(\nu^{(l)})}.$$

The *quantum weighted* (single) Hurwitz numbers are

$$H^d_{H_q}(\mu) := \sum_{k=1}^{d}\sum_{\substack{\mu^{(1)},\dots\mu^{(k)},\ |\mu^{(i)}|=N\\ \sum_{i=1}^k \ell^*(\mu^{(i)})=d}} \widetilde{\mathcal{W}}_{H_q}(\mu^{(1)},\dots,\mu^{(k)})H(\mu^{(1)},\dots,\mu^{k)},\mu), \tag{10}$$

where the *quantum weight factor* is

$$\widetilde{\mathcal{W}}_{H_q}(\mu^{(1)},\dots,\mu^{(k)}) := \frac{(-1)^{d-k}}{k!}\sum_{\sigma\in S_k}\prod_{j=1}^{k}\frac{1}{(1-q^{\sum_{i=1}^j \ell^*(\mu^{(\sigma(i))})})}.$$

2 Hypergeometric τ-Functions as Generating Functions for Weighted Hurwitz Numbers [3–6]

To construct a KP τ-function of hypergeometric type that serves as generating function for weighted Hurwitz numbers for a given weight generating function G,

choose a small parameter β and define coefficients $r_\lambda^{(G,\beta)}$ that are of *content product form*:

$$r_\lambda^{(G,\beta)} := \prod_{(ij)\in\lambda} r_{j-i}^{(G,\beta)} = \prod_{(ij)\in\lambda} G((j-i)\beta), \tag{11}$$

where

$$r_j^{(G,\beta)} := G(j\beta) = \frac{\rho_j^{(G,\beta)}}{\beta\rho_{j-1}^{(G,\beta)}}, \tag{12}$$

with

$$\begin{aligned} \rho_j^{(G,\beta)} &:= \beta^j \prod_{i=1}^{j} G(i\beta) =: e^{T_j^G(\beta)}, \quad \rho_0 = 1, = \frac{\rho_j^{(G,\beta)}}{\beta\rho_{j-1}^{(G,\beta)}}, \\ \rho_{-j}^{(G,\beta)}(&:= \beta^{-j} \prod_{i=1}^{j-1} \frac{1}{G(-i\beta)} =: e^{T_{-j}^G(\beta)}, \quad j = 1, 2, \ldots \end{aligned} \tag{13}$$

We then have [4, 6]:

Theorem 1 (Hypergeometric Toda τ-Functions Associated with Weight Generating Function $G(z)$) *The double Schur function series*

$$\tau^{(G,\beta)}(\mathbf{t}, \mathbf{s}) := \sum_\lambda \beta^{|\lambda|} r_\lambda^{(G,\beta)} s_\lambda(\mathbf{t}) s_\lambda(\mathbf{s}) \tag{14}$$

defines a 2D-Toda τ-function (at lattice value $n = 0$).

We now use the *Frobenius character formula*

$$s_\lambda(\mathbf{t}) = \sum_{\mu,|\mu|=|\lambda|} \frac{\chi_\lambda(\mu) p_\mu(\mathbf{t})}{z_\mu}, \quad s_\lambda(\mathbf{s}) = \sum_{\nu,|\nu|=|\lambda|} \frac{\chi_\lambda(\nu) p_\nu(\mathbf{s})}{z_\nu} \tag{15}$$

to change the basis of Schur functions to *power sum symmetric functions*

$$p_\mu(\mathbf{t}) := \prod_{i=1}^{\ell(\mu)} p_{\mu_i}(\mathbf{t}), \;\; p_j(\mathbf{t}) = j t_j, \quad p_\nu(\mathbf{s}) := \prod_{i=1}^{\ell(\nu)} p_{\nu_i}(\mathbf{s}), \;\; p_j(\mathbf{s}) = j s_j. \tag{16}$$

Theorem 2 (Hypergeometric Toda τ-Functions as Generating Function for weighted Double Hurwitz Numbers [4, 6]) *The τ-function $\tau^{(G,\beta)}(\mathbf{t}, \mathbf{s})$ can equivalently be expressed as a double infinite series in the bases of power sum symmetric functions as follows*

$$\tau^{(G,\beta)}(\mathbf{t},\mathbf{s}) = \sum_{d=0}^{\infty} \sum_{\substack{\mu,\nu,\\ |\mu|=|\nu|}} \beta^{|\mu|+d} H_G^d(\mu,\nu) p_\mu(\mathbf{t}) p_\nu(\mathbf{s}). \tag{17}$$

It is thus a generating function for the numbers $H_G^d(\mu,\nu)$ of weighted n-fold branched coverings of the sphere, with a pair of specified branch points having ramification profiles (μ,ν) and genus given by the Riemann–Hurwitz formula

$$2-2g = \ell(\mu)+\ell(\nu)-d, \quad d = \sum_{i=1}^{k} \ell^*(\mu^{(i)}). \tag{18}$$

Corollary 1 (Hypergeometric KP τ-Functions as Generating Functions for Weighted Single Hurwitz Numbers) *Set:* $\mathbf{s} = \beta^{-1}\mathbf{t}_0 := (\beta^{-1}, 0, 0, \dots)$. *Then the series*

$$\tau^{(G,\beta)}(\mathbf{t}, \beta^{-1}\mathbf{t}_0) := \tau^{(G,\beta)}(\mathbf{t}) = \sum_{\lambda} (h(\lambda))^{-1} r_\lambda^{(G,\beta)} s_\lambda(\mathbf{t})$$

$$= \sum_{d=0}^{\infty} \sum_{\mu} \beta^d H_G^d(\mu) p_\mu(\mathbf{t})$$

is a KP τ-function which is a generating function for weighted single numbers $H_G^d(\mu)$ for $|\mu|$-fold branched coverings of the sphere, with a branch point having ramification profile (μ) at Q_0 and genus given by the Riemann–Hurwitz formula.

$$2-2g = |\mu| + \ell(\mu) - d. \tag{19}$$

3 Wronskian and Matrix Integral Representation of $\tau^{(G,\beta)}([X])$

In [2, 7] new matrix integral representations were derived for the τ-functions that serve as generating functions for rationally and quantum weighted Hurwitz numbers. The main result is that, using Laurent series and Mellin-Barnes integral representations of the adapted bases for the respective elements of the infinite Grassmannian corresponding to these cases, the τ-functions may be expressed as Wronskian determinants or as matrix integrals.

3.1 Adapted Basis, Recursion Operators, Quantum Spectral Curve

Henceforth, we always set:

$$\mathbf{s} = \beta^{-1}\mathbf{t}_0 := (\beta^{-1}, 0, 0, \dots) \tag{20}$$

and

$$\tau^{(G,\beta)}(\mathbf{t}) := \tau^{(G,\beta)}(\mathbf{t}, \beta^{-1}\mathbf{t}_0) \tag{21}$$

is a KP τ-function of hypergeometric type.

For $k \in \mathbf{Z}$, define

$$\begin{aligned}\phi_k(x) &:= \frac{\beta}{2\pi i x^{k-1}} \oint_{|\zeta|=\epsilon} \rho^{(G,\beta)}(\zeta) e^{\beta^{-1}x\zeta} \frac{d\zeta}{\zeta^k}, \\ &= \beta x^{1-k} \sum_{j=0}^{\infty} \frac{\rho^{(G,\beta)}_{j-k}}{j!} \left(\frac{x}{\beta}\right)^j,\end{aligned} \tag{22}$$

where

$$\rho^{(G,\beta)}(\zeta) := \sum_{i=-\infty}^{k-1} \rho^{(G,\beta)}_{-i-1} \zeta^i. \tag{23}$$

Then $\{\phi_k(1/z)\}_{k\in\mathbf{N}^+}$ is a basis for the element $w^{(G,\beta)}$ of the Sato Grassmannian that determines the KP τ-function $\tau^{(G,\beta)}(\mathbf{t})$ [1].

3.2 Quantum and Classical Spectral Curve

Theorem 3 (Quantum Spectral Curve and Eigenvalue Equations [1]) *The functions $\phi_k(x)$ satisfy*

$$\mathcal{L}\phi_k(x) := (xG(\beta\mathcal{D}) - \mathcal{D})\,\phi_k(x) = (k-1)\phi_k(x), \tag{24}$$

where $\mathcal{D} := x\frac{d}{dx}$ is the Euler operator.

The classical spectral curve is

$$y = G(\beta x y). \tag{25}$$

Rational Weighting Case

For $G(z) = G_{\mathbf{c},\mathbf{d}}(z)$, denote $\phi_k(x) =: \phi_k^{(\mathbf{c},\mathbf{d},\beta)}(x)$. Then

$$\zeta \prod_{l=1}^{L}(\mathcal{D} + \frac{1}{\beta c_l})\phi_k^{(\mathbf{c},\mathbf{d},\beta)} + (\mathcal{D} + k - 1)\prod_{m=1}^{M}(\mathcal{D} - 1 - \frac{1}{\beta d_m})\phi_k^{(\mathbf{c},\mathbf{d},\beta)} = 0, \tag{26}$$

where

$$\zeta := -\kappa_{\mathbf{c},\mathbf{d}} x, \quad \kappa_{\mathbf{c},\mathbf{d}} := (-1)^M \frac{\prod_{l=1}^{L}\beta c_l}{\prod_{m=1}^{M}\beta d_m}. \tag{27}$$

Mellin-Barnes Integral Representation: Meijer *G*-Functions [2, 7]

It may be shown that $\phi_k^{(\mathbf{c},\mathbf{d},\beta)}$ has the Mellin-Barnes integral representation:

$$\begin{aligned}
\phi_k^{(\mathbf{c},\mathbf{d},\beta)} &= C_k^{(\mathbf{c},\mathbf{d},\beta)} G^{1,L}_{L,M+1}\left(\begin{matrix} 1 - \frac{1}{\beta c_1}, \cdots, 1 - \frac{1}{\beta c_L} \\ 1 - k, 1 + \frac{1}{\beta d_1}, \cdots, 1 + \frac{1}{\beta d_M}\end{matrix}\,\middle|\, -\kappa_{\mathbf{c},\mathbf{d}} x\right) \\
&= \frac{C_k^{(\mathbf{c},\mathbf{d},\beta)}}{2\pi i}\int_{\mathcal{C}_k} \frac{\Gamma(1-k-s)\prod_{\ell=1}^{L}\Gamma\left(s + \frac{1}{\beta c_\ell}\right)\left(-\kappa_{\mathbf{c},\mathbf{d}} x\right)^s}{\prod_{m=1}^{M}\Gamma\left(s - \frac{1}{\beta d_m}\right)} ds. \\
&\sim \frac{\beta \rho_{-k}(\mathbf{c},\mathbf{d})}{(\kappa x)^{k-1}} {}_L F_M\left(\begin{matrix} 1 - k + \frac{1}{\beta c_1}, \cdots, 1 - k + \frac{1}{\beta c_L} \\ 1 - k - \frac{1}{\beta d_1}, \cdots, 1 - k - \frac{1}{\beta d_M}\end{matrix}\,\middle|\, \kappa_{\mathbf{c},\mathbf{d}} x\right)
\end{aligned} \tag{28}$$

where

$$C_k^{(\mathbf{c},\mathbf{d},\beta)} := \frac{\prod_{j=1}^{M}\Gamma(-\frac{1}{\beta d_j})}{(-\beta)^{k-1}\prod_{\ell=1}^{L}\Gamma(\frac{1}{\beta c_\ell})}. \tag{29}$$

The contour $\mathcal{C}_k$ is chosen so that the poles at $1-k, 2-k, \cdots$ are to the right and the poles at $\{-i - \frac{1}{\beta c_j}\}_{j=1,\cdots L,\, i\in\mathbf{N}^+}$ to the left. (See Fig. 1.)

Quantum Case Expressed as Mellin-Barnes Integrals [7]

The following is an integral representation of $\phi_k^{(H_q,\beta)}(x)$, valid for all $x \in \mathbf{C}$,

$$\phi_k^{(H_q,\beta)} = \frac{1}{2\pi i}\int_{\mathcal{C}_k} A_{H_q,k}(s) x^s ds, \tag{30}$$

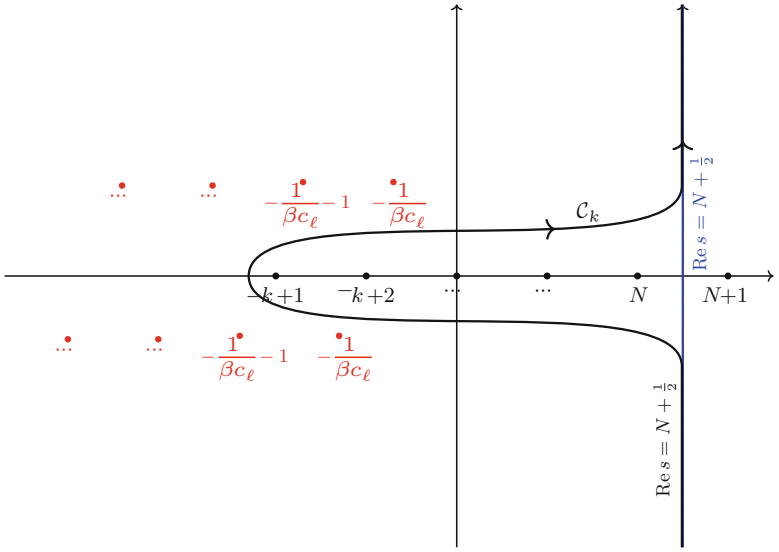


Fig. 1 The contours of integration for the function $\phi_k^{(\mathbf{c},\mathbf{d},\beta)}$ in the case $L > M+1$

where

$$A_{H_q,k}(z) := (-\beta)^{1-k}\Gamma(1-k-z)\prod_{m=0}^{\infty}\left((-\beta q^m)^{-z}\frac{\Gamma(-\beta^{-1}q^{-m})}{\Gamma(z-\beta^{-1}q^{-m})}\right). \tag{31}$$

The contour $\mathcal{C}_k$ is defined as starting at $+\infty$ immediately above the real axis, proceeding to the left above the axis, winding around the poles at the integers $s = -k, -k+1\ldots$ in a counterclockwise sense and continuing below the axis back to $+\infty$.

3.3 Determinantal Representation of $\tau^{(G,\beta)}(\mathbf{t})$

If $\tau^{(G,\beta)}(\mathbf{t})$ is evaluated at the trace invariants of diagonal $X \in \mathrm{Mat}^{n\times n}$

$$\begin{aligned}\mathbf{t} &= [X], \quad t_i = \frac{1}{i}\mathrm{tr}X^i,\\ X &:= \mathrm{diag}(x_1,\ldots,x_n),\end{aligned} \tag{32}$$

it is expressible as the ratio of $n\times n$ determinants

$$\tau^{(G,\beta)}([X]) = \frac{\prod_{i=1}^n x_i^{n-1}}{\prod_{i=1}^n \rho_{-i}}\frac{\det\left(\phi_i(x_j)\right)_{1\le i,j,\le n}}{\Delta(x)}, \tag{33}$$

where

$$\Delta(x) = \prod_{1 \le i < j \le n} (x_i - x_j) = \det(x_i^{n-j})_{1 \le i,j, \le n} \tag{34}$$

is the Vandermonde determinant.

Eulerian Wronskian Representation

It follows from the recursion relations

$$\beta(\mathcal{D} + k - 1)\phi_k = \phi_{k-1}, \quad k \in \mathbf{Z}, \tag{35}$$

that

$$\tau^{(G,\beta)}([X]) = \gamma_n \left(\prod_{i=1}^n x_i^{n-1} \right) \frac{\det\left(\mathcal{D}^{i-1}\phi_n(x_j)\right)_{1 \le i,j, \le n}}{\Delta(x)}, \tag{36}$$

where

$$\gamma_n := \frac{\beta^{\frac{1}{2}n(n-1)}}{\prod_{i=1}^n \rho_{-i}}. \tag{37}$$

3.4 Matrix Integral Representation of $\tau^{(G,\beta)}([X])$ [2, 7]

Wronskian Representation: Rational Case

For rational weight generating functions $G = G_{\mathbf{c},\mathbf{d}}$, and any $n \in \mathbf{N}^+$, let

$$\begin{aligned} \phi_n^{(\mathbf{c},\mathbf{d},\beta)}(e^y) &= \int_{\mathcal{C}_n} A_n^{(\mathbf{c},\mathbf{d},\beta)}(s) e^{ys} ds, \\ A_n^{(\mathbf{c},\mathbf{d},\beta)}(s) &:= \frac{C_n^{(\mathbf{c},\mathbf{d},\beta)} \Gamma(1-n-s) \prod_{l=1}^L \Gamma\left(s + \frac{1}{\beta c_l}\right) (-\kappa_{\mathbf{c},\mathbf{d}})^s}{2\pi i \prod_{m=1}^M \Gamma\left(s - \frac{1}{\beta d_m}\right)}. \end{aligned}$$

Define the diagonal matrix $Y = \mathrm{diag}(y_1, \ldots y_n)$

$$X = e^Y, \quad Y = \ln(X), \quad x_i = e^{y_i}, \quad i = 1, \ldots, n. \tag{38}$$

Then $\tau^{(G_{\mathbf{c},\mathbf{d},\beta})}\left([X]\right)$ becomes a ratio of Wronskian determinants

$$\tau^{(G_{\mathbf{c},\mathbf{d},\beta})}\left([X]\right) = \gamma_n \left(\prod_{i=1}^{n} x_i^{n-1}\right) \frac{\det\left((\phi_n^{(\mathbf{c},\mathbf{d},\beta)})^{(i-1)}(e^{y_j})\right)_{1\le i,j,\le n}}{\Delta(e^y)}. \tag{39}$$

Matrix Integral Representation of $\tau^{(G,\beta)}([X])$: Rational Case

It follows [2] that

$$\tau^{(G_{\mathbf{c},\mathbf{d},\beta})}([X]) = \frac{\beta^{\frac{1}{2}n(n-1)}(\prod_{i=1}^{n} x_i^{n-1})\Delta(\ln(x))}{(\prod_{i=1}^{n} i!)\Delta(x)} \mathbf{Z}_{d\mu_{(\mathbf{c},\mathbf{d},\beta,n)}}(X), \tag{40}$$

where

$$\mathbf{Z}_{d\mu_{(\mathbf{c},\mathbf{d},\beta,n)}}(X) = \int_{M\in \mathrm{Nor}_{\mathcal{C}_n}^{n\times n}} d\mu_{(\mathbf{c},\mathbf{d},\beta,n)}(M) e^{\mathrm{tr}YM} \tag{41}$$

and

$$d\mu_{(\mathbf{c},\mathbf{d},\beta,n)}(M) := (\Delta(\zeta)^2 \det(A_n^{(\mathbf{c},\mathbf{d},\beta)}(M)) d\mu_0(U) \prod_{j=1}^{n} d\zeta_i$$

is a conjugation invariant measure on the space of normal matrices

$$M = UZU^{\dagger} \in \mathrm{Nor}_{\mathcal{C}_n}^{n\times n}, \quad U \in U(n), \quad Z = \mathrm{diag}(\zeta_1, \ldots, \zeta_n) \tag{42}$$

with eigenvalues $\zeta_i \in \mathbf{C}$ supported on the contour $\mathcal{C}_n$.

Wronskian Representation: Quantum Case

For quantum weight generating functions $G = H_q$, and any $n \in \mathbf{N}^+$, let

$$\phi_n^{(H_q,\beta)}(e^y) = \int_{\mathcal{C}_n} A_n^{(\mathbf{c},\mathbf{d},\beta)}(s) e^{ys} ds,$$

$$A_{H_q,n}(z) := (-\beta)^{1-n}\Gamma(1-n-z) \prod_{m=0}^{\infty} \left((-\beta q^m)^{-z} \frac{\Gamma(-\beta^{-1}q^{-m})}{\Gamma(z-\beta^{-1}q^{-m})}\right).$$

Define the diagonal matrix $Y = \mathrm{diag}(y_1, \ldots y_n)$

$$X = e^Y, \quad Y = \ln(X), \quad x_i = e^{y_i}, \quad i = 1, \ldots, n, \tag{43}$$

Then $\tau^{(H_q,\beta)}\left([X]\right)$ becomes a ratio of Wronskian determinants

$$\tau^{(H_q,\beta)}\left([X]\right) = \gamma_n \left(\prod_{i=1}^{n} x_i^{n-1}\right) \frac{\det\left((\phi_n^{(\mathbf{c},\mathbf{d},\beta)})^{(i-1)}(e^{y_j})\right)_{1\le i,j,\le n}}{\Delta(e^y)}. \tag{44}$$

Matrix Integral Representation of $\tau^{(G,\beta)}([X])$: Quantum Case

It similarly follows [7] that

$$\tau^{(H_q,\beta)}([X]) = \frac{\beta^{\frac{1}{2}n(n-1)}(\prod_{i=1}^{n} x_i^{n-1})\Delta(\ln(x))}{(\prod_{i=1}^{n} i!)\Delta(x)} \mathbf{Z}_{d\mu_q}(\ln(X)), \tag{45}$$

$$\text{where} \quad \mathbf{Z}_{d\mu_{(q,n)}}(X) = \int_{M\in \mathrm{Nor}_{\mathcal{C}_n}^{n\times n}} d\mu_{(q,n)}(M) e^{\mathrm{tr}YM},$$

$$\text{and} \quad d\mu_{(q,n)}(M) := (\Delta(\zeta)^2 \det(A_{H_q,n}(M))$$

is a conjugation invariant measure on the space of normal matrices

$$M = UZU^\dagger \in \mathrm{Nor}_{\mathcal{C}_n}^{n\times n}, \quad U \in U(n), \quad Z = \mathrm{diag}(\zeta_1,\ldots,\zeta_n) \tag{46}$$

with eigenvalues $\zeta_i \in \mathbf{C}$ supported on the contour $\mathcal{C}_n$.

Acknowledgments This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec, Nature et technologies (FRQNT).

References

1. A. Alexandrov, G. Chapuy, B. Eynard, J. Harnad, Weighted Hurwitz numbers and topological recursion: an overview. J. Math. Phys. **59**(081102), 1–20 (2018)
2. M. Bertola, J. Harnad, Rationally weighted Hurwitz numbers, Meijer G-functions and matrix integrals. J. Math. Phys. **60**(103504), (2019)
3. M. Guay-Paquet, J. Harnad, 2D Toda τ-functions as combinatorial generating functions. Lett. Math. Phys. **105**, 827–852 (2015)
4. M. Guay-Paquet, J. Harnad, Generating functions for weighted Hurwitz numbers. J. Math. Phys. **58**, 083503 (2017)
5. J. Harnad, Weighted Hurwitz numbers and hypergeometric τ-functions: an overview, in *AMS Proceedings of Symposia in Pure Mathematics*, vol. 93, 289–333 (2016)
6. J. Harnad, A.Yu. Orlov, Hypergeometric τ-functions, Hurwitz numbers and enumeration of paths. Commun. Math. Phys. **338**, 267–284 (2015)

7. J. Harnad, B. Runov, Matrix model generating function for quantum weighted Hurwitz numbers. **53**(065201), (2020)
8. A. Okounkov, Toda equations for Hurwitz numbers. Math. Res. Lett. **7**, 447–453 (2000)
9. A. Okounkov, R. Pandharipande, Gromov-Witten theory, Hurwitz theory and completed cycles. Ann. Math. **163**, 517 (2006)

Constant Curvature Holomorphic Solutions of the Supersymmetric $G(2, 4)$ Sigma Model

Véronique Hussin, Marie Lafrance, and İsmet Yurduşen

Abstract We explore the constant curvature holomorphic solutions of the supersymmetric Grassmannian sigma model $G(M, N)$ using in particular the gauge invariance of the model. Supersymmetric invariant solutions are constructed by generalizing a known result for $\mathbb{C}P^{N-1}$. We show that other such solutions also exist. Considering the simplest case of $G(2, N)$ model, we give necessary and sufficient conditions for getting the constant curvature holomorphic solutions. Since, all the constant curvature holomorphic solutions of the non-supersymmetric $G(2, 4)$ sigma model are known, we treat this example in detail.

Keywords Supersymmetric · Grassmannian sigma model · Gauge invariance

PACS numbers: 12.60.Jv, 02.10.Ud, 02.10.Yn

1 Introduction

For a long time searching for exact solutions of integrable models has been a lively subject of great interest to the mathematics and physics communities. In particular, the integrable non-supersymmetric (non-susy) $\mathbb{C}P^{N-1}$ sigma model has found many applications in physics [1–4]. The solutions of this model have also been used to construct solutions of more general Grassmannian sigma models [5]. Another extension of these models is to consider the supersymmetric (susy) generalizations. The main motivation to study susy models is to include fermions into the theory [6]. Although, there exist many ways of including fermions into the Grassmannian models, the most interesting is the one which renders supersymmetry.

V. Hussin (✉) · M. Lafrance
Centre de Recherches Mathématiques, Montréal, QC, Canada
e-mail: veronique.hussin@umontreal.ca; marie.lafrance@umontreal.ca

İ. Yurduşen
Hacettepe University, Ankara, Turkey
e-mail: yurdusen@hacettepe.edu.tr

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_8

For the non-susy case, a general approach for constructing holomorphic maps of 2-sphere S^2 of constant curvature into $G(M, N)$ has been realized in [7, 8] and the cases $G(2, 4)$ and $G(2, 5)$ have been treated in detail [9, 10].

The natural question is to extend those results to susy $G(M, N)$ σ-models for $M > 1$. In the susy case, some results have already been known when $M = 1$, i.e. the $\mathbb{C}P^{N-1}$ σ-model [5, 6, 11, 12]. In particular, all the susy invariant solutions with constant curvature holomorphic (CCH) solutions of this model have been thoroughly discussed.

In a recent article [13], the present authors, together with W. J. Zakrzewski, introduced a general method for characterizing the constant curvature surfaces for the holomorphic solutions of the susy Grassmannian sigma models. The main tool there was to use the gauge invariance of these models and to our knowledge this was the first time in literature that this invariance is explicitly and effectively used in such a context. In this paper, first we give some criteria for having CCH solutions of the susy Grassmannian $G(M, N)$ σ-model by the help of gauge invariance and then apply this method on a specific example, namely the $G(2, 4)$ sigma model thoroughly. The problem splits into four cases and we separately investigate all of them. Whenever possible we give the canonical form of such constant curvature surfaces. Among these four types of solutions with different curvatures, only two of them produce the susy invariant solutions as the unique ones.

The structure of this paper is as follows; in Sect. 2, we discuss the necessary and sufficient conditions to get the CCH solutions of the general susy $G(M, N)$ σ-model. In Sect. 3 we give a detailed analysis of the susy $G(2, 4)$ σ-model. Taking into account the susy gauge invariance we present all the CCH solutions of this model. Finally, we end the article by giving some comments in Sect. 4.

2 CCH Solutions of the Susy $G(M, N)$ σ-Model

For the susy $G(M, N)$ σ-model [5], a general bosonic superfield has the following expansion $\Phi(x_\pm, \theta_\pm) = \Phi_0(x_\pm) + i\theta_+\Phi_1(x_\pm) + i\theta_-\Phi_2(x_\pm) - \theta_+\theta_-\Phi_3(x_\pm)$, where Φ_0 and Φ_3 are $N \times M$ bosonic complex matrices and Φ_1 and Φ_2 are $N \times M$ fermionic complex matrices. This bosonic superfield must satisfy $\Phi^\dagger\Phi = \mathbb{I}_M$. The energy action functional of the model is given by $S(\Phi) = \int\int_{S^2} dx_+dx_-d\theta_+d\theta_-L(\Phi)$, where $L(\Phi) = 2\,Tr\left(|\check{D}_+\Phi|^2 - |\check{D}_-\Phi|^2\right)$ and the supercovariant derivatives are defined by $\check{D}_\pm\tilde{\Lambda} = \check{\partial}_\pm\tilde{\Lambda} - \tilde{\Lambda}(\Phi^\dagger\check{\partial}_\pm\Phi)$, with $\check{\partial}_\pm = -i\partial_{\theta_\pm} + \theta_\pm\partial_\pm$ and $\partial_\pm \equiv \partial_{x_\pm}$. Using the principle of least action, it is found that the superfield Φ satisfies the Euler–Lagrange equations $\check{D}_+\check{D}_-\Phi + \Phi|\check{D}_-\Phi|^2 = 0$. As in the non-susy case, holomorphic solutions of the susy $G(M, N)$ σ-model are trivial solutions of the model [5, 13]. It has been shown that they take the form $\Phi = WL$, where W is an $N \times M$ matrix depending only on the coordinates (x_+, θ_+), while L is a non-singular $M \times M$ matrix that depends on the coordinates $(x_\pm, \theta_\pm)$. It means that the holomorphic superfield W takes the explicit form

$$W(x_+, \theta_+) = Z(x_+) + i\theta_+\eta A(x_+)\,, \tag{1}$$

where η is a fermionic constant and $Z(x_+)$ and $A(x_+)$ are usual $N \times M$ matrices that depend on x_+.

We now assume, as for the special case $M = 1$, that the susy Gaussian curvature of the surface corresponding to the susy holomorphic solution W is given by the formula

$$\tilde{\kappa} = -\frac{1}{\tilde{g}}\partial_+\partial_- \ln \tilde{g}, \tag{2}$$

where the susy expression of the metric is $\tilde{g} = \partial_+\partial_- \ln\left(\det\left(W^\dagger W\right)\right)$.

Thus asking for a CCH solution is equivalent to assuming that $\tilde{\kappa} = \kappa$ where κ is a purely bosonic constant (a strictly positive real number) and must be the curvature associated with the non-susy $G(M, N)$ solution Z involved in $W =$ (1).

Let us write explicitly the condition (2) using the expression of W in (1) and taking into account that $\tilde{\kappa} = \kappa$. In order to simplify the calculations, we take $T_1 = \theta_+\eta$ and $T_2 = \theta_-\eta^\dagger$. Notice that since T_1 and T_2 are both product of two fermionic quantities, we have $T_1^2 = 0$ and $T_2^2 = 0$. Moreover, they are bosonic and hence commute with all the other quantities.

We thus easily get

$$\begin{aligned}\det\left(W^\dagger W\right) &= (\det M_0)\det\left(\mathbb{I}_M + iT_1M_0^{-1}M_1 + iT_2M_0^{-1}M_2 - T_1T_2M_0^{-1}M_3\right)\\ &= (\det M_0)\left(1 + iT_1X_1 + iT_2X_2 - T_1T_2X_3\right),\end{aligned} \tag{3}$$

with $M_0 = Z^\dagger Z$, $M_1 = Z^\dagger A$, $M_2 = A^\dagger Z$ and $M_3 = A^\dagger A$. The expressions of X_1, X_2 and X_3 remain to be explicitly computed.

The metric $\tilde{g} = \partial_+\partial_- \ln\left(\det\left(W^\dagger W\right)\right)$ takes the form

$$\tilde{g} = g + \partial_+\partial_- \ln\left(1 + iT_1X_1 + iT_2X_2 - T_1T_2X_3\right), \tag{4}$$

with $g = \partial_+\partial_- \ln(\det M_0)$. Using the Taylor expansion of the logarithmic function $\ln(1 + x) = x - \frac{x^2}{2} + \mathcal{O}(x^3)$, we get

$$\tilde{g} = g + \partial_+\partial_-\left[iT_1X_1 + iT_2X_2 - T_1T_2\left(X_3 - X_1X_2\right)\right]. \tag{5}$$

By a similar procedure we can express the quantity $\partial_+\partial_- \ln \tilde{g}$ as

$$\begin{aligned}\partial_+\partial_- \ln \tilde{g} = {}& \partial_+\partial_- \ln g + iT_1\,\partial_+\partial_- Y_1 + iT_2\,\partial_+\partial_- Y_2\\ & -T_1\,T_2\,\partial_+\partial_-\left(Y_3 - Y_1Y_2\right),\end{aligned} \tag{6}$$

with

$$Y_1 \equiv \frac{\kappa}{2}(1+|x|^2)^2\partial_+\partial_- X_1\,, \quad Y_2 \equiv \frac{\kappa}{2}(1+|x|^2)^2\partial_+\partial_- X_2\,,$$
$$Y_3 \equiv \frac{\kappa}{2}(1+|x|^2)^2\partial_+\partial_-\big(X_3 - X_1\,X_2\big)\,. \tag{7}$$

Upon inserting these relations into (2) we get the following constraints

$$\partial_+\partial_-\ln g + \kappa g = 0\,, \tag{8}$$
$$\partial_+\partial_-\,(Y_1 + \kappa X_1) = 0\,, \qquad \partial_+\partial_-\,(Y_2 + \kappa X_2) = 0\,, \tag{9}$$
$$\partial_+\partial_-\,((Y_3 - Y_1\,Y_2) + \kappa(X_3 - X_1\,X_2)) = 0\,. \tag{10}$$

Notice that the two expressions in (9) are complex conjugate to each other and hence we have only one independent condition, say the one involving Y_1 and X_1. These are necessary and sufficient conditions for the susy holomorphic solutions to have a constant Gaussian curvature and will be the fundamental equations for our analysis.

2.1 Susy Invariant Solutions

Here we give a sufficient condition for obtaining CCH solutions. This result generalizes what we already proved in the case $M = 1$ [11]. We assume that the susy holomorphic solution is given by

$$W(x_+, \theta_+) = Z(x_+) + i\theta_+\eta\partial_+ Z(x_+), \tag{11}$$

i.e. $A(x_+) = \partial_+ Z(x_+)$ in (1), where Z is a CCH solution of the non-susy model. Using the MacFarlane parametrization [14], we can rewrite (11) as

$$W = \begin{pmatrix} \mathbb{I}_M \\ K + i\theta_+\eta\partial_+ K \end{pmatrix}. \tag{12}$$

Then we prove that $W =$ (12) is a CCH solution of the susy $G(M, N)$ model. Remember here that, $\det\left(Z^\dagger Z\right) = \det\left(\mathbb{I}_M + K^\dagger K\right) = \left(1+|x|^2\right)^r$, for some positive integer r and thus $\tilde{\kappa} = \kappa = \frac{2}{r}$.

3 CCH Solutions of the Susy $G(2, 4)$ σ-Model

The non-susy CCH solutions are given in [9] as

$$Z_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ x_+ & 0 \\ 0 & 0 \end{pmatrix}, \; Z_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ x_+^2 \cos 2t & \sqrt{2}x_+ \cos t \\ \sqrt{2}x_+ \sin t & 0 \end{pmatrix}, \quad t \in \mathbb{R}$$

$$Z_3 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \sqrt{3}x_+^2 & \sqrt{8/3}x_+ \\ 0 & \sqrt{1/3}x_+ \end{pmatrix}, \; Z_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2x_+^3 & \sqrt{3}x_+^2 \\ \sqrt{3}x_+^2 & 2x_+ \end{pmatrix}. \tag{13}$$

Since all the CCH solutions of the non-susy $G(2,4)$ σ-model are known, we use them to construct the CCH solutions of the corresponding susy model and investigate the constraints for them to satisfy (9) and (10).

The corresponding superfield takes the form

$$W_r(x_+) = Z_r(x_+) + i\theta_+\eta A_r(x_+), \qquad r = 1, 2, 3, 4, \tag{14}$$

where the different Z_r are given by (13). Using the gauge invariance of the susy model [13], we have

$$A_r(x_+) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \beta_{11}(x_+) & \beta_{12}(x_+) \\ \beta_{21}(x_+) & \beta_{22}(x_+) \end{pmatrix} = \begin{pmatrix} 0 \\ \beta(x_+) \end{pmatrix}. \tag{15}$$

Since the solutions $Z_r(x_+)$ are all real functions of x_+, we assume that it is also the case for $A_r(x_+)$. For each holomorphic solution $W_r(x_+)$ given in (14), the conditions (9) and (10) have to be satisfied.

3.1 The Case of Z_1

This is the simplest solution of the non-susy $G(2,4)$ model with $\det Z_1^\dagger Z_1 = \left(1 + |x|^2\right)$, i.e. $r = 1$ or $\kappa = 2$. It is easy to see that the condition (9) is trivially satisfied for W_1 given in (14). Hence we are left with the condition (10). It reads as

$$|x_+(\partial_+^2\beta_{22}) + 2(\partial_+\beta_{22})|^2 + |\partial_+^2\beta_{22}|^2 + |\partial_+^2\beta_{12}|^2 + |\partial_+^2\beta_{21}|^2 = 0. \tag{16}$$

Since β_{11} does not appear in this equation, it will remain arbitrary. Equation (16) implies that

$$\partial_+^2\beta_{12} = 0, \quad \partial_+^2\beta_{21} = 0, \quad \partial_+^2\beta_{22} = 0, \quad x_+(\partial_+^2\beta_{22}) + 2(\partial_+\beta_{22}) = 0, \tag{17}$$

which further fix the matrix A_1. We thus get susy CCH solutions of the form

$$W_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ x_+ & 0 \\ 0 & 0 \end{pmatrix} + i\theta_+\eta \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \beta_{11}(x_+) & b_1x_+ + b_0 \\ c_1x_+ + c_0 & d_0 \end{pmatrix}, \tag{18}$$

where b_1, b_0, c_1, c_0 and d_0 are arbitrary constants. Notice that when $b_0 = b_1 = c_0 = c_1 = d_0 = 0$, we get in particular the susy invariant solution. It is clear that we have more solutions than the susy invariant one in this case.

3.2 The Case of Z_2

We have a family of non-susy solutions, labeled by the parameter t:

$$Z_2(x_+, t) = \begin{pmatrix} \mathbb{I}_2 \\ K_2(t) \end{pmatrix}, \qquad K_2(x_+, t) = \begin{pmatrix} x_+^2 \cos 2t & \sqrt{2}x_+ \cos t \\ \sqrt{2}x_+ \sin t & 0 \end{pmatrix}. \tag{19}$$

Since $\det Z_2^\dagger Z_2 = \det\left(\mathbb{I}_2 + K_2^\dagger K_2\right) = \left(1 + |x|^2\right)^2$, the associated curvature is $\kappa = 1$. In [9], the parameter t can take any real values but due to the properties of the trigonometric functions, using a residual gauge invariance, we have been able to show that $t \in [0, \pi[$.

Considering now the corresponding susy holomorphic solution

$$W_2(x_+, \theta_+, t) = Z_2(x_+, t) + i\theta_+\eta A_2(x_+, t), \tag{20}$$

where $A_2(x_+, t)$ takes the form (15), the conditions (9) and (10) have to be satisfied in order to get a family of CCH solutions.

Introducing W_2 given in (20) into (9), we get two different cases:

1. The first case corresponds to $\cos 2t \neq 0$. Condition (9) implies $\beta_{11}(x_+, t) = x_+\left(\sqrt{2}\cos t\beta_{12}(x_+, t) - \sqrt{2}\sin t\beta_{21}(x_+, t) + x_+ \sin 2t\beta_{22}(x_+, t)\right)$. So we have only one condition (10) to resolve three unknown functions. Interestingly, starting with a polynomial form in x_+ of the unknown functions we get a pattern. Indeed, we find that

$$\beta_{12}(x_+, t) = c_0 + c_1x_+ + F(x_+)\,, \beta_{21}(x_+, t) = \left(c_0 + F(x_+)\right)\tan t + a_1x_+,$$
$$\beta_{22}(x_+, t) = \frac{\cos t}{\sqrt{2}}\left(a_1 - c_1 \tan t\right), \tag{21}$$

where a_1, c_0 and c_1 are constants, solve our problem. Thus the matrix $\beta(x_+, t)$ takes the form

$$\beta(x_+,t)=\frac{(c_0+F(x_+))}{\sqrt{2}\cos t}\begin{pmatrix}2x_+\cos 2t & \sqrt{2}\cos t\\ \sqrt{2}\sin t & 0\end{pmatrix}+a_1\begin{pmatrix}-\sqrt{2}x_+^2\sin^3 t & 0\\ x_+ & \frac{1}{\sqrt{2}}\cos t\end{pmatrix}$$
$$+c_1\begin{pmatrix}\sqrt{2}x_+^2\cos^3 t & x_+\\ 0 & -\frac{1}{\sqrt{2}}\sin t\end{pmatrix}. \tag{22}$$

The susy invariant solution is obtained when $a_1 = c_1 = 0$. Again the case Z_2 gives other solutions to our problem than the susy invariant ones.

2. The second case corresponds to $\cos 2t = 0$ or $t = \frac{\pi}{4}$ (the case $t = \frac{3\pi}{4}$ is gauge equivalent) so that $K_2(x_+, \frac{\pi}{4}) = \begin{pmatrix}0 & x_+\\ x_+ & 0\end{pmatrix}$. Since $K_2(x_+, \frac{\pi}{4})$ is symmetric, we assume that the matrix $\beta(x_+)$ is also symmetric, i.e. $\beta_{21}(x_+) = \beta_{12}(x_+)$. These quantities will remain arbitrary since the condition (10) depends only on β_{11} and β_{22} and the susy invariant solutions will be obtained when $\beta_{11} = \beta_{22} = 0$. The condition (10) may be written as follows, taking in particular $x_+ = x_- = x$:

$$(1+x^2)^2\Big(4(x^2-1)\big((\beta'_{11})^2+(\beta'_{22})^2\big)+(1+x^2)^2\big((\beta''_{11})^2+(\beta''_{22})^2\big)\Big)$$
$$-8x(1+x^2)(x^2-2)\big(\beta_{11}\beta'_{11}+\beta_{22}\beta'_{22}\big)+4x^2(1+x^2)^2\big(\beta_{11}\beta''_{11}+\beta_{22}\beta''_{22}\big)$$
$$+4(1-4x^2+x^4)(\beta_{11}^2+\beta_{22}^2)-4x(1+x^2)^3\big(\beta'_{11}\beta''_{11}+\beta'_{22}\beta''_{22}\big)=0\,. \tag{23}$$

Let us first mention the invariance of this equation with respect to the exchange $\beta_{11} \leftrightarrow \beta_{22}$. After some trials we first get a solution choosing $\beta_{22}(x) = x\beta_{11}(x)$. Condition (23) thus becomes very simple $(1+x^2)^5(\beta''_{11}(x))^2 = 0$, which implies that

$$\beta_{11}(x) = a_0 + d_2 x\,, \qquad \beta_{22}(x) = x(a_0 + d_2 x)\,. \tag{24}$$

Using this observation, we assume that $\beta_{11}(x)$ and $\beta_{22}(x)$ are real polynomial in x. We can easily show that they must be at most of degree 2. If we take $\beta_{11}(x) = a_2x^2+a_1x+a_0$, $\beta_{22}(x) = d_2x^2+d_1x+d_0$ and identify the coefficients of different powers of x in (23), we get three independent equations for the parameters a_i and d_i,

$$a_0^2 - a_1^2 + a_2^2 + d_0^2 - d_1^2 + d_2^2 = 0\,, \qquad a_0a_2 + d_0d_2 = 0\,,$$
$$a_0a_1 - a_1a_2 + d_1(d_0 - d_2) = 0\,. \tag{25}$$

Let us first assume that $a_0 \neq 0$, we then get

$$\beta_{11}(x) = a_0 + (d_2 - d_0)x - \frac{d_0d_2}{a_0}x^2,\ \beta_{22}(x) = d_0 + \left(a_0 + \frac{d_0d_2}{a_0}\right)x + d_2x^2,$$

where a_0, d_0 and d_2 remain arbitrary real parameters. Clearly the solution (24) is obtained when $d_0 = 0$.

Now, consider $a_0 = 0$. We then get different subcases ($\epsilon = \pm 1$)

- $d_2 \neq 0, d_0 = 0 \Longrightarrow \begin{cases} \beta_{11}(x) = a_2 x^2 + \epsilon d_2 x\,, \\ \beta_{22}(x) = d_2 x^2 - \epsilon a_2 x\,, \end{cases}$
- $d_0 \neq 0, d_2 = 0 \Longrightarrow \begin{cases} \beta_{11}(x) = a_2 x^2 + \epsilon d_0 x\,, \\ \beta_{22}(x) = \epsilon a_2 x + d_0\,. \end{cases}$

3.3 The Case of Z_3

In this case we have $\det Z_3^\dagger Z_3 = \left(1+|x|^2\right)^3$, i.e. $r = 3$ and $\kappa = \frac{2}{3}$. With the solution W_3 as in (14), the condition (9) becomes a third degree polynomial in x_-. Equating the coefficients of different powers of x_- to zero we obtain the following equations:

$$2x_+^3\left(\sqrt{2}\beta_{12}'' + 5\beta_{22}''\right) - x_+^2\left(3\beta_{11}'' + 8\sqrt{2}\beta_{12}' + 6\sqrt{2}\beta_{21}'' + 40\beta_{22}'\right)$$
$$+6x_+\left(3\beta_{11}' + 2\sqrt{2}\beta_{12} + 6\sqrt{2}\beta_{21}' + 10\beta_{22}\right) - 36\beta_{11} - 72\sqrt{2}\beta_{21} = 0\,, \quad (26)$$

$$x_+^2\left(-\beta_{11}'' + 8\sqrt{2}\beta_{12}' - 4\sqrt{2}\beta_{21}'' + 6x_+\beta_{22}'' + 4\beta_{22}'\right)$$
$$-x_+\left(4\beta_{11}' + 24\sqrt{2}\beta_{12} - 8\sqrt{2}\beta_{21}' + 48\beta_{22}\right) + 28\beta_{11} + 16\sqrt{2}\beta_{21} = 0\,, \quad (27)$$

$$x_+\left(\beta_{11}'' - 2\sqrt{2}x_+\beta_{12}'' + 8\sqrt{2}\beta_{12}' - 2\sqrt{2}\beta_{21}'' + 2x_+\beta_{22}'' + 16\beta_{22}'\right)$$
$$-10\beta_{11}' + 12\sqrt{2}\beta_{12} - 4\sqrt{2}\beta_{21}' + 12\beta_{22} = 0\,, \quad (28)$$

$$-3\beta_{11}'' + 4\sqrt{2}x_+\beta_{12}'' + 8\sqrt{2}\beta_{12}' + 2x_+\beta_{22}'' + 4\beta_{22}' = 0\,, \quad (29)$$

whose solution gives the final form of β_{11} and β_{22} as

$$\beta_{11}(x_+) = \frac{3x_+}{\sqrt{2}}\beta_{12}(x_+) + \frac{1}{\sqrt{2}}\beta_{21}(x_+), \quad (30)$$

$$\beta_{22}(x_+) = \frac{\sqrt{2}}{4}\beta_{12}(x_+) + \frac{3\sqrt{2}}{4x_+}\beta_{21}(x_+). \quad (31)$$

Introducing (30) and (31) into the condition (10) we obtain

$$\left(\frac{\left|\left(1+3|x|^2\right)\beta_{21} - x_+\left(1+|x|^2\right)\partial_+\beta_{21}\right|^2}{|x|^4\left(1+|x|^2\right)^2}\right) = f(x_+) + g(x_-)\,, \quad (32)$$

for arbitrary functions f and g of given variables. Requiring it to be satisfied when $x_+ = 0$ and $x_- = 0$ separately we obtain $\beta_{21}(x_+) = \gamma_1 x_+$, where γ_1 is an arbitrary constant. Upon introducing it into (32) we get $f(x_+) + g(x_-) = \frac{4|x|^2\gamma_1^2}{(1+|x|^2)^2}$, which immediately implies that $\gamma_1 = 0$ and hence $\beta_{21} = 0$.

The necessary and sufficient conditions (9) and (10) are thus satisfied and finally the CCH solution W_3 is given by the form

$$W_3 = Z_3 + i\theta_+\sqrt{3}\eta\beta_{22}(x_+)\partial_+ Z_3. \tag{33}$$

Hence in this case we have obtained the susy invariant solution as the unique CCH solution.

3.4 The Case of Z_4

In this case we have $\det Z_4^\dagger Z_4 = \left(1 + |x|^2\right)^4$, i.e. $r = 4$ and $\kappa = \frac{1}{2}$. Again the condition (9) becomes a third degree polynomial in x_- after introducing the solution W_4 given in (14). Similarly as what we did with W_3, we equate the coefficients of different powers of x_- to zero and now get

$$\beta_{11}(x_+) = 3x_+^2\beta_{22}(x_+), \qquad \beta_{21}(x_+) = -\beta_{12}(x_+) + 2\sqrt{3}x_+\beta_{22}(x_+). \tag{34}$$

In order to solve the last condition (10) we introduce (34) into it and find that $\beta_{21}(x_+) = \beta_{12}(x_+)$. Finally, the CCH solution W_4 is given as

$$W_4 = Z_4 + i\theta_+\frac{1}{2}\eta\beta_{22}(x_+)\partial_+ Z_4. \tag{35}$$

Thus, in the case of Z_4 once more we have obtained the susy invariant solution as the unique CCH solution.

4 Conclusions and Final Comments

In this article we give some criteria for having CCH solutions of the susy Grassmannian $G(M, N)$ σ-model. With the help of the susy gauge invariance of the model we first show that the susy holomorphic solution given in (11) (i.e., generalization of non-susy holomorphic solution) leads to a constant curvature surface. This kind of a solution is called a susy invariant one, in analogy with the discussion given in [11]. Then we restrict ourselves to the susy $G(2, N)$ σ-model and give the necessary and sufficient conditions to get such solutions. The case of $G(2, 4)$ is studied in detail taking into account the classification of non-susy solutions [9].

Acknowledgments İY thanks the Centre de Recherches Mathématiques, Université de Montréal, where part of this work was done, for the kind hospitality. VH acknowledges the support of research grants from NSERC of Canada. Several discussions with W J Zakrzewski have been very fruitful.

References

1. D.G. Gross, T. Piran, S. Weinberg, *Two Dimensional Quantum Gravity and Random Surfaces* (World Scientific, Singapore, 1992)
2. S. Safram, *Statistical Thermodynamics of Surfaces Interface and Membranes* (Addison-Wesley, Massachusetts, 1994)
3. R. Rajaraman, $\mathbb{C}P^N$ solitons in quantum Hall systems. Eur. Phys. **B 28**, 157–162 (2002)
4. N. Manton, P. Sutcliffe, *Topological Solitons* (Cambridge University Press, New York, 2004)
5. W.J. Zakrzewski, *Low Dimensional Sigma Models* (Adam Hilger, Bristol, 1989)
6. E. Witten, Supersymmetric form of the nonlinear σ model in two dimensions. Phys. Rev. D **16**, 2991–2994 (1977)
7. L. Delisle, V. Hussin, W.J. Zakrzewski, Constant curvature solutions of Grassmannian sigma models: (1) Holomorphic solutions. J. Geom. Phys. **66**, 24–36 (2013)
8. L. Delisle, V. Hussin, W.J. Zakrzewski, Constant curvature solutions of Grassmannian sigma models: (2) Non-holomorphic solutions. J. Geom. Phys **71**, 1–10 (2013)
9. Z.Q. Li, Z.H. Yu, Constant curved minimal 2-spheres in $G(2, 4)$. Manuscripta Math. **100**, 305–316 (1999)
10. X.X. Jiao, J.G. Peng, Classification of holomorphic spheres of constant curvature in complex Grassmann manifold $G(2, 5)$. Diff. Geom. Appl. **20**, 267–277 (2004)
11. L. Delisle, V. Hussin, İ. Yurduşen, W.J. Zakrzewski, Constant curvature surfaces of the supersymmetric $\mathbb{C}P^{N-1}$ sigma model. J. Math. Phys. **56**, 023506-1–023506-18 (2015)
12. L. Delisle, V. Hussin, W.J. Zakrzewski, General construction of solutions of the supersymmetric $\mathbb{C}P^{N-1}$ sigma model. J. Math. Phys **57**, 023506 (2016)
13. V. Hussin, M. Lafrance, İ. Yurduşen, W.J. Zakrzewski, Holomorphic solutions of the susy Grassmannian σ-model and gauge invariance. J. Phys. A Math. Theor. **51**, 185401-1–185401-8 (2018)
14. A.J. MacFarlane, Generalizations of σ-models and $\mathbb{C}P^N$ models and instantons. Phys. Lett. B **82**, 239–241 (1978)

How to Deal with Nonlocality and Pseudodifferential Operators. An Example: The Salpeter Equation

A. Lattanzi

Dedicated to Prof. Decio Levi on the occasion of his 70th birthday

Abstract The spinless (1+1)D free-particle Salpeter equation, a relativistic version of the Schrödinger equation, is presented focusing the attention on its nonlocality and its consequences on the structure of the solution.

Keywords Salpeter equation · Nonlocality · Operator method · Evolution operators

1 The Spinless Salpeter Equation: An Introduction

The spinless time-dependent Salpeter equation is a relativistic version of the time-dependent Schrödinger equation [1–9]. It is obtained by replacing the classical energy–momentum relation with the relevant relativistic relation, possibly including also some potential interactions. Contrary to the Schrödinger equation as well as to the other wave equations in relativistic quantum mechanics, i.e. the Klein–Gordon and the Dirac equations, the Salpeter equation has been the object of few analyses. This is definitely because of the mathematical complexity due to the presence of a pseudodifferential operator that implies a nonlocality increasing the difficulties to work in the coordinate space [10, 11].

Recent literature is revealing a greater interest in the Salpeter equation [12–29]. The reasons are mainly two: on the one hand, there are the advantages with respect to the Klein–Gordon and Dirac equations and on the other hand, there

A. Lattanzi (✉)
H. Niewodniczański Institute of Nuclear Physics, Polish Academy of Sciences, Kraków, Poland

ENEA Frascati Research Centre, FSN-FUSPHYS-TSM, Frascati (Rome), Rome, Italy
e-mail: ambra.lattanzi@ifj.edu.pl

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_9

is the fact that making suitable simplifications and approximations it stems from the Bethe–Salpeter equation [6], which is the most orthodox tool for discussing the relativistic two-body problem in quantum field theory [9]. Firstly proposed by Nambu [1] without derivation, the equation was then derived by Bethe and Salpeter [2] using Feynman-graphs and by Gell-Mann and Low [3] on the basis of the rigorous quantum field theory considerations. After some initial difficulties, the Bethe–Salpeter equation has been the object of intense theoretical studies [7, 9, 19–22]. As a general quantum field theoretical tool, the Bethe–Salpeter equation finds applications in several areas of quantum field theory, as, for instance, in connection with electron–positron pairs and excitons (i.e. bound states of electron–hole pairs).

A more detailed discussion on the approximations needed for obtaining the spinless Salpeter equation from the Bethe–Salpeter equation can be found in [12, 13] and a summary is delineated here. By eliminating the dependence on time-like variables through the assumption of static or instantaneous interactions, the Bethe–Salpeter equation [2] reduces to the Salpeter equation [4]. Neglecting furthermore all spin degrees of freedom and restricting it only to positive-energy solutions, one obtains the spinless Salpeter equation [6].

The spinless Salpeter equation is frequently employed in the phenomenological description of hadrons. Moreover, the agreement of the predictions of the spinless Salpeter equation with the experimental spectrum of mesonic atoms is as good as those of the Klein–Gordon equation.

In $(1+1)$D, the spinless Salpeter equation incorporates the relativistic expression for the energy of the particle, which in the presence of a potential $V(x,t)$ is

$$E = \sqrt{m^2c^4 + p^2c^2} + V(x,t), \tag{1}$$

where m and p denote, respectively, the rest mass and the momentum of the particle, while c is the speed of light in vacuum and x denotes the position.

In accord with the standard quantization rules

$$E \to i\hbar\partial_t, \qquad \vec{r} \to \vec{r}, \qquad \vec{p} \to -i\hbar\nabla,$$

the relativistic $(1+1)$D Hamiltonian reads

$$\hat{H} = \sqrt{m^2c^4 - c^2\hbar^2\frac{\partial^2}{\partial x^2}} + V(x,t), \tag{2}$$

whose main and most remarkable feature is the presence of the square-root operator.

In this work, we consider the free-particle Salpeter equation ($V(x,t) = 0$):

$$i\hbar\partial_t\ \psi(x,t) = \sqrt{m^2c^4 - c^2\hbar^2\frac{\partial^2}{\partial x^2}}\ \psi(x,t), \qquad \psi(x,0) = \psi_0(x). \tag{3}$$

The case of a linear potential for the Salpeter equation and related approximating equations can be found in [23, 25–29] and in references therein.

After a very short introduction in where the spinless Salpeter equation has been introduced in a physical and mathematical framework, Sect. 2 is devoted to the definition of its solutions in the coordinate and in the momentum space, respectively. It has been considered also the fundamental solution in its closed-analytical form. In Sect. 3, the attention has been focused on the nonlocality of the Hamiltonian of the spinless Salpeter equation and the consequent effects on its solution. It has been presented an application of the theory of evolution operator to define a recursive solution for the spinless Salpeter equation which allows to highlight the nonlocal nature via recursive series. The result shows the presence of a *space-memory*, or in other words the presence of a regular repeating spot-like structure in the light-cone which tends to fade as time elapses. Finally a comment on another way to deal with nonlocality has been presented exploiting the series expansion of the Hamiltonian in the spinless Salpeter equation. This second approach is based on the so-called Pearcey equation, a new equation introduced in [25–29] for describing what happen between the two theories: the classical quantum mechanics ruled by the Schrödinger equation and the relativistic quantum mechanics.

The square-modulus of the solutions in the Pearcey equation, and in particular of its fundamental solution, illustrates the *space-memory* hinted in the recursive series.

2 Solutions of the Salpeter Equation

The nonlocal nature of (3) makes it difficult to deal with directly in the coordinate space. It is then usually approached in the momentum space

$$i\hbar\partial_t\tilde{\psi}(p,t) = \sqrt{m^2c^4 + c^2p^2}\;\tilde{\psi}(p,t), \quad \tilde{\psi}(p,0) = \tilde{\psi}_0(p), \tag{4}$$

whose solution is

$$\tilde{\psi}(p,t) = e^{-\frac{it}{\hbar}\sqrt{m^2c^4+c^2p^2}}\tilde{\psi}_0(p). \tag{5}$$

Here $\tilde{\psi}(p,t)$ means the momentum wave function solution of the Salpeter equation. By Fourier transform it is possible to define $\psi(x,t)$:

$$\psi(x,t) = \frac{1}{\sqrt{2\pi\hbar}}\int_{-\infty}^{+\infty} e^{\frac{ipx}{\hbar}} e^{-\frac{it}{\hbar}\sqrt{m^2c^4+c^2p^2}}\tilde{\psi}_0(p)\,dp. \tag{6}$$

The same solution can be obtained considering the convolution of an initial condition with the fundamental solution $S(x,t)$ of the Salpeter equation that corresponds to a δ-function input, i.e. $\psi_0(x) = \delta(x)$:

$$\psi(x,t) = \int_{-\infty}^{+\infty} \frac{mc^2 t}{\pi\hbar} \frac{K_1\left(\frac{imc}{\hbar}\sqrt{c^2t^2-(x-x')^2}\right)}{\sqrt{c^2t^2-(x-x')^2}} \psi_0(x')dx'. \tag{7}$$

In the above equation, K_1 is the modified Bessel function of the second kind of first order, also known as the McDonald function [30].

3 A Recursive Solution: Nonlocality via Laplace Transform

The Hamiltonian (2) for $V(x,t) = 0$ is a nonlocal operator whose natural scale is inversely proportional to the mass and it is given by the (reduced) Compton wavelength $\lambda_C = \frac{\hbar}{mc}$. The Compton wavelength represents the cutoff between the quantum and the quantum field theory: below its value, the concept of single particle is no more applicable. This justifies the normalization used to define the following dimensionless variables:

$$\xi = \frac{x}{\lambda_C}, \qquad \tau = \frac{ct}{\lambda_C} \quad \text{and} \quad \kappa = \frac{p}{\hbar}\lambda_C, \tag{8}$$

which allows not only to simplify the analysis but also to bridge formally quantum mechanics and optics [25–29]. Accordingly, in dimensionless variables (8) the initial value problem (3) writes as

$$i\,\partial_\tau \psi(\xi,\tau) = \sqrt{1-\partial_\xi^2}\ \psi(\xi,\tau), \quad \psi(\xi,0) = \psi_0(\xi), \tag{9}$$

and its formal solution is then

$$\psi(\xi,\tau) = \frac{i\tau}{\pi} \int_{-\infty}^{+\infty} \frac{K_1(\sqrt{(\xi-\xi')^2-\tau^2})}{\sqrt{(\xi-\xi')^2-\tau^2}} \psi_0(\xi')d\xi'. \tag{10}$$

To illustrate the nonlocal nature of the Hamiltonian operator in (9) and in particular to emphasize its influence on the evolution of an initial input from a mathematical point of view, it is interesting to deal with the initial value problem using another mathematical approach based on the Laplace transform method [31–35]. This technique is an effective alternative method that allows us to treat fractional operators as the Hamiltonian of the Salpeter equation revealing its nonlocal nature via recursive series.

To this end, a key notion is the following Laplace-like identity [30]

$$\frac{1}{\sqrt{A^2+1}} = \int_0^\infty e^{-tA} J_0(t)dt. \tag{11}$$

In fact, once the parameter A in Eq. (11) gains the *status* of operator

$$\frac{1}{\sqrt{\hat{A}^2+1}} f(x) = \int_0^\infty e^{-t\hat{A}} J_0(t) f(x) dt, \tag{12}$$

where $f(x)$ is a given x-dependent differentiable function and J_0 is the Bessel function of the first kind of first order [30], it is possible to apply the theory of the evolution operator shown in [31, 32] and define the operator $\hat{A}$ to obtain a recursive series solution highlighting the nonlocal nature embedded in the Hamiltonian of the spinless Salpeter equation.

The formal structure of Eq. (12) is extremely attractive since it can be applied to the initial value problem (9) after rearranged it as:

$$i\frac{\partial}{\partial\tau}\psi(\xi,\tau) = \left(1-\frac{\partial^2}{\partial\xi^2}\right)\frac{1}{\sqrt{\left(1-\frac{\partial^2}{\partial\xi^2}\right)}}\psi(\xi,\tau). \tag{13}$$

Moreover, by definition

$$\bar{\psi}(\xi,\tau) = \frac{1}{\sqrt{1-\frac{\partial^2}{\partial\xi^2}}}\psi(\xi,\tau), \tag{14}$$

the initial value problem (13) can be written as

$$i\frac{\partial}{\partial\tau}\psi(\xi,\tau) = \left(1-\frac{\partial^2}{\partial\xi^2}\right)\bar{\psi}(\xi,\tau). \tag{15}$$

Assuming $\hat{A} = i\frac{\partial}{\partial\xi}$, $\bar{\psi}$ is immediately defined

$$\bar{\psi}(\xi,\tau) = \int_0^\infty J_0(y)\psi(\xi - iy,\tau)dy \quad \psi(\xi,0) = \psi_0(\xi), \tag{16}$$

where J_0 is a modified Bessel function of the first kind [30]. Then the solution of (16) reads

$$\psi(\xi,\tau) = \sum_{n=0}^{\infty}\frac{(i\tau)^n}{n!}\psi_n(\xi). \tag{17}$$

The initial condition $\psi(\xi,0)$ is the zero-order term, whereas the nth-term $\psi_n(\xi)$ can be defined in recursive way

$$\psi_n(\xi) = \left(1-\frac{\partial^2}{\partial\xi^2}\right)\int_0^{+\infty} J_0(y)\psi_{n-1}(\xi - iy)dy. \tag{18}$$

This solution expressed as a recursive series is a proof that the causality (and the light-cone structure) is not violated by the nonlocality of the Hamiltonian. Each solution term, ψ_n, is determined by recursion on the values assumed previously by ψ_{n-1}, till the evolution of the equation has been completely described. Mathematically, the recursive solution (16) allows to find approximations when the exact solution is not yet known and it is a common procedure to use it to evaluate the evolution equations with memory terms. In fact the properties of memory and causality induce a description by means of recursive equations and read-out maps involving input state and output variables.

In this work, the term "memory" referred to the kernel of the convolution risks to be too exotic since the convoluted variable is ξ, i.e. the dimensionless variable related with the space. The recursive solution of the Salpeter equation given by (18) allows an interesting comparison with the numerical solution and the solution defined via closed-analytical expression.

Here below, it is considered an example of the application of the previous procedure where the initial condition is the Bessel function of the second kind of first order:

$$\psi_0^K(\xi) = \left(\frac{1}{\pi}\right) \frac{K_1(\sqrt{1+\xi^2})}{\sqrt{1+\xi^2}}. \tag{19}$$

Replacing the above initial condition in Eq. (10) the solution is given by

$$\psi^K(\xi,\tau) = \left(\frac{1+i\tau}{\pi}\right) \frac{K_1[\sqrt{(1+i\tau)^2+\xi^2}\,]}{\sqrt{(1+i\tau)^2+\xi^2}}, \tag{20}$$

which expresses the free-evolution of the McDonald initial condition under the spinless Salpeter equation.

The choice of this initial condition is rooted on the fact that it generates a closed-analytical expression for the solution, so it means that the recursive series defined in Eq. (16) can be summed. Let us consider Eq. (18) to define the terms of the series. The first term reads:

$$\begin{aligned} \psi_1(\xi) &= \left(1 - \frac{\partial^2}{\partial\xi^2}\right) \int_0^{+\infty} J_0(y)\psi_0(\xi - iy)dy \\ &= \left(1 - \frac{\partial^2}{\partial\xi^2}\right) \int_0^{+\infty} J_0(y) \frac{K_1\sqrt{1+(\xi-iy)^2}}{\pi\sqrt{1+(\xi-iy)^2}} dy. \end{aligned} \tag{21}$$

The convolution in Eq. (21) can be solved applying formula **3.914.1** in [36]. Consequently, one has

$$\psi_1(\xi) = \frac{1}{\pi}\Big[- \frac{K_0(\sqrt{1+\xi^2})}{(1+\xi^2)} + \frac{(\xi^2-1)K_1(\sqrt{1+\xi^2})}{\sqrt{(1+\xi^2)^3}}\Big] \tag{22}$$

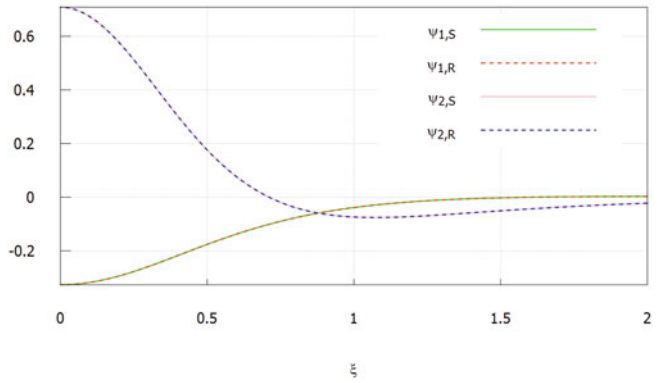


Fig. 1 Comparison between the first and the second order terms at $\tau = 1$ in the series expansion of the closed-analytical expression ($\psi_{1,S}$, $\psi_{2,S}$, respectively) and the recursive series ($\psi_{1,R}$, $\psi_{2,R}$, respectively)

and then, by recursive relation (18) the second term is

$$\psi_2(\xi) = \frac{1}{\pi}\Big[\frac{(1-3\xi^2)K_0(\sqrt{1+\xi^2})}{(1+\xi^2)^2} + \frac{(3-5\xi^2)K_1(\sqrt{1+\xi^2})}{\sqrt{(1+\xi^2)^5}}\Big]. \tag{23}$$

It is possible to check the correctness of the result obtained for the first and the second order considering the series expansion of the solution in Eq. (20) with respect to τ. Each term of the series expansion should correspond to the corresponding order-term in the recursive series. In Fig. 1, a direct comparison between the first order and the second order of both the series has been illustrated. A visual effect of the *space*-memory can be appreciated considering the fundamental solution of the Pearcey equation where the point-like structure is repeating within the light-cone on a regular lattice with descending intensity. [25–29]:

$$i\frac{\partial}{\partial\tau}\phi(\xi,\tau) = \left(1 - \frac{1}{2}\frac{\partial^2}{\partial\xi^2} - \frac{1}{8}\frac{\partial^4}{\partial\xi^4}\right)\phi(\xi,\tau), \qquad \phi(\xi,0) = \phi_0(\xi), \tag{24}$$

a quasi-relativistic evolution equation ruled by a Hamiltonian which is the fourth order series expansion of the Hamiltonian of the spinless Salpeter equation. In fact the squared modulus of the fundamental solution (see Fig. 2) presents a spot-like structure spreading always inside the light-cone structure reproducing a lattice embedding a sort of *space-memory* of the initial input which is fading as time elapses.

Acknowledgments The author was funded by the Polish National Agency for Academic Exchange NAWA project: Program im. Iwanowskiej PPN/IWA/2018/1/00098 and was supported

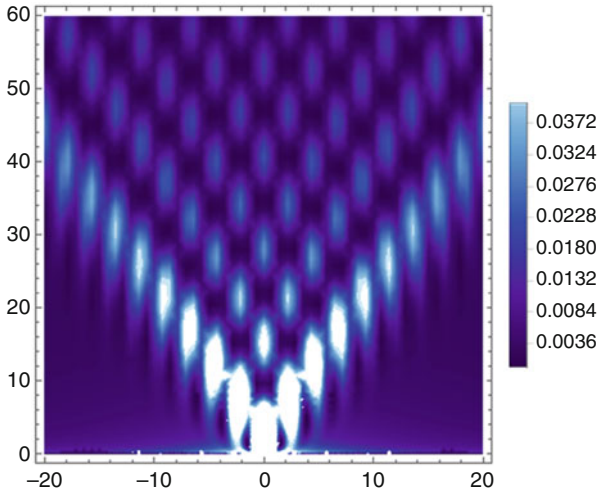


Fig. 2 (ξ, τ)-contour plot of the fundamental solution of the Pearcey equation

by the NCN research project OPUS 12 no. UMO-2016/23/B/ST3/01714. I would like to thank the two anonymous reviewers for their suggestions and comments.

References

1. Y. Nambu, Force potentials in quantum field theory. Prog. Theor. Phys. **5**, 614–633 (1950)
2. E.E. Salpeter, H.A. Bethe, A relativistic equation for bound-state problems. Phys. Rev. **84**, 1232–1242 (1951)
3. M. Gell-Mann, F. Law, Bound states in quantum theory. Phys. Rev. **84**, 350–354 (1951)
4. E.E. Salpeter, Mass corrections to the fine structure of hydrogen-like atoms. Phys. Rev. **87**, 328–343 (1952)
5. J.D. Bjorken, S.D. Drell, *Relativistic Quantum Mechanics* (Mc-Graw Hill, New York, 1964)
6. W. Greiner, J. Reinhardt, *Quantum Electrodynamics* (Springer, Berlin, 1994)
7. G.C. Wick, Properties of Bethe-Salpeter wave equations. Phys. Rev. **96**, 1124–1134 (1954)
8. J. Sucher, Relativistic invariance and the square-root Klein-Gordon equation. J. Math. Phys. **4**, 17–23 (1963)
9. N. Nakanishi, A general survey of the theory of the Bethe-Salpeter equation. Progr. Theor. Phys. Suppl. **45**, 1–81 (1969)
10. C. Lammerzahl, The pseudodifferential operator square root of the Klein-Gordon equation. J. Math. Phys. **34**, 3918–3932 (1993)
11. P.J. Olver, *Introduction to Partial Differential Equations* (Springer, New York, NY, 2014), Chap. 8, pp. 291–338
12. W. Lucha, F.F. Schöberl, *All around the spinless Salpeter equation*, arXiv preprint hep-ph/9410221 (1994)
13. W. Lucha, F.F. Schöberl, *Bound states by the Salpeter equation*, arXiv prep-rint hep-ph/9812526 (1998)
14. D. Babusci, G. Dattoli, M. Quattromini, Relativistic equations with fractional and pseudodifferential operators. Phys. Rev. A **83**(6), 062109 (2011)

15. G. Dattoli, E. Sabia, K. Górska, A. Horzela, K.A. Penson, Relativistic wave equations: an operational approach. J. Phys. A. Math. Theor. **48**, 125203 (2015)
16. K. Kowalski, J. Rembielińnski, The relativistic massless harmonic oscillator. Phys. Rev. A **81**, 012118 (2010)
17. K. Kowalski, J. Rembielińnski, The Salpeter equation and probability current in the relativistic Hamiltonian quantum mechanics. Phys. Rev. A **84**, 012108 (2011)
18. P. Garbaczewski, V. Stephanovich, Lévy flights and nonlocal quantum dynamics. J. Math. Phys. **54**(7), 072103 (2013)
19. T. Frederico, E. Pace, B. Pasquini, G. Salmè, Generalized parton distributions of the pion in a covariant Bethe-Salpeter model and light-front models. Nucl. Phys. Proc. Suppl. **199**, 264–269 (2010)
20. T. Frederico, G. Salmè, Projecting the Bethe-Salpeter equation onto the light-front and back: a short review. Few Body Syst. **49**, 163–175 (2011)
21. T. Frederico, G. Salmè, M. Viviani, Quantitative studies of the homogeneous Bethe-Salpeter equation in Minkowski space. Phys. Rev. D **89**, 016010 (2014)
22. G. Salmè, T. Frederico, M. Viviani, Quantitative studies of the homogeneous Bethe-Salpeter equation in Minkowski space. Few Body Syst. **55**, 693–696 (2014)
23. Y. Chargui, A. Dhabbi, L. Chetouni, A. Trabelsi, Solution of the spinless Salpeter equation with a time-dependent linear potential. Few-body Syst. **55**, 1233–1243 (2014)
24. D.J. Cirilo-Lombardo, Relativistic dynamics, Green function and pseudodifferential operators. J. Math. Phys. **57**(6), 063503 (2016)
25. A. Lattanzi, Thesis, University of Roma Tre (I) (2016)
26. A. Torre, A. Lattanzi, D. Levi, Time-dependent free-particle Salpeter equation: numerical and asymptotic analysis in the light of the fundamental solution. Ann. Phys. (Berlin) **529**, 1600231 (2017)
27. A. Torre, A. Lattanzi, D. Levi, Time-dependent free-particle Salpeter equation: features of the solutions, in *Quantum Theory and Symmetries with Lie Theory and Its Applications in Physics Volume 2*, ed. by V. Dobrev. Springer Proceeding in Mathematics & Statistics, vol. 255 (Singapore, Springer, 2018), pp. 297–307
28. A. Lattanzi, D. Levi, A. Torre, The missing piece: a new relativistic wave equation, the Pearcey equation. J. Phys. Conf. Ser. **1194**(1), 012065 (2019). IOP Publishing
29. A. Lattanzi, D. Levi, A. Torre, Evolution equations in a nutshell, in *Conference Proceedings CSPM 2018* (2019)
30. M. Abramowitz, I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th edn. (Dover Publication, Mineola, NY, 1970)
31. G. Dattoli, D. Babusci, E. Sabia, S. Licciardi, *Mathematical Methods for Physicists* (World Scientific, Singapore, 2019)
32. G. Dattoli, M. Quattromini, A. Torre, Miscellaneous results on the theory of evolution operators and generalized transforms. Nuovo Cimento Soc. Ital. Fis. Sezione **B 114**, 693–708 (1999)
33. G. Dattoli, D. Levi, Exponential operators and generalized difference equations, Nuovo Cimento Soc. Ital. Fis. Sezione **B 115**(6), 653–662 (2000)
34. G. Dattoli, B. Germano, M.R. Martinelli, P.E. Ricci, A novel theory of Legendre polynomials. Math. Comput. Model. **54**(1–2), 80–87 (2011)
35. G. Dattoli, P.E. Ricci, D. Sacchetti, Generalized shift operators and pseudo-polynomials of fractional order. Appl. Math. Comput. **141**(1), 215–224 (2013)
36. I.S. Gradshteyn, I.M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York, 2014)

A New Approach to Analysis of 2D Higher Order Quantum Superintegrable Systems

Bjorn K. Berntson, Ian Marquette, and Willard Miller, Jr.

Abstract We revise a method by Kalnins et al. (J Phys A Math Theor 43:265205, 2010) for constructing a canonical form for symmetry operators of arbitrary order for the Schrödinger eigenvalue equation $H\Psi \equiv (\Delta_2 + V)\Psi = E\Psi$ on any 2D Riemannian manifold, real or complex, that admits a separation of variables in some orthogonal coordinate system. Most of this paper is devoted to describing the method. Details will be provided elsewhere. As examples we revisit the Tremblay and Winternitz derivation of the Painlevé VI potential for a third order superintegrable flat space system that separates in polar coordinates and, as new results, we show that the Painlevé VI potential also appears for a third order superintegrable system on the 2-sphere that separates in spherical coordinates, as well as a third order superintegrable system on the 2-hyperboloid that separates in spherical coordinates and one that separates in horocyclic coordinates. The purpose of this project is to develop tools for analysis and classification of higher order superintegrable systems on any 2D Riemannian space, not just Euclidean space.

Keywords Quantum superintegrable systems · Painlevé VI equation · Weierstrass equation

B. K. Berntson
Department of Mathematics, KTH Royal Institute of Technology, Stockholm, Sweden
e-mail: bbernts@kth.se

I. Marquette
School of Mathematics and Physics, The University of Queensland, Brisbane, QLD, Australia
e-mail: i.marquette@uq.edu.au; https://smp.uq.edu.au/profile/211/ian-marquette

W. Miller, Jr. (✉)
School of Mathematics, University of Minnesota, Minneapolis, MN, USA
e-mail: mille003@umn.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_10

1 Introduction

In the paper [1] the authors constructed a canonical form for symmetry operators of any order in 2D and used it to give the first proof of the superintegrability of the quantum Tremblay, Turbiner, and Winternitz (TTW) system [2] in polar coordinates, for all rational values of the parameter k. In the original method the various potentials were given and the problem was the construction of higher order symmetry operators that would verify superintegrability. The method was highly algebraic and required the solution of systems of difference equations on a lattice. Here, we consider an arbitrary space admitting a separation in some orthogonal coordinate system (hence admitting a 2nd order symmetry operator), and search for all potentials V for which the Schrödinger equation admits an additional independent symmetry operator of order higher than 2. Now the problem reduces to solving a system of partial differential equations.

We give a brief introduction to the method and then specialize it to third order superintegrable systems where we treat a few examples. We revisit the Tremblay and Winternitz derivation of the Painlevé VI potential for a third order superintegrable flat space system that separates in polar coordinates [3], and we show among other new results that the Painlevé VI potential also appears for a third order superintegrable system on the 2-sphere that separates in spherical coordinates, as well as a third order superintegrable system on the 2-hyperboloid that separates in spherical coordinates.

2 The Canonical Form for a Symmetry Operator

We consider a Schrödinger equation on a 2D real or complex Riemannian manifold with Laplace–Beltrami operator Δ_2 and potential V:

$$H\Psi \equiv \left(-\frac{\hbar^2}{2}\Delta_2 + V\right)\Psi = E\Psi \tag{1}$$

that also admits an orthogonal separation of variables. If $\{u_1, u_2\}$ is the orthogonal separable coordinate system, the corresponding Schrödinger operator can always be put in the form

$$\begin{aligned} H &= -\frac{\hbar^2}{2}\Delta_2 + V(u_1, u_2) \\ &= \frac{1}{f_1(u_1) + f_2(u_2)}\left(-\frac{\hbar^2}{2}\partial_{u_1}^2 - \frac{\hbar^2}{2}\partial_{u_2}^2 + v_1(u_1) + v_2(u_2)\right) \end{aligned} \tag{2}$$

and, due to the separability, there is the second-order symmetry operator

$$L_2 = \frac{f_2(u_2)}{f_1(u_1)+f_2(u_2)}\left(-\frac{\hbar^2}{2}\partial_{u_1}^2 + v_1(u_1)\right) - \frac{f_1(u_1)}{f_1(u_1)+f_2(u_2)}\left(-\frac{\hbar^2}{2}\partial_{u_2}^2 + v_2(u_2)\right),$$

i.e., $[H, L_2] = 0$. We look for a partial differential symmetry operator of arbitrary order $\tilde{L}(H, L_2, u_1, u_2)$ that satisfies

$$[H, \tilde{L}] = 0. \tag{3}$$

We require that the symmetry operator takes the standard form

$$\tilde{L} = \sum_{j,k}\Big(A^{j,k}(u_1,u_2)\partial_{u_1u_2} - B^{j,k}(u_1,u_2)\partial_{u_1} - C^{j,k}(u_1,u_2)\partial_{u_2} + D^{j,k}(u_1,u_2)\Big) H^j L_2^k. \tag{4}$$

This can always be done. Note that if the formal operator $\tilde{L}$ contained partial derivatives in u_1 and u_2 of orders $\geqslant 2$, we could rearrange terms to achieve the unique standard form (4).

Details of the derivation can be found in [1].

Note that condition (4) makes sense, at least formally, for infinite order differential equations. Indeed, one can consider H, L_2 as parameters in these equations. Then once $\tilde{L}$ is expanded as a power series in these parameters, the terms are reordered so that the powers of the parameters are on the right, before they are replaced by explicit differential operators. Of course (4) is defined rigorously for finite order symmetry operators.

In this view we can write

$$\tilde{L}(H, L_2, u_1, u_2) = A(u_1,u_2)\partial_{u_1u_2} - B(u_1,u_2)\partial_{u_1} - C(u_1,u_2)\partial_{u_2} + D(u_1,u_2), \tag{5}$$

and consider $\tilde{L}$ as an at most second-order differential operator in u_1, u_2 that is analytic in the parameters H, L_2. Then the above system of equations can be written in the more compact form

$$\partial_{u_1}^2 A + \partial_{u_2}^2 A - 2\partial_{u_2} B - 2\partial_{u_1} C = 0, \tag{6}$$

$$\frac{\hbar^2}{2}(\partial_{u_1}^2 B + \partial_{u_2}^2 B) - 2\partial_{u_2} A\, v_2 - \hbar^2\partial_{u_1} D - A v_2' + (2\partial_{u_2} A\, f_2 + A f_2')H - 2\partial_{u_2} A\, L_2 = 0, \tag{7}$$

$$\frac{\hbar^2}{2}(\partial_{u_1}^2 C + \partial_{u_2}^2 C) - 2\partial_{u_1} A v_1 - \hbar^2\partial_{u_2} D - A v_1' + (2\partial_{u_1} A\, f_1 + A f_1')H + 2\partial_{u_1} A\, L_2 = 0, \tag{8}$$

$$-\frac{\hbar^2}{2}(\partial_{u_1}^2 D + \partial_{u_2}^2 D) + 2\partial_{u_1} B\, v_1 + 2\partial_{u_2} C\, v_2 + B v_1' + C v_2' \tag{9}$$
$$-(2\partial_{u_1} B\, f_1 + 2\partial_{u_2} C\, f_2 + B f_1' + C f_2')H + (-2\partial_{u_1} B + 2\partial_{u_2} C)\, L_2 = 0.$$

We can view (6) as an equation for A, B, C and (7), (8) as the defining equations for $\partial_{u_1} D, \partial_{u_2} D$. Then $\tilde{L}$ is $\hat{L}$ with the terms in H and L_2 interpreted as (4) and considered as partial differential operators.

We can simplify this system by noting that there are two functions $F(u_1, u_2, H, L_2), G(u_1, u_2, H, L_2)$ such that (6) is satisfied by

$$A = F, \qquad B = \frac{1}{2}\partial_{u_2} F + \partial_{u_1} G, \qquad C = \frac{1}{2}\partial_{u_1} F - \partial_{u_2} G. \tag{10}$$

Then the integrability condition for (7), (8) is (with the shorthand $\partial_{u_j} F = F_j$, $\partial_{u_j}\partial_{u_\ell} F = F_{j\ell}$, etc., for F and G),

$$-\hbar^2 G_{1222} - \frac{1}{4}\hbar^2 F_{2222} + 2F_{22}(v_2 - f_2 H + L_2) + 3F_2(v_2' - f_2' H) + F(v_2'' - f_2'' H) =$$
$$\hbar^2 G_{1112} - \frac{1}{4}\hbar^2 F_{1111} + 2F_{11}(v_1 - f_1 H - L_2) + 3F_1(v_1' - f_1' H) + F(v_1'' - f_1'' H), \tag{11}$$

and Eq. (9) becomes

$$\frac{1}{4}\hbar^2 F_{1112} - 2F_{12}(v_1 - f_1 H) - F_1(v_2' - f_2' H) + \frac{1}{4}\hbar^2 G_{1111}$$
$$- 2G_{11}(v_1 - f_1 H - L_2) - G_1(v_1' - f_1' H)$$
$$= -\frac{1}{4}\hbar^2 F_{1222} + 2F_{12}(v_2 - f_2 H) \tag{12}$$
$$+ F_2(v_1' - f_1' H) + \frac{1}{4}\hbar^2 G_{2222} - 2G_{22}(v_2 - f_2 H + L_2) - G_2(v_2' - f_2' H).$$

We remark that any solution of (11), (12) with A, B, C not identically 0 corresponds to a symmetry operator that does not commute with L_2, hence is algebraically independent of the symmetries H, L_2.

3 Third Order Superintegrability

To illustrate how Eqs. (11) and (12) can be used to find potentials for superintegrable systems, we provide detailed derivations of the determining equations for third order superintegrability. First we note that the most general third order operator must be of the form (4) with

$$A = A^0(x, y), \quad B = B^0(x, y) + B^H(x, y)H + B^L(x, y)L,$$
$$C = C^0(x, y) + C^H(x, y)H + C^L(x, y)L,$$
$$D = D^0(x, y) + D^H(x, y)H + D^L(x, y)L,$$

or, in view of (10),

$$F(x, y) = F^0(x, y), \quad G(x, y) = G^0(x, y) + G^H(x, y)H + G^L(x, y)L. \tag{13}$$

Substituting (13) into (11), (12) and noting that the coefficients of independent powers of H and L in these expressions must vanish, we obtain nine equations (the first three from (11) and the next six from (12)):

$$\begin{aligned}
0 &= -6v_1' F_1^0 + 6v_2' F_2^0 - 4v_1 F_{11}^0 + 4v_2 F_{22}^0 - 2\hbar^2 G_{1112}^0 - 2\hbar^2 G_{1222}^0 \\
&\quad +2F^0 v_2'' - 2F^0 v_1'', \\
0 &= F_{11}^0 + F_{22}^0, \\
0 &= -\hbar^2 G_{1112}^H - \hbar^2 G_{1222}^H + 3f_1' F_1^0 - 3f_2' F_2^0 + 2f_1 F_{11}^0 - 2f_2 F_{22}^0 - F^0 f_2'' + F^0 f_1'', \\
0 &= v_2' F_1^0 + v_1' F_2^0 + v_1' G_1^0 - v_2' G_2^0 + 2F_{12}^0 v_2 + 2F_{12}^0 v_1 + 2v_1 G_{11}^0 - 2v_2 G_{22}^0 \\
&\quad -\frac{1}{4}\hbar^2 G_{1111}^0 + \frac{1}{4}\hbar^2 G_{2222}^0, \\
0 &= v_1' G_1^L - v_2' G_2^L + 2v_1 G_{11}^L - 2G_{11}^0 - 2v_2 G_{22}^L - 2G_{22}^0, \\
0 &= G_{11}^L + G_{22}^L, \\
0 &= -f_2' F_1^0 - f_1' F_2^0 + v_1' G_1^H - f_1' G_1^0 - v_2' G_2^H + f_2' G_2^0 - 2F_{12}^0 f_2 - 2F_{12}^0 f_1 \\
&\quad +2v_1 G_{11}^H - 2f_1 G_{11}^0 - 2v_2 G_{22}^H + 2f_2 G_{22}^0 - \frac{1}{4}\hbar^2 G_{1111}^H + \frac{1}{4}\hbar^2 G_{2222}^H, \\
0 &= -f_1' G_1^L + f_2' G_2^L + 2f_2 G_{22}^L - 2f_1 G_{11}^L - 2G_{11}^H - 2G_{22}^H, \\
0 &= -f_1' G_1^H + f_2' G_2^H + 2f_2 G_{22}^H - 2f_1 G_{11}^H.
\end{aligned}$$

4 Some Examples (Mostly New)

We are particularly interested in potentials with nonlinear defining equations. First, we show that we get the result of Tremblay and Winternitz [3] that the quantum system separating in polar coordinates in 2D Euclidean space admits potentials that are expressed in terms of the sixth Painlevé transcendent or in terms of the

Weierstrass elliptic function. To do this we must put the system in the canonical form (2). The separable polar coordinates are $(x, y) = (r\cos(\theta), r\sin(\theta))$. For the canonical form we use the coordinates $\{u_1, u_2\}$, where $r = \exp(u_1)$, $\theta = u_2$. Thus, $f_1(u_1) = \exp(2u_1)$ and $f_2(u_2) = 0$. We know that these extreme potentials can appear only if the potential depends on the angular variable alone, so we set $v_1(u_1) = 0$. Since we want only systems that satisfy nonlinear equations alone, whenever an explicit linear equation for the potential appears, we require that it vanish identically. We have the freedom to replace the angular variable u_2 by $u_2 + c$ for some real constant c to simplify the expressions, Also we can rescale the answer.

We obtain a solution

$$F^0 = -4\hbar^2 \exp(-u_1)\sin(u_2), \quad G^L = -8\exp(-u_1)\cos(u_2) + a_4 u_2 + a_3,$$
$$G^0 = -U_1(u_2)\exp(-u_1) + U_2(u_2), \quad G^H = a_5,$$

subject to the conditions

$$0 = a_4 \frac{dv_2}{du_2} + 2\frac{d^2U_2}{du_2^2}, \tag{14}$$

$$0 = \hbar^2 \frac{d^4U_2}{du_2^4} + 4a_4 \frac{dv_2}{du_2} v_2 - 4\frac{dv_2}{du_2}\frac{dU_2}{du_2}, \tag{15}$$

$$0 = 8v_2\cos(u_2) + 4\frac{dv_2}{du_2}\sin(u_2) - \frac{d^2U_1}{du_2^2} - U_1, \tag{16}$$

$$0 = \frac{dv_2}{du_2}\frac{dU_1}{du_2} - \hbar^2\frac{d^3v_2}{du_2^3}\sin(u_2) - 4\hbar^2\frac{d^2v_2}{du_2^2}\cos(u_2) \tag{17}$$
$$+2\sin(u_2)(5\hbar^2 + 4v_2)\frac{dv_2}{du_2} + 2v_2\left(6\hbar^2\cos(u_2) + 8v_2\cos(u_2) - U_1\right).$$

There are basically two cases to consider:

1. $a_4 = 0$.

Then condition (14) says that U_2 is linear in u_2. Thus condition (15) is a linear equation for $v_2(u_2)$ which must vanish. Then condition (16) can be solved for $U_1(y)$ and the result substituted into condition (17) to obtain an equation for $v_2(u_2)$. After some manipulation we obtain an equation characterizing Painlevé VI, in agreement with [3, Eq. (4.27)]:

$$\hbar^2\left(\sin(u_2)\frac{d^4W}{du_2^4} + 4\cos(u_2)\frac{d^3W}{du_2^3} - 6\sin(u_2)\frac{d^2W}{du_2^2} - 4\cos(u_2)\frac{dW}{du_2}\right) \tag{18}$$

$$-12\sin(u_2)\frac{dW}{du_2}\frac{d^2W}{du_2^2}-4\cos(u_2)W\frac{d^2W}{du_2^2}-4(\beta_1\sin(u_2)-\beta_2\cos(u_2))\frac{d^2W}{du_2^2}$$
$$-16\cos(u_2)\left(\frac{dW}{du_2}\right)^2+8\sin(u_2)W\frac{dW}{du_2}-8(\beta_1\cos(u_2)+\beta_2\sin(u_2))\frac{dW}{du_2}=0$$

Here $v_2(u_2)=\frac{dW(u_2)}{du_2}$.

2. $a_4\neq 0$.

Solving condition (14) for $v_2(u_2)$ and substituting the result and (14) into (15) we obtain the equation that characterizes the Weierstrass $\wp$-function (in fact it is a translated and rescaled version):

$$\hbar^2\frac{d^3v_2}{du_2^3}+12\frac{dv_2}{du_2}v_2-8a_1\frac{dv_2}{du_2}=0. \tag{19}$$

Thus $v_2(u_2)=\wp(\hbar u_2;g_2,g_3)+2a_1/3$, where g_2 and g_3 are arbitrary constants. As shown in [3] this solution is subject to the compatibility conditions (16) and (17), which leads to a complicated nonlinear differential equation for $v_2(u_2)$.

With this verification out of the way, we consider the analogous system on the 2-sphere, separable in spherical coordinates. Here $s_1=\sin(\theta)\cos(\phi)$, $s_2=\sin(\theta)\sin(\phi)$, $s_3=\cos(\theta)$ with $s_1^2+s_2^2+s_3^2=1$. This system is in canonical form with coordinates u_1,u_2 where

$$\sin(\theta)=(\cosh(u_1))^{-1},\ \phi=u_2,\ f_1(u_1)=(\cosh(u_1))^{-2},\ f_2(u_2)=0. \tag{20}$$

As before we look for solutions such that $v_1(u_1)=0$ and v_2 satisfies a nonlinear equation only.

The computation is very similar to that for the Euclidean space example. We obtain the solution

$$F^0=-4\hbar^2\cosh(u_1)\sin(u_2),\ G^L=8\sinh(u_1)\cos(u_2)+a_4u_2+a_3,$$
$$G^0=\sinh(u_1)\,U_1(u_2)+U_2(u_2),\ G^H=a_5,$$

subject to the conditions (14)–(17), exactly the same as for Euclidean space. Thus the system on the 2-sphere also admits Painlevé VI and special Weierstrass potentials for third order superintegrability. It is clear from these results that these systems in Euclidean space can be obtained as Bôcher contractions, [4, chapter 15], of the corresponding systems on the 2-sphere.

Next we consider spherical coordinates on the hyperboloid $s_1^2-s_2^2-s_3^2=1$,

$$s_1=\cosh(x),\ s_2=\sinh(x)\cos(\phi),\ s_3=\sinh(x)\sin(\phi).$$

For the canonical form we find

$$\tanh\left(\frac{u_1}{2}\right) = \exp(x),\ u_2 = \phi, \quad f_1(u_1) = (\sinh(u_1))^{-2},\ f_2(u_2) = 0,$$

and we look for solutions such that $v_1(u_1) = 0$ and $v_2(u_2)$ satisfies only a nonlinear equation. We obtain the solution

$$F^0 = -4\hbar^2 \sinh(u_1)\sin(u_2),\ G^L = 8\cosh(u_1)\cos(u_2) + a_4u_2 + a_3,$$
$$G^0 = \cosh(u_1)\ U_1(u_2) + U_2(u_2),\ G^H = a_5,$$

subject to the conditions (14)–(17), again exactly the same as for flat space. Thus the system on the 2-hyperboloid admits Painlevé VI and special Weierstrass potentials for third order superintegrability.

For our next example we consider horocyclic coordinates $\{u_1, u_2\}$ on the hyperboloid $s_1^2 - s_2^2 - s_3^2 = 1$, e.g. [4, section 7.7]:

$$s_1 = \frac{1}{2}\left(u_1 + \frac{u_2^2+1}{u_1}\right),\ s_2 = \frac{1}{2}\left(u_1 + \frac{u_2^2-1}{u_1}\right),\ s_3 = \frac{u_2}{u_1}. \tag{21}$$

These coordinates are separable and the canonical system is defined by $f_1(u_1) = 1/u_1^2$, $f_2(u_2) = 0$. We look for systems such that $v_1(u_1) = 0$, in analogy with our first three examples.

We obtain the solution

$$F^0 = -\frac{1}{2}a_8\hbar^2u_1, \quad G^L = \frac{u_1^2(a_8u_2 + a_9)}{2} - \frac{a_8u_2^3}{6} - \frac{a_9u_2^2}{2} + a_{10}u_2,$$
$$G^0 = \frac{u_1^2}{2}U_1(u_2) + U_2(u_2), \quad G^H = a_7,$$

subject to the conditions

$$0 = a_8\frac{dv_2}{du_2} + 2\frac{d^2U_1}{du_2^2}, \tag{22}$$

$$0 = \frac{1}{2}\hbar^2a_8\frac{d^3v_2}{du_2^3} - 4a_8\frac{dv_2}{du_2}v_2 + 4\frac{dv_2}{du_2}\frac{dU_1}{du_2}, \tag{23}$$

$$0 = (2a_{10} - 2a_9u_2 - a_8u_2^2)\frac{dv_2}{du_2} - 4(a_9 + a_8u_2)v_2 + 4U_1 + 4\frac{d^2U_2}{du_2^2}, \tag{24}$$

$$0 = 4u_2^2\frac{dv_2}{du_2} + 16u_2v_2^2 + 8a_9u_2\frac{dv_2}{du_2}v_2 + 16a_9v_2^2 - 8a_{10}\frac{dv_2}{du_2}v_2 \tag{25}$$
$$+\hbar^2a_8\frac{d^3v_2}{du_2^3} - 16v_2U_1 + 8\frac{dv_2}{du_2}\frac{dU_2}{du_2}.$$

There are again two basic cases here:

1. $a_8 = 0$.
Then conditions (22) and (23) say that U_1 is a constant: $U_1(u_2) = d_1$. Then condition (24) can be solved for $U_2(u_2)$ and the result substituted into condition (25) to obtain an equation for $v_2(u_2)$:

$$-4a_9\left(\frac{dW}{du_2}\right)^2 + \left((-3a_9u_2 + 3a_{10})\frac{d^2W}{du_2^2} + 4d_1\right)\frac{dW}{du_2}$$
$$+ (-a_9W + 2d_1u_2 - 2d_3)\frac{d^2W}{du_2^2}$$
$$+ \hbar^2 a_9\frac{d^3W}{du_2^3} - \frac{1}{4}\hbar^2(-a_9u_2 + a_{10})\frac{d^4W}{du_2^4} = 0, \text{ where } v_2(u_2) = \frac{dW(u_2)}{du_2}. \tag{26}$$

2. $a_8 \neq 0$.
Here we can solve (22) for $v_2(u_2)$ and substitute the result into (23) to obtain the equation

$$\hbar^2\frac{d^3v_2}{du_2^3} - 12v_2\frac{dv_2}{du_2} + q\frac{dv_2}{du_2} = 0. \tag{27}$$

Solutions of (27) are further subject to the requirement that a solution $U_2(u_2)$ of Eqs. (24) and (25) exists. Setting $v_2(u_2) = w(u_2) + q_1/12$ in (27) leads to

$$\hbar^2\frac{d^3w}{du_2^3} - 12w\frac{dw}{du_2} = 0 \tag{28}$$

and it follows that $v_2(u_2) = \wp(\hbar u_2; g_2, g_3) + q/12$, where g_2 and g_3 are arbitrary constants.

Acknowledgments We thank Pavel Winternitz for helpful discussions and Adrian Escobar for pointing out the relevance of the paper [1] to classification of third order superintegrable systems. W.M. was partially supported by a grant from the Simons Foundation (# 412351 to Willard Miller, Jr.). I.M. was supported by the Australian Research Council Discovery Grant DP160101376 and Future Fellowship FT180100099.

References

1. E.G. Kalnins, J.M. Kress, W.Jr. Miller, Superintegrability and higher order integrals for quantum systems. J. Phys. A Math. Theor. **43**, 265205 (2010)
2. F. Tremblay, V.A. Turbiner, P. Winternitz, An infinite family of solvable and integrable quantum systems on a plane. J. Phys. A Math. Theor. **42**, 242001 (2009)
3. F. Tremblay, P. Winternitz, Third order superintegrable systems separating in polar coordinates. J. Phys. **A43**, 175206 (2010)
4. E.G. Kalnins, J.M. Kress, W.Jr. Miller, *Separation of Variables and Superintegrability: The Symmetry of Solvable Systems* (Institute of Physics, London, 2018), ISBN: 978-0-7503-1314-8, http://iopscience.iop.org/book/978-0-7503-1314-8, e-book

Ladder Operators and Rational Extensions

David Gómez-Ullate, Yves Grandati, Zoé McIntyre, and Robert Milson

Abstract This note presents the classification of ladder operators corresponding to the class of rational extensions of the harmonic oscillator. We show that it is natural to endow the class of rational extensions and the corresponding intertwining operators with the structure of a category $\mathbb{REXT}$. The combinatorial data for this interpretation is realized as a functor $\mathbb{MD} \to \mathbb{REXT}$, where $\mathbb{MD}$ refers to the set of Maya diagrams appropriately endowed with categorical structure. Our formalism allows us to easily reproduce and extend earlier results on ladder operators.

Keywords Rational extensions · Ladder operators · Maya diagrams

1 Introduction

Supersymmetric quantum mechanics (SUSYQM) has proven to be a key technique in the construction of exactly solvable potentials and in the understanding of shape-invariance. The supersymmetric partners of the harmonic oscillators are known as rational extensions because the corresponding potentials have the form of a harmonic oscillator plus a rational term that vanishes at infinity.

D. Gómez-Ullate
Escuela Superior de Ingeniería, U. Cádiz, Puerto Real, Spain

Departamento de Física Teórica, U. Complutense, Madrid, Spain
e-mail: david.gomezullate@uca.es

Y. Grandati
L. Physique et Chimie Théoriques, U. de Lorraine, Metz, France
e-mail: yves.grandati@univ-lorraine.fr

Z. McIntyre
Department of Physics, McGill U., Montréal, QC, Canada
e-mail: zoe.mcintyre@mail.mcgill.ca

R. Milson (✉)
Department of Mathematics and Statistics, Dalhousie U., Halifax, NS, Canada
e-mail: rmilson@dal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_11

There has been some recent interest in rational extensions possessing ladder operators, which may be thought of as higher order analogues of the classical creation and annihilation operators. There are applications of such ladder operators to superintegrable systems [9, 10], rational solutions of Painlevé equations [11], and coherent states [7].

In this note we classify the ladder operators corresponding to the class of rational extensions of the harmonic oscillator. Rational extensions are naturally associated with combinatorial objects called Maya diagrams. We show that any two rational extensions are related by an intertwining relation. It therefore makes sense to endow both Maya diagrams and rational extensions with the structure of a category, and to interpret the relation Maya diagram $\mapsto$ rational extension as a functor between these categories. This approach allows us to classify ladder operators and syzygies of ladder operators, and thereby to generalize the results of [9, 10].

2 Maya Diagrams

A Maya diagram is a set of integers $M \subset \mathbb{Z}$ containing a finite number of positive integers, and excluding a finite number of negative integers. We visualize a Maya diagram as a horizontally extended sequence of $\boxdot$ and $\square$ symbols, with the filled symbol $\boxdot$ in position m indicating membership $m \in M$. The defining assumption now manifests as the condition that a Maya diagram begins with an infinite filled $\boxdot$ segment and terminates with an infinite empty $\square$ segment.

A Maya diagram may also be regarded as a strictly decreasing sequence of integers $m_1 > m_2 > \cdots$, subject to the constraint that $m_{i+1} = m_i - 1$ for i sufficiently large. It follows that there exists a unique integer σ, called the index of M, such that $m_i = -i + \sigma$ for i sufficiently large.

Let $\mathcal{M}$ denote the set of all Maya diagrams. The flip at position $k \in \mathbb{Z}$ is the involution $f_k : \mathcal{M} \to \mathcal{M}$ defined by

$$f_k : M \mapsto \begin{cases} M \cup \{k\}, & \text{if} \quad k \notin M, \\ M \setminus \{k\}, & \text{if} \quad k \in M. \end{cases} \qquad M \in \mathcal{M}. \tag{1}$$

In the first case, we say that the flip acts on M by a state-deleting transformation ($\square \to \boxdot$), and in the second case, by a state-adding transformation ($\boxdot \to \square$).

Let $\mathcal{Z}_p$ denote the set of subsets of $\mathbb{Z}$ having cardinality p, and $\mathcal{Z} = \bigcup_p \mathcal{Z}_p$ the set of all finite subsets of $\mathbb{Z}$. For $K \in \mathcal{Z}_p$ consisting of distinct $k_1, \ldots, k_p \in \mathbb{Z}$ we define the multi-flip $f_K : \mathcal{M} \to \mathcal{M}$ by

$$f_K(M) = (f_{k_1} \circ \cdots \circ f_{k_p})(M), \quad M \in \mathcal{M}. \tag{2}$$

Since flips commute, the action of f_K does not depend upon the order of $k_1, \ldots, k_p$.

It is useful to regard $\mathcal{M}$ as a complete graph whose edges are multi-flips. For Maya diagrams $M_1, M_2 \in \mathcal{M}$, the symmetric difference

$$M_1 \ominus M_2 = (M_1 \setminus M_2) \cup (M_2 \setminus M_1)$$

is precisely the edge that connects M_1 and M_2. More precisely, if

$$K = M_1 \ominus M_2 = M_2 \ominus M_1,$$

then $f_K(M_1) = M_2$ and $f_K(M_2) = M_1$.

Multi-flips can also be used to define a bijection $\mathcal{Z} \to \mathcal{M}$ given by $K \mapsto f_K(M_\emptyset)$, where $M_\emptyset := \mathbb{Z}_-$ denotes the trivial Maya diagram. We refer to $K \in \mathcal{Z}$ as the index set of the Maya diagram $f_K(M_\emptyset)$.

The additive group $\mathbb{Z}$ acts on $\mathcal{M}$, because for $M \in \mathcal{M}$ and $n \in \mathbb{Z}$, the set

$$M + n = \{m + n : m \in M\}$$

is also a Maya diagram. Moreover, we have

$$\sigma_{M+n} = \sigma_M + n. \tag{3}$$

We will refer to an equivalence class of Maya diagrams related by translations as an *unlabelled Maya diagram*, and denote the set of all unlabelled Maya diagrams by $\mathcal{M}/\mathbb{Z}$. One can visualize the passage from an unlabelled to a labelled Maya diagram as choosing the placement of the origin.

For $B \in \mathcal{Z}_p$, where $p = 2g + 1$ is odd, we define the Maya diagram

$$\Xi(B) = (-\infty, b_0) \cup [b_1, b_2) \cup \cdots \cup [b_{2g-1}, b_{2g}), \tag{4}$$

where $b_0 < b_1 < \cdots < b_{2g}$ is an increasing enumeration of B and where $[m, n) = \{j \in \mathbb{Z} : m \le j < n\}$. Every Maya diagram has a unique representation of the form $\Xi(B)$ for some $B \in \mathcal{Z}_{2g+1}$. We will call the corresponding $g \ge 0$ the genus of $M = \Xi(B)$ and refer to $(b_0, \ldots, b_{2g})$ as the block coordinates of M. The block coordinates may also be characterized as the unique set $B \in \mathcal{Z}$ such that $f_B(M) = M + 1$.

After removal of the initial infinite ◙ segment and the trailing infinite □ segment, a Maya diagram consists of alternating empty □ and filled ◙ segments of variable length. The genus g counts the number of such pairs. The even block coordinates b_{2i} indicate the starting positions of the empty segments, and the odd block coordinates b_{2i+1} indicate the starting positions of the filled segments.

3 Rational Extensions

For $n \in \mathbb{Z}$, set

$$\psi_n(x) = \begin{cases} e^{-\frac{x^2}{2}} H_n(x) & \text{if } n \geq 0 \\ e^{\frac{x^2}{2}} \tilde{H}_{-n-1}(x) & \text{if } n < 0 \end{cases}$$

where

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}, \quad n = 0, 1, 2, \ldots$$

are the Hermite polynomials, and

$$\tilde{H}_n(x) = (-\mathrm{i})^n H_n(\mathrm{i}x)$$

are the conjugate Hermite polynomials. We then have

$$-\psi_n''(x) + x^2 \psi_n(x) = (2n+1)\psi_n(x), \quad n \in \mathbb{Z}.$$

For $n \geq 0$, the above solutions correspond to the bound states of the quantum harmonic oscillator. The solutions for $n < 0$ do not satisfy the boundary conditions at $\pm\infty$ and therefore represent virtual states.

For $M \in \mathcal{M}$ with index set $K \in \mathcal{Z}_p$, let $s_1 > \cdots > s_r \geq 0$ and $t_1 > \cdots > t_q \geq 0$ be the uniquely specified lists of natural numbers such that

$$K = \{-1 - s_1, \ldots, -1 - s_r, t_q, \ldots, t_1\}, \quad p = q + r.$$

We will refer to $(s_1, \ldots, s_r \mid t_q, \ldots, t_1)$ as the *Frobenius symbol* of M. It is easy to check that the index of M is given by $\sigma = q - r$.

Let us now define

$$H_M(x) = e^{\sigma_M \frac{x^2}{2}} \operatorname{Wr}[\psi_{k_1}, \ldots, \psi_{k_p}], \tag{5}$$

where $k_1 < \cdots < k_p$ is an increasing enumeration of K, where $\sigma_M \in \mathbb{Z}$ is the index, and Wr is the usual Wronskian determinant. The polynomial nature of $H_M(x)$ becomes evident in the following pseudo-Wronskian [4] realization:

$$H_M = \begin{vmatrix} \tilde{H}_{s_1} & \tilde{H}_{s_1+1} & \ldots & \tilde{H}_{s_1+r+q-1} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{H}_{s_r} & \tilde{H}_{s_r+1} & \ldots & \tilde{H}_{s_r+r+q-1} \\ H_{t_q} & H'_{t_q} & \ldots & H^{(r+q-1)}_{t_q} \\ \vdots & \vdots & \ddots & \vdots \\ H_{t_1} & H'_{t_1} & \ldots & H^{(r+q-1)}_{t_1} \end{vmatrix}. \tag{6}$$

A suitably normalized pseudo-Wronskian is a translation invariant of the underlying Maya diagram. The following result was proved in [4]. Set

$$\widehat{H}_M = \frac{(-1)^{rq} H_M}{\prod_{i<j} 2(s_j - s_i) \prod_{i<j} 2(t_i - t_j)}. \tag{7}$$

Then for $M \in \mathcal{M}$ and $n \in \mathbb{Z}$ we have

$$\widehat{H}_M = \widehat{H}_{M+n}. \tag{8}$$

The potential

$$\begin{aligned} U_M(x) &= x^2 - 2\frac{d^2}{dx^2} \log \mathrm{Wr}[\psi_{k_1}, \ldots, \psi_{k_p}], \\ &= x^2 + 2\left(\frac{H_M'}{H_M}\right)^2 - \frac{2H_M''}{H_M} - 2\sigma_M \end{aligned} \tag{9}$$

is known as a rational extension [3] of the harmonic oscillator. The corresponding Hamiltonian operators

$$T_M = -\frac{d^2}{dx^2} + U_M \tag{10}$$

are exactly solvable with

$$T_M[\psi_{M,k}] = (2k+1)\psi_{M,k},$$

where

$$\psi_{M,k} = e^{\frac{\epsilon x^2}{2}} \frac{H_{f_k(M)}}{H_M}, \qquad \epsilon = \begin{cases} +1 & \text{if } k \in M \\ -1 & \text{if } k \notin M \end{cases}.$$

Note that, as a consequence of (3) and (8), T_M is translation covariant:

$$T_{M+n} = T_M + 2n, \quad n \in \mathbb{Z}. \tag{11}$$

Let $(b_0, b_1, \ldots, b_{2g})$ be the block coordinates of M. By the Krein–Adler theorem [1, 3, 8], the polynomial H_M has no real zeros if and only if $b_{2j} - b_{2j-1}$ is even for all $j = 1, \ldots, g$, i.e., if all the finite ◘ segments of M have even size. For such Maya diagrams, the potential U_M is non-singular and hence T_M corresponds to a self-adjoint operator. The bound states of the operator correspond to the empty boxes of M, i.e., to $k \notin M$. It is precisely for such $M \in \mathcal{M}$ and $k \notin M$ that the eigenfunction $\psi_{M,k}$ is square-integrable. For such M and k, the polynomial part of $\psi_{M,k}$ is known as an exceptional Hermite polynomial [3].

4 Categorical Structure

In this section, we define $\mathbb{MD}$, a category whose objects are Maya diagrams and whose arrows are multi-flips, and $\mathbb{REXT}$, another category whose objects are rational extensions and whose arrows are intertwining operators (definition given below). We then exhibit a functor $\mathbb{MD} \to \mathbb{REXT}$ that we use to classify ladder operators.

In order to define composition of arrows, it will first be necessary to generalize the notion of a multi-flip. A multi-set is a generalized set object that allows for multiple instances of each of its elements. Let $\widehat{\mathcal{Z}}_p$ denote the set of integer multi-sets of cardinality p and $\widehat{\mathcal{Z}} = \bigcup_p \widehat{\mathcal{Z}}_p$ the set of finite integer multi-sets. We express a multi-set $K \in \widehat{\mathcal{Z}}$ as

$$K = \{k_1^{p_1}, \ldots, k_q^{p_q}\} \tag{12}$$

where $k_1, \ldots, k_q \in \mathbb{Z}$ are distinct, and where $p_i > 0$ indicate the multiplicity of element k_i. The cardinality is then given by $p = p_1 + \cdots + p_q$. The notion of a multi-flip extends naturally from sets to multi-sets. Indeed, for $K \in \widehat{\mathcal{Z}}$, we re-use (2) to define the multi-flip $f_K : \mathcal{M} \to \mathcal{M}$.

We say that K is an even multi-set if all of its elements have an even multiplicity. Since flips are involutions, f_K is the identity transformation if and only if K is even. If K is an even multi-set, then it has the unique decomposition $K = K_1 \cup K_1$ where K_1 has the same elements as K but with the multiplicities divided by 2. More generally, every multi-set $K \in \widehat{\mathcal{Z}}$ has a unique decomposition of the form

$$K = K_0 \cup K_1 \cup K_1, \quad K_0 \in \mathcal{Z},\ K_1 \in \widehat{\mathcal{Z}}, \tag{13}$$

where K_0 is the set of integers that occur in K with an odd multiplicity. Again, since flips are involutions, we have $f_K = f_{K_0}$.

The objects of $\mathbb{MD}$ are labelled Maya diagrams $\mathcal{M}$, and the arrows are pairs $(M, K) \in \mathcal{M} \times \widehat{\mathcal{Z}}$. The source of (M, K) is M, and the target is $f_K(M)$. Composition of morphisms is given by the union of multi-sets:

$$(M_2, K_2) \circ (M_1, K_1) = (M_1, K_1 \cup K_2),$$

where $M_1 \in \mathcal{M},\ K_1, K_2 \in \widehat{\mathcal{Z}},\ M_2 = f_{K_1}(M_1)$.

For differential operators A, T_1, T_2, we say that A intertwines T_1, T_2 if

$$AT_1 = T_2 A.$$

The objects of $\mathbb{REXT}$ are the rational extensions T_M, $M \in \mathcal{M}$, and the arrows are monic differential operators that intertwine two rational extensions. Observe that if A intertwines T_1, T_2, then so does $A \circ p(T_1)$, where $p(x)$ is an arbitrary

polynomial. Given T_1, T_2, we say that A is a primitive intertwiner if it does not include a nontrivial right factor $p(T_1)$.

For a Maya diagram $M \in \mathcal{M}$ and a set $K \in \mathcal{Z}_p$, we define the operator

$$A_{M,K}[y] = \frac{\mathrm{Wr}[\psi_{M,k_1}, \ldots, \psi_{M,k_p}, y]}{\mathrm{Wr}[\psi_{M,k_1}, \ldots, \psi_{M,k_p}]}.$$

By construction, $A_{M,K}$ is a monic differential operator of order p. These intertwining operators have their origin in SUSYQM (supersymmetric quantum mechanics), and obey the intertwining relation

$$A_{M_1,K} T_{M_1} = T_{M_2} A_{M_1,K}, \quad M_2 = f_K(M_1), \quad M_1, M_2 \in \mathcal{M}, \ K \in \mathcal{Z}.$$

It is possible to show that $A_{M,K}$ is a primitive intertwiner between T_M and $T_{f_K(M)}$. Moreover, it is possible to show [2, Proof of Theorem 3.10] that every arrow in $\mathbb{REXT}$ has the form $A_{M,K} \circ p(T_M)$, where $A_{M,K}$ is primitive (i.e., K is a set), and $p(x)$ is a polynomial. We also note that these intertwiners are translation invariant:

$$A_{M+n,K+n} = A_{M,K}, \quad n \in \mathbb{Z}. \tag{14}$$

In order to describe the composition of intertwiners, we need to extend the above definition to include multi-sets. For $K \in \widehat{\mathcal{Z}}$, let $K_0 \in \mathcal{Z}$ and $K_1 \in \widehat{\mathcal{Z}}$ be as per (13). For $M \in \mathcal{M}$, we now define

$$A_{M,K} = A_{M,K_0} \circ \prod_{k \in K_1} (2k + 1 - T_M). \tag{15}$$

In other words, if $K \in \widehat{\mathcal{Z}}$ contains elements of higher multiplicity, then $A_{M,K}$ is no longer primitive. The arrows of $\mathbb{REXT}$ are the operators $A_{M,K}, \ M \in \mathcal{M}, \ K \in \widehat{\mathcal{Z}}$. Composition of arrows is just the usual composition of differential operators.

Theorem 1 *The correspondence $M \mapsto T_M, \ M \in \mathcal{M}$ and $(M, K) \mapsto A_{M,K}, \ K \in \widehat{\mathcal{Z}}$ is a covariant functor $\mathbb{MD} \to \mathbb{REXT}$.*

Proof It suffices to observe that for $M_1 \in \mathcal{M}, \ K_1, K_2 \in \widehat{\mathcal{Z}}$ we have

$$A_{M_2,K_2} \circ A_{M_1,K_1} = A_{M_1,K_1 \cup K_2}, \quad M_2 = f_{K_1}(M_1).$$

5 Ladder Operators

We define a ladder operator to be an intertwiner A such that

$$A T_M = (T_M + \lambda) A$$

for some $M \in \mathcal{M}$ and constant λ. Since $T_{M+n} = T_M + 2n$, Theorem 1 implies that for every rational extension T_M, $M \in \mathcal{M}$, and $n \in \mathbb{Z}$, there exists a ladder operator $A_{M,K}$, where $K = (M + n) \ominus M$. By (14) no generality is lost if we index such ladder operators in terms of unlabelled Maya diagrams $[M] \in \mathcal{M}/Z$.

A recent result provides a characterization of translational multi-flips [5] in terms of cyclic Maya diagrams. This characterization makes it possible to establish the order of a ladder operator [6].

Theorem 2 *Let $M \in \mathcal{M}$ and $n = 1, 2, \ldots$. Then,*

$$|(M + n) \ominus M| = n + 2 \sum_{i=0}^{n-1} g_i, \tag{16}$$

where g_i is the genus of the Maya diagram

$$M_i = \{m \in \mathbb{Z} \colon mn + i \in M\}, \quad i = 0, 1, \ldots, n - 1.$$

Proof Let $B_i \in \mathcal{Z}_{2g_i+1}$ be the block coordinates of M_i, and set

$$B = \bigcup_{i=0}^{n-1} (nB_i + i) = \bigcup_{i=0}^{n-1} \{nb + i \colon b \in B_i\}.$$

Since B_i is the unique set such that $f_{B_i}(M_i) = M_i + 1$, it follows that B is the unique set such that $f_B(M) = M + n$. Therefore $B = (M + n) \ominus M$.

Fix a Maya diagram $M \in \mathcal{M}$. An immediate consequence of Theorem 2 is the existence of a primitive ladder operator that intertwines T_M and $T_M + 2n$ for every $n \in \mathbb{Z}$. The ladder operator in question is $L_n := A_{M,K}$, where $K = (M + n) \ominus M$. The order of L_n is given by (16). If $n > 0$, then both L_n and L_1^n intertwine T_M and $T_M + 2n$; it follows that there must be a syzygy of the form

$$L_1^n = L_n \circ p(T_M),$$

where the roots of the polynomial p are determined by (15).

The action of ladder operators on states is that of a lowering or raising operator according to

$$L_n[\psi_{M,k}] = C_{M,n,k} \psi_{M,k-n}, \quad k \notin M,$$

where $C_{M,n,k}$ is]zero if $\psi_{M,k-n}$ is not a bound state, i.e., if $k - n \in M$. Otherwise, $C_{M,n,k}$ is a rational number whose explicit form can be derived on the basis of (7). As a particular example, suppose that the index set of M consists of positive integers $0 < k_1 < \cdots < k_p$, that $n > 0$, and that $k \notin M$. In this case,

$$C_{M,n,k} = \begin{cases} \prod_{i \in M \setminus (M+n)} (2i - 2j) \times (k - n + 1)_n \, 2^n & \text{if } k - n \notin M \\ 0 & \text{otherwise.} \end{cases}$$

6 Examples

The articles [9, 10] considered a particular class of ladder operators corresponding to Maya diagrams obtained by a single state-adding transformation. Fix some $n = 1, 2, \ldots$, and let $\tilde{M}_n$ be the Maya diagram with index set $\{-n\}$, i.e., let $\tilde{M}_n = \mathbb{Z}_- \setminus \{-n\}$. We set

$$\hat{M}_n = \tilde{M}_n + n = \mathbb{Z}_- \cup \{1, \ldots, n-1\},$$

and observe that $\hat{M}_n$ has index set $\{1, \ldots, n-1\}$. Hence,

$$L_n := A_{\tilde{M}_n, \{-n, 1, \ldots, n-1\}},$$

is an nth order ladder operator that intertwines $T_{\tilde{M}_n}$ and $T_{\hat{M}_n}$. Ordering the flips in ascending order, we obtain the following factorization into first-order intertwiners:

$$L_n = A_{\hat{M}_{n-1}, \{n-1\}} \cdots A_{\hat{M}_2, \{2\}} A_{\hat{M}_1, \{1\}} A_{\tilde{M}_n, \{-n\}};$$

each flip corresponds to a state-deleting transformation.

Let us also observe that $\tilde{M}_n$ is a genus 1 Maya diagram. It follows that

$$L_1 := A_{\tilde{M}_n, \{-n, -n+1, 0\}}$$

is a third-order ladder operator that intertwines $\tilde{M}_n$ and $\tilde{M}_n + 1$.

The composition L_1^n is represented by the multi-set

$$\bigcup_{j=0}^{n-1} \{-n+j, -n+j+1, j\} = \{-n, 1, \ldots, n-1\} \cup \{(-n+1)^2, \ldots, (-1)^2, (0)^2\},$$

where the superscripts indicate repetition (and not a square). The syzygy between L_n and L_1 is therefore

$$L_1^n = L_n \prod_{j=-n+1}^{0} (2j + 1 - T_{\tilde{M}_n}).$$

References

1. V.E. Adler, A modification of Crum's method. Theoret. Math. Phys. **101**, 1381–1386 (1994)
2. M.A. García-Ferrero, D. Gómez-Ullate, R. Milson, A Bochner type characterization theorem for exceptional orthogonal polynomials. J. Math. Anal. Appl. **472**, 584–626 (2019)
3. D. Gómez-Ullate, Y. Grandati, R. Milson, Rational extensions of the quantum harmonic oscillator and exceptional Hermite polynomials. J. Phys. A **47**, 015203 (2013)

4. D. Gómez-Ullate, Y. Grandati, R. Milson, Durfee rectangles and pseudo-Wronskian equivalences for Hermite polynomials. Stud. Appl. Math. **141**, 596–625 (2018)
5. D. Gomez-Ullate, Y. Grandati, S. Lombardo, R. Milson, Rational solutions of dressing chains and higher order Painleve equations (2018). arXiv:1811.10186
6. P. A. Clarkson, D. Gómez-Ullate, Y. Grandati, R. Milson, Cyclic Maya diagrams and rational solutions of higher order Painlevé systems. Stud. Appl. Math. **144**, 357–385 (2020)
7. S.E. Hoffmann, V. Hussin, I. Marquette, Y.Z. Zhang, Coherent states for ladder operators of general order related to exceptional orthogonal polynomials. J. Phys. A **51**, 315203 (2018)
8. M.G. Krein, On a continual analogue of a Christoffel formula from the theory of orthogonal polynomials. Dokl. Akad. Nauk SSSR (N.S.) **113**, 970–973 (1957)
9. I. Marquette, C. Quesne, New ladder operators for a rational extension of the harmonic oscillator and superintegrability of some two-dimensional systems. J. Math. Phys. **54**, 102102 (2013)
10. I. Marquette, C. Quesne, New families of superintegrable systems from Hermite and Laguerre exceptional orthogonal polynomials. J. Math. Phys. **54**, 042102 (2013)
11. J. Mateo, J. Negro, Third-order differential ladder operators and supersymmetric quantum mechanics. J. Phys. A **41**, 045204 (2008)

Tachyons and Representations of Sp(2, $\mathbb{R}$)

P. Moylan

Abstract Lacking in the mathematical physics literature is a detailed treatment of tachyonic representations of the Poincaré group along lines similar to that for its real mass, positive and negative energy representations. Such representations Wigner did not consider in any detail in his 1939 paper on the unitary representations of the inhomogeneous Lorentz group (Wigner, Ann Math 40:149–204, 1939), and Bargmann and Wigner in their paper on the group theoretical classification of relativistic wave equations did not consider them either because "they are . . . unlikely to have a simple physical interpretation" (Bargmann and Wigner, Proc Nat Acad Sci (USA) 34(5):211–223, 1948). We are making a detailed study of tachyonic representations of the Poincaré group in four space-time dimensions and we describe some of our results here. In particular, we relate tachyonic representations of the Poincaré group to representations of the anti-de Sitter group, in a way analogous to the way in which positive energy, real mass representations of the Poincaré group are related to unitary principal series representations of the de Sitter group via group contraction and deformation.

Keywords Tachyons · Representation theory · Poincaré group · de Sitter groups

1 Introduction

The connection between the unitary representation theory of the universal covering groups of the de Sitter group, $SO_0(1, 4)$, and the Poincaré group is well understood and much of it even in explicit detail. Unfortunately, the same cannot be said for the connection between the unitary representations of the anti-de Sitter group, $SO_0(2, 3)$, and the Poincaré group, even though $SO_0(2, 3)$ is more interesting from the point of view of physical applications.

P. Moylan (✉)
Physics Department, Pennsylvania State University, The Abington College, Abington, PA, USA
e-mail: pjm11@psu.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_12

For $SO_0(1,4)$, its representations are described in terms of various realizations or parallelizations [1]. The spherical, flat, and hyperbolic parallelizations are as follows. Let KAN be the Iwasawa decomposition of $SO_0(1,4)$, and let M be the centralizer of A in K, then the "compact picture" [2] of its representations (spherical parallelization) is given by vector-valued fields over $K/M \cong S^3$, with S^3 being the three sphere. The Bruhat decomposition $SO_0(1,4) \cong' \widetilde{N}MAN$ leads to the non-compact picture (or flat parallelization) which describes representations in terms of vector-valued functions on $\widetilde{N} \cong \mathbb{R}^3$. (By $\cong'$ we mean isomorphic to a dense open subset of $SO_0(1,4)$.) Finally the Hannabuss decomposition, which is $SO_0(1,4) \cong' HAN$, where H is $SO(1,3)$, gives the hyperbolic picture (hyperbolic parallelization) describing $SO_0(1,4)$ representations on $SO_0(1,3)/M \cong \mathbb{T}^3$, the two-sheeted momentum hyperboloid. Now unitary representations of the inhomogeneous Euclidean group in four dimensions, the inhomogeneous Galilean group, and the Poincaré group have realizations in terms of vector-valued fields on S^3, $\mathbb{R}^3$, and $\mathbb{T}^3$, respectively, and the relationship via contraction and deformation of these representations to representations of $SO_0(1,4)$, at least for the case of unitary principal series representations of $SO_0(1,4)$, is very well understood. We would like to obtain an analogous description of this situation for $SO_0(2,3)$ and it is to this goal that the current article contributes. Compared to $SO_0(1,4)$ the situation with respect to $SO_0(2,3)$ is much more complicated: instead of S^3 we now have $S^1 \times S^2$; there are more Bruhat-like decompositions, so there are more non-compact pictures; and finally there are two analogs of the Hannabuss decomposition [1].

2 $SO_0(2,3)$, $Sp(2,\mathbb{R})$, the Poincaré Group and Their Lie Algebras

Let $\beta_0 = \mathrm{diag}(1,\ 1,\ -1,\ -1,\ -1)$, where the right-hand side of this equation denotes a diagonal matrix with diagonal entries as shown inside the parentheses. $SO_0(2,3)$ is the component connected to identity of the group

$$\mathrm{SO}(2,3) = \{g \in SL(5,\mathbb{R}) \mid g\,\beta_0\,g^\dagger = \beta_0\}.$$

(† denotes transpose of a matrix.) Denote by $\mathfrak{so}(2,3)$ the Lie algebra of $SO_0(2,3)$. A realization of $\mathfrak{so}(2,3)$ is provided by the set of all matrices $(a_{i,j})$ $(-1 \leqslant i,j \leqslant 3)$ such that $a_{i,i} = 0$ $(-1 \leqslant i \leqslant 3)$, $a_{i,j} = -a_{j,i}$ $(1 \leqslant i \leqslant j \leqslant 3)$, $a_{0,j} = a_{j,0}$ $(1 \leqslant j \leqslant 3)$, $a_{-1,j} = a_{j,-1}$ $(1 \leqslant j \leqslant 3)$ and $a_{-1,0} = -a_{0,-1}$. Let $E_{i,j}$ be the matrix such that the (i,j) component is equal to 1 and the other components are all equal to 0. Let $\mathbf{L}_{-1,0} = -E_{-1,0} + E_{0,-1}$, $\mathbf{L}_{i,j} = E_{i,j} - E_{j,i}$ $(1 \leqslant i,j \leqslant 3, i \neq j)$, $\mathbf{L}_{0,i} = E_{i,0} + E_{0,i}$ $(1 \leqslant i \leqslant 3)$, $\mathbf{L}_{-1,i} = E_{i,-1} + E_{-1,i}$ $(1 \leqslant i \leqslant 3)$. The $\mathbf{L}_{a,b}$ $(a,b = -1,0,1,2,3)$, viewed abstractly, are a basis for $\mathfrak{so}(2,3)$. The commutation relations of the $\mathbf{L}_{a,b}$ are

$$[\mathbf{L}_{a,b}, \mathbf{L}_{b,c}] = -e_b \mathbf{L}_{a,c} \tag{1}$$

with $e_{-1} = e_0 = -e_1 = -e_2 = -e_3 = 1$.

Denote the Lie algebra of the Poincaré group by $\mathfrak{p}$. A basis for $\mathfrak{p}$ is given by the $\mathbf{L}_{0,i}$ and $\mathbf{L}_{i,j}$ $(i, j = 1, 2, 3)$ of $\mathfrak{so}(1, 3)$, the Lie algebra of the Lorentz group, together with a Lorentz vector operator $\mathbf{P}_\mu$ $(\mu = 0, 1, 2, 3)$, the components of which mutually commute. (By "Lorentz vector operator" we mean that the $\mathbf{P}_\mu$ satisfy the same commutation relations with the generators $\mathbf{L}_{\mu,\nu}$ of $\mathfrak{so}(1, 3)$ as the $\mathbf{L}_{-1,\mu}$ $(\mu = 0, 1, 2, 3)$.)

We let $\mathcal{U}(\mathfrak{p})$ be the universal enveloping algebra of $\mathfrak{p}$. Since $\mathcal{U}(\mathfrak{p})$ is an integral domain, it has no zero divisors. Hence it has a skew field (or "Lie field") of fractions which we denote by $\mathfrak{D}(\mathfrak{p})$. Similarly we denote the universal enveloping algebra of $\mathfrak{g} = \mathfrak{so}(2, 3)$ by $\mathcal{U}(\mathfrak{g})$. For the same reason as for $\mathcal{U}(\mathfrak{p})$, it also has a skew field of fractions, i.e. Lie field, which we denote by $\mathfrak{D}(\mathfrak{g})$.

We introduce the following elements of

$$\begin{aligned}
\mathcal{U}(\mathfrak{g}) : \mathbf{L}^2 &= \mathbf{L}_{1,2}^2 + \mathbf{L}_{2,3}^2 + \mathbf{L}_{3,1}^2,\ \mathbf{Q}_2 = \mathbf{L}_{0,1}^2 + \mathbf{L}_{0,2}^2 + \mathbf{L}_{0,3}^2 - \mathbf{L}^2, \\
\mathbf{Q}_4 &= \left(\mathbf{L}_{1,2}\mathbf{L}_{3,0} + \mathbf{L}_{2,3}\mathbf{L}_{1,0} + \mathbf{L}_{3,1}\mathbf{L}_{2,0}\right)^2, \\
\mathbf{D}_2 &= -\mathbf{L}_{-1,0}^2 + \mathbf{L}_{-1,1}^2 + \mathbf{L}_{-1,2}^2 + \mathbf{L}_{-1,3}^2 + \mathbf{Q}_2 \text{ and} \\
\mathbf{D}_4 &= \left(\sum_{ijk=1}^{3} \frac{1}{2}\epsilon_{ijk}\mathbf{L}_{-1,i}\mathbf{L}_{j,k}\right)^2 \\
&\quad + \sum_{ijk\ell m=1}^{3} \left(\epsilon_{ijk}\left\{\frac{1}{2}\mathbf{L}_{-1,0}\,\mathbf{L}_{j,k} + \mathbf{L}_{-1,k}\mathbf{L}_{0,j}\right\}\right) \\
&\quad \times \left(\epsilon_{i\ell m}\left\{\frac{1}{2}\mathbf{L}_{-1,0}\mathbf{L}_{\ell,m} + \mathbf{L}_{-1,\ell}\mathbf{L}_{0,m}\right\}\right).
\end{aligned}$$

The center $\mathbb{Z}(\mathfrak{g})$ of $\mathcal{U}(\mathfrak{g})$ is generated by $\mathbf{D}_2$ and $\mathbf{D}_4$. The center $\mathbb{Z}(\mathfrak{p})$ of $\mathcal{U}(\mathfrak{p})$ is generated by the following set of elements: $\mathbf{P}^2 = \mathbf{P}_0^2 - \mathbf{P}_1^2 - \mathbf{P}_2^2 - \mathbf{P}_3^2$ and $\mathbf{W} = \sum_{\mu\nu\rho=0}^{3} \left(\mathbf{P}_\mu\,\mathbf{P}^\nu\,\mathbf{L}_{\nu,\rho}\,\mathbf{L}^{\rho,\mu} - \frac{1}{2}\mathbf{P}_\rho\,\mathbf{P}^\rho\,\mathbf{L}_{\mu,\nu}\,\mathbf{L}^{\nu,\mu}\right)$ where we use Einstein summation convention with metric tensor β_0.

Now to the symplectic group, Sp(2, $\mathbb{R}$), which is defined as

$$\mathrm{Sp}(2, \mathbb{R}) = \{g \in \mathrm{GL}(4, \mathbb{R}) \mid g^\dagger J g = J\} \quad \text{where} \quad J = \begin{pmatrix} 0 & \mathbb{I}_2 \\ -\mathbb{I}_2 & 0 \end{pmatrix} \in \mathrm{GL}(4, \mathbb{R}).$$

The Lie algebra of Sp(2, $\mathbb{R}$) is given by

$$\mathfrak{sp}(2, \mathbb{R}) = \{X \in \mathrm{End}(\mathbb{R}^4) \mid JX + X^\dagger J = 0\}.$$

We have the short exact sequence

$$\mathbb{Z}_2 \to \mathrm{Sp}(2,\mathbb{R}) \to \mathrm{SO}_0(2,3). \tag{2}$$

An explicit isomorphism φ of the respective Lie algebras is given by

$$\varphi(\mathbf{L}_{-1,0}) = \tfrac{1}{2}\begin{bmatrix} 0 & -\mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{bmatrix}, \varphi(\mathbf{L}_{1,2}) = \tfrac{1}{2}\begin{bmatrix} \widehat{\sigma}_2 & 0 \\ 0 & \widehat{\sigma}_2 \end{bmatrix},$$
$$\varphi(\mathbf{L}_{1,3}) = \tfrac{1}{2}\begin{bmatrix} 0 & -\sigma_1 \\ \sigma_1 & 0 \end{bmatrix}, \varphi(\mathbf{L}_{2,3}) = \tfrac{1}{2}\begin{bmatrix} 0 & -\sigma_3 \\ \sigma_3 & 0 \end{bmatrix},$$
$$\varphi(\mathbf{L}_{-1,1}) = \tfrac{1}{2}\begin{bmatrix} 0 & \sigma_1 \\ \sigma_1 & 0 \end{bmatrix}, \varphi(\mathbf{L}_{-1,2}) = \tfrac{1}{2}\begin{bmatrix} 0 & \sigma_3 \\ \sigma_3 & 0 \end{bmatrix}, \varphi(\mathbf{L}_{-1,3}) = \tfrac{1}{2}\begin{bmatrix} \mathbb{I}_2 & 0 \\ 0 & -\mathbb{I}_2 \end{bmatrix},$$
$$\varphi(\mathbf{L}_{0,1}) = \tfrac{1}{2}\begin{bmatrix} -\sigma_1 & 0 \\ 0 & \sigma_1 \end{bmatrix}, \varphi(\mathbf{L}_{0,2}) = \tfrac{1}{2}\begin{bmatrix} -\sigma_3 & 0 \\ 0 & \sigma_3 \end{bmatrix}, \varphi(\mathbf{L}_{03}) = \tfrac{1}{2}\begin{bmatrix} 0 & \mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{bmatrix}$$

where $\mathbb{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\widehat{\sigma}_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. The linear span of the four matrices in the top two rows of the above equation is $\mathfrak{k}$, the Lie algebra of the maximal compact subgroup K of $\mathrm{Sp}(2,\mathbb{R})$. With θ the Cartan involution on $\mathfrak{sp}(2,\mathbb{R})$ specified by $\theta(X) = -X^\dagger$ for $X \in \mathfrak{sp}(2,\mathbb{R})$, we have that $\mathfrak{sp}(2,\mathbb{R}) = \mathfrak{k} \oplus \mathfrak{p}$ where $\mathfrak{p}$ is the real linear span of the six matrices in the bottom two rows of the above equation.

A maximal parabolic subgroup $P = MAN$ of $\mathrm{Sp}(2,\mathbb{R})$ is: ($t, y_0, y_1, y_2 \in \mathbb{R}$)

$$M = \left\{ m = \begin{pmatrix} (\sigma_3)^\epsilon \ell & 0 \\ 0 & (\sigma_3)^\epsilon \ell^{\dagger\,-1} \end{pmatrix} \;\middle|\; \ell = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}),\ \epsilon \in \{0,1\} \right\};$$

$$A = \left\{ a(t) = \begin{pmatrix} t & 0 & 0 & 0 \\ 0 & t & 0 & 0 \\ 0 & 0 & t^{-1} & 0 \\ 0 & 0 & 0 & t^{-1} \end{pmatrix} \right\}; N = \left\{ n(t) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ y_0 + y_2 & y_1 & 1 & 0 \\ y_1 & y_0 - y_2 & 0 & 1 \end{pmatrix} \right\}.$$

Then $P = MAN$ is a maximal proper parabolic subgroup of $\mathrm{Sp}(2,\mathbb{R})$.

3 Algebraic Results

We define commutative algebraic extensions of $\mathfrak{D}(\mathfrak{p})$ and $\mathfrak{D}(\mathfrak{g})$ as [3–6]: $\widetilde{\mathfrak{D}}(\mathfrak{p}) = \left\{ a + bY \;\middle|\; a,\ b \in \mathfrak{D}(\mathfrak{p}) \right\}$, where Y commutes with all elements of $\mathfrak{D}(\mathfrak{p})$ and satisfies the equation $Y^2 = \mathbf{P}^2$; $\widetilde{\mathfrak{D}}(\mathfrak{g}) = \left\{ a + b\,\widetilde{Y} + c\,\widetilde{Y}^2 + d\,\widetilde{Y}^3 \;\middle|\; a,\ b,\ c,\ d \in \right.$

$\mathfrak{D}(\mathfrak{g})$ $\Big\}$ where $\widetilde{Y}$ commutes with all elements of $\mathfrak{D}(\mathfrak{g})$ and satisfies the equation $\widetilde{Y}^4 + \mathbf{D}_2' \widetilde{Y}^2 + \mathbf{D}_4' = 0$ with $\mathbf{D}_2' = \left(\mathbf{D}_2 + \frac{5}{2}\mathbf{I}\right)$ and $\mathbf{D}_4' = \left(\mathbf{D}_4 + \frac{1}{4}\mathbf{D}_2 + \frac{9}{16}\mathbf{I}\right)$. ($\mathbf{I}$ is the identity in $\widetilde{\mathfrak{D}}(\mathfrak{g})$.) Now define a mapping τ_λ ($\lambda \in \mathbb{R}$) from $\mathfrak{g}$ to $\widetilde{\mathfrak{D}}(\mathfrak{p})$ by

$$\tau_\lambda(\widetilde{\mathbf{L}}_{\mu,\nu}) = \mathbf{L}_{\mu,\nu} \ , \ \tau_\lambda(\widetilde{\mathbf{L}}_{-1,\mu}) = \frac{i\,\lambda}{2\,Y}\Big[\mathbf{Q}_2 \ , \ \mathbf{P}_\mu\Big] + \mathbf{P}_\mu. \tag{3}$$

The $\lambda^{-1}\tau_\lambda(\widetilde{\mathbf{L}}_{-1,\mu})$ and $\tau_\lambda(\widetilde{\mathbf{L}}_{\mu,\nu})$ satisfy the commutation relations of the generators of Sp(2, $\mathbb{R}$). The $\tau_\lambda(\widetilde{\mathbf{L}}_{-1,\mu})$ and $\tau_\lambda(\widetilde{\mathbf{L}}_{\mu,\nu})$ are a basis for an isomorphic copy $\mathfrak{g}_\lambda$ of $\mathfrak{g}$, which differs from $\mathfrak{g}$ by a scaling factor λ in the $\mathbf{L}_{-1,\mu}$ directions, and hence generate Sp(2, $\mathbb{R}$)$_\lambda$. We henceforth consider for simplicity the case $\lambda = 1$ and let $\tau_{\lambda=1}(\widetilde{Y}) = Y$, then $\tau = \tau_{\lambda=1}$ can be extended to a homomorphism of $\widetilde{\mathfrak{D}}(\mathfrak{g})$ into $\widetilde{\mathfrak{D}}(\mathfrak{p})$ in an obvious way, which turns out to be surjective because of Theorem 2. Denote this extension also by τ. Elements of $\widetilde{\mathfrak{D}}(\mathfrak{g})$ are denoted with a tilde to keep them distinct from elements of $\widetilde{\mathfrak{D}}(\mathfrak{p})$.

Theorem 1 *Let* $\mathfrak{g}$ *be the deformation of* $\mathfrak{p}$ *having basis elements* $\mathbf{L}_{i,j} \in \mathfrak{g}$ *and* $\mathbf{L}_{-1,\mu} \in \widetilde{\mathfrak{D}}(\mathfrak{p})$ *defined by Eqs.* (3). *Then (for* $\lambda = 1$*) the following holds:*

$$\mathbf{D}_2 = -Y^2 - \left[\frac{\mathbf{W}}{Y^2} + \frac{9}{4}\mathbf{I}\right] \ , \ \mathbf{D}_4 = \left[Y^2 + \frac{1}{4}\right]\frac{\mathbf{W}}{Y^2}. \tag{4}$$

Now we view the second set of equations in Eqs. (3) as algebraic equations in $\widetilde{\mathfrak{D}}(\mathfrak{p})$ and solve them for the $\mathbf{P}_\mu$.

Theorem 2 *Solutions* $\mathbf{P}_\mu$ *to Eqs.* (3) *(*$\lambda = 1$*) are given by:*

$$\mathbf{P}_\mu = \mathbf{D}^{-1}\,\mathbf{A}_\mu{}^\nu \mathbf{L}_{\nu,4} \tag{5}$$

with $\mathbf{A}_\mu{}^\nu = -\mathbf{D}_4'\,\delta_\mu^\nu + \frac{i}{2}\left[\left(\mathbf{Q}_2 + \frac{1}{4}\right)\delta_\mu^\nu - \frac{3}{2}\mathbf{L}_\mu^\nu - \mathbf{L}_{\mu,\rho}\mathbf{L}^{\rho,\nu} - \mathbf{Q}_4\epsilon_{\mu\rho\tau}^{\nu}\mathbf{L}^{\rho\tau}\right]Y -$ $\left[\left(\mathbf{Q}_2 + \frac{1}{4} - \mathbf{D}_2'\right)\delta_\mu^\nu - \mathbf{L}_\mu^\nu - \mathbf{L}_{\mu,\rho}\mathbf{L}^{\rho,\nu}\right]Y^2 + i\left(\frac{1}{2}\delta_\mu^\nu - \mathbf{L}_\mu^\nu\right)Y^3$ *and* $\mathbf{D} = \mathbf{Q}_4 +$ $\frac{1}{4}\mathbf{Q}_2 - \mathbf{D}_4' + \frac{3}{16}\mathbf{I} + i\left(\mathbf{Q}_2 + \frac{1}{2}\right)Y - \left(\mathbf{Q}_2 - \mathbf{D}_2' - \frac{1}{2}\right)Y^2 + 2iY^3$. *Furthermore* Y^2 *satisfies the equation*

$$Y^4 + \mathbf{D}_2' Y^2 + \mathbf{D}_4' = 0\,. \tag{6}$$

The proof of this theorem involves straightforward, tedious calculation. First one shows that $\mathbf{P}_0 = \mathbf{D}^{-1}\,\mathbf{A}_0{}^\nu \mathbf{L}_{-1,\nu}$ satisfies Eqs. (3), and then use $\mathbf{P}_i = \Big[\mathbf{L}_{0i} \ , \ \mathbf{P}_0\Big]$ to easily show the same is true for other components. For more details on the proofs of both theorems see Ref. [5].

Now we introduce a $*$ structure on $\widetilde{\mathfrak{D}}(\mathfrak{p})$ which in view of Eqs. (3) and Theorem 2 induces a $*$ structure on $\widetilde{\mathfrak{D}}(\mathfrak{g})$ via the homomorphism τ_λ. Let † be the adjoint associated with this $*$ structure, then we have

Theorem 3 *If* $\widetilde{\mathbf{L}}_{\mu,\nu}^\dagger = -\widetilde{\mathbf{L}}_{\mu,\nu}$, $\widetilde{\mathbf{L}}_{4,\mu}^\dagger = -\widetilde{\mathbf{L}}_{4,\mu}$ *and if* $\widetilde{Y}^\dagger = \widetilde{Y}$, *then* $\widetilde{\mathbf{P}}_\mu = \left(\widetilde{\mathbf{D}}^{-1}\widetilde{\mathbf{A}}_\mu^\rho \widetilde{\mathbf{L}}_{-1,\rho}\right)$ *and also* $\widetilde{\mathbf{P}}_\mu^\dagger = \left(\widetilde{\mathbf{D}}^{-1}\widetilde{\mathbf{A}}^{\mu,\rho}\widetilde{\mathbf{L}}_{-1,\rho}\right)^\dagger = \widetilde{\mathbf{L}}_{-1,\rho}^\dagger \widetilde{\mathbf{A}}^{\mu,\rho\dagger}\left(\widetilde{\mathbf{D}}^\dagger\right)^{-1} = -\widetilde{\mathbf{P}}_\mu$. *Furthermore* $\left[\hat{\widetilde{\mathbf{P}}}_\mu \, , \, \hat{\widetilde{\mathbf{P}}}_\nu\right] = 0$.

A proof of this theorem can be found in [5].

Although a representation of $\mathfrak{g}$ always gives rise to a representation of the enveloping algebra $\mathcal{U}(\mathfrak{g})$, it does not necessarily give a representation of the skew field. We have instead the following [5, 6]:

Theorem 4 *Let* $(d\pi, \mathcal{H})$ *be an infinitesimally unitarizable representation of* $\mathfrak{g}$ *on an Hilbert space* $\mathcal{H}$, *and let* $\widetilde{Y}$ *be a self-adjoint operator on* $\mathcal{H}$ *which satisfies* $\widetilde{Y}^4 + d\pi(\widetilde{\mathbf{D}}_2')\widetilde{Y}^2 + d\pi(\widetilde{\mathbf{D}}_4') = 0$. *Then, if both* $d\pi(\widetilde{\mathbf{D}})^{-1}$ *and* $d\pi(\widetilde{\mathbf{D}})^{\dagger^{-1}}$ *exist on a suitable, dense domain in* $\mathcal{H}$, *there exists a skew symmetric representation* $d\widetilde{\pi}$ *of* $\mathfrak{p}$ *on* $\mathcal{H}$ *defined by:* $d\widetilde{\pi}(\mathbf{L}_{i,j}) = d\pi(\widetilde{\mathbf{L}}_{i,j})$, $d\widetilde{\pi}(\mathbf{P}_0) = d\pi(\widetilde{\mathbf{D}})^{-1}\, d\pi(\sum_{\mu=0}^{3} \widetilde{\mathbf{A}}_0^\mu \, \widetilde{\mathbf{L}}_{-1,\mu})$, *and* $d\widetilde{\pi}(\mathbf{P}_i) = [d\widetilde{\pi}(\mathbf{L}_{i,0}), d\widetilde{\pi}(\mathbf{P}_0)](i = 1, 2, 3)$.

Note that it is necessary for both of the operators $d\pi(\widetilde{\mathbf{D}})^{-1}$ and $d\pi(\widetilde{\mathbf{D}})^{\dagger^{-1}}$ to exist on the suitable dense domain in $\mathcal{H}$ postulated in this theorem for the existence of mutually commuting translation operators $d\widetilde{\pi}(\mathbf{P}_\mu)$ on the representation space $\mathcal{H}$ of the representation $(d\pi, \mathcal{H})$.

4 Representations

Let $\chi_\nu \in \mathbb{C}$ be a complex character of A, i.e. $\chi_\nu(a(t)) = e^{(\nu+3/2)t}$ with $\nu \in \mathbb{C}$ and let $(\pi^{(\sigma,\epsilon)}, V^{(\sigma,\epsilon)})$ be the representation $\pi^\sigma \otimes \chi_\epsilon$ of M where $\chi_\epsilon \in \widehat{\{1, \sigma_3\}}$ $(\epsilon = 0, 1)$ and π^σ is a unitary representation of $\mathrm{SL}(2, \mathbb{R})$ on the complex vector space $V^{(\rho,\epsilon)}$. Consider $\pi^{(\sigma,\epsilon)} \otimes \chi_\nu : MA \to V^{(\sigma,\epsilon)}$ and extend this to a representation from P to $V^{(\sigma,\epsilon)}$ by requiring that it act trivially on N.

Definition 1 Let $G = \mathrm{Sp}(2, \mathbb{R})$ and consider the space

$$\mathcal{H}^{(\sigma,\epsilon,\nu)} := Ind_P^G(\pi^{(\sigma,\epsilon)} \otimes \chi_\nu \otimes 1_N) = \Big\{ f : G \to V^{(\sigma,\epsilon)} \mid f \in C^\infty(G) \ni$$

$$f(gman) = \pi^{(\sigma,\epsilon)}(m)\chi_\nu(a) f(g) \text{ for } g \in G, m \in M, a \in A, n \in N \Big\}.$$

For $f \in Ind_P^G(\pi^{(\sigma,\epsilon)} \otimes \chi_\nu \otimes 1_N)$ and $g \in G$ we define a representation $\pi^{(\sigma,\epsilon,\nu)}$ of G on $\mathcal{H}^{(\sigma,\epsilon,\nu)}$ by $\pi^{(\sigma,\epsilon,\nu)}(g) f(g') = f(gg')$ with $g' \in G$.

Let H be the lift-up of $SO_0(1,3)$ specified by the exact sequence in Eq. (2) and concretely determined by the Lie algebra homomorphism φ given in Sect. 1. Then, according to Sekiguchi, we have $Sp(2,\mathbb{R}) \cong HAN$ as a C^∞ isomorphism of HAN onto a dense open subset of $G = Sp(2,\mathbb{R})$. Furthermore, the Cartan decomposition $H \cong VM$ where $V \cong H/M$ gives us the isomorphism $Sp(2,\mathbb{R}) \cong VMAN$ again onto a dense open subset of $Sp(2,\mathbb{R})$. From this it follows that any f in $\mathcal{H}^{(\sigma,\epsilon,\nu)}$ is essentially uniquely specified by its values on $V \cong H/M$. Thus the representation space $\mathcal{H}^{(\sigma,\epsilon,\nu)}$ can equivalently be viewed essentially as the space of C^∞ functions on V with values in $V^{(\sigma,\epsilon)}$. I do not state the required asymptotic conditions on $f \in \mathcal{H}^{(\sigma,\epsilon,\nu)}$. They are determined in a way analogous to that described in Ref. [7] for SU(2, 2) and in Ref. [8] for $S0_0(2,4) \cong SU(2,2)/\mathbb{Z}_2$.)

Now we come to a description of tachyonic representations of the Poincaré group. Consider the double cover $H \ltimes T^4$ of the Poincaré group, $SO_0(1,3) \ltimes T^4$. The translation subgroup $T^4 = \{e^{a^\mu \mathbf{P}_\mu} | a_\mu \in \mathbb{R}^4\}$. T^4 is an additive vector group and so every unitary irreducible representation (UIR) of T^4 is one-dimensional and of the form [15, 16]

$$\chi_p : T^4 \to \mathbb{C}, \quad a \to \chi_p(a) = \exp\{(ip \cdot a)\},$$

where $p, a \in \mathbb{R}^4$ and $p \cdot a$ is the $SO_0(1,3)$ invariant scalar product of the two vectors p and a. It follows that we can characterize the equivalence classes of the UIR's of T^4 by elements p of the vector space dual $\hat{T}^4$ to T^4. The coadjoint action of $SO_0(1,3)$ on $\hat{T}^4$ is given by $p \to \Lambda^{-1} p$. Let $\mathcal{O}_{p_0}$ be the orbit in $\hat{T}^4$ of a point $p_0 \in \hat{T}^4$ under the action of H and let M_{p_0} be the isotropy subgroup (stabilizer subgroup) of the point p_0. Clearly M_{p_0} is a closed subgroup of H and so $\mathcal{O}_{p_0} \cong H/M_{p_0}$. Let $\gamma : \mathcal{O}_{p_0} \to H$ be a smooth cross-section such that for any point $p \in \mathcal{O}_{p_0}$ we have $\gamma(p)p_0 = p$. For tachyonic representations of $H \ltimes T^4$ it suffices to consider p_0 of the form $p_0 = (0,0,0,\mu)$ where $\mu \in \mathbb{R}$. For such p_0 we have $M_{p_0} \cong M$ and $\mathcal{O}_{p_0} = H/M \cong \{p \in T^4 \mid p \cdot p = -|\mu|^2 < 0\}$. We use the same representation $\pi^{(\sigma,\epsilon)}$ of M on the space $V^{(\sigma,\epsilon)}$ as for $Sp(2,\mathbb{R})$, and we extend it to a representation of the semidirect product $B = M \ltimes T^4$ by requiring $\pi^{(\sigma,\epsilon)} \otimes \chi_{p_0}(m,a) = \pi^{(\sigma,\epsilon)}(m)\ \chi_{p_0}(a)$ where $(m,a) \in B$ with $m \in M$ and $a \in T^4$. The representation of $H \ltimes T^4$ induced from $\pi^{(\sigma,\epsilon)} \otimes \chi_{p_0}$ is defined as follows:

Definition 2 Let $p_0 = (0,0,0,\mu)$ where $\mu \in \mathbb{R}$ and let

$$\mathcal{H}^{(\sigma,\epsilon,p_0)} := Ind_B^{H\ltimes T^4}(\pi^{(\sigma,\epsilon)} \otimes \chi_{p_0}) = \Big\{ f : H \ltimes T^4 \to V^{(\sigma,\epsilon)} \mid f \in C^\infty(H \ltimes T^4) \ni$$

$$f(gmp) = \pi^{(\sigma,\epsilon)}(m)\chi_{p_0}(a) f(g) \ for\ g \in H \ltimes T^4, m \in M, a \in T^4 \Big\}.$$

For $f \in Ind_B^{H\ltimes T^4}(\pi^{(\sigma,\epsilon)} \otimes \chi_{p_0})$ and $g \in H \ltimes T^4$ we define the representation $\pi^{(\sigma,\epsilon,p_0)}$ of $H \ltimes T^4$ on $\mathcal{H}^{(\sigma,\epsilon,p_0)}$ by $\pi^{(\sigma,\epsilon,\nu)}(g) f(g') = f(gg')$ with $g' \in H \ltimes T^4$.

As in the $\mathrm{Sp}(2,\mathbb{R})$ case any $f \in \mathcal{H}^{(\sigma,\epsilon,p_0)}$ is uniquely specified by its values on $V \cong H/M$.

From Definitions 1 and 2 it is clear that tachyonic representations of $H \ltimes T^4$ are associated, in the sense of Sect. 3, with the representations $\pi^{(\sigma,\epsilon,\nu)}$ of $\mathrm{Sp}(2,\mathbb{R})$ and that these representations should go over into tachyonic representations of $H \ltimes T^4$ in the contraction limit, where by contraction limit we mean in the (global) sense of Ref. [9]. In fact, the method given in Ref. [9] for the contraction limit of the principal series unitary representations of a non-compact semisimple Lie group G into its associated Cartan motion group $K \ltimes V$ for (G, K) a non-compact Riemannian symmetric pair should carry over to the case considered here and this should provide an example of their contraction process for a non-compact semi-Riemannian symmetric pair not of rank one, namely, $(\mathrm{Sp}(2,\mathbb{R}), H)$.

For the action of the two Casimir operators $\mathbf{D}_2$ and $\mathbf{D}_4$ in the representation $\pi^{(\sigma,\epsilon,\nu)}$ we obtain the following: ($\nu = i\rho$, ρ real and I is the identity)

$$\pi^{(\sigma,\epsilon,\nu)}(\mathbf{D}_2) = -\left(\rho^2 + \sigma(\sigma+1) + \frac{9}{4}\right) I \ , \ \pi^{(\sigma,\epsilon,\nu)}(\mathbf{D}_4) = \left(\rho^2 + \frac{1}{4}\right)\sigma(\sigma+1) I.$$

Substitution of these values into Eqs. (4) and solving for $Y^2 (= \mathbf{P}^2)$ and $\mathbf{W}$ gives the following possibilities for the action of $\mathbf{P}^2$ and $\mathbf{W}$ in this representation

$$\tilde{\pi}^{(\sigma,\epsilon,\nu)}(\mathbf{P}^2) = \rho^2 \ , \ \tilde{\pi}^{(\sigma,\epsilon,\nu)}(\mathbf{W}) = \rho^2 \sigma(\sigma+1)$$

or

$$\tilde{\pi}^{(\sigma,\epsilon,\nu)}(\mathbf{P}^2) = \left(\sigma + \frac{1}{2}\right)^2 - \frac{1}{2} \ , \ \tilde{\pi}^{(\sigma,\epsilon,\nu)}(\mathbf{W}) = \left(\rho^2 + \frac{1}{4}\right)\left(\left(\sigma + \frac{1}{2}\right)^2 - \frac{1}{2}\right).$$

Since the $\mathbf{P}_\mu$ are skew symmetric translation generators, this result implies, for the first possibly, that representations of $H \ltimes T^4$ obtained out of Theorem 4 are tachyonic. A study of the unitary representations of $\mathrm{SL}(2,\mathbb{R})$ shows that we also obtain imaginary mass with the second possibility for the case of $\mathrm{Sp}(2,\mathbb{R})$ representations induced from discrete series unitary representations of $\mathrm{SL}(2,\mathbb{R})$. We leave it to the reader to show that the hypotheses of Theorems 3 and 4 hold true for at least some of these representations, so that we get skew symmetric representations of $\mathfrak{p}$ (cf. Refs. [5, 6] where some special cases are worked out).

Finally, we conclude with a few remarks about possible physical relevance of such tachyonic representations of the Poincaré group. Some of the tachyonic representations described in the previous paragraph occur as contractions of his case 4 representations in Ehrman's classification of the unitary representations of the universal covering group of $\mathrm{SO}_0(2,3)$ [10, 11]. They include contractions of principal series representations of $\mathrm{SO}_0(2,3)$ to representations of the Poincaré group [10]. In addition to calling attention to recent radical proposals for their possible use in dark matter [12], we think it should also be possible to use

such tachyonic representations for describing constituents in the construction of elementary particles like the muon [13]. In fact, since tachyons cannot exist as physically observable particles, it provides a method of confinement similar to the kinematical confinement of Flato and Fronsdal in their description of massless particles as tensor products of Di's and Rac's [14].

References

1. P. Moylan, Parallelizations for induced representations of SO(2,q), in *Proceedings in Mathematics and Statistics*, ed. by V.K. Dobrev, vol. 263 (Springer, Berlin, 2018). https://doi.org/10.1007/978-981-13-2715-5_12
2. A.W. Knapp, *Representation Theory of Semisimple Groups: An Overview Based on Examples* (Princeton University Press, Princeton, 2001)
3. P. Bozek, M. Havlíček, O. Navrátil, Charles Univ. Preprint (1985)
4. M. Havlíček, P. Moylan, J. Math. Phys. **34**(11), 5320–5332 (1993)
5. P. Moylan, Representations of classical Lie algebras from their quantum deformations, in *Institute of Physics Conference Series*, ed. by J.P. Gazeau, R. Kerner, vol. 173 (CRC Press, Boca Raton, FL, 2003), pp. 683–687
6. P. Moylan, Tachyons and representations of the de Sitter groups (Chapter 2), in *Studies in Mathematical Physics Research*, ed. by C.V. Benton (Nova Science, Piscataway, NJ, 2004)
7. V.K. Dobrev, P. Moylan, Elementary intertwining operators for SU(2,2): II. Fortschr. der Phys. **42**, 339–392 (1994)
8. J. Hebda, P. Moylan, A topological criterion for group decompositions. Math. Proc. Camb. Phil. Soc. **103**, 285–298 (1988)
9. A.H. Dooley, J.W. Rice, On contractions of semisimple Lie groups. Trans. Am. Math. Soc. **289**, 185–202 (1985)
10. J.B. Ehrman, On the unitary irreducible representations of the universal covering group of the 3+2 DeSitter group. Math. Proc. Cambr. Phil. Soc. **53**, 290 (1956)
11. N.T. Evans, Discrete series for the universal covering group of the 3 + 2 de Sitter group. J. Math. Phys. **8**, 170 (1967)
12. P.C.W. Davies, Tachyonic dark matter, arXiv:astro-ph/0403048 (2004). https://arxiv.org/abs/astro-ph/0403048
13. P. Moylan, Mass formulae for the Pion and Muon, Preprint: 88-0038 Sofia, BG (1988). http://inspirehep.net/record/22955/
14. M. Flato, C. Fronsdal, Interacting singletons. Lett. Math. Phys. **44**, 249–259 (1998)
15. E.P. Wigner, On unitary representations of the inhomogeneous Lorentz group. Ann. Math. **40**, 149–204 (1939)
16. V. Bargmann, E.P. Wigner, Group theoretical discussion of relativistic wave equations. Proc. Nat. Acad. Sci. (USA) **34**(5), 211–223 (1948). https://doi.org/10.1073/pnas.34.5.211

A Confined Quasi-Maximally Superintegrable N-dimensional System, Classical and Quantum, in a Space with Variable Curvature

Orlando Ragnisco

Abstract In the present paper I will briefly summarize some recent results about the solvability of the classical and quantum version of a (hyper-)spherically symmetric N-dimensional system living on a curved manifold characterized by a conformally flat metric. The system appears as a generalization of the so-called Taub–NUT system. We call it Quasi-Maximally Superintegrable (QMS) since it is endowed with $2N-2$ constants of the motion (with $2N-1$ it would have been Maximally Superintegrable (MS)) functionally independent and Poisson commuting in the Classical case, algebraically independent and commuting as operators in the Quantum case. The eigenvalues and eigenfunctions of the quantum system are explicitly given, while for the classical version we provide the analytic solution of the radial equation of motion. A few comments about the connection between exact solvability and superintegrability are made in the final part of the paper.

Keywords Hamiltonian systems · Superintegrability · Exact solvability

1 The Classical Model: General Features

We consider the two-parameter family of N-dimensional (ND) classical Hamiltonian systems (a degenerate case of the Perlick's II system, introduced by V. Perlick in the seminal paper [1], where a highly nontrivial extension of Bertrand theorem [2] to curved spacetimes was proven), given by

$$\mathcal{H} = \mathcal{T}(\mathbf{q},\mathbf{p}) + \mathcal{U}(\mathbf{q}) = \left(1 - \frac{|\mathbf{q}|}{\xi}\right)\frac{|\mathbf{q}|}{\xi}\,\frac{\mathbf{p}^2}{2} + \frac{k}{\xi - |\mathbf{q}|} \tag{1}$$

O. Ragnisco (✉)
Roma TRE University, Rome, Italy
e-mail: ragnisco@fis.uniroma3.it; http://webusers.fis.uniroma3.it/~ragnisco

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_13

where ξ and k are positive real numbers, $\mathbf{q}, \mathbf{p} \in \mathbb{R}^N$ are conjugate coordinates and momenta with canonical Poisson bracket $\{q_i, p_j\} = \delta_{ij}$ and

$$\mathbf{q}^2 = \sum_{i=1}^{N} q_i^2, \qquad \mathbf{p}^2 = \sum_{i=1}^{N} p_i^2, \qquad |\mathbf{q}| = (\mathbf{q}^2)^{\frac{1}{2}}.$$

To have a positive definite Hamiltonian, the position variables have to be restricted to the (punctured) open ball $(0 < |\mathbf{q}| < \xi)$.

The Hamiltonian $\mathcal{H}$ can also be written in terms of hyperspherical coordinates r, θ_j, (and canonical momenta p_r, p_{θ_j}), $(j = 1, \ldots, N-1)$ defined by

$$q_j = r \cos\theta_j \prod_{k=1}^{j-1} \sin\theta_k, \quad 1 \le j < N, \qquad q_N = r \prod_{k=1}^{N-1} \sin\theta_k \tag{2}$$

so that

$$r = |\mathbf{q}|, \qquad \mathbf{p}^2 = p_r^2 + r^{-2}\mathbf{L}^2, \qquad \mathbf{L}^2 = \sum_{j=1}^{N-1} p_{\theta_j}^2 \prod_{k=1}^{j-1} \frac{1}{\sin^2\theta_k}.$$

Thus, for a given value of $\mathbf{L}$, the Hamiltonian (1) becomes a 1D radial system:

$$\mathcal{H}(r, p_r) = \mathcal{T}(r, p_r) + \mathcal{U}(r) = \left(1 - \frac{r}{\xi}\right)\frac{r}{\xi}\left(\frac{p_r^2}{2} + \frac{\mathbf{L}^2}{2r^2}\right) + \frac{k}{\xi - r} \tag{3}$$

where $r \in (0, \xi)$.

1.1 Metrics and Scalar Curvature

The classical Hamiltonian (we introduce an explicit dependence upon ξ in the notation):

$$\mathcal{H}_\xi = \mathcal{T}_\xi(r, p_r) + \mathcal{U}_\xi(r) = \frac{1}{f_\xi^2(r)}\left(\frac{p_r^2}{2} + \frac{\mathbf{L}^2}{2r^2}\right) + \mathcal{U}_\xi(r) \tag{4}$$

describes a particle (with unit mass) on an ND hyperspherically symmetric space under the action of the central potential $\mathcal{U}_\xi(r) = \frac{k}{\xi - r}$, with $k, \xi > 0$.

In radial coordinates the ND hyperspherically symmetric metrics reads:

$$ds_N^2 = f_\xi^2(r)(dr^2 + r^2 d\Omega_{N-1}^2) \tag{5}$$

where $d\Omega_{N-1}^2 \doteq \sum_{j=1}^{N-1} d\theta_j^2 \prod_{k=1}^{j-1} \sin^2(\theta_k)$, which is the metrics of the unit hypersphere $\mathbb{S}^{N-1}$.

In (5) $f_\xi(r) = \frac{\xi}{\sqrt{(\xi - r)r}} > 0$.

The metrics is singular in the limits $r \to 0$ and $r \to \xi$, and the scalar curvature turns out to be:

$$R_\xi^{(N)}(r) = -(N-1)\frac{(N-4)f_\xi'^2(r) + f_\xi(r)(2f_\xi''(r) + \frac{2(N-1)}{r}f_\xi'(r))}{f_\xi^4(r)},$$

i.e.:

$$R_\xi^{(N)}(r) = (N-1)\frac{(N-2)(4r^2+3\xi^2) - 4\xi r(2N-3)}{4\xi^2 r(\xi - r)} \quad (0 < r < \xi)\,. \tag{6}$$

1.2 Solution of the Radial Equation of Motion

We do not provide the complete solution (time evolution and trajectory) to the dynamics of the classical system but focus our attention on the time behavior of the radial variable. Indeed, starting from the expression of the radial Hamiltonian, introducing the rescaled variable $\rho = \frac{r}{\xi}$, and taking into account that consequently $p_\rho = \xi p_r$, we readily see that the radial equation of motion can be written as the following first order ordinary quadratic differential equation (on the energy surface $\mathcal{H} = E$):

$$(\dot{\rho})^2 = -\rho^2(\varepsilon + l^2) + \rho(-\kappa + \varepsilon + 2l^2) - l^2) \tag{7}$$

where $\varepsilon = \frac{2E}{\xi^2}$, $l^2 = \frac{\mathbf{L}^2}{\xi^4}$, $\kappa = \frac{2k}{\xi^3}$.

Denoting by $\rho_\pm$ the (real and positive) roots of the above quadratic polynomial with $\rho_+ > \rho_-$, $\rho_+ < 1$, and by a the negative quantity $-(\varepsilon + l^2)$, (7) can be cast in the form:

$$\dot{\rho} = \pm\sqrt{|a|(\rho_+ - \rho)(\rho - \rho_-)}. \tag{8}$$

By setting (Euler substitution):

$$|a|(\rho_+ - \rho)(\rho - \rho_-) = |a|y\frac{(\rho_+ - \rho_-)}{|a| + y^2},$$

the differential equation (8) can be integrated for the variable y, whence the following expression for the variable ρ can be finally obtained:

$$\rho(t) = \rho_- \sin^2(\omega t + \alpha) + \rho_+ \cos^2(\omega t + \alpha)$$

$$= \frac{\rho_+ - \rho_-}{2} \cos(2(\omega t + \alpha)) + \frac{\rho_+ + \rho_-}{2} \tag{9}$$

with $2\omega = |a|^{\frac{1}{2}}$, and α an arbitrary phase.

So, we have a simple harmonic motion with frequency given by $|a|^{\frac{1}{2}}$. We notice that both the frequency and the amplitude are (algebraic) functions of the constants of motion ε and l^2, as well as of the coupling constant k.

Remark 1 One could ask about the possibility of manufacturing superintegrable examples leading to a radial time evolution such that $(\dot{r})^2$ be given by a higher degree polynomial. The simplest, and possibly most interesting case, would be a third degree polynomial, entailing its solvability in terms of Weierstrass $\mathcal{P}$ function. Work is in progress in that direction.

2 The Quantum Model

Hereafter, we will use the standard definitions for the quantum position $\hat{\mathbf{q}}$ and momentum $\hat{\mathbf{p}}$ operators:

$$\hat{q_i}\,\psi(\mathbf{q}) = q_i, \quad \hat{p_i}\,\psi(\mathbf{q}) = -\mathrm{i}\hbar \frac{\partial\,\psi(\mathbf{q})}{\partial \mathrm{q_i}}, \quad [\hat{\mathrm{q}}_\mathrm{i}, \hat{\mathrm{p}}_\mathrm{j}] = \mathrm{i}\hbar\delta_{\mathrm{ij}}, \qquad \mathrm{i,j} = 1, \ldots, \mathrm{N},$$

together with the conventions

$$\nabla = \left(\frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_N}\right), \quad \Delta = \nabla^2 = \frac{\partial^2}{\partial^2 q_1} + \cdots + \frac{\partial^2}{\partial^2 q_N}, \quad \mathbf{q}\cdot\nabla = \sum_{i=1}^{N} q_i \frac{\partial}{\partial q_i}.$$

Note that the operator $|\hat{\mathbf{q}}|$ is defined as $|\hat{\mathbf{q}}|\,\psi(\mathbf{q}) = |\mathbf{q}|\,\psi(\mathbf{q})$.

We will apply the so-called direct Schrödinger quantization prescription, and will take the hyperspherical coordinates (2) together with the usual definition of the linear momentum operators, namely

$$\hat{p}_r = -\mathrm{i}\hbar\frac{\partial}{\partial \mathrm{r}}, \qquad \hat{\mathrm{p}}_{\theta_\mathrm{j}} = -\mathrm{i}\hbar\frac{\partial}{\partial \theta_\mathrm{j}}, \qquad \mathrm{j} = 1, \ldots, \mathrm{N}-1, \tag{10}$$

so that the quantum radial Hamiltonian $\hat{H}_r$ reads

$$\hat{H}_r = \frac{1}{2}\left(1 - \frac{r}{\xi}\right)\frac{r}{\xi}\left(\frac{1}{\hat{r}^{N-1}}\,\hat{p}_r\,\hat{r}^{N-1}\,\hat{p}_r + \frac{\hat{\mathbf{L}}^2}{\hat{r^2}}\right) + \frac{k}{\xi - \hat{r}} \tag{11}$$

where $\hat{\mathbf{L}}^2$ is the square of the total quantum angular momentum operator, i.e.

$$\hat{\mathbf{L}^2} = \sum_{j=1}^{N-1} \left(\prod_{k=1}^{j-1} \frac{1}{\sin^2 \theta_k} \right) \frac{1}{(\sin \theta_j)^{N-1-j}} \, \hat{p}_{\theta_j} (\sin \theta_j)^{N-1-j} \, \hat{p}_{\theta_j}.$$

After reordering terms, we arrive at the following Schrödinger equation

$$\left(1 - \frac{r}{\xi}\right) \frac{r}{\xi} \left(-\frac{\hbar^2}{2} \partial_r^2 - \frac{\hbar^2 (N-1)}{2r} \, \partial_r + \frac{\hat{\mathbf{L}}^2}{2r^2} \right) \Psi(r, \boldsymbol{\theta}) + \frac{k}{\xi - r} \Psi(r, \boldsymbol{\theta}) = E \, \Psi(r, \boldsymbol{\theta}), \tag{12}$$

with $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_{N-1})$. By taking into account that the hyperspherical harmonics $Y(\boldsymbol{\theta})$ are such that

$$\hat{\mathbf{L}}^2 Y(\boldsymbol{\theta}) = \hat{C}_{(N)} Y(\boldsymbol{\theta}) = \hbar^2 l (l + N - 2) \, Y(\boldsymbol{\theta}), \quad l = 0, 1, 2 \ldots$$

where l is the angular momentum quantum number, the Eq. (12) admits a complete set of factorized solutions of the form

$$\Psi(r, \boldsymbol{\theta}) = \Phi(r) Y(\boldsymbol{\theta}), \tag{13}$$

and, moreover,

$$\hat{C}_{(m)} \Psi = c_m \Psi, \quad m = 2, \ldots, N \tag{14}$$

where c_m are the eigenvalues of the "Casimir" operators $\hat{C}_{(m)}$ $(m = 2, \ldots, N)$:

$$\hat{C}_{(m)} = \sum_{N-m<i<j\leq N} (\hat{q}_i \hat{p}_j - \hat{q}_j \hat{p}_i)^2, \qquad \hat{C}_{(N)} = \hat{\mathbf{L}}^2. \tag{15}$$

We notice that, being a hyperspherically symmetric system, our system is quasi-maximally superintegrable, since it possesses further $N - 1$ commuting operators, having the same expression as the "Casimir" operators (15), up to a reshuffling of the summations. Technically, together with the $right$ Casimirs , we have the $left$ Casimirs, defined as follows $(m = 0, \ldots, N - 2)$:

$$\tilde{\hat{C}}_{(N-m)} = \sum_{m<i<j\leq N} (\hat{q}_i \hat{p}_j - \hat{q}_j \hat{p}_i)^2, \qquad \tilde{\hat{C}}_{(N)} = \hat{\mathbf{L}}^2. \tag{16}$$

As $\hat{C}_{(N)}$ and $\tilde{C_{(N)}}$ coincide, we have $2N - 3$ commuting operators. So, the set $\mathcal{H}, \hat{C}_{(m)}, \tilde{\hat{C}}_{(m)}$ consists of $2N - 2$ independent commuting operators, related to the $(N - 1)$ quantum numbers of the angular observables, namely:

$$c_k \leftrightarrow l_{k-1}, \quad k = 2, \ldots, N - 1, \qquad c_N \leftrightarrow l.$$

Hence it follows

$$Y(\boldsymbol{\theta}) \equiv Y^{c_N}_{c_{N-1},..,c_2}(\theta_1, \theta_2, \ldots, \theta_{N-1}) \equiv Y^{l}_{l_{N-2},..,l_1}(\theta_1, \theta_2, \ldots, \theta_{N-1}), \tag{17}$$

Accordingly, the radial Schrödinger equation associated with $\hat{\mathcal{H}}$ reads

$$(\xi - r)r\left(-\frac{\hbar^2}{2\xi^2}\left(\frac{\mathrm{d}^2}{\mathrm{d}r^2} + \frac{N-1}{r}\frac{\mathrm{d}}{\mathrm{d}r} - \frac{l(l+N-2)}{r^2}\right)\right)\Phi(r) + \frac{k}{\xi - r}\Phi(r) = E\,\Phi(r), \tag{18}$$

with $k, \xi > 0$ and $0 < r < \xi$.

So we can formulate the following *Proposition*:

Proposition 1 *The solution of the spectral problem associated with the radial Schrödinger equation exists, and has the explicit closed form given by:*

$$\Phi_{n,l}(r) = c_{n,l}(\xi)\left(1 - \frac{r}{\xi}\right)^{\frac{1}{2}\sqrt{1+\frac{8k\xi}{\hbar^2}}+\frac{1}{2}}\left(\frac{r}{\xi}\right)^{l}\mathcal{P}_n^{\left(\sqrt{1+\frac{8k\xi}{\hbar^2}},\,2l+N-2\right)}\left(\frac{2r}{\xi} - 1\right)$$

$$E_{n,l} = \frac{\hbar^2}{8\xi^2}\left(1+2n+\sqrt{1+\frac{8k\xi}{\hbar^2}}\right)\left(1+2n+4\left(l+\frac{N-2}{2}\right)+\sqrt{1+\frac{8k\xi}{\hbar^2}}\right).$$

In the formula for $\Phi_{n,l}(r)$, $c_{n,l}(\xi)$ are normalization factors and $\mathcal{P}_n^{(\alpha,\beta)}(x)$ are the Jacobi orthogonal polynomials with parameters $\alpha \doteq \sqrt{1+\frac{8k\xi}{\hbar^2}}$ and $\beta \doteq 2l+N-2$.

The Jacobi orthogonal polynomials are defined for $\alpha, \beta > -1$, meaning $N + 2l > 1$, i.e. $N > 1$. Moreover, the eigenfunctions $\Phi_{n,l}(r)$ are square integrable in $r \in (0, \xi)$.

Finally, the (unnormalized) eigenfunctions of the confined Hamiltonian read:

$$\Psi_{n,l,l_{N-2},\ldots,l_1}(r, \boldsymbol{\theta}) \propto Y^{l}_{l_{N-2},..,l_1}(\theta_1, \theta_2, \ldots, \theta_{N-1})\Phi_{n,l}(r)\,, \tag{19}$$

and they are orthogonal with respect to the measure

$$d\mu(r, \boldsymbol{\theta}) \doteq \frac{r^{N-1}}{r(\xi - r)}\sin^{N-2}(\theta_1)\sin^{N-3}(\theta_2)\ldots\sin(\theta_{N-2})dr\,d\theta_1\,d\theta_2\ldots d\theta_{N-1}\,,$$

with

$$\theta_1, \ldots, \theta_{N-2} \in [0, \pi), \qquad \theta_{N-1} \in [0, 2\pi), \quad r \in (0, \xi).$$

In particular, the Jacobi orthogonal polynomials yield a $\delta_{nn'}$ at fixed $l = l'$ (for $l \neq l'$ the orthogonality comes from the $\delta_{ll'}$arising from the hyperspherical harmonics).

Note that the eigenfunctions $\Phi_{n,l}(r)$ are identically zero at $r = \xi$ for each $l \geq 0$. At the origin the situation is different since they are zero for $l \neq 0$ and approach a constant value for $l = 0$.

Remark 2 We point out that the spectrum cannot be written in terms of a single combination of the quantum numbers n and l, so it is not fully degenerate. This is a clear indication that the system is not maximally superintegrable. However, Quasi-Maximal Superintegrability (QMS) is ensured due to the hyperspherical symmetry inside the punctured (open) hyperball $\mathcal{B}_{\xi}^{N} := \{x_j > 0 | \sum_{j=1}^{N} x_j^2 < \xi^2\}$.

3 Concluding Remarks and Future Perspectives

In the paper we have presented an example of a dynamically confined system, i.e. a system whose eigenfunctions are square integrable and whose energy spectrum is discrete by virtue of the functional form of the metrics and of the potential, not because of external boundary conditions. In this sense, it is much similar to the harmonic oscillator or to the Sutherland model. An interesting feature is that the system is exactly solvable, its eigenfunctions being expressed in terms of polynomials (up to an algebraic pre-factor), though not being maximally superintegrable, but just quasi-maximally superintegrable. Actually, the radial system does not seem to be amenable neither to an intrinsic Kepler nor to an intrinsic oscillator. Although a deeper investigation on this delicate point is certainly needed, at the present stage we do not expect extra dynamical symmetries of Laplace–Runge–Lenz or Demkov–Fradkin type [3–6]. Of course, we do not claim that exact solvability and maximal superintegrability are unrelated notions [7, 8]: we just claim that on this topic there is still something that has to be better understood.

Notes and Comments I want to mention that most of the results described in this paper have been obtained in collaboration with my former student Danilo Latini, who got his Ph.D.a couple of years ago. Unfortunately at the moment he has not got any position whatsoever.

References

1. V. Perlick, Bertrand spacetimes. Class. Quantum Grav. **9** 1009–1021 (1992)
2. J. Bertrand, Theoreme relatif au mouvement d'un point attire vers un centre fixe. C. R. Acad. Sci. Paris **77** (1873) 849–853
3. A. Ballesteros, A. Enciso, F.J. Herranz, O. Ragnisco, Bertrand spacetimes as Kepler/oscillator potentials. Class. Quantum Grav. **25**, 165005 (2008)
4. A. Ballesteros, A. Enciso, F.J. Herranz, O. Ragnisco, Hamiltonian systems admitting a Runge-Lenz vector and an optimal extension of Bertrand's theorem to curved manifolds. Commun. Math. Phys. **290**, 1033–1049 (2009)

5. A. Ballesteros, A. Enciso, F.J. Herranz, O. Ragnisco, Superintegrability on N-dimensional curved spaces: central potentials, centrifugal terms and monopoles. Ann. Phys. **324**, 1219–1233 (2009)
6. D. Latini, O. Ragnisco, The classical Taub-Nut system: factorization, spectrum generating algebra and solution to the equations of motion. J. Phys. A Math. Theor. **48**, 175201 (2015)
7. F. Tremblay, A.V. Turbiner, P. Winternitz, Periodic orbits for an infinite family of classical superintegrable systems. J. Phys. A Math. Theor. **43**, 015202 (2010)
8. W. Miller Jr., S. Post, P. Winternitz, Classical and quantum superintegrability with applications. J. Phys. A Math. Theor. **46**, 423001 (2013)

Conditional Discretization of a Generalized Reaction–Diffusion Equation

Decio Levi, Miguel A. Rodríguez, and Zora Thomova

Abstract A PDE modeling a reaction–diffusion physical system is discretized using its conditional symmetries. Discretization is carried out using two specific conditional symmetries. Explicit solutions of the difference equation are constructed when the symmetry is projective.

Keywords Symmetry · Integrable systems · Difference equations · Conditional symmetry · Invariant discretization

1 Introduction

Partial differential equations (PDE) modeling interesting models in Physics are generically hard to solve and exact solutions are rare and difficult to obtain. The Lie symmetries of these equations provide, in many cases, explicit solutions, or at least hints on how to find those solutions. This is a well-known topic, see, for example, the following monographs dedicated to it [1, 3, 15, 16].

A particularly useful application of this method is the reduction of the order of the equation or of the number of variables. The invariants of the vector field corresponding to the infinitesimal symmetry provide a new set of coordinates and the PDE, written in this new system, can be usually solved and solutions of the original equation can be obtained. The well-known example is the construction of the fundamental solution of the heat equation [2].

D. Levi
INFN, Sezione Roma Tre, Roma, Italy
e-mail: decio.levi@roma3.infn.it

M. A. Rodríguez
Depto. de Física Teórica, Universidad Complutense de Madrid, Madrid, Spain
e-mail: rodrigue@ucm.es

Z. Thomova (✉)
Department of Mathematics and Physics, SUNY Polytechnic Institute, Utica, NY, USA
e-mail: zora.thomova@sunypoly.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_14

However, some reductions cannot be obtained by just studying the Lie point symmetries of the equation, as was shown in [5] for the Boussinesq equation. Additional reductions and new solutions can be obtained by considering conditional symmetries [2], as done in [11]. Many articles are dedicated to this class of symmetries, for example [17]. To find conditional symmetries of the given PDE, one adds a condition (first order PDE) to the equation. The prolongation of the vector field of the conditional symmetry annihilates the equation when the condition and its differential consequences are simultaneously satisfied.

Given a PDE with a group of symmetries, the equation can be written in terms of the invariants of these symmetries. We have recently shown [13] that this is also the case for conditional symmetries when the characteristic equation and its differential consequences are satisfied.

In a recent work [14] we have extended these ideas to the discrete case, using the notion of invariant discretization. This concept has been introduced in [12] for Lie point symmetries and applied in a great number of particular examples proving its usefulness in the computation of approximate solutions and, for instance, in the study of the behavior of the solutions in a neighborhood of a singularity. In the case of conditional symmetries, the discrete equation can be written in terms of the discrete invariants of the vector fields and the discretized characteristic equation.

In this work we apply these techniques to a particular equation used in reaction–diffusion models and related to the Hodgkin–Huxley model of action potentials in neurons [4]:

$$u_t = u_{xx} + k(x)u^2(1-u). \tag{1}$$

For particular values of $k(x)$ the conditional symmetries of this equation have been computed in [6–8].

We will study the Eq. (1) with $k(x) = \frac{2}{x^2}$

$$u_t = u_{xx} + 2\frac{u^2}{x^2}(1-u), \quad u = u(x,t), \tag{2}$$

whose conditional symmetries (cases $\hat{Q}_1$ and $\hat{Q}_2$ were given in [7, 8]) are

$$\hat{Q}_1 = \partial_t + \frac{3}{x}(u-1)\partial_x - \frac{3}{x^2}u(u-1)^2\partial_u, \tag{3}$$

$$\hat{Q}_2 = \partial_x + \frac{u^2-1}{x}\partial_u, \tag{4}$$

$$\hat{Q}_3 = \partial_x - \frac{(u-1)(u-3)}{x}\partial_u, \tag{5}$$

$$\hat{Q}_4 = \partial_x - \frac{u(u-1)}{x}\partial_u. \tag{6}$$

We will discuss two of the conditional symmetries, $\hat{Q}_1$ (3), corresponding to the case when the coefficient of ∂_t in the vector field is equal to 1, and Q_4 (6), when the same coefficient is equal to 0.

2 Case $\hat{Q}_1$

Let us discuss here the symmetry reduction and discretization provided by the vector field $\hat{Q}_1$.

2.1 Symmetry Reduced Equations and Solutions

The symmetry variables, that is, the invariants of the vector field, are

$$v = \frac{x(u-1)}{u}, \quad y = x^2\frac{u-3}{u-1} + 18t, \quad v = v(y) \tag{7}$$

for $u \neq 0$ and $u \neq 1$ Then (2) reduces to the ODE

$$vv_{yy} - 2v_y^2 = 0,$$

whose solution is

$$v = \frac{c_1}{y + c_2},$$

where c_1 and c_2 are two integration constants. Then from (7) a solution of (2), different from the trivial constant $u = 1$, is

$$u(x,t) = \frac{x\left(3x^2 + 18t + c_2\right)}{x(x^2 + 18t + c_2) - c_1}.$$

2.2 Construction of the Discretized Equation

To be able to construct the conditionally invariant discretization of (2) we need at first to construct the invariant lattice.

The relevant necessary invariants in the point nm are

$$\mathcal{I}_1 = x_{nm} - \frac{x_{nm}}{u_{nm}}, \quad \mathcal{I}_2 = x_{nm}^2\frac{u_{nm}-3}{u_{nm}-1} + 18t_{nm}.$$

We look for a lattice invariant under the vector field $\hat{Q}_1$. Details on construction of invariant lattices can be found in [9, 10, 14]. Introducing the distances and angles of the discrete net

$$x_{n+1,m} = x_{nm} + h^x_{nm}, \quad x_{n-1,m} = x_{nm} + \sigma^x_{nm},$$

and of the fields

$$\begin{aligned} u_{n+1,m} =& u_{nm} + h^x_{nm} D_x u, \\ u_{n,m+1} =& u_{nm} + \sigma^x_{nm} D_x u + k D_t u, \end{aligned}$$

we get, after some nontrivial calculations, that h^x_{nm} satisfies the polynomial equation

$$\begin{aligned} &(h^x_{nm})^3 (u_{nm} - 1) D_x u + (u-1)(2x_{nm} D_x u + u - 3)(h^x_{nm})^2 \\ &+ \left[\left(2x^2_{nm} - h(u_{nm} - 1) \right) D_x u + 2x_{nm}(u_{nm} - 3)(u_{nm} - 1) \right] h^x_{nm} \\ &- h(u_{nm} - 1)^2 = 0. \end{aligned}$$

This expression will provide h^x_{nm} in terms of x, u and the differences of u in each point. This a cubic equation for h^x_{nm} and we cannot get explicit expressions for the point distances simple enough to achieve a complete solution of the problem.

This is a problem we have not found in previous works and it is due to the fact that the vector field is not projective, that is, the coefficients of the partial derivatives with respect to x (or t, the independent variables) depend on the dependent variable u.

3 Case $\hat{Q}_4$

3.1 Symmetry Reduced Equations and Solutions

The symmetry variables corresponding to the vector field $\hat{Q}_4$ (6) are

$$v = \frac{x(u-1)}{u}, \quad y = t, \quad v = v(y) \tag{8}$$

and then (2) reduces to the ODE

$$v_y = 0,$$

whose solution is $v = c$ where c is a constant. Then, from (8) a family of solutions for (2) different from the trivial constant $u = 1$ is

$$u = \frac{x}{x - c}. \tag{9}$$

Invariants and Reconstruction of the Equation The invariants of the vector field $\hat{Q}_4$ are easily computed

$$I_1 = t, \quad I_2 = x - \frac{x}{u}, \quad I_3 = \frac{x}{u^2} u_t, \tag{10}$$

$$I_4 = \frac{x}{u^2} u_x - \frac{1}{u}, \quad I_5 = \frac{x}{u^2} u_{xx} + \frac{2}{u^2} u_x - \frac{2x}{u^3} u_x^2. \tag{11}$$

The condition (the characteristic equation) of the conditional symmetry $\hat{Q}_4$ is given by

$$\mathcal{C} = u_x + \frac{u}{x}(u - 1) = \frac{u^2}{x}(I_4 + 1) \tag{12}$$

and its x derivative (differential consequence) is

$$\mathcal{C}_x = u_{xx} + \frac{2u - 1}{x} u_x - \frac{u(u - 1)}{x^2}.$$

Equation (2) is obtained in terms of the invariants, the condition, and the differential consequences of the condition:

$$u_t - u_{xx} + \frac{2}{x^2} u^2 (u - 1) = \frac{u^2}{x} \left(I_3 - \frac{x}{u^2} \mathcal{C}_x + \frac{2u - 1}{u^2} \mathcal{C} \right) = 0$$

as it can be done for any conditional symmetry [13].

3.2 Construction of the Discretized Equation

Lattice Construction Due to the form of $\hat{Q}_4$ (6) the lattice will be orthogonal and constant in both directions x and t. From (6) and the results on the construction of invariant lattices (see [9, 10, 14] for details) we get that with no loss of generality we can choose

$$h_{nm}^{(t)} = k, \quad \sigma_{nm}^{(t)} = 0, \quad h_{nm}^{(x)} = h, \quad \sigma_{nm}^{(x)} = 0,$$

where k and h are constants, the lattice spacing in the t and x directions. Then [9] the discrete derivatives are

$$\mathcal{D}_x = \frac{\Delta_n}{h}, \quad \mathcal{D}_t = \frac{\Delta_m}{k}.$$

Conditional Discretization of the Generalized Reaction–Diffusion Eq. (2) Using the prolongation of the discrete vector field [9] we get the following invariants:

$$\mathcal{I}_1 = t_{nm}, \quad \mathcal{I}_2 = x_{nm} - \frac{x_{nm}}{u_{nm}}$$

$$\mathcal{I}_3 = \frac{x_{nm}\mathcal{D}_t u}{u_{nm}(u_{nm} + k\mathcal{D}_t u)}, \quad \mathcal{I}_4 = \frac{x_{nm}D_x u - u_{nm}}{u_{nm}(u_{nm} + h\mathcal{D}_x u)}$$

$$\mathcal{I}_5 = \frac{2u_{nm}D_x u - 2x_{nm}(D_x u)^2 - hx_{nm}D_x u D_{xx}u + u_{nm}(x_{nm} + 2h)D_{xx}u}{u_{nm}(u_{nm} + hD_x u)\left(u_{nm} + 2hD_x u + h^2 D_{xx}u_{nm}\right)}$$

whose continuous limits are the corresponding continuous invariants (10), (11). The discrete characteristic equation of the discrete conditional symmetry is given by

$$\mathcal{C} = \frac{u_{nm}^2}{x_{nm}}(\mathcal{I}_4 + 1) = \frac{u_{nm}\left((x_{nm} + hu_{nm})D_x u + u_{nm}(u_{nm} - 1)\right)}{x_{nm}(u_{nm} + hD_x u)},$$

with continuous limit the characteristic (12) of $\hat{Q}_4$, and with x-difference:

$$\begin{aligned}
\mathcal{D}_x\mathcal{C} = &\frac{1}{x_{nm}(x_{nm} + h)(u_{nm} + hD_x u)\left(u_{nm} + 2hD_x u + h^2 D_{xx}u\right)} \\
&\Big(2h^3 x_{nm}(D_x u)^4 + hx_{nm}(x_{nm} + 7hu_{nm})(D_x u)^3 \\
&+ hu_{nm}((7u_{nm} - 3)x_{nm} - 2hu_{nm})(D_x u)^2 \\
&+ u_{nm}^2((2u_{nm} - 1)x_{nm} - h(3u_{nm} - 2))D_x u \\
&+ \left(h^4 x_{nm}(D_x u)^3 + h^2 x_{nm}(x_{nm} + h + 3hu_{nm})(D_x u)^2\right. \\
&+ hu_{nm}(x_{nm}^2 - h^2 u_{nm} + hx_{nm}(2u_{nm} + 1))D_x u \\
&\left.+ u_{nm}^2(x_{nm}^2 - h^2(u_{nm} - 1) + 2hx_{nm})\right)D_{xx}u - u_{nm}^3(u_{nm} - 1)\Big).
\end{aligned}$$

The discretized equation (2) is written in terms of the discrete invariants and the discrete condition and its differences as:

$$\frac{u_{nm}^2}{x_{nm}}\left(\mathcal{I}_3 + \frac{2u_{nm} - 1}{u_{nm}^2}\mathcal{C} - \frac{x_{nm}}{u_{nm}^2}\mathcal{D}_x\mathcal{C}\right) = 0,$$

with Eq. (2) as its continuous limit. Explicitly this equation is

$$\begin{aligned}
u_{n+2,m} = &ku_{n,m+1}u_{n+1,m}^3 x_{nm}^2(x_{nm} + 2h) \\
&\times\Big[u_{nm}^2(x_{nm} + h)(ku_{n,m+1}(hu_{n+1,m}(h - 3x_{nm}) + x_{nm}^2 - h^2)
\end{aligned} \tag{13}$$

$$+h^2u_{n+1,m}x_{nm}^2)-u_{n,m+1}u_{n+1,m}u_{nm}x_{nm}(x_{nm}+h)$$
$$\times\left(\left(h^2+k\right)x_{nm}-hk\right)-2hku_{n,m+1}u_{nm}^3(x_{nm}+h)$$
$$\times(hu_{n+1,m}-x_{nm}-h)+ku_{n,m+1}u_{n+1,m}^2x_{nm}^2(hu_{n+1,m}+x_{nm}+h)\Big]^{-1}.$$

The existence of this conditional symmetry allows, as in the continuous case, to simplify the equation—reducing the number of variables in the continuous case and reducing the number of indices in the discrete case. In this case we can assume that u_{nm} depends only on the first index (that is, $u_{n,m+1}=u_{nm}$). The difference equation (13) becomes

$$\begin{aligned}u_n=&\big[(x_{n-2}+h)\left(x_{n-2}^2-h(3x_{n-2}-h)u_{n-1}-h^2\right)u_{n-2}^2\\&-x_{n-2}\left(x_{n-2}^2-h^2\right)u_{n-1}u_{n-2}-2h(x_{n-2}+h)(h(u_{n-1}-1)-x_{n-2})u_{n-2}^3\\&+x_{n-2}^2(h(u_{n-1}+1)+x_{n-2})u_{n-1}^2\big]^{-1}x_{n-2}^2(x_{n-2}+2h)u_{n-1}^3.\end{aligned}\tag{14}$$

Equation (2) has the exact conditionally invariant solution (9) depending only on x. It turns out that in this case, the discrete function and lattice:

$$u_n=\frac{x_n}{x_n-c},\quad x_{n+1}=x_n+h\tag{15}$$

are an exact solution of the difference equation (14) for any constant c and any step h. The proof is obtained by a direct substitution of (15) into (14). This implies that the **discrete scheme is exact**. We present in Fig. 1 two plots of these expressions (the continuous and discrete solutions) for $c=-9$ and $c=\pi$, respectively, to graphically describe this situation. In the second plot, the discrete solution fits the continuous one in spite of the singularity.

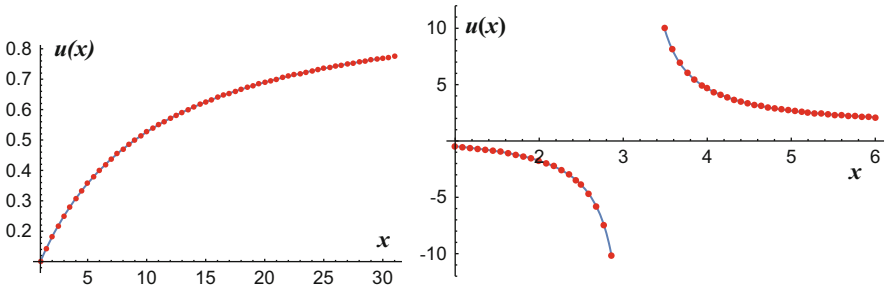


Fig. 1 Solutions of Eqs. (2) and (14), for $c=-9$ (left) and $c=\pi$ (right). The solid curves correspond to the continuous exact solution, the dots to the discrete exact solution

4 Conclusions

In this contribution we have applied our recent work on conditional symmetries of continuous and discrete equations to a reaction–diffusion equation. We have shown how to carry out the necessary computations for one conditional symmetry and the difficulties which appear when the symmetry is not projective. In the solved case we have also shown that the exact solution of the continuous equation becomes an exact solution of the difference equation.

Acknowledgments DL has been supported by INFN IS-CSN4 Mathematical Methods of Nonlinear Physics. MAR was supported by the Spanish MINECO under project PGC2018-094898-B-I00. All authors thank the hospitality of CRM, Montreal (Canada).

References

1. G.W. Bluman, S. Anco, *Symmetry and Integration Methods for Differential Equations* (Springer, New York, 2002)
2. G.W. Bluman, J.D. Cole, The general similarity solutions of the heat equation. J. Math. Mech. **18**, 1025–1042 (1969)
3. G.W. Bluman, S. Kumei, *Symmetries of Differential Equations* (Springer, New York, 1989)
4. B.H. Bradshaw-Hajek, M.P. Edwards, P. Broadbridge, G.H. Williams, Nonclassical symmetry solutions for reaction-diffusion equations with explicit spatial dependence. Nonlinear Anal. **67**, 2541–2552 (2007)
5. P.A. Clarkson, M.D. Kruskal, New similarity reductions of the Boussinesq equation. J. Math. Phys. **30** 2201–2213 (1989)
6. P.G. Estévez, Non-classical symmetries and the singular manifold method: the Burgers and the Burgers-Huxley equations. J. Phys. A: Math. Gen. **27**, 2113–2127 (1994)
7. N.M. Ivanova, On Lie symmetries of a class of reaction-diffusion equations, in *Proceedings of the IV Workshop "Group Analysis of Differential Equations and Integrable Systems"* (2009), pp. 84–86
8. N.M. Ivanova, C. Sophocleous, On nonclassical symmetries of generalized Huxley equations, in *Proceedings of the V Workshop "Group Analysis of Differential Equations and Integrable Systems"* 91–98 (2009), arXiv:1010.2388v1
9. D. Levi, M.A. Rodríguez, Construction of partial difference schemes: I. The Clairaut, Schwarz, Young theorem on the lattice. J. Phys. A Math. Theor. **46**, 295203 (2013)
10. D. Levi, M.A. Rodríguez, On the construction of partial difference schemes II: discrete variables and invariant schemes. Acta Polytech. **56**, 236–244 (2014)
11. D. Levi, P. Winternitz, Nonclassical symmetry reduction: example of the Boussinesq equation. J. Phys. A Math. Gen. **22**, 2915–2924 (1989)
12. D. Levi, P. Winternitz, Continuous symmetries of difference equations. J. Phys. A Math. Gen. **39**, R1 (2006)
13. D. Levi, M.A. Rodríguez, Z. Thomova, Differential equations invariant under conditional symmetries J. Nonlinear Math. Phys. **26** 281–293 (2019)
14. D. Levi, M.A. Rodríguez, Z. Thomova, The discretized Boussinesq equation and its conditional symmetry reduction. J. Phys. A Math. Theor. **53** 045201 (2019)
15. P.J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1993)
16. H. Stephani, *Differential Equations, their Solution using Symmetries* (Cambridge University Press, Cambridge, 1989)
17. R.Z. Zhdanov, I.M. Tsyfra, R.O. Popovich, A precise definition of reduction of partial differential equations. J. Math. Anal. Appl. **238**, 101–123 (1999)

Discrete Curve Flows in Two-Dimensional Cayley–Klein Geometries

Joseph Benson and Francis Valiquette

Abstract Using the method of equivariant moving frames, we study geometric flows of discrete curves in the nine Cayley–Klein planes. We show that, under a certain arc-length preserving flow, the curvature invariant κ_n evolves according to the differential-difference equation $\frac{\partial \kappa_n}{\partial t} = (1 + \epsilon \kappa_{n+1}^2)(\kappa_{n+1} - \kappa_{n-1})$, where the value of $\epsilon \in \{-1, 0, 1\}$ is linked to the geometry of the Cayley–Klein plane.

Keywords Discrete curve flows · Cayley–Klein geometry

1 Introduction

Invariant submanifold flows, particularly curve and surface flows, arise in a wide range of applications, including geometric optics, computer vision, visual tracking and control, and much more. Given a geometric submanifold flow, one of the typical problems consists of determining the induced evolution on the geometric invariants of the submanifold. For many geometric flows, this leads to completely integrable evolution equations.

A similar phenomenon occurs in discrete geometry. For example, discrete geometric curve flows in the Euclidean plane have been considered in [1–3], and it was shown that the curvature evolves, under a certain arc-length preserving flow, according to the differential-difference mKdV equation, which is completely integrable. In this paper, we extend some of the work done in [1–3] by considering discrete geometric curve flows in all nine 2-dimensional Cayley–Klein geometries, [4]. Using the method of equivariant moving frames, our computations are performed symbolically, which allows us to tackle the nine geometries simultaneously.

J. Benson
Macalester College, Saint Paul, MN, USA
e-mail: jbenson4@macalester.edu

F. Valiquette (✉)
Monmouth University, West Long Branch, NJ, USA
e-mail: fvalique@monmouth.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_15

Our computations will closely follow the exposition found in [1]. In this paper, we will not be able to recall and explain all the results used from [1], and therefore recommend that the reader reviews [1] for a more detailed exposition.

The main result of this note is that the curvature invariant of a discrete curve in 2-dimensional Cayley–Klein geometries evolves, under a certain arc-length preserving flow, according to the integrable differential-difference equation [5],

$$\frac{\partial \kappa_n}{\partial t} = (1 + \epsilon \kappa_{n+1}^2)(\kappa_{n+1} - \kappa_{n-1}),$$

where the value of $\epsilon \in \{-1, 0, 1\}$ is related to the geometry of the Cayley–Klein plane.

2 Cayley–Klein Planes

Let $\epsilon_1, \epsilon_2 \in \{-1, 0, 1\}$, and consider the 3-dimensional Cayley–Klein Lie algebra $\mathfrak{so}_{\epsilon_1,\epsilon_2}(3)$ spanned by three vectors P_1, P_2, J_{12}, with nonzero commutators

$$[J_{12}, P_1] = P_2, \qquad [J_{12}, P_2] = -\epsilon_2 P_1, \qquad [P_1, P_2] = \epsilon_1 J_{12}.$$

Exponentiating $\mathfrak{so}_{\epsilon_1,\epsilon_2}(3)$ yields the special orthogonal Cayley–Klein group $SO_{\epsilon_1,\epsilon_2}(3)$ with 1-parameter subgroups

$$H_1 = \{e^{\varepsilon P_1} : \varepsilon \in \mathbb{R}\}, \qquad H_2 = \{e^{\varepsilon P_2} : \varepsilon \in \mathbb{R}\}, \qquad H_{12} = \{e^{\varepsilon J_{12}} : \varepsilon \in \mathbb{R}\}.$$

Definition 1 The *two-dimensional Cayley–Klein planes* are defined as the homogeneous spaces

$$S^2_{[\epsilon_1],\epsilon_2} := SO_{\epsilon_1,\epsilon_2}(3)/H_{12}.$$

The nine geometries that one obtains appear in Table 1.

Table 1 Two-dimensional Cayley–Klein geometries

	Measure of lengths		
Measure of angles	Elliptic $\epsilon_1 = 1$	Parabolic $\epsilon_1 = 0$	Hyperbolic $\epsilon_1 = -1$
Elliptic $\epsilon_2 = 1$	Elliptic geometry	Euclidean geometry	Hyperbolic geometry
Parabolic $\epsilon_2 = 0$	Oscillating Newton–Hooke spacetime	Galilean spacetime	Expanding Newton–Hooke spacetime
Hyperbolic $\epsilon_2 = -1$	Anti-de Sitter spacetime	Minkowski spacetime	de Sitter spacetime

Definition 2 For $\epsilon_2 \in \{-1, 0, 1\}$, the *generalized imaginary number* i_{ϵ_2} is a formal symbol satisfying the equation

$$i_{\epsilon_2}^2 = -\epsilon_2.$$

Using generalized imaginary numbers, a point in $S^2_{[\epsilon_1],\epsilon_2}$ may be represented by the complex number

$$z = x + i_{\epsilon_2} y \quad \in \quad \mathbb{C}_{\epsilon_2}.$$

The induced action of the Cayley–Klein group $SO_{\epsilon_1,\epsilon_2}(3)$ on the point $z \in \mathbb{C}_{\epsilon_2}$ is then given by

$$Z = \frac{\alpha z + \beta}{\overline{\alpha} - \epsilon_1 \overline{\beta} z} \qquad \text{where} \qquad \alpha, \beta \in \mathbb{C}_{\epsilon_2} \qquad \text{with} \qquad \alpha\overline{\alpha} + \epsilon_1 \beta\overline{\beta} = 1, \tag{1}$$

and $\overline{z} = x - i_{\epsilon_2} y$ denotes the complex conjugate of z. Geometrically, (1) corresponds to the isometry group of the metric

$$g = \frac{dz d\overline{z}}{(1 + \epsilon_1 z\overline{z})^2}.$$

Computing the infinitesimal generators of the group action (1), and using the isomorphism $\mathbb{C}_{\epsilon_2} \simeq \mathbb{R}^2$, we obtain

$$\begin{aligned} \mathbf{v}_{12} &= -\epsilon_2 y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}, \qquad \mathbf{v}_1 = [1 + \epsilon_1 (x^2 - \epsilon_2 y^2)] \frac{\partial}{\partial x} + 2\epsilon_1 xy \frac{\partial}{\partial y}, \\ \mathbf{v}_2 &= 2\epsilon_1 \epsilon_2 xy \frac{\partial}{\partial x} + [1 - \epsilon_1 (x^2 - \epsilon_2 y^2)] \frac{\partial}{\partial y}. \end{aligned} \tag{2}$$

3 Moving Frames

Consider a discrete curve $z_n = x_n + i_{\epsilon_2} y_n \in \mathbb{C}_{\epsilon_2}$, where $n \in \mathbb{Z}$. The Cayley–Klein group acts on the curve via the product action

$$Z_n = X_n + i_{\epsilon_2} Y_n = \frac{\alpha z_n + \beta}{\overline{\alpha} - \epsilon_1 \overline{\beta} z_n}.$$

To define a moving frame, we consider the *second order discrete jet space*

$$\mathbf{J}^{[2]} = \mathbb{Z} \times \mathbb{C}^3_{i_{\epsilon_2}}$$

with local coordinates $(n, z_{n-1}, z_n, z_{n+1})$, where the fiber $\mathbb{C}^3_{i_{\epsilon_2}}$ consists of three neighboring points on the curve. We refer to $\mathrm{J}^{[2]}$ as a discrete jet space since the points (z_{n-1}, z_n, z_{n+1}) are sufficient to provide finite difference approximations of the order 2 jet of a differentiable curve. A moving frame is then an equivariant map $\rho\colon \mathrm{J}^{[2]} \to SO_{\epsilon_1,\epsilon_2}(3)$ from the discrete jet space into the Cayley–Klein group. Choosing the cross-section

$$\mathcal{K} = \{z_n = 0,\, y_{n+1} = 0\} \subset \mathrm{J}^{[2]},$$

a moving frame is obtained by requiring that $\rho_n \cdot (z_{n-1}, z_n, z_{n+1}) \in \mathcal{K}$, which yields the *normalization equations* $Z_n = Y_{n+1} = 0$. Solving the normalization equations for the group parameters, we obtain the moving frame

$$\beta = -\sqrt{1 + |z_n|^2}\, e^{i_{\epsilon_2}\theta} z_n, \qquad \alpha = \sqrt{1 + |z_n|^2}\, e^{i_{\epsilon_2}\theta},$$

where $e^{i_{\epsilon_2}\theta} = C_{\epsilon_2}(\theta) + i_{\epsilon_2} S_{\epsilon_2}(\theta)$ is the generalized complex exponential function with

$$C_{\epsilon_2}(\theta) = \begin{cases} \cos\bigl(\sqrt{\epsilon}x\bigr) & \epsilon > 0 \\ 1 & \epsilon = 0 \\ \cosh\bigl(\sqrt{-\epsilon}x\bigr) & \epsilon < 0 \end{cases}, \qquad S_{\epsilon_2}(\theta) = \begin{cases} \dfrac{1}{\sqrt{\epsilon}}\sin\bigl(\sqrt{\epsilon}x\bigr) & \epsilon > 0 \\ x & \epsilon = 0 \\ \dfrac{1}{\sqrt{-\epsilon}}\sinh\bigl(\sqrt{-\epsilon}x\bigr) & \epsilon < 0 \end{cases},$$

denoting the generalized cosine and sine functions, and where the angle θ is determined by the equation

$$T_{\epsilon_2}(2\theta) = -\frac{\Delta x_n \eta + \Delta y_n \xi}{\Delta x_n \xi - \epsilon_2 \Delta y_n \eta},$$

where $T_{\epsilon_2}(\theta) = \dfrac{S_{\epsilon_2}(\theta)}{C_{\epsilon_2}(\theta)}$ denotes the generalized tangent function, and

$$\Delta x_n = x_{n+1} - x_n, \qquad \xi = \mathfrak{Re}(1 + \epsilon_1 z_n \overline{z}_{n+1}),$$
$$\Delta y_n = y_{n+1} - y_n, \qquad \eta = \mathfrak{Im}(1 + \epsilon_1 z_n \overline{z}_{n+1}).$$

Given a moving frame, there is a systematic procedure, known as *invariantization*, for constructing invariant functions. For example, the invariantization of a coordinate function z_k is the invariant $\iota_n(z_k) = \rho_n \cdot z_k$. Invariantizing x_{n+1} and z_{n-1}, we obtain the invariants

$$K_n = \iota_n(x_{n+1}) = \frac{|\Delta z_n|}{|1+\epsilon_1 z_n \overline{z}_{n+1}|},$$
$$I_n + i_{\epsilon_2} J_n = \iota_n(z_{n-1}) = -\frac{\Delta\overline{z}_n \Delta z_{n-1}(1+\epsilon_1\overline{z}_n z_{n+1})(1+\epsilon_1 z_n\overline{z}_{n-1})}{|\Delta z_n||1+\epsilon_1 z_n\overline{z}_{n+1}||1+\epsilon_1 z_n\overline{z}_{n-1}|^2}, \tag{3}$$

where I_n, J_n are the real and imaginary parts of $\iota_n(z_{n-1})$, respectively. The invariantization map ι_n extends to one-forms [1]. For example, the invariantization of $dz_k = dx_k + i_{\epsilon_2} dy_k$ is the invariant one-form

$$\varpi_n^k = \omega_n^k + i_{\epsilon_2}\sigma_n^k = \iota_n(dz_k) = \frac{\Delta\overline{z}_n(1+\epsilon_1 z_n\overline{z}_n)(1+\epsilon_1\overline{z}_n z_{n+1})(1+\epsilon_1 z_n\overline{z}_k)^2 dz_k}{|\Delta z_n||1+\epsilon_1 z_n\overline{z}_{n+1}||1+\epsilon_1 z_n\overline{z}_k|^4},$$

where ω_n^k, σ_n^k are the real and imaginary parts of $\iota_n(dz_k)$, respectively.

4 Recurrence Relations

We now compute the recurrence relations for the exterior derivative and the shift map. These formulas encapsulate the fact that the invariantization map does not, in general, commute with the exterior derivative and the shift map. The general recurrence formulas can be found in [1]. Below, we specialize these formulas to our problem.

4.1 Exterior Derivative

Let μ_1, μ_2, and μ_{12} be a basis of Maurer–Cartan forms dual to the infinitesimal generators (2). Then the recurrence relations for the exterior derivative are

$$\begin{aligned} d[\iota_n(x_k)] &= \omega_n^k + [1+\epsilon_1(\iota_n(x_k)^2 - \epsilon_2\iota_n(y_k)^2]\nu_1 \\ &\quad + 2\epsilon_1\epsilon_2\iota_n(x_k)\iota_n(y_k)\nu_2 - \epsilon_2\iota_n(y_k)\nu_{12}, \\ d[\iota_n(y_k)] &= \sigma_n^k + 2\epsilon_1\iota_n(x_k)\iota_n(y_k)\nu_1 + [1-\epsilon_1(\iota_n(x_k)^2 - \epsilon_2\iota_n(y_k)^2]\nu_2 + \iota_n(x_k)\nu_{12}, \end{aligned} \tag{4}$$

where $\nu_1 = \rho^*\mu_1$, $\nu_2 = \rho^*\mu_2$, $\nu_{12} = \rho^*\mu_{12}$ denote the pull-back of the Maurer–Cartan forms via the moving frame. The recurrence relations for the phantom invariants $\iota_n(x_n) = \iota_n(y_n) = \iota_n(y_{n+1}) = 0$ yield the normalized Maurer–Cartan forms

$$\nu_1 = -\omega_n^n, \qquad \nu_2 = -\sigma_n^n, \qquad \nu_{12} = \frac{\sigma_n^n - \sigma_n^{n+1}}{K_n} - \epsilon_1 K_n \sigma_n^n.$$

Substituting these expressions into the recurrence relations for the normalized invariants $I_n = \iota_n(x_{n-1})$, $J_n = \iota_n(y_{n-1})$, $K_n = \iota_n(x_{n+1})$ introduced in (3), we obtain

$$
\begin{aligned}
dI_n &= \omega_n^{n-1} - \omega_n^n + \frac{\epsilon_2 J_n}{K_n}(\sigma_n^{n+1} - \sigma_n^n) + \epsilon_1\big[(\epsilon_2 J_n^2 - I_n^2)\omega_n^n + \epsilon_2 J_n(K_n - 2I_n)\sigma_n^n\big],\\
dJ_n &= \sigma_n^{n-1} - \sigma_n^n + \frac{I_n}{K_n}(\sigma_n^n - \sigma_n^{n+1}) + \epsilon_1\big[-2I_n J_n \omega_n^n + (I_n^2 - \epsilon_2 J_n^2 - I_n K_n)\sigma_n^n\big],\\
dK_n &= \omega_n^{n+1} - \omega_n^n - \epsilon_1 K_n^2 \omega_n^n.
\end{aligned}
$$

4.2 Shift Map

Let

$$
\mathfrak{m}_n = \rho_{n+1}\,\rho_n^{-1} = \begin{bmatrix} a_n & b_n \\ -\epsilon_1 \overline{b}_n & \overline{a}_n \end{bmatrix}, \qquad a_n\overline{a}_n + \epsilon_1 b_n \overline{b}_n = 1,
$$

denote the Maurer–Cartan invariant matrix, which is an element of the Cayley–Klein group $SO_{\epsilon_1,\epsilon_2}(3)$. Using the recurrence relations

$$
\begin{aligned}
I_{n+1} + i_{\epsilon_2} J_{n+1} = \iota_{n+1}(z_n) = \mathfrak{m}_n \cdot \iota_n(z_n) = \mathfrak{m}_n \cdot 0 = \frac{b_n}{\overline{a}_n},\\
0 = \iota_{n+1}(z_{n+1}) = \mathfrak{m}_n \cdot \iota_n(z_{n+1}) = a_n K_n + b_n,
\end{aligned}
$$

we find that

$$
a_n = \frac{e^{i_{\epsilon_2}\phi_n}}{\sqrt{1+\epsilon_1 K_n^2}}, \qquad b_n = \frac{K_n^2\, e^{-i_{\epsilon_2}\phi_n}}{\sqrt{1+\epsilon_1 K_n^2}}, \tag{5}
$$

where

$$
T_{\epsilon_2}(2\phi_n) = \frac{J_{n+1}}{I_{n+1}}.
$$

We also obtain the syzygy

$$
K_n = \sqrt{I_{n+1}^2 + \epsilon_2 J_{n+1}^2}.
$$

For the invariant one-forms, we have the recurrence relation

$$\mathbf{S}[\varpi_n^n] = \iota_{n+1}(dz_{n+1}) = \mathfrak{m}_n \cdot \iota_n(dz_{n+1}) = -\frac{I_{n+1} + i_{\epsilon_2} J_{n+1}}{K_n(1+\epsilon_1 K_n^2)}\, \varpi_n^{n+1},$$

so that

$$\begin{aligned}
(\omega_n^{n+1} + i_{\epsilon_2}\sigma_n^{n+1}) &= \varpi_n^{n+1} \\
&= -\frac{1}{K_n}(1+\epsilon_1 K_n^2)(I_{n+1} - i_{\epsilon_2} J_{n+1})\mathbf{S}[\varpi_n^n] \\
&= -\frac{1+\epsilon_1 K_n^2}{K_n}\Big[(I_{n+1}\mathbf{S}(\omega_n^n) + \epsilon_2 J_{n+1}\mathbf{S}(\sigma_n^n)) + i_{\epsilon_2}(-J_{n+1}\mathbf{S}(\omega_n^n) + I_{n+1}\mathbf{S}(\sigma_n^n))\Big].
\end{aligned}$$

Similarly, for the backward shift,

$$\begin{aligned}
\omega_n^{n-1} + i_{\epsilon_2}\sigma_n^{n-1} &= \varpi_n^{n-1} = \mathfrak{m}_{n-1} \cdot \mathbf{S}^{-1}[\varpi_n^n] \\
&= -\frac{1+\epsilon_1 K_{n-1}^2}{K_{n-1}}(I_n + i_{\epsilon_2} J_n)\mathbf{S}^{-1}[\varpi_n^n] \\
&= -\frac{1+\epsilon_1 K_{n-1}^2}{K_{n-1}}\Big[(I_n\mathbf{S}^{-1}(\omega_n^n) - \epsilon_2 J_n\mathbf{S}^{-1}(\sigma_n^n)) + i_{\epsilon_2}(J_n\mathbf{S}^{-1}(\omega_n^n) + I_n\mathbf{S}^{-1}(\sigma_n^n))\Big].
\end{aligned}$$

5 Invariant Linearization Operators

Definition 3 The *invariant linearization* of an invariant I_n, is the invariant difference operator $\mathcal{A}_{I_n}$ satisfying the equality

$$dI_n = \mathcal{A}_{I_n}\left(\begin{bmatrix}\omega_n^n \\ \sigma_n^n\end{bmatrix}\right).$$

To compute the invariant linearization of I_n, compute its differential using the recurrence relations (4) for the exterior derivative. The result is a linear combination of the invariant one-forms ω_n^k, σ_n^k. Then, use the recurrence relation for the shift map to express ω_n^k, σ_n^k in terms of ω_n^n, σ_n^n and their shifts. We note that these computations can be done symbolically, without requiring the coordinate expressions for the invariant I_n and the one-forms ω_n^k, σ_n^k.

For the normalized invariants $I_n = \iota_n(x_{n-1})$, $J_n = \iota_n(y_{n-1})$, and $K_n = \iota_n(x_{n+1})$, the components of the invariant linearization operators are

$$
\begin{aligned}
\mathcal{A}^x_{I_n} &= -\frac{(1+\epsilon_1 K_{n-1}^2) I_n}{K_{n-1}} \mathbf{S}^{-1} + \epsilon_1(\epsilon_2 J_n^2 - I_n^2) - \mathbb{1} + \frac{\epsilon_2 J_n J_{n+1}(1+\epsilon_1 K_n^2)}{K_n^2} \mathbf{S}, \\
\mathcal{A}^y_{I_n} &= \frac{\epsilon_2(1+\epsilon_1 K_{n-1}^2) J_n}{K_{n-1}} \mathbf{S}^{-1} + \epsilon_1 \epsilon_2 J_n (K_n - 2I_n) - \frac{\epsilon_2 J_n}{K_n} \\
&\quad - \frac{\epsilon_2 J_n I_{n+1}(1+\epsilon_1 K_n^2)}{K_n^2} \mathbf{S}, \\
\mathcal{A}^x_{J_n} &= -\frac{(1+\epsilon_1 K_{n-1}^2) J_n}{K_{n-1}} \mathbf{S}^{-1} - 2\epsilon_1 I_n J_n - \frac{I_n J_{n+1}(1+\epsilon_1 K_n^2)}{K_n^2} \mathbf{S}, \\
\mathcal{A}^y_{J_n} &= -\frac{(1+\epsilon_1 K_{n-1}^2) I_n}{K_{n-1}} \mathbf{S}^{-1} + \frac{I_n}{K_n} - \mathbb{1} + \epsilon_1(I_n^2 - \epsilon_2 J_n^2 - I_n K_n) \\
&\quad + \frac{I_n I_{n+1}(1+\epsilon_1 K_n^2)}{K_n^2} \mathbf{S}, \\
\mathcal{A}^x_{K_n} &= -\frac{I_{n+1}(1+\epsilon_1 K_n^2)}{K_n} \mathbf{S} - (\mathbb{1} + \epsilon_1 K_n^2), \\
\mathcal{A}^y_{K_n} &= -\frac{\epsilon_2 J_{n+1}(1+\epsilon_1 K_n^2)}{K_n} \mathbf{S}.
\end{aligned}
\tag{6}
$$

6 Geometric Flows

In the Euclidean plane, where $\epsilon_1 = 0$ and $\epsilon_2 = 1$, the authors of [1–3] define a discrete curve to be parametrized by arc-length if $K_n = |\Delta z_n| = 1$ for all $n \in \mathbb{Z}$. We extend the notion of arc-length parametrized discrete curves to the remaining Cayley–Klein planes as follows.

Definition 4 A discrete curve z_n is said to be parametrized by *arc-length* if

$$K_n = K$$

is constant for all $n \in \mathbb{Z}$.

In light of the Maurer–Cartan invariants (5), we notice that when $\epsilon_1 = -1$, we cannot set $K = 1$. This explains why the value of the constant in Definition 4 remains unspecified. From now on, we restrict our considerations to arc-length parametrized discrete curves. For such curves, we introduce the *discrete curvature*

$$\kappa_n = \frac{J_n}{I_n - K} = \frac{\widetilde{J}_n}{\widetilde{I}_n - 1},$$

where $\widetilde{I}_n = \frac{I_n}{K}$, $\widetilde{J}_n = \frac{J_n}{K}$. Using the syzygy $K^2 = I_n^2 + \epsilon_2 J_n^2$, we have that

$$\widetilde{I}_n = \frac{I_n}{K} = \frac{\epsilon_2\kappa_n^2 - 1}{\epsilon_2\kappa_n^2 + 1} \qquad \text{and} \qquad \widetilde{J}_n = \frac{J_n}{K} = -\frac{2\kappa_n}{\epsilon_2\kappa_n^2 + 1}. \tag{7}$$

For an arc-length parametrized curve, the invariant linearization operators (6) can be expressed in term of the quantities (7):

$$\begin{aligned}
\mathcal{A}^x_{I_n} &= (1+\epsilon_1 K^2)\big[\epsilon_2 \widetilde{J}_n \widetilde{J}_{n+1}\mathbf{S} - \mathbb{1} - \widetilde{I}_n \mathbf{S}^{-1}\big] + 2\epsilon_1\epsilon_2 K^2 \widetilde{J}_n^2,\\
\mathcal{A}^y_{I_n} &= (1+\epsilon_1 K^2)\epsilon_2\big[- \widetilde{J}_n \widetilde{I}_{n+1}\mathbf{S} - \widetilde{J}_n + \widetilde{J}_n \mathbf{S}^{-1}\big] + 2\epsilon_1\epsilon_2 K^2 \widetilde{J}_n(1 - \widetilde{I}_n),\\
\mathcal{A}^x_{J_n} &= (1+\epsilon_1 K^2)\big[- \widetilde{I}_n \widetilde{J}_n \mathbf{S} - \widetilde{J}_n \mathbf{S}^{-1}\big] - 2\epsilon_1 K^2 \widetilde{I}_n \widetilde{J}_n,\\
\mathcal{A}^y_{J_n} &= (1+\epsilon_1 K^2)\big[\widetilde{I}_n \widetilde{I}_{n+1}\mathbf{S} + \widetilde{I}_n - \mathbb{1} - \widetilde{I}_n \mathbf{S}^{-1}\big] + 2\epsilon_1 \widetilde{I}_n(\widetilde{I}_n - 1),\\
\mathcal{A}^x_{K_n} &= (1+\epsilon_1 K^2)[-\widetilde{I}_{n+1}\mathbf{S} - \mathbb{1}\big],\\
\mathcal{A}^y_{K_n} &= (1+\epsilon_1 K^2)\big[- \epsilon_2 \widetilde{J}_{n+1}\mathbf{S}\big].
\end{aligned}$$

Computing the differential of the curvature, we obtain

$$d\kappa_n = \frac{\epsilon_2\kappa_n^2 + 1}{2} \cdot (\kappa_n\, d\widetilde{I}_n - d\widetilde{J}_n).$$

Therefore, the components of the invariant linearization operator for κ_n are

$$\begin{aligned}
\mathcal{A}^x_{\kappa_n} &= \frac{(1+\epsilon_1 K^2)(\epsilon_2\kappa_n^2 + 1)}{2K}\left(- \kappa_n \mathbf{S}^{-1} - \kappa_n + \frac{2\kappa_{n+1}}{\epsilon_2\kappa_{n+1}^2 + 1}\mathbf{S}\right) + 2\epsilon_1 K\kappa_n,\\
\mathcal{A}^y_{\kappa_n} &= \frac{(1+\epsilon_1 K^2)(\epsilon_2\kappa_n^2 + 1)}{2K}\left(-\mathbf{S}^{-1} + 2 + \frac{\epsilon_2\kappa_{n+1}^2 - 1}{\epsilon_2\kappa_{n+1}^2 + 1}\mathbf{S}\right) - 2\epsilon_1 K.
\end{aligned}$$

Next, let $\mathbf{T}_n$ and $\mathbf{N}_n$ be vectors in $\mathbb{C}_{i_{\epsilon_2}}$ defined by the pairings

$$\langle \omega_n^n, \mathbf{T}_n\rangle = 1, \qquad \langle \sigma_n^n, \mathbf{T}_n\rangle = 0, \qquad \langle \omega_n^n, \mathbf{N}_n\rangle = 0, \qquad \langle \sigma_n^n, \mathbf{N}_n\rangle = 1.$$

We now investigate the induced evolution equation of the curvature κ_n when the curve z_n evolves according to the geometric flow

$$\frac{\partial z_n}{\partial t} = \alpha_n \mathbf{T}_n + \beta_n \mathbf{N}_n, \tag{8}$$

where α_n and β_n are functions of the curvature κ_n and its shifts.

Theorem 1 *Let I_n be an invariant. Under the geometric curve flow (8), the invariant I_n evolves according to the differential-difference equation*

$$\frac{\partial I_n}{\partial t} = \mathcal{A}_{I_n}\left(\begin{bmatrix}\alpha_n\\ \beta_n\end{bmatrix}\right),$$

where $\mathcal{A}_{I_n}$ is the invariant linearization of I_n.

To preserve the fact that our curves are parametrized by arc-length, we consider flows that are arc-length preserving. Thus, we require that

$$\frac{\partial K_n}{\partial t} = \mathcal{A}^x_{K_n}(\alpha_n) + \mathcal{A}^y_{K_n}(\beta_n) = 0.$$

This leads to the finite difference equation

$$\alpha_n + \widetilde{I}_{n+1}\alpha_{n+1} + \epsilon_2 \widetilde{J}_{n+1}\beta_{n+1} = 0. \tag{9}$$

If $\epsilon_2 \neq 0$, then

$$\beta_n = -\frac{1}{\epsilon_2 \widetilde{J}_n}(\widetilde{I}_n\alpha_n + \alpha_{n-1}), \tag{10}$$

where α_n is arbitrary. If $\epsilon_2 = 0$, Eq. (9) reduces to

$$\alpha_{n+1} = -\frac{1}{\widetilde{I}_{n+1}}\alpha_n$$

whose solution is

$$\alpha_n = (-1)^n \boldsymbol{\pi}_k\left(\frac{1}{\widetilde{I}_{k+1}}, 0, n\right)\alpha_0, \tag{11}$$

where

$$\boldsymbol{\pi}_k(f_k, n_0, n) = \begin{cases} \displaystyle\prod_{k=n_0}^{n-1} f_k & n > n_0 \\ 1 & n = n_0 \\ \displaystyle\prod_{k=n}^{n_0-1} \frac{1}{f_k} & n < n_0 \end{cases}.$$

In this case, we note that β_n is arbitrary. We observe that both constraints (10) and (11) are satisfied when the components of the flow are

$$\alpha_n = \frac{2K}{1+\epsilon_1 K^2}, \qquad \beta_n = \frac{2K\kappa_n}{1+\epsilon_1 K^2}.$$

For such an arc-length preserving flow, the evolution of curvature is governed by the completely integrable differential-difference equation

$$\frac{\partial \kappa_n}{\partial t} = \mathcal{A}^x_{\kappa_n}\left(\frac{2K}{1+\epsilon_1 K^2}\right) + \mathcal{A}^y_{\kappa_n}\left(\frac{2K\kappa_n}{1+\epsilon_1 K^2}\right) = (1+\epsilon_2\kappa^2_{n+1})(\kappa_{n+1}-\kappa_{n-1}).$$

References

1. J. Benson, F. Valiquette, Invariant discrete flows. Stud. Appl. Math. **143**, 81–119 (2019)
2. A.I. Bobenko, Geometry II: discrete differential geometry. Lecture Notes, TU Berlin (2015), page.math.tu-berlin.de/bobenko/Lehre/Skripte/DDG_Lectures.pdf
3. T. Hoffmann, N. Kutz, Discrete curves in $\mathbb{C}P^1$ and the Toda lattice. Stud. Appl. Math. **113**, 31–55 (2004)
4. I.M. Yaglom, *A Simple Non-Euclidean Geometry and Its Physical Basis* (Springer, New York, 1979)
5. R. Yamilov, Symmetries as integrability criteria for differential difference equations. J. Phys. A Math. Gen. **39**, R541–623 (2006)

Zernike System Stems from Free Motion on the 3-Sphere

Kurt Bernardo Wolf, Natig M. Atakishiyev, George S. Pogosyan, and Alexander Yakhno

Abstract Systems that stem from projection of free motion on a manifold are the best candidates to exhibit remarkable symmetry properties. This is the case of free motion on the 3-sphere which, properly projected on the 2-dimensional manifold of a disk, yields the Zernike system. This exhibits separability in a variety of coordinate systems, polynomial solutions, and interbasis expansion coefficients that are special Clebsch–Gordan coefficients and Hahn orthogonal polynomials.

Keywords Spherical geometry · Zernike system · Separation of variables · Clebsch–Gordan coefficients

1 Introduction: The so(4) Algebra

The Lie algebras of the orthogonal groups have a basis of generators $K_{i,j} = -K_{j,i}$, whose commutation relations are

$$[K_{i,j}, K_{k,\ell}] = \delta_{j,k} K_{i,\ell} + \delta_{i,\ell} K_{j,k} + \delta_{k,i} K_{\ell,j} + \delta_{\ell,j} K_{k,i}, \tag{1}$$

K. B. Wolf (✉)
Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico
e-mail: bwolf@icf.unam.mx; http://www.fis.unam.mx/~bwolf/

N. M. Atakishiyev
Instituto de Matemáticas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

G. S. Pogosyan
Yerevan State University, Yerevan, Armenia

Joint Institute for Nuclear Research, Dubna, Russian Federation

A. Yakhno
Departamento de Matemáticas, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Mexico

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_16

and whose range of indices, $i, j, k, \ell \in \{1, 2, \dots N\}$ determines the Lie algebra $\mathsf{so}(N)$ of dimension $\frac{1}{2}N(N-1)$.

We shall work in particular with the 4-dimensional orthogonal algebra $\mathsf{so}(4)$, which has six generators. This is the only orthogonal algebra that *splits* into a direct sum of two algebras:

$$\mathsf{so}(4) = \mathsf{so}(3)^{(1)} \oplus \mathsf{so}(3)^{(2)}. \tag{2}$$

To prove this, it is sufficient to build the generators

$$J_i^{(1)} := \tfrac{1}{2}(K_{j,k} + K_{i,4}), \qquad J_i^{(2)} := \tfrac{1}{2}(K_{j,k} - K_{i,4}), \tag{3}$$

for $i, j, k \in \{1, 2, 3\}$ cyclic. These two sets mutually commute,

$$[J_i^{(1)}, J_j^{(1)}] = -J_k^{(1)}, \quad [J_i^{(2)}, J_j^{(2)}] = -J_k^{(2)}, \quad [J_i^{(1)}, J_j^{(2)}] = 0. \tag{4}$$

On the other hand, as all orthogonal algebras, it contains a Gel'fand–Zetlin or *canonical* chain of subalgebras,

$$\mathsf{so}(4) \supset \mathsf{so}(3) \supset \mathsf{so}(2), \tag{5}$$

whose generators $K_{i,j}$ have their skew-symmetric pairs of indices restricted to $i, j \leqslant 4$, 3 or 2, respectively. We are very familiar with the $\mathsf{SO}(3)$ representation theory, eigenvectors, and spectra $\sim \ell(\ell+1)$, so we have reason to expect that the Clebsch–Gordan coefficients $C^{\ell,m}_{j^{(1)},m^{(1)};j^{(2)},m^{(2)}} = \langle j^{(1)}, m^{(1)}; j^{(2)}, m^{(2)} | \ell, m \rangle$ will appear when we introduce specific realizations of the $\mathsf{so}(4)$ Lie algebra generators.

The Zernike system can be presented as a quantum mechanical problem, with a Schrödinger equation and non-standard Hamiltonian

$$\widehat{Z}(x, y)\, \Psi(\mathbf{r}) = -E_J\, \Psi(\mathbf{r}), \qquad \widehat{Z}(x, y) := \nabla^2 - (\mathbf{r} \cdot \nabla)^2 - 2\, \mathbf{r} \cdot \nabla, \tag{6}$$

on a space of functions $\Psi(\mathbf{r})$ on the closed unit disk $\mathsf{D} := \{(x, y) \,|\, x^2 + y^2 \leqslant 1\}$, that are *finite* on its boundary, $|\Psi(\mathbf{r})_{|\mathbf{r}|=1}| < \infty$ [1, 2]. The spectrum of E_J in (6) is then found to be $J(J+2)$, for $J \in \{0, 1, \dots\} =: \mathsf{Z}_0^+$. We recognize this as the spectrum of a $\mathsf{so}(4)$ Casimir invariant that is the Laplace–Beltrami operator on a 3-sphere. These are of course not coincidences, as we shall now detail, but bring in the Zernike system as one of the fundamental proto-systems of quantum mechanics such as the harmonic oscillator and the Bohr atom.

In our research into the classical [3] and quantum [4] Zernike systems, the interbasis coefficients [5, 6] between two solutions sets with different separation coordinates were found to be a special type of Clebsch–Gordan coefficient. The reason for this appearance was laid out in Ref. [7], from which this proceedings contribution is a concentrate.

In Sect. 2 we present the well-known Lie algebra realization of the orthogonal groups on spheres. We are using the algebra $\mathsf{so}(4)$ and the 3-sphere as its homogeneous space, so in Sect. 3 we introduce two coordinate systems, where the Laplace–Beltrami operator appears in different differential forms. The eigenfunctions there already relate through Clebsch–Gordan coefficients, and continue to do so when the 3-sphere is projected on a two-sphere in Sect. 4, resulting in a restricted set of Clebsch–Gordan's as interbasis coefficients between solutions of the Zernike system in the two coordinate systems, shown in Sect. 5. In the concluding Sect. 6 we add some remarks on the significance of free motions on conics that project to remarkable physical systems.

2 Realization of $\mathsf{SO}(4)$ on the 3-Sphere

Lie algebras, when exponentiated to the group, can act faithfully and transitively, on any of its homogeneous *coset spaces*. Corresponding to cosets by the group identity $\{1\}$, the action is on the group itself, a manifold of dimension $\frac{1}{2}N(N-1)$. One may have spaces of cosets by $\mathsf{SO}(2)$, $\mathsf{SO}(3)$, etc. up to cosets by $\mathsf{SO}(N-1)$ [8]. The last is a privileged space because $\mathsf{SO}(N)/\mathsf{SO}(N-1) = \mathsf{S}^{N-1}$ is the $(N-1)$-dimensional manifold of a sphere. In this space one can *realize* the generators as $K_{i,j} = s_i\partial_j - s_j\partial_i$ ($\partial_j := \partial/\partial s_j$), that generate rotations of S^{N-1}; the s_i are the Cartesian coordinates restricted to the sphere by $\sum_{i=1}^{N} s_i^2 = 1$. Thus we realize $\mathsf{so}(4)$ as the generators of rotations of the 3-dimensional manifold of S^3.

While the Lie algebra $\mathsf{so}(3)$ has one well-known invariant $J^2 := \sum_{i,j=1}^{3} K_{i,j}^2$ with eigenvalues $j(j+1)$ where $j \in \mathsf{Z}_0^+$, the Lie algebra $\mathsf{so}(4)$ has *two* second-degree invariant Casimir operators. The first is the sum of all squares, i.e., the Laplace–Beltrami operator on the 3-sphere, $\Delta_{LB}^{(3)} = \sum_{i,j=1}^{4} K_{i,j}^2$ with spectrum $J(J+2)$, $J \in \mathsf{Z}_0^+$ on S^3; the second invariant, $\sum_{i,j,k=1}^{3} \varepsilon_{i,j,k} K_{i,j} K_{k,4} = 0$ vanishes in the coset space of the sphere. This implies that

$$J^{(1)\,2} = \tfrac{1}{4} \sum_{i,j,k=1}^{3} (K_{j,k} \pm K_{i,4})^2 = J^{(2)\,2} \quad \Rightarrow \quad j^{(1)} = j^{(2)} =: j, \tag{7}$$

$$\left.\begin{aligned} \Delta_{LB}^{(3)} = 2J^{(1)\,2} + 2J^{(2)\,2} \quad & \text{spectrum } J(J+2) \\ 4J^{(1)\,2} = 4J^{(2)\,2} \quad & \text{spectrum } 4\,j(j+1) \end{aligned}\right\} \quad \Rightarrow \quad J = 2j \in \mathsf{Z}_0^+. \tag{8}$$

Finally, in (5) the $\mathsf{so}(3)$ Casimir invariant $\frac{1}{2}\sum_{i,j=1}^{3} K_{i,j}^2$ has the spectrum $\ell(\ell+1)$, $\ell \in \mathsf{Z}_0^+$, with the range of the branching rule $0 \leqslant \ell \leqslant J$.

3 Two Coordinate Systems for S^3

To parametrize the S^3 sphere embedded in an ambient 4-space $(s_1, s_2, s_3, s_4) \in \mathsf{R}^4$, there exist six distinct orthogonal coordinates, listed in Refs. [9, 10] as spherical, cylindrical, sphero-elliptic, oblate and prolate elliptic, and ellipsoidal. Whereas in Ref. [7] three coordinate systems were considered, in the present report we shall consider only the cylindrical and the spherical systems. These are tailored, respectively, for the split and the canonical subalgebra chains (3) and (5),

$$\begin{array}{ll}
\text{System I: cylindrical} & \text{System II: spherical} \\
\mathsf{so}(4) \supset \mathsf{so}(2)^{(1)} \oplus \mathsf{so}(2)^{(2)} & \mathsf{so}(4) \supset \mathsf{so}(3) \supset \mathsf{so}(2) \\
s_1 = \cos\gamma\cos\phi_1, & s_1 = \sin\chi\sin\theta\cos\phi, \\
s_2 = \cos\gamma\sin\phi_1, & s_2 = \sin\chi\sin\theta\sin\phi, \\
s_3 = \sin\gamma\cos\phi_2, & s_3 = \sin\chi\cos\theta, \\
s_4 = \sin\gamma\sin\phi_2, & s_4 = \cos\chi \\
0 < \gamma < \tfrac{1}{2}\pi, & 0 < \theta, \chi < \pi, \\
0 \leqslant \phi_1, \phi_2 < 2\pi, & 0 \leqslant \phi < 2\pi.
\end{array} \tag{9}$$

In these two coordinate systems, the $\mathsf{so}(4)$ Laplace–Beltrami operator $\Delta^{(3)}_{\mathrm{LB}}$, of spectrum $J(J+2)$, is realized as two corresponding forms of second-order differential operators [11–13],

$$\Delta^{(3)\,I}_{\mathrm{LB}} = \frac{\partial^2}{\partial\gamma^2} + (\cot\gamma - \tan\gamma)\frac{\partial}{\partial\gamma} + \frac{1}{\cos^2\gamma}\frac{\partial^2}{\partial\phi_1^2} + \frac{1}{\sin^2\gamma}\frac{\partial^2}{\partial\phi_2^2}, \tag{10}$$

$$\Delta^{(3)\,II}_{\mathrm{LB}} = \frac{\partial^2}{\partial\chi^2} + 2\cot\chi\frac{\partial}{\partial\chi} + \frac{1}{\sin^2\chi}\left(\frac{\partial^2}{\partial\theta^2} + \cot\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right). \tag{11}$$

These determine the eigen-spaces $J \in \mathsf{Z}_0^+$ of solutions, $\Phi^{\mathrm{I}}_{J,m_1,m_2}(\gamma,\phi_1,\phi_2)$ and $\Phi^{II}_{J,\ell,m}(\chi,\theta,\phi)$, where their further specification by labels, m_1, m_2 and ℓ, m is done by the realization of the following Lie algebra generators,

$$\begin{array}{lll}
K_{1,2}\Phi^{\,I}_{J,m_1,m_2} = \mathrm{i}m_1\Phi^{\,I}_{J,m_1,m_2}, & K_{3,4}\Phi^{\,I}_{J,m_1,m_2} & = \mathrm{i}m_2\Phi^{\,I}_{J,m_1,m_2}, \\
\sum_{i,j=1}^{3} K^2_{i,j}\Phi^{\,II}_{J,\ell,m} = -\ell(\ell+1)\Phi^{\,II}_{J,\ell,m}, & K_{1,2}\Phi^{\,II}_{J,\ell,m} & = \mathrm{i}m\Phi^{\,II}_{J,\ell,m}.
\end{array} \tag{12}$$

The differential operators (10) and (11), having implemented (12), lead to Pöschl–Teller quantum mechanical Schödinger equations in the angle γ with their quadratic spectrum, and parameters determined by m_1, m_2 and ℓ, m. These potentials have hypergeometric polynomial and/or trigonometric solutions:

$$\Phi^{\mathrm{I}}_{J,m_1,m_2}(\gamma,\phi_1,\phi_2) = (\cos\gamma)^{|m_1|}(\sin\gamma)^{|m_2|}\; P^{(|m_2|,|m_1|)}_{\frac{1}{2}(J-|m_1|-|m_2|)}(\cos 2\gamma) \times e^{\mathrm{i}(m_1\phi_1+m_2\phi_2)} =: e^{\mathrm{i}m_1\phi_1}\; \Xi^{\mathrm{I}}_{J,m_1,m_2}(\gamma,\phi_2), \quad (13)$$

$$\Phi^{\mathrm{II}}_{J,\ell,m}(\chi,\theta,\phi) = (\sin\chi)^{\ell}\; C^{\ell+1}_{J-\ell}(\cos\chi) \times P^{m}_{\ell}(\cos\theta)e^{\mathrm{i}m\phi} =: e^{\mathrm{i}m\phi}\; \Xi^{\mathrm{II}}_{J,\ell,m}(\chi,\theta). \quad (14)$$

Here P^m_ℓ, C^λ_μ, and $P^{(\alpha,\beta)}_n$ are the associated Legendre, Gegenbauer, and Jacobi polynomials. In (13), m_1, m_2 are restricted by $J - |m_1| - |m_2| =$ even. We thus expect that the *overlaps* of the two solution sets (with $\phi_1 = \phi$ and $m = m_1$),

$$\Phi^{\mathrm{I}}_{J,m_1,m_2}(\gamma,\phi_1,\phi_2) = \sum_{\ell=|m|}^{J} W^{\ell,m}_{J,m_1,m_2}\Phi^{\mathrm{II}}_{J,\ell,m}(\chi,\theta,\phi), \quad (15)$$

once properly normalized, are Clebsch–Gordan coefficients; but not generic ones, because the two coupled angular momenta are equal,

$$W^{\ell,m}_{J,m_1,m_2} \sim C^{\ell,|m_1|}_{\frac{1}{2}J,\frac{1}{2}(|m_1|+|m_2|);\,\frac{1}{2}J,\frac{1}{2}(|m_1|-|m_2|)}. \quad (16)$$

4 Projection on the 2-Sphere S^2

In both the I and II coordinate systems, (13) and (14), the solutions factorize into a phase of one coordinate, and functions $\Xi_{J,\circ,\circ}$ of the two remaining angles on a 2-sphere S^2. It is indeed serendipitous that this reduction to the 2-sphere reveals the Zernike system written in (6), and contained in the formulation of free motion on the three-sphere [7].

Consider a change of coordinates $(s_1, s_2, s_3, s_4) \mapsto (\xi_1, \xi_2, \xi_3, \varphi)$,

$$s_1 = \xi_3\cos\varphi, \quad s_2 = \xi_3\sin\varphi, \quad s_3 = \xi_2, \quad s_4 = \xi_1, \quad (17)$$

that maps the 3-sphere $\sum_{i=1}^4 s_i^2 = 1$ on the 2-sphere $\sum_{i=1}^3 \xi_i^2 = 1$ and $\varphi \in \mathsf{S}^1$ on the circle; over this angle we shall integrate over. In these coordinates, the Laplace–Beltrami operator on S^3, $\Delta^{(3)}_{LB}$ contains the *two*-dimensional $\Delta^{(2)}_{LB}$ in the (s_3, s_4) subspace, plus derivatives in ξ_3 and φ,

$$\Delta^{(3)}_{LB} = \Delta^{(2)}_{LB} - \sum_{i=1}^{3}\xi_i\frac{\partial}{\partial\xi_i} + \frac{1}{\xi_3}\frac{\partial}{\partial\xi_3} + \frac{1}{\xi_3^2}\frac{\partial^2}{\partial\varphi^2}. \quad (18)$$

The functions $\Xi_{J,\circ,\circ}$ in (13) and (14) of angles in the 2-sphere S^2 can be isolated through integrating, respectively, over the circles of ϕ_1 and of ϕ; this will also set to zero the corresponding labels m_1 and m in those equations.

We perform this integration for the solutions in the coordinate System I of cylindrical coordinates (13), without regard for normalization at this stage,

$$\begin{aligned}\Psi^{\mathrm{I}}_{n,m}(x,y) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}\phi_1\, \Phi^{\mathrm{I}}_{J,m_1,m_2}(\gamma,\phi_1,\phi_2)\\ &= e^{-\mathrm{i}\frac{1}{2}\pi m}(x^2{+}y^2)^{\frac{1}{2}|m|} P^{(|m|,0)}_{n_r}\Big(1{-}2(x^2{+}y^2)\Big)e^{\mathrm{i}m\phi}. \end{aligned} \tag{19}$$

The index ranges are: $n = J \in \mathsf{Z}_0^+$ is the principal quantum number, we have now fixed $m_1 = 0$ so we write $m := -m_2$; and we have the *radial* quantum number $n_r := \frac{1}{2}(n{-}|m|) \in \mathsf{Z}_0^+$. For $\phi = \frac{1}{2}\pi{-}\phi_2$ and the ranges $\gamma|_0^{\pi/2}$ and $\phi|_{-\pi}^{\pi}$, the Cartesian coordinates on the disk $(x,y) \in \mathsf{D}$ and positive *half*-sphere S^2_+ are

$$x = \xi_1 = \sin\gamma\sin\phi_2, \quad y = \xi_2 = \sin\gamma\cos\phi_2, \quad \xi_3 = \cos\gamma \geqslant 0. \tag{20}$$

These are polar coordinates with radius $\sin\gamma \leqslant 1$ and angle ϕ_2 over a circle.

The integration for the System II solutions in polar coordinates on the sphere (14) yields

$$\begin{aligned}\Psi^{\mathrm{II}}_{m_1,m_2}(x,y) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{d}\phi\, \Phi^{\mathrm{II}}_{J,\ell,\mu}(\chi,\theta,\phi)\\ &= (1-x^2)^{\frac{1}{2}m_1} C^{m_1+1}_{m_2}(x)\, P_{m_1}\!\left(\frac{y}{\sqrt{1{-}x^2}}\right). \end{aligned} \tag{21}$$

Here the principal quantum number is also $n = m_1{+}m_2 = J \in \mathsf{Z}_0^+$, while we can set $\ell := m_1$. The coordinates on the disk and half-sphere for the ranges $\chi|_0^{\pi}$ and $\theta|_0^{\pi}$, are

$$x = \xi_1 = \cos\chi, \quad y = \xi_2 = \sin\chi\cos\theta, \quad \xi_3 = \sin\chi\sin\theta \geqslant 0. \tag{22}$$

This coordinate system can be visualized as polar coordinates on a sphere, projected on a plane that contains its poles.

We identify this construction to pertain the Zernike system because the defining Hamiltonian, the first quadratic invariant of the Lie algebra $\mathsf{so}(4)$, is the Casimir operator that serves also to classify hyper-spherical harmonics and generate free evolution for ideal quantum systems given in (18), surprisingly contains the Zernike system Hamiltonian (6). For functions $\Phi_{J,\nu,\mu}(\alpha,\beta,\phi)$ such as (13) and (14) with a factor $\exp(\mathrm{i}\mu\phi)$, the Laplace–Beltrami operator on S^3 is

$$\Delta^{(3)}_{LB}\Phi_{J,\nu,\mu} = \left(\Delta^{(2)}_{LB} + \frac{1}{\xi_3}\frac{\partial}{\partial\xi_3} + \frac{1}{\xi_3^2}\frac{\partial^2}{\partial\varphi^2}\right)\Phi_{J,\nu,\mu} = \left(\widehat{Z}(x,\,y) - \frac{\mu^2}{\xi_3^2}\right)\Phi_{J,\nu,\mu}, \tag{23}$$

and yields the Zernike Hamiltonian $\widehat{Z}(x,\,y)$ in (6), as the integration over ϕ restricts $\mu = 0$, and thus leaves only two labels for the original Zernike solutions $\Psi^{\mathrm{I}}_{n,m}(x,\,y)$, and for the solutions $\Psi^{II}_{m_1,m_2}(x,\,y)$ found in [4], as well as all separated solutions in other coordinate systems, including the solutions that separate in elliptic coordinates on the sphere [14, 15].

5 Interbasis Expansion Coefficients

The interbasis expansion coefficients are the analogues of (15) and (16), for the reduced indices,

$$\Psi^{\mathrm{I}}_{n,m}(x,\,y) = \sum_{m_1=0}^{n} W^{m_1,m_2}_{n,m}\,\Psi^{II}_{m_1,m_2}(x,\,y), \tag{24}$$

where $m_2 = n - m_1$ and, up to phases ω, the coefficients are a more special subset of Clebsch–Gordan coefficients, and where $m \in \{-n, -n{+}2, \ldots, n\}$,

$$W^{m_1,m_2}_{n,m} = \omega\, C^{m_1,0}_{\frac{1}{2}n,-\frac{1}{2}m;\,\frac{1}{2}n,\frac{1}{2}m}. \tag{25}$$

A property of these special Clebsch–Gordan coefficients is that they are special hypergeometric be Saalschutzian ${}_3F_2(\cdots|1)$ terminating series, known as Hahn polynomials $Q_n(x; a, a, b)$ in the Askey scheme [6],

$$\begin{aligned} C^{m_1,0}_{\frac{1}{2}n,-\frac{1}{2}m;\,\frac{1}{2}n,\frac{1}{2}m} &= \frac{n!}{\left(\frac{1}{2}(m_1-m_2-m)\right)!\left(\frac{1}{2}(n+m)\right)!}\sqrt{\frac{2m_1+1}{m_2!\,(n+m_1+1)!}} \\ &\quad\times {}_3F_2\left(\begin{matrix}-m_2, & m_1+1, & -\frac{1}{2}(n+m)\\ -n, & \frac{1}{2}(m_1-m_2-m)+1\end{matrix}\,\middle|\,1\right) \qquad (26)\\ &= \frac{(n!)^2}{\left(\frac{1}{2}(n-m)\right)!\left(\frac{1}{2}(n+m)\right)!}\sqrt{\frac{2m_1+1}{m_2!\,(n+m_1+1)!}} \\ &\quad\times Q_{m_2}\left(\tfrac{1}{2}(n+m);\,-n-1,\,-n-1,\,n\right) \qquad (27)\end{aligned}$$

with $m_2 = n - m_1$.

6 Concluding Remarks

We have here abstracted some of the results in Ref. [7] to exhibit the Zernike system as a projection of the evidently highly symmetrical system of free motion on a conic. In the case of the 3-sphere as homogeneous space for $\mathsf{so}(4)$, we have the benefit of additional Lie-theoretical properties, such as Schrödinger equations with potentials of Pöschl–Teller type.

The algebra $\mathsf{so}(4)$ and its special property of splitting (3) also serves for finite and discrete image analysis, between Cartesian- and polar-pixellated arrays [16]. In polar pixellation, the normal modes factorize and the radial functions are also Clebsch–Gordan coefficients, although of the more general type (16) rather than (25). Under various guises, the $\mathsf{so}(4)$ is an algebra that may contain other physical or optical systems in their various realizations.

In this report we have used the Schrödinger representation of the Zernike "wavefunctions" $\Psi^{\mathrm{I}}_{n,m}(x, y)$ and $\Psi^{II}_{m_1,m_2}(x, y)$ in their separated bases. It is then natural to label kets $|n, m\rangle_I$ and $|m_1, m_2\rangle_{II}$ as a short and equally good realization for the states of the system, and useful for computations. Following common Dirac notation, we could bind the two realizations through stating $\Psi^{\mathrm{I}}_{n,m}(x, y) = (x, y|n, m\rangle_I$ and $\Psi^{II}_{m_1,m_2}(x, y) = (x, y|m_1, m_2\rangle_{II}$, provided a proper definition exists for a Dirac basis $\{|x, y)\}_{x,y\in\mathsf{D}}$ over a finite disk. This subject has been addressed in a recent paper by Celeghini et al. [17], through the construction of a Hilbert space on a *closed subset* of the R^2 plane. We have been accustomed to use Hilbert spaces and Gel'fand triplets for functions over the whole plane R^2; the Zernike model necessitates also function-theoretic analyses. Finally, we are aware that similar constructions and projections can also be done with planes and hyperbolas—not only spheres, and that a Lie algebra can project out a superintegrable Higgs algebra [18, 19].

Acknowledgments N.M.A. and K.B.W. thanks project AG-100119 awarded by the Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México. A.Y. thanks the support of project PRO-SNI-2019 (Universidad de Guadalajara).

References

1. F. Zernike, Beugungstheorie des Schneidenverfahrens und Seiner Verbesserten Form der Phasenkontrastmethode. Physica **1**, 689–704 (1934)
2. F. Zernike, H.C. Brinkman, Hypersphärische Funktionen und die in sphärischen Bereichen orthogonalen Polynome. Verh. Akad. Wet. Amst. (Proc. Sec. Sci.) **38**, 161–170 (1935)
3. G.S. Pogosyan, K.B. Wolf, A. Yakhno, Superintegrable classical Zernike system. J. Math. Phys. **58**, 072901 (2017)
4. G.S. Pogosyan, C. Salto-Alegre, K.B. Wolf, A. Yakhno A, Quantum superintegrable Zernike system. J. Math. Phys. **58**, 072101 (2017)
5. G.S. Pogosyan, K.B. Wolf, A. Yakhno, New separated polynomial solutions to the Zernike system on the unit disk and interbasis expansion. J. Opt. Soc. Am. A **34**, 1844–1848 (2017)

6. N.M. Atakishiyev, G.S. Pogosyan, K.B. Wolf, A. Yakhno, Interbasis expansions in the Zernike system. J. Math. Phys. **58**, 103505 (2017)
7. N.M. Atakishiyev, G.S. Pogosyan, K.B. Wolf, A. Yakhno, Spherical geometry, Zernike's separability, and interbasis expansion coefficients. J. Math. Phys. **60**, 101701 (2019)
8. R.L. Anderson, K.B. Wolf, Complete sets of functions on homogeneous spaces with compact stabilizers. J. Math. Phys. **11**, 3176–3183 (1970)
9. C. Grosche, Kh.H. Karayan, G.S. Pogosyan, A.N. Sissakian, Quantum motion on the three-dimensional sphere: the ellipso-cylindrical basis. J. Phys. A Math. Gen. **30**, 1629–1657 (1997)
10. A.A. Izmest'ev, G.S. Pogosyan, A.N. Sissakian, P. Winternitz, Contraction of Lie algebras and separation of variables. N-dimensional sphere. J. Math. Phys. **40**, 1549–1573 (1999)
11. E.G. Kalnins, W. Miller Jr., Separation of variables on n-dimensional Riemannian manifolds. I. The n-sphere S_n and Euclidean n-space R_n. J. Math. Phys. **27**, 1721–1736 (1986). https://doi.org/10.1063/1.527088
12. G.S. Pogosyan, A.N. Sissakian, P. Winternitz, Separation of variables and Lie algebra contractions. Applications to special functions. Phys. Part. Nuclei **33**(Suppl. 1), S123–S144 (2002)
13. W. Miller Jr., E.G. Kalnins, G.S. Pogosyan, Exact and quasi-exact solvability of second-order superintegrable systems. I. Euclidean space preliminaries. J. Math. Phys. **47**, 033502 (2006)
14. N.M. Atakishiyev, G.S. Pogosyan, K.B. Wolf, A. Yakhno, Elliptic basis for the Zernike system: Heun function solutions. J. Math. Phys. **59**, 073503 (2018). https://doi.org/10.1063/1.5030759
15. N.M. Atakishiyev, G.S. Pogosyan, K.B. Wolf, A. Yakhno, On elliptic trigonometric form of the Zernike system and polar limits. Phys. Scr. **94**, 045202 (2019)
16. K.B. Wolf, Discrete systems and signals on phase space. Appl. Math. Inf. Sci. **4**, 141–181 (2010)
17. E. Celeghini, M. Gadella, M.A. del Olmo, Zernike functions, rigged Hilbert spaces, and potential applications. J. Math. Phys. **60**, 083508 (2019). https://doi.org/10.1063/1.5093488
18. P.W. Higgs, Dynamical symmetries in a spherical geometry. J. Phys. A: Math. Gen. **12**, 309–323 (1979)
19. E.G. Kalnins, J.M. Kress, W. Miller Jr., *Separation of Variables and Superintegrability. The Symmetry of Solvable Systems.* (IOP Publishing, Bristol, 2018)

Part II
Algebraic and Non-perturbative Methods

W-Algebras via Lax Type Operators

Daniele Valeri

Abstract W-algebras are certain algebraic structures associated to a finite-dimensional Lie algebra $\mathfrak{g}$ and a nilpotent element f via Hamiltonian reduction. In this note we give a review of a recent approach to the study of (classical affine and quantum finite) W-algebras based on the notion of Lax type operators.

For a finite-dimensional representation of $\mathfrak{g}$ a Lax type operator for W-algebras is constructed using the theory of generalized quasideterminants. This operator carries several pieces of information about the structure and properties of the W-algebras and shows the deep connection of the theory of W-algebras with Yangians and integrable Hamiltonian hierarchies of Lax type equations.

Keywords $\mathcal{W}$-algebras · Lax type operators · Generalized quasideterminants · Integrable Hamiltonian hierarchies · (Twisted) Yangians

1 Introduction

The first *quantum affine W-algebra*, the so-called Zamolodchikov W_3-algebra [50], appeared in the physics literature in the study of 2-dimensional Conformal Field Theory. Further generalizations of this algebra were provided soon after [25, 40]. Physicists thought of these algebras as "non-linear" infinite dimensional Lie algebras extending the Virasoro Lie algebra. In [27] the affine W-algebras $W_\kappa(\mathfrak{g}, f)$ (κ is called the level), for a principal nilpotent element $f \in \mathfrak{g}$, were described as vertex algebras obtained via a quantization of the Drinfeld–Sokolov Hamiltonian reduction, which was used in [24] to construct *classical affine W-algebras*. In particular, for $\mathfrak{sl}_2$ one gets the Virasoro vertex algebra, and for $\mathfrak{sl}_3$ the Zamolodchikov's W_3 algebra. The construction was finally generalized to arbitrary nilpotent element f in [36–38]. In these papers, affine W-algebras were applied to

D. Valeri (✉)
School of Mathematics and Statistics, University of Glasgow, Glasgow, UK
e-mail: daniele.valeri@glasgow.ac.uk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_17

representation theory of superconformal algebras. Quantum affine W-algebras may be also considered as an affinization of *quantum finite W-algebras* [46] which are a natural quantization of Slodowy slices [34].

W-algebras are at the cross roads of representation theory and mathematical physics and play important roles (just to cite some of them) in applications to integrable systems [15, 24], to Gromov–Witten theory and singularity theory [3, 44], the geometric Langlands program [28, 31–33], four-dimensional gauge theories [1, 8, 48].

In this note we survey the recent approach to (quantum finite and classical affine) W-algebras based on the notion of Lax type operators [18–21]. For a review of the approach to (classical) W-algebras via generators and relations we refer to [13].

Throughout the paper the base field $\mathbb{F}$ is a field of characteristic zero.

2 What Is a W-Algebra?

W-algebras are a rich family of algebraic structures associated to a pair $(\mathfrak{g}, f)$ consisting of a finite-dimensional reductive Lie algebra $\mathfrak{g}$ and a nilpotent element $f \in \mathfrak{g}$. They are obtained via Hamiltonian reduction in different categories: Poisson algebras, associative algebras and (Poisson), vertex algebras. We should think of them as algebraic structures underlying some physical theories with "extended symmetries."

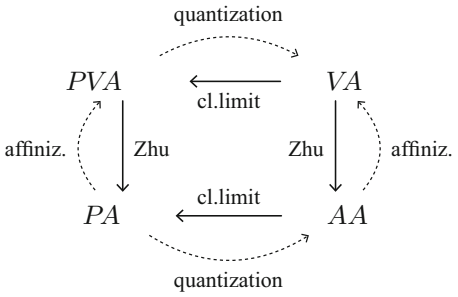
2.1 Fundamental Physical Theories and Corresponding Fundamental Algebraic Structures

In Classical Mechanics the phase space, describing the possible configurations of a physical system, is a Poisson manifold. The physical observables are the smooth functions on the manifold, and they thus form a *Poisson algebra* (PA).

By quantizing this theory we go to Quantum Mechanics. The observables become noncommutative objects, elements of an *associative algebra* (AA). Hence, the Poisson bracket is replaced by the usual commutator and the phase space is described as a representation of this associative algebra.

Going from a finite to an infinite number of degrees of freedom, we pass from Classical and Quantum Mechanics to Classical and Quantum Field Theory, respectively. The algebraic structure corresponding to an arbitrary Quantum Field Theory is still to be understood, but in the special case of chiral quantum fields of a 2-dimensional Conformal Field Theory (CFT) the adequate algebraic structure is a *vertex algebra* (VA) [6], and its quasi-classical limit is known as *Poisson vertex algebra* (PVA) [14].

Hence, the algebraic counterparts of the four fundamental frameworks of physical theories can be put in the following diagram:



(1)

The straight arrows in the above diagram correspond to canonical functors and have the following meaning. Given a filtered AA (respectively, VA), its associated graded algebra is a PA (respectively, PVA) called its *classical limit*. Moreover, starting from a positive energy VA (respectively, PVA) we can construct an AA (resp. PA) governing its representation theory, known as its *Zhu algebra* [51]. The processes of going from a classical theory to a quantum theory ("quantization") or from finitely many to infinitely many degrees of freedom ("affinization") are not functorial and they are thus represented with dotted arrows.

(Poisson) Vertex Algebras

PVAs provide a convenient framework to study Hamiltonian partial differential equations. Recall from [4] that a PVA is a differential algebra, i.e. a unital commutative associative algebra with a derivation ∂, endowed with a λ-*bracket*, i.e. a bilinear (over $\mathbb{F}$) map $\{\cdot\,_\lambda\,\cdot\}: \mathcal{V}\times\mathcal{V}\to\mathcal{V}[\lambda]$, satisfying the following axioms ($a, b, c \in \mathcal{V}$):

(i) sesquilinearity: $\{\partial a_\lambda b\} = -\lambda\{a_\lambda b\}$, $\{a_\lambda \partial b\} = (\lambda+\partial)\{a_\lambda b\}$;
(ii) skewsymmetry: $\{b_\lambda a\} = -\{a_{-\lambda-\partial} b\}$;
(iii) Jacobi identity: $\{a_\lambda\{b_\mu c\}\} - \{b_\mu\{a_\lambda c\}\} = \{\{a_\lambda b\}_{\lambda+\mu} c\}$;
(iv) (left) Leibniz rule: $\{a_\lambda bc\} = \{a_\lambda b\}c + \{a_\lambda c\}b$.

Applying skewsymmetry to the left Leibniz rule we get

(v) right Leibniz rule: $\{ab_\lambda c\} = \{a_{\lambda+\partial} c\}_{\to} b + \{b_{\lambda+\partial} c\}_{\to} a$.

In (ii) and (iv) we use the following notation: if $\{a_\lambda b\} = \sum_{n\in\mathbb{Z}_+}\lambda^n\alpha_n \in \mathcal{V}[\lambda]$, then $\{a_{\lambda+\partial}b\}_{\to}c = \sum_{n\in\mathbb{Z}_+}\alpha_n(\lambda+\partial)^n c \in \mathcal{V}[\lambda]$ and $\{a_{-\lambda-\partial}b\} = \sum_{n\in\mathbb{Z}_+}(-\lambda-\partial)^n\alpha_n \in \mathcal{V}[\lambda]$ (if there is no arrow, we move ∂ to the left).

We denote by $\int : \mathcal{V}\to\mathcal{V}/\partial\mathcal{V}$ the canonical quotient map of vector spaces. Recall that, if $\mathcal{V}$ is a PVA, then $\mathcal{V}/\partial\mathcal{V}$ carries a well-defined Lie algebra structure given by $\{\int f, \int g\} = \int\{f_\lambda g\}|_{\lambda=0}$, and we have a representation of the Lie algebra $\mathcal{V}/\partial\mathcal{V}$ on $\mathcal{V}$ given by $\{\int f, g\} = \{f_\lambda g\}|_{\lambda=0}$. A *Hamiltonian equation* on $\mathcal{V}$ associated

to a *Hamiltonian functional* $\int h \in \mathcal{V}/\partial\mathcal{V}$ is the evolution equation

$$\frac{du}{dt} = \{\textstyle\int h, u\}\ , \quad u \in \mathcal{V}\,. \tag{2}$$

The minimal requirement for *integrability* is to have an infinite collection of linearly independent integrals of motion in involution:

$$\textstyle\int h_0 = \int h,\ \int h_1,\ \int h_2,\ \ldots \ \text{ s.t. } \{\int h_m, \int h_n\} = 0 \ \text{ for all } \ m, n \in \mathbb{Z}_{\geq 0}\,.$$

In this case, we have the *integrable hierarchy* of Hamiltonian equations

$$\frac{du}{dt_n} = \{\textstyle\int h_n, u\}\ , \quad n \in \mathbb{Z}_{\geq 0}\,.$$

Example 1 The Virasoro–Magri PVA on the algebra of differential polynomials $\mathcal{V} = \mathbb{C}[u, u', u'', \ldots]$ is defined by letting

$$\{u_\lambda u\} = (2\lambda + \partial)u + \lambda^3\,,$$

and extending it to a λ-bracket for the whole $\mathcal{V}$ using sesquilinearity and Leibniz rules. Let $\int h = \int \frac{u^2}{2}$. Then the corresponding Hamiltonian equation (2) is the famous *KdV equation*:

$$\frac{du}{dt} = u''' + 3uu'\,.$$

Using the Lenard–Magri scheme of integrability [42] it can be shown that it belongs to an integrable hierarchy.

Vertex Algebras

VAs were introduced in [6]. Following [14], we provide here a "Poisson-like" definition using λ-brackets. A VA is a (not necessarily commutative nor associative) unital algebra V with a derivation ∂ endowed with a λ-bracket $[\cdot_\lambda \cdot] : V \times V \longrightarrow V[\lambda]$ satisfying sesquilinearity, skewsymmetry, Jacobi identity, and, moreover ($a, b, c \in V$):

1. quasicommutativity: $ab - ba = \int_{-\partial}^{0} [a_\lambda b] d\lambda$;
2. quasiassociativity: $(ab)c - a(bc) = (|_{\lambda=\partial} a) \int_0^\lambda [b_\mu c] d\mu + (|_{\lambda=\partial} b) \int_0^\lambda [a_\mu c] d\mu$;
3. noncommutative Wick formula: $[a_\lambda bc] = [a_\lambda b]c + b[a_\lambda c] + \int_0^\lambda [[a_\lambda b]_\mu c] d\mu$.

We refer to [14] for explanations about the notation. As before, we denote by $\int : V \to V/\partial V$ the canonical quotient map of vector spaces. If V is a VA, then $V/\partial V$ carries a well-defined Lie algebra structure given by $[\int f, \int g] = \int [f_\lambda g]|_{\lambda=0}$,

and we have a representation of the Lie algebra $V/\partial V$ on V given by $[\int f, g] = [f_\lambda g]|_{\lambda=0}$. A quantum integrable system consists in a collection of infinitely many linearly independent elements $\int h_m \in V/\partial V$, $m \in \mathbb{Z}_{\geq 0}$, in involution.

Example 2 A VA is commutative if $[a_\lambda b] = 0$, for every $a, b \in V$. It follows immediately from the definition that the category of commutative VAs is the same as the category of differential algebras.

Remark 1 The (not necessarily commutative nor associative) product in a VA corresponds to the *normally ordered product* of quantum fields in a CFT, while the λ-bracket encodes the singular part of their *operator product expansion* (OPE). We give a naive explanation of the latter sentence in a particular case. Consider the VA λ-bracket of a Virasoro element u (recall Example 1 for its PVA analogue)

$$[u_\lambda u] = (2\lambda + \partial)u + \frac{c}{12}\lambda^3 ,$$

where $c \in \mathbb{C}$ is called the central charge. Replace, in the above relation, u by a quantum field, say $T(w)$, ∂ by ∂_w and λ by ∂_w acting on the rational function $\frac{1}{z-w}$. Then we get

$$[T(w)_{\partial_w} T(w)]_{\to} \frac{1}{z-w} = \frac{\partial_w T(w)}{z-w} + \frac{2T(w)}{(z-w)^2} + \frac{c/2}{(z-w)^4} ,$$

which is the singular part of the OPE of the stress-energy tensor in CFT.

2.2 A Toy Model

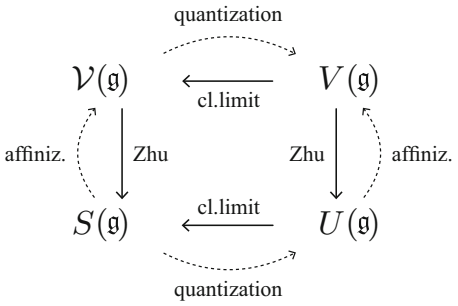
The simplest example when all four objects in diagram (1) can be constructed is obtained starting with a finite-dimensional Lie algebra $\mathfrak{g}$, with Lie bracket $[\cdot\,,\,\cdot]$, and with a non-degenerate invariant symmetric bilinear form $(\cdot\,|\,\cdot)$.

The *universal enveloping algebra* of $\mathfrak{g}$, usually denoted by $U(\mathfrak{g})$, is an associative algebra, and its classical limit is the symmetric algebra $S(\mathfrak{g})$, with the Kirillov–Kostant Poisson bracket.

Furthermore, we have also a Lie conformal algebra $\operatorname{Cur}\mathfrak{g} = (\mathbb{F}[\partial] \otimes \mathfrak{g}) \oplus \mathbb{F}K$, with the following λ-bracket:

$$[a_\lambda b] = [a, b] + (a|b)K\lambda \ , \quad [a_\lambda K] = 0 \ , \quad \text{for } \ a, b \in \mathfrak{g} \, . \tag{3}$$

The universal enveloping vertex algebra of $\operatorname{Cur}\mathfrak{g}$ is the so-called *universal affine vertex algebra* $V(\mathfrak{g})$, and its classical limit is the algebra of differential polynomials $\mathcal{V}(\mathfrak{g}) = S(\mathbb{F}[\partial]\mathfrak{g})$, with the PVA λ-bracket defined by (3). We refer to [14] for the definition of the latter structures and the construction of the corresponding Zhu maps. Thus, we get the following basic example of diagram (1):



(4)

2.3 Hamiltonian Reduction

All the four algebraic structures in diagram (1) admit a Hamiltonian reduction. We review here only the case for associative algebras. Recall that the Hamiltonian reduction of a unital associative algebra A by a pair (B, I), where $B \subset A$ is a unital associative subalgebra and $I \subset B$ is a two sided ideal, is the following unital associative algebra:

$$W = W(A, B, I) = \left(A/AI\right)^{\mathrm{ad}\, B} \tag{5}$$

where $\mathrm{ad}\, B$ denotes the usual adjoint action given by the commutator in an associative algebra (note that B acts on A/AI both by left and right multiplication). It is not hard to show that the obvious associative product on W is well defined.

Now, let $\{e, 2x, f\} \subset \mathfrak{g}$ be an $\mathfrak{sl}_2$-triple, and let

$$\mathfrak{g} = \bigoplus_{\substack{j=-d \\ j\in\frac{1}{2}\mathbb{Z}}}^{d} \mathfrak{g}_j\,, \tag{6}$$

be the $\mathrm{ad}\, x$-eigenspace decomposition. We can perform the Hamiltonian reduction of $A = U(\mathfrak{g})$ as follows. Let $B = U(\mathfrak{g}_{>0})$ and $I \subset B$ be the two sided ideal generated by the set

$$\left\{m - (f|m) \,\middle|\, m \in \mathfrak{g}_{\geq 1}\right\}. \tag{7}$$

Applying the Hamiltonian reduction (5) with the above data we get the so-called *quantum finite W-algebra* (it first appeared in [46])

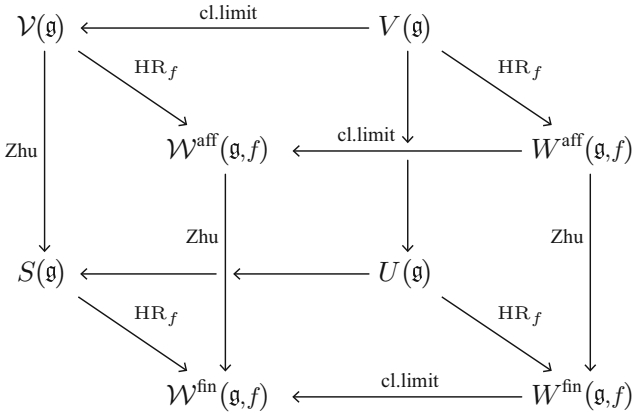
$$W^{\mathrm{fin}}(\mathfrak{g}, f) = \left(U(\mathfrak{g})/U(\mathfrak{g})\{m - (f|m) \,|\, m \in \mathfrak{g}_{\geq 1}\}\right)^{\mathrm{ad}\, \mathfrak{g}_{>0}}.$$

The Hamiltonian reduction (5) still makes sense if we replace associative algebras with PVAs (respectively, PAs), and we can perform it with $A = \mathcal{V}(\mathfrak{g})$, $B = \mathcal{V}(\mathfrak{g}_{>0})$ and $I \subset B$ the differential algebra ideal generated by the set (7) (respectively, $A = S(\mathfrak{g})$, $B = S(\mathfrak{g}_{>0})$, and $I \subset B$ the ideal generated by the set (7)). As a result we get the so-called *classical affine* W-algebra $\mathcal{W}^{\mathrm{aff}}(\mathfrak{g}, f)$ (respectively, *classical finite* W-algebra $\mathcal{W}^{\mathrm{aff}}(\mathfrak{g}, f)$), see [17] for further details.

Unfortunately, a similar construction of a Hamiltonian reduction for vertex algebras is not known, and the *quantum affine* W*-algebra* $W^{\mathrm{aff}}(\mathfrak{g}, f)$ is constructed using a cohomological approach [27, 36].

2.4 From the Toy Model to W-Algebras

Let $\mathfrak{g}$ be a finite-dimensional reductive Lie algebra, and let $f \in \mathfrak{g}$ be a nilpotent element. By the Jacobson–Morozov Theorem it can be embedded in an $\mathfrak{sl}_2$-triple $\{e, 2x, f\} \subset \mathfrak{g}$. Applying the machinery described in Sect. 2.3 we thus obtain a Hamiltonian reduction of the whole diagram (4):



(8)

It is a convention to use the calligraphic $\mathcal{W}$ to denote objects appearing in the "classical" column of diagram (8) and the block letter W to denote objects appearing in the "quantum" column of the same diagram. Also the upper label "fin" (resp. "aff") is used to denote objects appearing in the "finite" (resp. "affine") row of diagram (8), corresponding to physical theories with a finite (resp. infinite) number of degrees of freedom.

Hence, as we can see from diagram (8), *W-algebras* provide a very rich family of examples which appear in all the four fundamental aspects in diagram (1). Each of these classes of algebras was introduced and studied separately, with different applications in mind. The relations between them became fully clear later, see [14, 17, 34] for further details.

Classical Finite W-Algebras

The classical finite W-algebra $\mathcal{W}^{\mathrm{fin}}(\mathfrak{g}, f)$ is a PA, which can be viewed as the algebra of functions on the so-called *Slodowy slice* $\mathcal{S}(\mathfrak{g}, f)$, introduced by Slodowy while studying the singularities associated to the coadjoint nilpotent orbits of $\mathfrak{g}$ [49].

Finite W-Algebras

The first appearance of the finite W-algebras $W^{\mathrm{fin}}(\mathfrak{g}, f)$ was in a paper of Kostant [39]. He constructed the finite W-algebra for principal nilpotent $f \in \mathfrak{g}$ (in which case it is commutative), and proved that it is isomorphic to the center of the universal enveloping algebra $U(\mathfrak{g})$. The construction was then extended in [41] for even nilpotent element $f \in \mathfrak{g}$. The general definition of finite W-algebras $W^{\mathrm{fin}}(\mathfrak{g}, f)$, for an arbitrary nilpotent element $f \in \mathfrak{g}$, appeared later in a paper by Premet [46]. Finite W-algebras have deep connection with geometry and representation theory of simple finite-dimensional Lie algebras, with the theory of primitive ideals, and the Yangians, see [9, 43, 46, 47].

Classical Affine W-Algebras

The classical affine W-algebras $\mathcal{W}^{\mathrm{aff}}(\mathfrak{g}, f)$ were introduced, for principal nilpotent element f, in the seminal paper of Drinfeld and Sokolov [24]. They were introduced as Poisson algebras of functions on an infinite dimensional Poisson manifold, and they were used to study KdV-type integrable bi-Hamiltonian hierarchies of PDE's, nowadays known as Drinfeld–Sokolov hierarchies. Later, there have been several papers aimed at the construction of generalized Drinfeld–Sokolov hierarchies [10, 12, 23, 26, 29, 30]. In [15], the classical W-algebras $\mathcal{W}^{\mathrm{aff}}(\mathfrak{g}, f)$ were described as PVA, and the theory of generalized Drinfeld–Sokolov hierarchies was formalized in a more rigorous and complete way [16, 18, 20].

Quantum Affine W-Algebras

They have been extensively discussed in the Introduction. A review of the subject up to the early 1990s may be found in the collection of a large number of reprints on W-algebras [7]. Recently, it has been shown that they are at the base of an unexpected connections of vertex algebras with the geometric invariants called the Higgs branches in the four-dimensional $N = 2$ superconformal field theories [2, 5].

3 Linear Algebra Intermezzo

3.1 Set Up

Let $\mathfrak{g}$ be a finite-dimensional reductive Lie algebra, let $\{f, 2x, e\} \subset \mathfrak{g}$ be an $\mathfrak{sl}_2$-triple and let (6) be the corresponding $\operatorname{ad} x$-eigenspace decomposition. In Sects. 4 and 5 we will use the projection map $\pi_{\leq \frac{1}{2}} : \mathfrak{g} \to \mathfrak{g}_{\leq \frac{1}{2}} = \oplus_{k \leq \frac{1}{2}} \mathfrak{g}_k$ with kernel $\mathfrak{g}_{> \frac{1}{2}} = \oplus_{k > \frac{1}{2}} \mathfrak{g}_k$.

Let $\varphi : \mathfrak{g} \to \operatorname{End} V$ be a faithful representation of $\mathfrak{g}$ on an N-dimensional vector space V. Throughout the paper we shall often use the following convention: we denote by lowercase Latin letters elements of the Lie algebra $\mathfrak{g}$, and by the same uppercase letters the corresponding (via φ) elements of $\operatorname{End} V$. For example, $F = \varphi(f)$ is a nilpotent endomorphism of V. Moreover, $X = \varphi(x)$ is a semisimple endomorphism of V with half-integer eigenvalues. The corresponding X-eigenspace decomposition of V is

$$V = \bigoplus_{k \in \frac{1}{2}\mathbb{Z}} V[k] \,. \tag{9}$$

Note that $\frac{d}{2}$ is the largest X-eigenvalue in V.

Recall that the trace form on $\mathfrak{g}$ associated to the representation V is, by definition,

$$(a|b) = \operatorname{tr}_V(AB) \,, \qquad a, b \in \mathfrak{g} \,, \tag{10}$$

and we assume that it is non-degenerate. Let $\{u_i\}_{i \in I}$ be a basis of $\mathfrak{g}$ compatible with the $\operatorname{ad} x$-eigenspace decomposition (6), i.e. $I = \sqcup_k I_k$ where $\{u_i\}_{i \in I_k}$ is a basis of $\mathfrak{g}_k$. We also denote $I_{\leq \frac{1}{2}} = \sqcup_{k \leq \frac{1}{2}} I_k$. Moreover, we shall also need, in Sect. 5, that $\{u_i\}_{i \in I}$ contains a basis $\{u_i\}_{i \in I_f}$ of $\mathfrak{g}^f = \{a \in \mathfrak{g} \mid [a, f] = 0\}$, the centralizer of f in $\mathfrak{g}$. Let $\{u^i\}_{i \in I}$ be the basis of $\mathfrak{g}$ dual to $\{u_i\}_{i \in I}$ with respect to the form (10), i.e. $(u_i|u^j) = \delta_{i,j}$. According to our convention, we denote by $U_i = \varphi(u_i)$ and $U^i = \varphi(u^i)$, $i \in I$, the corresponding endomorphisms of V.

In Sects. 4 and 5 we will consider the following important element:

$$U = \sum_{i \in I} u_i U^i \in \mathfrak{g} \otimes \operatorname{End} V \,. \tag{11}$$

Here and further we are omitting the tensor product sign.

Furthermore, the following endomorphism of V, which we will call the *shift matrix*, will play an important role in Sect. 4

$$D = -\sum_{i \in I_{\geq 1}} U^i U_i \in \operatorname{End} V \,. \tag{12}$$

Finally, we denote by $\Omega_V \in \mathrm{End}\, V \otimes \mathrm{End}\, V$ the permutation map:

$$\Omega_V(v_1 \otimes v_2) = v_2 \otimes v_1 \quad \text{for all} \quad v_1, v_2 \in V\,. \tag{13}$$

Using Sweedler's notation we write $\Omega_V = \Omega'_V \otimes \Omega''_V$ to denote, as usual, a sum of monomials in $\mathrm{End}\, V \otimes \mathrm{End}\, V$. Suppose that V has a non-degenerate bilinear form $\langle \cdot\,|\,\cdot \rangle : V \times V \to \mathbb{F}$, which is symmetric or skewsymmetric:

$$\langle v_1 | v_2 \rangle = \epsilon \langle v_2 | v_1 \rangle\,, \quad v_1, v_2 \in V\,, \quad \text{where } \epsilon \in \{\pm 1\}\,. \tag{14}$$

Then, we denote by

$$\Omega_V^\dagger = (\Omega'_V)^\dagger \otimes \Omega''_V\,, \tag{15}$$

where $A^\dagger$ is the adjoint of $A \in \mathrm{End}\, V$ with respect to (14).

3.2 The "Identity" Notation

Let $U \subset V$ be a subspace of V, and assume that there is "natural" splitting $V = U \oplus U'$. We shall denote, with an abuse of notation, by $\mathbb{1}_U$ both the identity map $U \xrightarrow{\sim} U$, the inclusion map $U \hookrightarrow V$, and the projection map $V \twoheadrightarrow U$ with kernel U'. The correct meaning of $\mathbb{1}_U$ should be clear from the context.

3.3 Generalized Quasideterminants

Let R be a unital associative algebra and let V be a finite-dimensional vector space with direct sum decompositions $V = U \oplus U' = W \oplus W'$. Assume that $A \in R \otimes \mathrm{End}(V)$ and $\mathbb{1}_W A^{-1} \mathbb{1}_U \in R \otimes \mathrm{Hom}(U, W)$ are invertible. The (generalized) *quasideterminant* of A with respect to U and W, cf. [18, 35], is defined as

$$|A|_{U,W} := (\mathbb{1}_W A^{-1} \mathbb{1}_U)^{-1} \in R \otimes \mathrm{Hom}(W, U)\,. \tag{16}$$

Remark 2 Provided that both A and $\mathbb{1}_{U'} A \mathbb{1}_{W'}$ are invertible, it is possible to write the generalized quasideterminant (16) in the more explicit form $|A|_{U,W} = \mathbb{1}_U A \mathbb{1}_W - \mathbb{1}_U A \mathbb{1}_{W'} (\mathbb{1}_{U'} A \mathbb{1}_{W'})^{-1} \mathbb{1}_{U'} A \mathbb{1}_W$.

4 Quantum Finite W-Algebras and (Twisted) Yangians

4.1 Lax Type Operators for Quantum Finite W-Algebras

We introduce some important $\mathrm{End}\, V$-valued polynomials in z, and Laurent series in z^{-1}, with coefficients in $U(\mathfrak{g})$. The first one is (cf. (11))

$$A(z) = z\mathbb{1}_V + U = z\mathbb{1}_V + \sum_{i\in I} u_i U^i \ \in U(\mathfrak{g})[z] \otimes \mathrm{End}(V)\,. \tag{17}$$

(As in Sect. 3, we are dropping the tensor product sign.) Another important operator is (keeping the same notation as in [19])

$$A^\rho(z) = z\mathbb{1}_V + F + \pi_{\leq \frac{1}{2}} U = z\mathbb{1}_V + F + \sum_{i\in I_{\leq \frac{1}{2}}} u_i U^i \ \in U(\mathfrak{g})[z] \otimes \mathrm{End}\, V\,. \tag{18}$$

Now we introduce the Lax operator $L(z)$. Consider the generalized quasideterminant (cf. (16))

$$\widetilde{L}(z) = |A^\rho(z) + D|_{V[\frac{d}{2}], V[-\frac{d}{2}]} = \left(\mathbb{1}_{V[-\frac{d}{2}]}\big(z\mathbb{1}_V + F + \pi_{\leq \frac{1}{2}} U + D\big)^{-1} \mathbb{1}_{V[\frac{d}{2}]}\right)^{-1}, \tag{19}$$

where $\mathbb{1}_{V[-\frac{d}{2}]}$ and $\mathbb{1}_{V[\frac{d}{2}]}$ are defined in Sect. 3.2 (using the obvious splittings of V given by the grading (9)), $A^\rho(z)$ is defined in Eq. (18) and D is the "shift matrix" (12).

Let us denote by $\bar{1}$ the image of $1 \in U(\mathfrak{g})$ in the quotient $U(\mathfrak{g})\big/U(\mathfrak{g})\{m - (f|m) \,\big|\, m \in \mathfrak{g}_{\geq 1}\}$. The Lax operator $L(z)$ is defined as the image of $\widetilde{L}(z)$ in this quotient:

$$L(z) = L_{\mathfrak{g},f,V}(z) := \widetilde{L}(z)\bar{1}\,. \tag{20}$$

The first main result in [21] can be summarized as follows.

Theorem 1

(a) The operator $A^\rho(z) + D$ is invertible in $U(\mathfrak{g})((z^{-1})) \otimes \mathrm{End}\, V$, and the operator $\mathbb{1}_{V[-\frac{d}{2}]}(A^\rho(z)+D)^{-1}\mathbb{1}_{V[\frac{d}{2}]}$ is invertible in $U(\mathfrak{g})((z^{-1}))\otimes \mathrm{Hom}\left(V\big[-\frac{d}{2}\big], V\big[\frac{d}{2}\big]\right)$. Hence, the quasideterminant defining $\widetilde{L}(z)$ (cf. (19)) exists and lies in $U(\mathfrak{g})((z^{-1})) \otimes \mathrm{Hom}\left(V\big[-\frac{d}{2}\big], V\big[\frac{d}{2}\big]\right)$.

(b) The entries of the coefficients of the operator $L(z)$ defined in (20) lie in the W-algebra $W(\mathfrak{g}, f)$:

$$L(z) := |z\mathbb{1}_V + F + \pi_{\leq \frac{1}{2}} U + D|_{V[\frac{d}{2}], V[-\frac{d}{2}]}\bar{1}$$

$$\in W(\mathfrak{g}, f)((z^{-1})) \otimes \mathrm{Hom}\left(V[-\frac{d}{2}], V[\frac{d}{2}]\right).$$

Remark 3 For $\mathfrak{g} = \mathfrak{gl}_N$ and $V = \mathbb{F}^N$ the standard representation, Eq. (19) may be used to find a generating set (in the sense of PBW Theorem) for the quantum finite W-algebra, see [22] for more details.

4.2 The Generalized Yangian Identity

Let $\alpha, \beta, \gamma \in \mathbb{F}$. Let R be a unital associative algebra, and let V be an N-dimensional vector space. For $\beta \neq 0$, we also assume, as in Sect. 3.1, that V is endowed with a non-degenerate bilinear form $\langle \cdot \,|\, \cdot \rangle : V \times V \to \mathbb{F}$ which we assume to be symmetric or skewsymmetric, and we let $\epsilon = +1$ and -1, respectively. Again, when denoting an element of $R \otimes \mathrm{End}(V)$ or of $R \otimes \mathrm{End}(V) \otimes \mathrm{End}(V)$, we omit the tensor product sign on the first factor, i.e. we treat elements of R as scalars.

The *generalized* (α, β, γ)*-Yangian identity* for $A(z) \in R((z^{-1})) \otimes \mathrm{End}(V)$ is the following identity, holding in $R[[z^{-1}, w^{-1}]][z, w] \otimes \mathrm{End}(V) \otimes \mathrm{End}(V)$:

$$\begin{aligned} &(z - w + \alpha\Omega_V)(A(z) \otimes \mathbb{1}_V)(z + w + \gamma - \beta\Omega_V^\dagger)(\mathbb{1}_V \otimes A(w)) \\ &= (\mathbb{1}_V \otimes A(w))(z + w + \gamma - \beta\Omega_V^\dagger)(A(z) \otimes \mathbb{1}_V)(z - w + \alpha\Omega_V)\,. \end{aligned} \tag{21}$$

Recall that Ω_V and $\Omega_V^\dagger$ are defined by Eqs. (13) and (15), respectively.

Remark 4 In the special case $\alpha = 1$, $\beta = \gamma = 0$, Eq. (21) coincides with the so-called RTT presentation of the Yangian of $\mathfrak{gl}(V)$, cf. [19, 45]. Moreover, in the special case $\alpha = \beta = \frac{1}{2}$, $\gamma = 0$, Eq. (21) coincides with the so-called RSRS presentation of the extended twisted Yangian of $\mathfrak{g} = \mathfrak{so}(V)$ or $\mathfrak{sp}(V)$, depending on whether $\epsilon = +1$ or -1, cf. [45]. Hence, if $A(z) \in R((z^{-1})) \otimes \mathrm{End}\, V$ satisfies the generalized $(\frac{1}{2}, \frac{1}{2}, 0)$-Yangian identity we automatically have an algebra homomorphism from the extended twisted Yangian $X(\mathfrak{g})$ to the algebra R. If, moreover, $A(z)$ satisfies the symmetry condition (required in the definition of twisted Yangian in [45])

$$A^\dagger(-z) - \epsilon A(z) = -\frac{A(z) - A(-z)}{4z}\,,$$

then we have an algebra homomorphism from the twisted Yangian $Y(\mathfrak{g})$ to the algebra R.

4.3 Quantum Finite W-Algebras and (Extended) Twisted Yangians

Let $\mathfrak{g}$ be one of the classical Lie algebras $\mathfrak{gl}_N$, $\mathfrak{sl}_N$, $\mathfrak{so}_N$ or $\mathfrak{sp}_N$, and let $V = \mathbb{F}^N$ be its standard representation (endowed, in the cases of $\mathfrak{so}_N$ and $\mathfrak{sp}_N$, with a non-degenerate symmetric or skewsymmetric bilinear form, respectively). Then, the operator $A(z)$ defined in Eq. (17) satisfies the generalized Yangian identity (21), where α, β, γ are given by the following table:

$\mathfrak{g}$	α	β	γ
$\mathfrak{gl}_N$ or $\mathfrak{sl}_N$	1	0	0
$\mathfrak{so}_N$ or $\mathfrak{sp}_N$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{\epsilon}{2}$

Note that $V[\frac{d}{2}] \cong V[-\frac{d}{2}]$. Fix and isomorphism $\chi : V[\frac{d}{2}] \xrightarrow{\cong} V[-\frac{d}{2}]$. Then, $\chi \circ L(z) \in W(\mathfrak{g}, f)((z^{-1})) \otimes \mathrm{End}(V[-\frac{d}{2}])$. By an abuse of notation, we still denote this operator by $L(z)$. We also let $n = \dim V[-\frac{d}{2}]$.

The second main result in [21] states that, for classical Lie algebras, the Lax operator defined in (20) also satisfies a generalized Yangian identity.

Theorem 2 *The operator $L(z) \in W(\mathfrak{g}, f)((z^{-1})) \otimes \mathrm{End}(V[-\frac{d}{2}])$ defined by (19) and (20) (cf. Theorem 1) satisfies the generalized Yangian identity (21) with the values of α, β, γ as in the following table:*

$\mathfrak{g}$	α	β	γ
$\mathfrak{gl}_N$ *or* $\mathfrak{sl}_N$	1	0	0
$\mathfrak{so}_N$ *or* $\mathfrak{sp}_N$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{\epsilon-N+n}{2}$

By Theorem 2 and Remark 4 we have an algebra homomorphism from the extended twisted Yangian $X(\bar{\mathfrak{g}})$ ($\bar{\mathfrak{g}}$ depends on the pair $(\mathfrak{g}, f)$) to the quantum finite W-algebra $W(\mathfrak{g}, f)$. A stronger result has been obtained for $\mathfrak{g} = \mathfrak{gl}_N$ by Brundan and Kleshchev in [9] where quantum finite W-algebras were constructed as truncated shifted Yangians (which are subquotients of the Yangian for $\mathfrak{gl}_N$).

5 Classical Affine W-Algebras and Integrable Hierarchies of Lax Type Equations

5.1 Lax Type Operators for Classical Affine W-Algebras

For classical affine W-algebras the discussion is similar to the one in Sect. 4 but in a different setting: we need to substitute polynomials and Laurent series with

differential operators and pseudodifferential operators, respectively (see [20] for a review of their basic properties).

Consider the differential operators

$$A(\partial) = \partial \mathbb{1}_V + U = \partial \mathbb{1}_V + \sum_{i \in I} u_i U^i \ \in \mathcal{V}(\mathfrak{g})[\partial] \otimes \mathrm{End}(V)$$

and

$$A^\rho(\partial) = \partial \mathbb{1}_V + F + \pi_{\leq \frac{1}{2}} U = \partial \mathbb{1}_V + F + \sum_{i \in I_{\leq \frac{1}{2}}} u_i U^i \ \in \mathcal{V}(\mathfrak{g}_{\leq \frac{1}{2}})[\partial] \otimes \mathrm{End}\, V\,.$$

Recall from [15] that in the classical affine case we have $\mathcal{W}(\mathfrak{g}, f) \subset \mathcal{V}(\mathfrak{g}_{\leq \frac{1}{2}})$ and that there exists a differential algebra isomorphism $w : \mathcal{V}(\mathfrak{g}^f) \xrightarrow{\sim} \mathcal{W}(\mathfrak{g}, f)$. Consider the generalized quasideterminant (cf. (16))

$$L(\partial) = |A^\rho(\partial)|_{V[\frac{d}{2}], V[-\frac{d}{2}]} = \left(\mathbb{1}_{V[-\frac{d}{2}]} \left(\partial \mathbb{1}_V + F + \pi_{\leq \frac{1}{2}} U \right)^{-1} \mathbb{1}_{V[\frac{d}{2}]} \right)^{-1}. \tag{22}$$

The following result has been proved in [20].

Theorem 3 *$L(\partial) \in \mathcal{W}(\mathfrak{g}, f)((\partial^{-1})) \otimes \mathrm{Hom}\left(V\left[-\frac{d}{2}\right], V\left[\frac{d}{2}\right]\right)$ and*

$$L(\partial) = \left(\mathbb{1}_{V[-\frac{d}{2}]} \left(\partial \mathbb{1}_V + F + \sum_{i \in I_f} w(u_i) U^i \right)^{-1} \mathbb{1}_{V[\frac{d}{2}]} \right)^{-1}. \tag{23}$$

The above theorem consists of two statements. First, it claims that $L(\partial)$ is well defined, i.e. both inverses in formula (22) can be carried out in the algebra of pseudodifferential operators with coefficients in $\mathcal{V}(\mathfrak{g}_{\leq \frac{1}{2}})$, and that the coefficients of $L(\partial)$ lie in the $\mathcal{W}$-algebra $\mathcal{W}(\mathfrak{g}, f)$. Then, it gives a formula, Eq. (23), for $L(\partial)$ in terms of the generators $w(u_i)$, $i \in I_f$, of the $\mathcal{W}$-algebra $\mathcal{W}(\mathfrak{g}, f)$.

5.2 *Integrable Hierarchies of Lax Type Equation*

Let $\mathfrak{g}$ be one of the classical Lie algebras $\mathfrak{gl}_N$, $\mathfrak{sl}_N$, $\mathfrak{so}_N$ or $\mathfrak{sp}_N$, and let $V = \mathbb{F}^N$ be its standard representation (endowed, in the cases of $\mathfrak{so}_N$ and $\mathfrak{sp}_N$, with a non-degenerate symmetric or skewsymmetric bilinear form, respectively). Then, we can use the operator $L(\partial)$ in (23) to get explicit formulas for the λ-brackets among the generators of $\mathcal{W}(\mathfrak{g}, f)$ and construct integrable hierarchies of Hamiltonian equations, see [20].

Theorem 4

(1) $L(\partial)$ *satisfies the generalized Adler type identity*

$$\begin{aligned}\{L(z)_\lambda L(w)\} &= \alpha(\mathbb{1}_V \otimes L(w+\lambda+\partial))(z-w-\lambda-\partial)^{-1}\\ &\times (L^*(\lambda-z)\otimes \mathbb{1}_V)\Omega_V\\ &- \alpha\Omega_V\left(L(z)\otimes(z-w-\lambda-\partial)^{-1}L(w)\right)\\ &- \beta(\mathbb{1}_V\otimes L(w+\lambda+\partial))\Omega_V^\dagger(z+w+\partial)^{-1}(L(z)\otimes\mathbb{1}_V)\\ &+ \beta(L^*(\lambda-z)\otimes\mathbb{1}_V)\Omega_V^\dagger(z+w+\partial)^{-1}(\mathbb{1}_V\otimes L(w))\\ &+ \gamma\left(\mathbb{1}_V\otimes\left(L(w+\lambda+\partial)-L(w)\right)\right)(\lambda+\partial)^{-1}\\ &\times\left(\left(L^*(\lambda-z)-L(z)\right)\otimes\mathbb{1}_V\right),\end{aligned} \tag{24}$$

for the following values of $\alpha, \beta, \gamma \in \mathbb{F}$*:*

$\mathfrak{g}$	α	β	γ
$\mathfrak{gl}_N$	1	0	0
$\mathfrak{sl}_N$	1	0	$\frac{1}{N}$
$\mathfrak{so}_N$ *or* $\mathfrak{sp}_N$	$\frac{1}{2}$	$\frac{1}{2}$	0

In Eq. (24) L^* *denotes the formal adjoint of pseudodifferential operators, and* Ω_V *and* $\Omega_V^\dagger$ *are defined by Eqs.* (13) *and* (15)*, respectively.*

(2) For $B(\partial)$ *a* K*-th root of* $L(\partial)$ *(i.e.,* $L(\partial) = B(\partial)^K$ *for* $K \geq 1$*) define the elements* $h_{n,B} \in \mathcal{W}(\mathfrak{g}, f)$*,* $n \in \mathbb{Z}_{\geq}0$*, by (*$\mathrm{tr} = 1 \otimes \mathrm{tr}$*)*

$$h_{n,B} = \frac{-K}{n}\,\mathrm{Res}_z\,\mathrm{tr}(B^n(z)) \text{ for } n > 0\,,\ \ h_0 = 0\,.$$

Then, all the elements $\int h_{n,B}$ *are Hamiltonian functionals in involution and we have the corresponding integrable hierarchy of Lax type Hamiltonian equations*

$$\frac{dL(w)}{dt_{n,B}} = \{\textstyle\int h_{n,B}, L(w)\} = [\alpha(B^n)_+ - \beta((B^n)^{*^\dagger})_+, L](w)\,,\ \ n \in \mathbb{Z}_{\geq 0}\,. \tag{25}$$

(In the RHS of (25) *we are taking the symbol of the commutator of matrix pseudodifferential operators.)*

Remark 5 For $\beta = 0$ solutions to the integrable hierarchy (25) can be obtained by reductions of solutions to the multicomponent KP hierarchy, see [11].

Acknowledgments This review is based on the talk I gave at the XIth International Symposium Quantum Theory and Symmetry in Montréal. I wish to thank the organizers for the invitation and the hospitality.

References

1. L.F. Alday, D. Gaiotto, Y. Tachikawa, Liouville correlation functions from four-dimensional gauge theories. Lett. Math. Phys. **91**(2), 167–197 (2010)
2. T. Arakawa, Representation theory of *W-algebras and Higgs branch conjecture*, in Proceedings of the International Congress of Mathematicians - Rio de Janeiro 2018. Invited Lectures, vol. II (World Scientific Publishing, Hackensack), pp. 1263–1281
3. Bakalov B., Milanov T., *$\mathcal{W}$-constraints for the total descendant potential of a simple singularity*. Compos. Math. **149**(5), 840–888 (2013)
4. A. Barakat, A. De Sole, V.G. Kac, Poisson vertex algebras in the theory of Hamiltonian equations. Jpn. J. Math. **4**(2), 141–252 (2009)
5. C. Beem, L. Rastelli, Vertex operator algebras, Higgs branches, and modular differential equations. J. High Energ. Phys. **2018**, 114 (2018)
6. R. Borcherds, Vertex algebras, Kac-Moody algebras and the Monster. Proc. Natl. Acad. Sci. U. S. A. **83**, 3068–3071 (1986)
7. P. Bouwknegt, K. Schoutens, *W-symmetry*. Advanced Series in Mathematical Physics, vol. 22 (World Scientific, Singapore, 1995)
8. A. Braverman, M. Finkelberg, H. Nakajima, Instanton moduli spaces and W-algebras. Astérisque **385**, pp. vii+128, (2016)
9. J. Brundan, A. Kleshchev, Shifted Yangians and finite W-algebras Adv. Math. **200**(1), 136–195 (2006)
10. N. Burruoughs, M. de Groot, T. Hollowood, L. Miramontes, Generalized Drinfeld-Sokolov hierarchies II: the Hamiltonian structures. Commun. Math. Phys. **153**, 187–215 (1993)
11. S. Carpentier, A. De Sole, V.G. Kac, D. Valeri, J. van de Leur, $\underline{p}$-reduced multicomponent KP hierarchy and classical $\mathcal{W}$-algebras $\mathcal{W}(\mathfrak{gl}_N, \underline{p})$ (2019). arXiv:1909.03301
12. M. de Groot, T. Hollowood, L. Miramontes, Generalized Drinfeld-Sokolov hierarchies. Commun. Math. Phys. **145**, 57–84 (1992)
13. A. De Sole, On classical finite and affine W-algebras, in *Advances in Lie Superalgebras*. Springer INdAM Series, vol. 7 (Springer, Berlin, 2014), pp. 51–66
14. A. De Sole, V.G. Kac, Finite vs. affine W-algebras. Jpn. J. Math. **1**(1), 137–261 (2006)
15. A. De Sole, V.G. Kac, D. Valeri, Classical $\mathcal{W}$-algebras and generalized Drinfeld-Sokolov bi-Hamiltonian systems within the theory of Poisson vertex algebras. Commun. Math. Phys. **323**(2), 663–711 (2013)
16. A. De Sole, V.G. Kac, D. Valeri, Classical $\mathcal{W}$-algebras and generalized Drinfeld-Sokolov hierarchies for minimal and short nilpotents. Commun. Math. Phys. **331**(2), 623–676 (2014)
17. A. De Sole, V.G. Kac, D. Valeri, Structure of classical (finite and affine) $\mathcal{W}$-algebras. J. Eur. Math. Soc. **18**(9), 1873–1908 (2016)
18. A. De Sole, V.G. Kac, D. Valeri, Classical affine $\mathcal{W}$-algebras for $\mathfrak{gl}_N$ and associated integrable Hamiltonian hierarchies. Commun. Math. Phys. **348**(1), 265–319 (2016)
19. A. De Sole, V.G. Kac, D. Valeri, Finite $\mathcal{W}$-algebras for $\mathfrak{gl}_N$. Adv. Math. **327**, 173–224 (2018)
20. A. De Sole, V.G. Kac, D. Valeri, Classical affine W-algebras and the associated integrable hierarchies for classical Lie algebras. Commun. Math. Phys. **360**(3), 851–918 (2018)
21. A. De Sole, V.G. Kac, D. Valeri, A Lax type operator for quantum finite W-algebras. Sel. Math. New Ser. **24**(5), 4617–4657 (2018)
22. A. De Sole, L. Fedele, D. Valeri, Generators of the quantum finite W-algebras in type A. J. Algebra Appl. **19**(9), 2050175, p. 76

23. F. Delduc, L. Fehér, Regular conjugacy classes in the Weyl group and integrable hierarchies. J. Phys. A **28**(20), 5843–5882 (1995)
24. V.G. Drinfeld, V.V. Sokolov, Lie algebras and equations of KdV type. Soviet J. Math. **30**, 1975–2036 (1985)
25. V.A. Fateev, S.L. Lukyanov, The models of two-dimensional conformal quantum field theory with $\mathbb{Z}_n$ symmetry. Int. J. Mod. Phys. A **3**(2), 507–520 (1988)
26. L. Fehér, J. Harnad, I. Marshall, Generalized Drinfeld-Sokolov reductions and KdV type hierarchies. Commun. Math. Phys. **154**(1), 181–214 (1993)
27. B.L. Feigin, E. Frenkel, Quantization of Drinfeld-Sokolov reduction. Phys. Lett. B **246**, 75–81 (1990)
28. B.L. Feigin, E. Frenkel, Quantization of soliton systems and Langlands duality, in *Exploring New Structures and Natural Constructions in Mathematical Physics*. Advanced Studies in Pure Mathematics, vol. 61 (Mathematical Society Japan, Tokyo, 2011), pp. 185–274
29. C. Fernández-Pousa, M. Gallas, L. Miramontes, J. Sánchez Guillén, $\mathcal{W}$-algebras from soliton equations and Heisenberg subalgebras. Ann. Phys. **243**(2), 372–419 (1995)
30. C. Fernández-Pousa, M. Gallas, L. Miramontes, J. Sánchez Guillén, *Integrable Systems and $\mathcal{W}$-Algebras*. VIII J. A. Swieca Summer School on Particles and Fields (Rio de Janeiro, 1995), pp. 475–479
31. E. Frenkel, Affine algebras, Langlands duality and Bethe ansatz, in *XIth International Congress of Mathematical Physics (Paris, 1994)* (International Press, Cambridge, 1995), pp. 606–642
32. E. Frenkel, T. Arakawa, Quantum Langlands duality of representations of W-algebras. Compos. Math. **155**(12), 2235–2262 (2019)
33. E. Frenkel, D. Hernandez, Spectra of quantum KdV Hamiltonians, Langlands duality, and affine opers. Comm. Math. Phys. **362**(2), 361–414 (2018)
34. W.L. Gan, V. Ginzburg, Quantization of Slodowy slices. Int. Math. Res. Not. **5**, 243–255 (2002)
35. I.M. Gelfand, S.I. Gelfand, V. Retakh, R.I. Wilson, Quasideterminants. Adv. Math. **193**(1), 56–141 (2005)
36. V.G. Kac, M. Wakimoto, Quantum reduction and representation theory of superconformal algebras. Adv. Math. **185**, 400–458 (2004). Corrigendum, Adv. Math. **193**, 453–455 (2005)
37. V.G. Kac, M. Wakimoto, Quantum reduction in the twisted case. Progress Math. **237**, 85–126 (2005)
38. V.G. Kac, S.-S. Roan, M. Wakimoto, Quantum reduction for affine superalgebras. Commun. Math. Phys. **241**, 307–342 (2003)
39. B. Kostant, On Whittaker vectors and representation theory. Inv. Math **48**, 101–184 (1978)
40. S.L. Lukyanov, V.A. Fateev, Exactly soluble models of conformal quantum field theory associated with the simple Lie algebra D_n. Yadernaya Fiz. **49**(5), 1491–1504 (1989)
41. T.E. Lynch, Generalized Whittaker vectors and representation theory, Thesis (Ph.D.), Massachusetts Institute of Technology, 1979
42. F. Magri, A simple model of the integrable Hamiltonian equation. J. Math. Phys. **19**(5), 1156–1162 (1978)
43. H. Matumoto, Whittaker modules associated with highest weight modules. Duke Math. J. **60**, 59–113 (1990)
44. T. Milanov, D. Lewanski, $\mathcal{W}$-algebra constraints and topological recursion for A_N-singularity (with an appendix by Danilo Lewanski). Int. J. Math. **27**(13), 1650110 (2016)
45. A. Molev, *Yangians and Classical Lie Algebras*. Mathematical Surveys and Monographs, vol. 143 (American Mathematical Society, Providence, 2007)
46. A. Premet, Special transverse slices and their enveloping algebras. Adv. Math. **170**, 1–55 (2002)
47. A. Premet, Enveloping algebras of Slodowy slices and the Joseph ideal. J. Eur. Math. Soc. (JEMS) **9**(3), 487–543 (2007)
48. O. Schiffmann, E. Vasserot, Cherednik algebras, W-algebras and the equivariant cohomology of the moduli space of instantons on A_2. Publ. Math. Inst. Hautes Etudes Sci. **118**, 213–342 (2013)

49. P. Slodowy, *Simple Singularities and Simple Algebraic Groups*. Lecture Notes in Mathematics, vol. 815 (Springer, Berlin, 1980)
50. A. Zamolodchikov, Infinite extra symmetries in two-dimensional conformal quantum field theory. Teor. Mat. Fiz. **65**(3),347–359 (1985)
51. Y. Zhu, Modular invariance of characters of vertex operator algebras. J. AMS **9**, 237–302 (1996)

Color Algebraic Extension of Supersymmetric Quantum Mechanics

Naruhiko Aizawa, Kosuke Amakawa, and Shunya Doi

Abstract In the recent paper, Bruce and Duplij introduced a $\mathbb{Z}_2^2$-graded version of supersymmetric quantum mechanics (SQM). It is an extension of Lie superalgebraic nature of $\mathcal{N} = 1$ SQM to a $\mathbb{Z}_2^2$-graded color superalgebra. We present three extensions of the result of Bruce and Duplij. Namely, $\mathbb{Z}_2^2$-graded SQM with higher values of $\mathcal{N}$, $\mathbb{Z}_2^2$-graded version of superconformal mechanics, and $\mathbb{Z}_2^3$-graded SQM. All these were done by realizations of color superalgebra in terms of ordinary Lie superalgebra.

Keywords $\mathbb{Z}_2^2$-graded Lie algebras · Supersymmetric quantum mechanics · Superconformal mechanics

1 Introduction

Supersymmetric and superconformal quantum mechanics have been discussed in surprisingly wide variety of problems in physics. Even in some of modern problems such as curved extra dimension or M-theory they play fundamental and important roles, see, for instance [1, 2] and references therein. It is, therefore, natural that there exist many considerations on possible extensions of supersymmetric quantum mechanics. Supersymmetric quantum mechanics (SQM) is a quantum mechanical realization of the super-Poincaré algebra in $(0 + 1)$-dimensional spacetime. Thus many extensions of SQM discuss possible replacement of Lie superalgebraic nature of super-Poincaré algebra with more general setting (also in connection with the no-go theorem of Coleman and Mandula).

One of the most recent works in this direction is due to Bruce [3] where $\mathbb{Z}_2$-grading of the super-Poincaré algebra in $(3 + 1)$ dimensional Minkowski space are replaced with $\mathbb{Z}_2^n$-grading. $\mathbb{Z}_2^n$ denotes a direct product of n copies of the

N. Aizawa (✉) · K. Amakawa · S. Doi
Department of Physical Science, Osaka Prefecture University, Sakai, Osaka, Japan
e-mail: aizawa@p.s.osakafu-u.ac.jp

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_18

abelian group $\mathbb{Z}_2$ and Lie algebras with this kind of grading are referred to as *color* (super)algebras in the literatures. Soon after this work, Bruce and Duplij presented a model of $\mathbb{Z}_2^2$-graded SQM based on the reduction of the $\mathbb{Z}_2^2$-graded super-Poincaré algebra of [3] to $(0+1)$-dimension [4] (references for other works on extension of SQM are found in [4], too). We call the model discussed in [4] Bruce–Duplij model.

In the present work, we interpret the Bruce–Duplij model from more general perspective. Namely, we provide a realization of $\mathbb{Z}_2^2$-graded color superalgebra by ordinary Lie superalgebra. Then it can be seen that Bruce–Duplij model is a special case of this realization. Moreover, the realization allows us further generalizations of SQM. The Bruce–Duplij model may be regarded as a $\mathbb{Z}_2^2$-graded version of $\mathcal{N} = 1$ SQM. By using the realization, one may easily construct models of $\mathbb{Z}_2^2$-graded version of SQM with higher values of $\mathcal{N}$. It is also possible to include conformal invariance, since one may apply the realization to many models of superconformal mechanics (SCM).

We remark that color superalgebras attract some physical interests in connection with symmetries of non-relativistic Dirac equation (Lévy-Leblond equation) and parastatistics [5, 6]. The present work, as well as [4], provides a new example of deep connection of such algebras and physics.

This paper is organized as follows: In the next section, we give a definition of $\mathbb{Z}_2^n$-graded color superalgebra and an algebraic basis of the Bruce–Duplij model. In Sect. 3, it is shown that if a matrix Lie superalgebra satisfies a certain condition, then one may obtain a $\mathbb{Z}_2^2$-graded color superalgebra with the same structure constants. This result is used to extend the result in [4] to extended supersymmetry and conformal supersymmetry. We also show in Sect. 4 that it is possible to construct $\mathbb{Z}_2^3$-graded SQM from the ordinary SQM as we did for $\mathbb{Z}_2^2$-graded case.

2 $\mathbb{Z}_2^n$-Graded Color Superalgebra and $\mathbb{Z}_2^2$-Graded SQM

We start with the definition of $\mathbb{Z}_2^n$-graded color superalgebra. Let $\mathfrak{g}$ be a vector space over $\mathbb{C}$ or $\mathbb{R}$ which is a direct sum of 2^n subspaces labelled by an element of the group $\mathbb{Z}_2^n$:

$$\mathfrak{g} = \bigoplus_{\boldsymbol{\alpha} \in \mathbb{Z}_2^n} \mathfrak{g}_{\boldsymbol{\alpha}}. \tag{1}$$

Regarding an element $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ of $\mathbb{Z}_2^n$ as an n dimensional vector, we define an inner product of two elements of $\mathbb{Z}_2^n$ by

$$\boldsymbol{\alpha} \cdot \boldsymbol{\beta} = \sum_{i=1}^{n} \alpha_i \beta_i. \tag{2}$$

Definition 1 If $\mathfrak{g}$ admits a bilinear form $[\![\ ,\ ]\!] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfying the following three relations, then $\mathfrak{g}$ is called a $\mathbb{Z}_2^n$-graded color superalgebra:

1. $[\![\mathfrak{g}_\alpha, \mathfrak{g}_\beta]\!] \subseteq \mathfrak{g}_{\alpha+\beta}$,
2. $[\![X_\alpha, X_\beta]\!] = -(-1)^{\alpha\cdot\beta}\,[\![X_\beta, X_\alpha]\!]$,
3. $[\![X_\alpha, [\![X_\beta, X_\gamma]\!]]\!](-1)^{\alpha\cdot\gamma} + \text{cyclic perm.} = 0$,

where $X_\alpha \in \mathfrak{g}_\alpha$ and the third relation is called the graded Jacobi identity.

It is easily recognized that the bilinear form $[\![X_\alpha, X_\beta]\!]$ is realized by commutator and anticommutator:

$$[\![X_\alpha, X_\beta]\!] = X_\alpha X_\beta - (-1)^{\alpha\cdot\beta} X_\beta X_\alpha. \tag{3}$$

The $n = 1$ case ($\mathbb{Z}_2$ grading) corresponds to the ordinary Lie superalgebras. The first non-trivial example is the $\mathbb{Z}_2^2$-graded color superalgebra consisting of four sectors labelled by $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$.

The $\mathbb{Z}_2^2$-graded version of SQM considered in [4] (Bruce–Duplij model) is a realization of $\mathbb{Z}_2^2$-graded super-Poincaré algebra ($\mathbb{Z}_2^2$-SPA) in the $\mathbb{Z}_2^2$-graded Hilbert space $L^2(\mathbb{R}) \otimes \mathbb{C}^4$. $\mathbb{Z}_2^2$-SPA is spanned by H_{00}, Q_{01}, Q_{10}, Z_{11} with the indicated $\mathbb{Z}_2^2$ grading and their non-vanishing relations are given by

$$\{Q_{01}, Q_{01}\} = \{Q_{10}, Q_{10}\} = H_{00}, \qquad [Q_{01}, Q_{10}] = i Z_{11}. \tag{4}$$

H_{00} is a diagonal matrix operator interpreted as a quantum mechanical Hamiltonian. Q_{01} and Q_{10} play the role of supercharges, however, they have different degree. As a consequence, they close by commutator (instead of anticommutator) into Z_{11}, the central element of the algebra.

3 Extensions of Bruce–Duplij Model

3.1 From Superalgebra to $\mathbb{Z}_2^2$-Graded Color Superalgebra

As is seen from (4), in the $\mathbb{Z}_2^2$-graded SQM of [4], each subspaces of degree $(0, 1)$ and $(1, 0)$ has only one supercharge. So one may say that it is a $\mathbb{Z}_2^2$-version of $\mathcal{N} = 1$ SQM. We would like to have quantum mechanical models which have more than one supercharges in each subspace. We also want models of $\mathbb{Z}_2^2$-graded version of SCM. These will be done by using the theorem shown below which relates an ordinary superalgebra to its $\mathbb{Z}_2^2$-graded version [7].

Let $\mathfrak{s}$ be an ordinary Lie superalgebra ($\mathbb{Z}_2$-graded Lie algebra) spanned by the elements T_i^a with $a \in \mathbb{Z}_2 = \{0, 1\}$. The defining relations may be written as

$$[T_i^0, T_j^0] = if_{ij}^k T_k^0, \qquad [T_i^0, T_j^1] = ih_{ij}^k T_k^1, \qquad \{T_i^1, T_j^1\} = g_{ij}^k T_k^0, \tag{5}$$

where the summation over the repeated indices is understood.

Suppose that we have a representation of $\mathfrak{s}$ in which odd (degree 1) elements are represented by block-antidiagonal Hermitian matrix of dimensions $2m \times 2m$. Suppose further that there exists a Hermitian block-diagonal matrix Γ of the same dimension which satisfies the relations

$$\{\Gamma, T_i^1\} = 0, \qquad \Gamma^2 = \mathbb{I}_{2m}, \tag{6}$$

where T_i^1 denotes the matrix representation of $\mathfrak{s}$ (slight abuse of notation) and $\mathbb{I}_{2m}$ denotes the $2m \times 2m$ identity matrix. It then follows that $[\Gamma, T_i^0] = 0$.

Now we define a set of Hermitian matrices:

$$\mathcal{T}_i^a = \mathbb{I}_2 \otimes T_i^a, \qquad \tilde{\mathcal{T}}_i^a = \sigma_1 \otimes i^a T_i^a \Gamma. \tag{7}$$

With these setting we have the followings:

Theorem 1 *Let $\hat{\mathfrak{s}}$ be the complex vector space spanned by the matrices (7). By the assignment of the $\mathbb{Z}_2^2$-degree*

$$\deg(\mathcal{T}_i^0) = (0,0), \quad \deg(\mathcal{T}_i^1) = (0,1), \quad \deg(\tilde{\mathcal{T}}_i^1) = (1,0), \quad \deg(\tilde{\mathcal{T}}_i^0) = (1,1), \tag{8}$$

$\hat{\mathfrak{s}}$ forms a $\mathbb{Z}_2^2$-graded color superalgebra with the defining relations:

$$\begin{aligned}
&[\mathcal{T}_i^0, \mathcal{T}_j^0] = if_{ij}^k \mathcal{T}_k^0, && [\mathcal{T}_i^0, \mathcal{T}_j^1] = ih_{ij}^k \mathcal{T}_k^1,\\
&[\mathcal{T}_i^0, \tilde{\mathcal{T}}_j^1] = ih_{ij}^k \tilde{\mathcal{T}}_k^1, && [\mathcal{T}_i^0, \tilde{\mathcal{T}}_j^0] = if_{ij}^k \tilde{\mathcal{T}}_k^0,\\
&\{\mathcal{T}_i^1, \mathcal{T}_j^1\} = \{\tilde{\mathcal{T}}_i^1, \tilde{\mathcal{T}}_j^1\} = g_{ij}^k \mathcal{T}_k^0, && [\tilde{\mathcal{T}}_i^0, \tilde{\mathcal{T}}_j^0] = if_{ij}^k \mathcal{T}_k^0,\\
&[\mathcal{T}_i^1, \tilde{\mathcal{T}}_j^1] = ig_{ij}^k \tilde{\mathcal{T}}_k^0, && \{\tilde{\mathcal{T}}_i^0, \mathcal{T}_j^1\} = -h_{ij}^k \tilde{\mathcal{T}}_k^1,\\
&\{\tilde{\mathcal{T}}_i^0, \tilde{\mathcal{T}}_j^1\} = h_{ij}^k \mathcal{T}_k^1. &&
\end{aligned} \tag{9}$$

If there exist models of SQM or SCM satisfying the condition of Theorem 1, then one may obtain their $\mathbb{Z}_2^2$-graded version immediately. As we see below, such models of SQM and SCM indeed exist.

3.2 $\mathcal{N}$ Extension of $\mathbb{Z}_2^2$-Graded SQM

In order to have a model of $\mathcal{N}$-extended version of Bruce–Duplij model, let us see that the model of $\mathcal{N}$-extended SQM by Akulov and Kudinov [8] satisfies the

condition of Theorem 1. The model is formulated by using matrix representations of the Clifford algebra. For a given positive integer n we introduce the $\mathcal{N} = 2n$ Hermitian block-antidiagonal matrices subject to the relations:

$$\{\gamma_I, \gamma_J\} = 2\delta_{IJ}\,\mathbb{I}_{2^n}, \qquad \gamma_I^\dagger = \gamma_I, \tag{10}$$

where I, J run from 1 to $\mathcal{N}$. We mainly work on an alternative choice of the basis of γ-matrices:

$$\gamma_a^\pm = \frac{1}{2}(\gamma_{2a-1} \pm i\gamma_{2a}), \quad a = 1, 2, \ldots, n. \tag{11}$$

In this basis the relation (10) reads as follows:

$$\{\gamma_a^\pm, \gamma_b^\pm\} = 0, \qquad \{\gamma_a^+, \gamma_b^-\} = \delta_{ab}\,\mathbb{I}_{2^n}. \tag{12}$$

We also consider n Hermitian block-diagonal matrices given by a product of γ_I's:

$$\Gamma_a = i^a \gamma_1 \gamma_2 \ldots \gamma_{2a}, \quad a = 1, 2, \ldots, n. \tag{13}$$

It is then immediate to verify that

$$\Gamma_a^2 = \mathbb{I}_{2^n}, \qquad [\Gamma_a, \Gamma_b] = 0 \tag{14}$$

and

$$[\gamma_k^\pm, \Gamma_a] = 0 \quad (k > a), \qquad \{\gamma_k^\pm, \Gamma_a\} = 0 \quad (k \le a). \tag{15}$$

The $\mathcal{N}$ supercharges of Akulov–Kudinov model are defined by the matrices $\gamma_a^\pm$ and Γ_a as follows:

$$Q_a^+ = \frac{1}{\sqrt{2}}\gamma_a^+(p + iW_a^{(n)}(x, \Gamma_1, \ldots, \Gamma_n)), \qquad Q_a^- = (Q_a^+)^\dagger. \tag{16}$$

The superpotentials $W_a^{(n)}$ are defined recursively. For instance, for $n = 1$ $W^{(1)}$ is chosen to be $W^{(1)} = w_0(x)$ and for $n = 2$

$$W_1^{(2)} = w_0(x) + \Gamma_2 w_1(x), \quad w_1(x) = \frac{\partial_x w_0(x)}{2w_0(x)} \tag{17}$$

and so on. It is seen from (15) that Γ_n anticommutes with all the supercharges:

$$\{Q_a^\pm, \Gamma_n\} = 0, \quad \forall a. \tag{18}$$

Thus one may apply Theorem 1 to Akulov–Kudinov model.

Proposition 1 *Let H be Hamiltonian of Akulov–Kudinov model: $\{Q_a^+, Q_b^-\} = \delta_{ab}H$ and $A = \begin{pmatrix} 0 & \Gamma_n \\ \Gamma_n & 0 \end{pmatrix}$. The complex vector space spanned by the matrix differential operators*

$$\mathcal{Q}_a^{\pm} = \mathbb{I}_2 \otimes Q_a^{\pm}, \quad \tilde{\mathcal{Q}}_a^{\pm} = i\mathcal{Q}_a^{\pm}A, \quad \mathcal{H} = \mathbb{I}_2 \otimes H, \quad \tilde{\mathcal{H}} = \mathcal{H}A \tag{19}$$

forms a $\mathbb{Z}_2^2$-graded superalgebra having the following non-vanishing relations:

$$\{\mathcal{Q}_a^+, \mathcal{Q}_b^-\} = \{\tilde{\mathcal{Q}}_a^+, \tilde{\mathcal{Q}}_b^-\} = \delta_{ab}\mathcal{H}, \qquad [\mathcal{Q}_a^{\pm}, \tilde{\mathcal{Q}}_b^{\mp}] = i\delta_{ab}\tilde{\mathcal{H}}. \tag{20}$$

The assignment of $\mathbb{Z}_2^2$-degree is

$$\deg(\mathcal{H}) = (0,0), \quad \deg(\mathcal{Q}_a^{\pm}) = (0,1), \quad \deg(\tilde{\mathcal{Q}}_a^{\pm}) = (1,0), \quad \deg(\tilde{\mathcal{H}}) = (1,1). \tag{21}$$

Thus (19) gives a $\mathcal{N}$-extended version of Bruce–Duplij model. The Bruce–Duplij model is recovered from (19) by setting $n = 1$ ($\mathcal{N} = 2$) and the identification

$$Q_{01} = \frac{1}{\sqrt{2}}(\mathcal{Q}_a^+ + \mathcal{Q}_a^-), \qquad Q_{10} = \frac{1}{\sqrt{2}}(\tilde{\mathcal{Q}}_a^+ + \tilde{\mathcal{Q}}_a^-). \tag{22}$$

More detailed analysis of $\mathcal{N} = 2$ case is found in [7].

3.3 $\mathbb{Z}_2^2$-Graded SCM

Now we consider $\mathbb{Z}_2^2$-graded version of SCM by using Theorem 1. Many models of SCM have been obtained so far (see, for instance, [2]). Some of the models, e.g. the ones in [9, 10], satisfy the condition (6) so that we may have models of $\mathbb{Z}_2^2$-graded SCM of $\mathcal{N} = 2, 4, 8$ and so on.

As an example, we here present $\mathcal{N} = 1$ model with $osp(1|2)$ symmetry:

$$Q = \frac{1}{\sqrt{2}}\left(\sigma_1 p - \sigma_2 \frac{\beta}{x}\right), \qquad S = \frac{x}{\sqrt{2}}\sigma_1,$$
$$H = \frac{1}{2}\left(p^2 + \frac{\beta^2}{x^2}\right)\mathbb{I}_2 + \frac{\beta}{2x^2}\sigma_3, \quad D = -\frac{1}{4}\{x, p\}\,\mathbb{I}_2, \quad K = \frac{x^2}{2}\,\mathbb{I}_2, \tag{23}$$

where σ_i is Pauli matrices and $\beta \in \mathbb{R}$ is a coupling constant. Conformal subalgebra $so(1,2)$ is given by $\langle\ H,\ D,\ K\ \rangle$. For this realization of $osp(1|2)$ one may immediately see that $\Gamma = \sigma_3$ commute with Q and S. Thus Theorem 1 gives us the following operators which is a model of $\mathbb{Z}_2^2$-graded SCM:

$$
\begin{aligned}
&(0,0):\ \mathcal{H}=\begin{pmatrix} H & 0\\ 0 & H\end{pmatrix}, && \mathcal{D}=\begin{pmatrix} D & 0\\ 0 & D\end{pmatrix}, && \mathcal{K}=\begin{pmatrix} K & 0\\ 0 & K\end{pmatrix},\\
&(0,1):\ \mathcal{Q}=\begin{pmatrix} Q & 0\\ 0 & Q\end{pmatrix}, && \mathcal{S}=\begin{pmatrix} S & 0\\ 0 & S\end{pmatrix},\\
&(1,0):\ \tilde{\mathcal{Q}}=i\begin{pmatrix} 0 & Q\sigma_3\\ Q\sigma_3 & 0\end{pmatrix}, && \tilde{\mathcal{S}}=i\begin{pmatrix} 0 & S\sigma_3\\ S\sigma_3 & 0\end{pmatrix},\\
&(1,1):\ \tilde{\mathcal{H}}=\begin{pmatrix} 0 & H\sigma_3\\ H\sigma_3 & 0\end{pmatrix}, && \tilde{\mathcal{D}}=\begin{pmatrix} 0 & D\sigma_3\\ D\sigma_3 & 0\end{pmatrix}, && \tilde{\mathcal{K}}=\begin{pmatrix} 0 & K\sigma_3\\ K\sigma_3 & 0\end{pmatrix}.
\end{aligned}
$$

One may analyze this model in a standard way of conformal mechanics. Namely, one consider eigenvalue problem of $\mathcal{R} = \mathcal{H} + \mathcal{K}$, instead of $\mathcal{H}$, by creation–annihilation operator. Details of the analysis are presented in [7].

4 $\mathbb{Z}_2^3$-Graded SQM

Encouraged by the results obtained so far, we try to build a model of $\mathbb{Z}_2^3$-graded version of SQM. Our strategy is similar to $\mathbb{Z}_2^2$ case, i.e., find a realization of $\mathbb{Z}_2^3$-graded super-Poincaré algebra by ordinary $\mathcal{N} = 1$ SUSY algebra which is defined by the supercharge Q and Hamiltonian H satisfying the relations:

$$\{Q, Q\} = 2H, \qquad [H, Q] = 0. \tag{24}$$

The $\mathbb{Z}_2^3$-SPA obtained from the result in [3] by dimensional reduction has the following elements:

$$
\begin{aligned}
&\mathcal{Q}_1\ (0,0,1),\ \ \mathcal{Q}_2\ \ (0,1,0),\ \mathcal{Q}_3\ (1,0,0),\ \mathcal{Q}_4\ (1,1,1),\\
&\mathcal{H}\ \ (0,0,0),\ \mathcal{Z}_{ab},
\end{aligned}
\tag{25}
$$

and the non-vanishing relations (cf. [11]):

$$\{\mathcal{Q}_a, \mathcal{Q}_a\} = 2\mathcal{H}, \quad [\mathcal{Q}_i, \mathcal{Q}_j] = 2i\mathcal{Z}_{ij}, \quad \{\mathcal{Q}_i, \mathcal{Q}_4\} = 2\mathcal{Z}_{i4}, \tag{26}$$

where a, b take a value from $1, 2, 3, 4$ and i, j are restricted to $1, 2, 3$. The $\mathbb{Z}_2^3$-degrees of $\mathcal{H}$ and $\mathcal{Q}_a$ are indicated in (25) and $\deg(\mathcal{Z}_{ab}) = \deg(\mathcal{Q}_a) + \deg(\mathcal{Q}_b)$.

In order to realize (25) in terms of (24) we introduce a complex representation of the Clifford algebra $Cl(4)$:

$$\gamma_1 = \sigma_1 \otimes \sigma_1, \quad \gamma_2 = \sigma_1 \otimes \sigma_2, \quad \gamma_3 = \sigma_1 \otimes \sigma_3, \quad \gamma_4 = \sigma_2 \otimes \mathbb{I}_2. \tag{27}$$

It is then straightforward to verify the following:

Proposition 2 *Let Γ be an Hermitian operator anticommuting with Q. We assume further that Γ^2 is the identity operator. Then*

$$\begin{aligned} \mathcal{Q}_1 &= \mathbb{I}_2 \otimes \mathbb{I}_4 \otimes Q, & \mathcal{Q}_2 &= i\sigma_1 \otimes \gamma_1 \otimes Q\Gamma, \\ \mathcal{Q}_3 &= i\sigma_1 \otimes \gamma_3 \otimes Q\Gamma, & \mathcal{Q}_4 &= i\sigma_3 \otimes \gamma_2\gamma_4 \otimes Q \end{aligned} \tag{28}$$

realizes the $\mathbb{Z}_2^3$-SPA (25).

For completeness, we give the formula of the Hamiltonian and the central elements:

$$\begin{aligned} \mathcal{H} &= \mathbb{I}_8 \otimes H, \\ \mathcal{Z}_{12} &= \sigma_1 \otimes \gamma_1 \otimes H\Gamma, & \mathcal{Z}_{13} &= \sigma_1 \otimes \gamma_3 \otimes H\Gamma, \\ \mathcal{Z}_{14} &= i\sigma_3 \otimes \gamma_2\gamma_4 \otimes H, & \mathcal{Z}_{23} &= -i\mathbb{I}_2 \otimes \gamma_1\gamma_3 \otimes H, \\ \mathcal{Z}_{24} &= -i\sigma_2 \otimes \gamma_1\gamma_2\gamma_4 \otimes H\Gamma, & \mathcal{Z}_{34} &= i\sigma_2 \otimes \gamma_2\gamma_3\gamma_4 \otimes H\Gamma. \end{aligned} \tag{29}$$

By (28) any model of (24) may be converted to the corresponding $\mathbb{Z}_2^3$-graded version if the operator Γ exists. Such Γ would exist in many cases. The simplest example is a single particle moving in one-dimensional space. In this case, Q consists of σ_1, σ_2 and differential operator, so that one may take $\Gamma = \sigma_3$.

In the present work, we showed that Bruce–Duplij model of $\mathbb{Z}_2^2$-graded SQM is easily extended to higher values of $\mathcal{N}$ and to superconformal setting. It was also shown that $\mathcal{N} = 1$ SQM is possible to generalize $\mathbb{Z}_2^3$-graded SQM. These were done by finding a realization of a color superalgebra by an ordinary Lie superalgebra. Therefore, we expect that it is possible to obtain models of $\mathbb{Z}_2^n$-graded SQM and SCM in a similar and a systematic way. If it is the case, then color superalgebras would be a quite natural object in analysis of physical problems.

References

1. I. Ueba, Extended supersymmetry with central charges in Dirac action with curved extra dimensions (2019). arXiv:1905.11673 [math-ph]
2. T. Okazaki, Superconformal quantum mechanics from M2-branes, Ph.D thesis (2015). arXiv:1503.03906 [hep-th]
3. A.J. Bruce, On a $\mathbb{Z}_2^n$-graded version of supersymmetry. Symmetry **11**, 116 (2019). https://doi.org/10.3390/sym11010116
4. A.J. Bruce, S. Duplij, Double-graded supersymmetric quantum mechanics (2019). arXiv:1904.06975 [math-ph]
5. N. Aizawa, Z. Kuznetsova, H. Tanaka, F. Toppan, $\mathbb{Z}_2 \times \mathbb{Z}_2$-graded Lie Symmetries of the Lévy-Leblond Equations. Prog. Theor. Exp. Phys. **2016**, 123A01 (2016). https://doi.org/10.1093/ptep/ptw176
6. V.N. Tolstoy, Once more on parastatistics. Phys. Part. Nucl. Lett. **11**, 933 (2014). https://doi.org/10.1134/S1547477114070449

7. N. Aizawa, K. Amakawa, S. Doi, $\mathcal{N}$-Extension of double-graded supersymmetric and superconformal quantum mechanics (2019). arXiv:1905.06548 [math-ph]
8. V. Akulov, M. Kudinov, Extended supersymmetric quantum mechanics. Phys. Lett. B **460**, 365 (1999). https://doi.org/10.1016/S0370-2693(99)00773-X
9. N. Aizawa, Z. Kuznetsova, F. Toppan, The quasi-nonassociative exceptional $F(4)$ deformed quantum oscillator. J. Math. Phys. **59**, 022101 (2018). https://doi.org/10.1063/1.5016915
10. N. Aizawa, I.E. Cunha, Z. Kuznetsova, F. Toppan, On the spectrum-generating superalgebras of the deformed one-dimensional quantum oscillators. J. Math. Phys. **60**, 042102 (2019). https://doi.org/10.1063/1.5085164
11. B. Le Roy, $\mathbb{Z}_n^3$-Graded colored supersymmetry. Czech. J. Phys. **47**, 47 (1997). https://doi.org/10.1023/A:1021491927893

The Racah Algebra and $\mathfrak{sl}_n$

Hendrik De Bie, Luc Vinet, and Wouter van de Vijver

Abstract We conjecture the existence of an embedding of the Racah algebra into the universal enveloping algebra of $\mathfrak{sl}_n$. Evidence of this conjecture is offered by realizing both algebras using differential operators and giving an embedding in this realization.

Keywords Racah algebra · Embedding · Lie algebra $\mathfrak{sl}_n$

1 Introduction

The Racah algebra synthesizes the properties of the Racah polynomials [8, 13], which are the most complicated univariate discrete orthogonal polynomials in the Askey scheme [11].

Multivariate Racah polynomials were introduced by Tratnik in [12]. These polynomials also have a solid algebraic underpinning, as was recently established in [3] using the higher rank Racah algebra. This higher rank Racah algebra was initially introduced in [9, 10] in the context of superintegrability and later in [4] as a subalgebra of intermediate Casimir elements in the n-fold tensor product of $\mathfrak{su}(1,1)$.

Although the initial motivation to introduce the (higher rank) Racah algebra was to establish a connection with the multivariate Racah polynomials, the algebra has now become an independent object of study. In particular its relation with other algebraic structures is part of ongoing investigations. We refer the reader to [1] for

H. De Bie (✉) · W. van de Vijver
Department of Electronics and Information Systems, Faculty of Engineering and Architecture, Ghent University, Gent, Belgium
e-mail: hendrik.debie@ugent.be; wouter.vandevijver@ugent.be

L. Vinet
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: vinet@crm.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_19

connections with Brauer and Temperley-Lieb algebras, and to [6] for connections with Howe duality.

The present paper aims to provide evidence for the following conjecture:

Conjecture 1 There exists an embedding of the higher rank Racah algebra into the universal enveloping algebra of the Lie algebra $\mathfrak{sl}_n$.

Indeed, we construct this embedding for a differential operator realization of $\mathcal{R}_n$ (recently introduced in [5]) in the enveloping algebra of a differential operator realization of $\mathfrak{sl}_n$. Note that the embedding in the rank one case was already constructed in [7].

2 Definition of the Higher Rank Racah Algebra

The algebra $\mathfrak{su}(1,1)$ is generated by three elements $A_\pm$ and A_0 with following relations:

$$[A_-, A_+] = 2A_0, \qquad [A_0, A_\pm] = \pm A_\pm.$$

Its universal enveloping algebra $\mathcal{U}(\mathfrak{su}(1,1))$ contains the Casimir element of $\mathfrak{su}(1,1)$:

$$C := A_0^2 - A_0 - A_+A_-.$$

We define the following elements of $\mathcal{U}(\mathfrak{su}(1,1))^{\otimes n}$ for $1 \leq k \leq n$:

$$A_{0,k} := 1^{\otimes(k-1)} \otimes A_0 \otimes 1^{\otimes(n-k)},$$
$$A_{\pm,k} := 1^{\otimes(k-1)} \otimes A_\pm \otimes 1^{\otimes(n-k)}.$$

For any non-empty subset $K \subset [n] := \{1, \ldots, n\}$ we define similarly

$$A_{0,K} := \sum_{k\in K} A_{0,k}, \quad A_{\pm,K} := \sum_{k\in K} A_{\pm,k}.$$

The three operators $A_{0,K}$ and $A_{\pm,K}$ generate an algebra isomorphic to $\mathfrak{su}(1,1)$. Its Casimir is given by

$$C_K := A_{0,K}^2 - A_{0,K} - A_{+,K}A_{-,K}.$$

These operators generate the higher rank Racah algebra.

Definition 1 The higher rank Racah algebra $\mathcal{R}_n$ is the subalgebra of $\mathcal{U}(\mathfrak{su}(1,1))^{\otimes n}$ generated by the set of operators

$$\{C_A \mid A \subset \{1, \ldots, n\} \text{ and } A \neq \emptyset\}.$$

A full account of $\mathcal{R}_n$ is presented in [4]. We mention one fact here: the operators C_A are not linearly independent. By formula (17) in [4] we have

$$C_A = \sum_{\{i,j\}\subset A} C_{ij} - (|A| - 2)\sum_{i\in A} C_i.$$

Hence, if one wants to present realizations of $\mathcal{R}_n$, it suffices to give expressions for the operators C_{ij} and C_i. In Sect. 4 we will present the higher rank Racah algebra as given in [5] this way.

3 Realizing $\mathfrak{sl}_n$ in $n-1$ Variables

Let $\mathfrak{sl}_n(\mathbb{R})$ be the algebra of $n \times n$ matrices whose trace equals zero and with the commutator as Lie bracket. Let E_{ij} be the matrix whose entries are equal to 0 except for the entry on the ith row and jth column which equals 1. Then the Lie algebra $\mathfrak{sl}_n(\mathbb{R})$ is generated by the set

$$\{E_{ij} | 1 \leq i, j \leq n \text{ and } i \neq j\} \cup \{E_{ii} - E_{nn} | 1 \leq i \leq n-1\}.$$

Let $u_i, i \in \{1 \ldots n-1\}$ be real variables. We introduce the differential operators:

$$\begin{aligned} T_{ij} &:= -u_j \partial_i \qquad i \neq j \text{ and } i, j < n \\ T_{in} &:= -\partial_i \qquad\qquad i < n \\ T_{nj} &:= u_j \tilde{\mathbb{E}} \qquad\qquad j < n \\ \tilde{T}_d &:= -u_d \partial_d - \tilde{\mathbb{E}} \qquad d < n, \end{aligned}$$

where the operator $\tilde{\mathbb{E}}$ is defined as

$$\tilde{\mathbb{E}} := \sum_{i=1}^{n-1} u_i \partial_i - k.$$

Using again the commutator as Lie bracket we denote by $\mathfrak{D}_n$ the Lie algebra spanned by all the $\tilde{T}_d$ and T_{ij}. The real number k is a deformation parameter that leaves the algebra relations invariant. One observes that $\mathfrak{sl}_n(\mathbb{R})$ and $\mathfrak{D}_n$ are isomorphic. The isomorphism σ is given by

$$\begin{aligned} \sigma(E_{ij}) &= T_{ij} \\ \sigma(E_{dd} - E_{nn}) &= \tilde{T}_d. \end{aligned}$$

Note that this isomorphism does not extend to their universal enveloping algebras.

3.1 Some Operators in $\mathcal{U}(\mathfrak{D}_n)$

We introduce a number of operators we need later on. Observe that the operator $\tilde{\mathbb{E}}$ is in the universal enveloping algebra $\mathcal{U}(\mathfrak{D}_n)$ but not in $\mathfrak{D}_n$ because $\mathfrak{D}_n$ lacks the identity:

$$\tilde{\mathbb{E}} = \frac{-1}{n}\left(k + \sum_{d=1}^{n-1} \tilde{T}_d\right).$$

Let $u_B := \sum_{k\in B} u_k$. We have

$$u_B\tilde{\mathbb{E}} := \sum_{j\in B} T_{nj}.$$

We will also express $u_B\partial_\alpha$ in function of the generators:

$$u_B\partial_\alpha := -\delta_{\alpha B}(\tilde{T}_\alpha + \tilde{\mathbb{E}}) - \sum_{j\in B\setminus\alpha} T_{\alpha j}, \tag{1}$$

where we introduced a new symbol standing for:

$$\delta_{\alpha B} := \begin{cases} 0, & \text{if } \alpha \notin B \\ 1, & \text{if } \alpha \in B. \end{cases}$$

It is then easy to check the following Lemma.

Lemma 1 *The following holds*

$$[u_B\tilde{\mathbb{E}}, \partial_\alpha] = -u_B\partial_\alpha - \delta_{\alpha B}\tilde{\mathbb{E}}$$
$$[u_A, \partial_\alpha] = -\delta_{\alpha A}.$$

4 Realization of $\mathcal{R}_n$ in $n-2$ Variables

In [5] an explicit differential operator realization of $\mathcal{R}_n$ was given in Theorem 5. We repeat this theorem here.

Theorem 1 *The space Π_k^{n-2} of all polynomials of degree k in $n-2$ variables carries a realization of the rank $n-2$ Racah algebra $\mathcal{R}_n$. This realization is given explicitly by*

$$\widetilde{C}_i = \nu_i(\nu_i - 1), \qquad i \in [n]$$

and, for $i, j \in \{3, \dots, n\}$,

$$\begin{aligned}
\widetilde{C_{12}} &= -\left(k-1-\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)\left(-k-\partial_{u_1}+\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)+2\nu_2\left(k-\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)\\
&\quad -2\nu_1\left(-k-\partial_{u_1}+\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)+(\nu_1+\nu_2)(\nu_1+\nu_2-1)\\
\widetilde{C_{1j}} &= -\left(1-\sum_{\ell=1}^{j-2} u_\ell\right)^2\left(k-1-\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +2\nu_j\left(1-\sum_{\ell=1}^{j-2} u_\ell\right)\left(k-\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)-2\nu_1\left(1-\sum_{\ell=1}^{j-2} u_\ell\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +(\nu_1+\nu_j)(\nu_1+\nu_j-1)\\
\widetilde{C_{2j}} &= -\left(\sum_{\ell=1}^{j-2} u_\ell\right)^2\left(1-k-\partial_{u_1}+\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +2\nu_j\left(\sum_{\ell=1}^{j-2} u_\ell\right)\left(k+\partial_{u_1}-\sum_{\ell=1}^{n-2} u_\ell \partial_{u_\ell}\right)+2\nu_2\left(\sum_{\ell=1}^{j-2} u_\ell\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +(\nu_2+\nu_j)(\nu_2+\nu_j-1)\\
\widetilde{C_{ij}} &= -\left(\sum_{\ell=j-1}^{i-2} u_\ell\right)^2\left(\partial_{u_{i-2}}-\partial_{u_{i-1}}\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +2\nu_j\left(\sum_{\ell=j-1}^{i-2} u_\ell\right)\left(\partial_{u_{i-2}}-\partial_{u_{i-1}}\right)-2\nu_i\left(\sum_{\ell=j-1}^{i-2} u_\ell\right)\left(\partial_{u_{j-2}}-\partial_{u_{j-1}}\right)\\
&\quad +(\nu_i+\nu_j)(\nu_i+\nu_j-1),
\end{aligned}$$

where we assume $i > j$ *and with* $u_{n-1} = 0$ *whenever it appears.*

We want to express these operators as elements in $\mathcal{U}(\mathfrak{D}_{n-1})$.

4.1 The Differential Embedding

To show that each generator of $\mathcal{R}_n$ is in $\mathcal{U}(\mathfrak{D}_{n-1})$, we will express each generator in function of ∂_α, $\tilde{\mathbb{E}}$, $u_B\partial_\alpha$, and $u_B\tilde{\mathbb{E}}$. Let us start with the operator $\widetilde{C_{12}}$. We can express this operator as follows:

$$\begin{aligned}\widetilde{C_{12}} = &-\left(-\tilde{\mathbb{E}}-1\right)\left(-\partial_1+\tilde{\mathbb{E}}\right)+2\nu_2\left(-\tilde{\mathbb{E}}\right)-2\nu_1\left(-\partial_1+\tilde{\mathbb{E}}\right)\\ &+(\nu_1+\nu_2)(\nu_1+\nu_2-1).\end{aligned}$$

As $\partial_1 = -T_{1n-1}$ and $\tilde{\mathbb{E}}$ are in $\mathcal{U}(\mathfrak{D}_{n-1})$ so is $\widetilde{C_{12}}$.

Consider the first term of the operator $\widetilde{C_{1j}}$:

$$\begin{aligned}&-\left(1-u_{[j-2]}\right)^2\left(-1-\tilde{\mathbb{E}}\right)\left(\partial_{j-2}-\partial_{j-1}\right)\\ &=\left(1-u_{[j-2]}\right)^2\left(\partial_{j-2}-\partial_{j-1}\right)\tilde{\mathbb{E}}\\ &=\left(1-u_{[j-2]}\right)\left(\left(\partial_{j-2}-\partial_{j-1}\right)\left(1-u_{[j-2]}\right)+1\right)\tilde{\mathbb{E}}\\ &=\left(1-u_{[j-2]}\right)\left(\partial_{j-2}-\partial_{j-1}\right)\left(1-u_{[j-2]}\right)\tilde{\mathbb{E}}+\left(1-u_{[j-2]}\right)\tilde{\mathbb{E}}.\end{aligned}$$

In line 3 we used Lemma 1. Let

$$\begin{aligned}\mathbb{L}_1^{(j)} &:= \left(1-u_{[j-2]}\right)\left(\partial_{j-2}-\partial_{j-1}\right)\\ \mathbb{L}_2^{(j)} &:= \left(1-u_{[j-2]}\right)\tilde{\mathbb{E}}.\end{aligned}$$

Both $\mathbb{L}_1^{(j)}$ and $\mathbb{L}_2^{(j)}$ can be expressed in function of the generators of $\mathcal{U}(\mathfrak{D}_{n-1})$, because of expression (1). The operator $\widetilde{C_{1j}}$ can be expressed as follows:

$$\widetilde{C_{1j}} = \mathbb{L}_1^{(j)}\mathbb{L}_2^{(j)} - (2\nu_j-1)\mathbb{L}_2^{(j)} - 2\nu_1\mathbb{L}_1^{(j)} + (\nu_1+\nu_j)(\nu_1+\nu_j-1).$$

This means that $\widetilde{C_{1j}}$ is also in $\mathcal{U}(\mathfrak{D}_{n-1})$.

Consider the first term of the operator $\widetilde{C_{2j}}$:

$$\begin{aligned}&-u_{[j-2]}^2\left(1-\partial_1+\tilde{\mathbb{E}}\right)\left(\partial_{j-2}-\partial_{j-1}\right)\\ &=-u_{[j-2]}^2\left(\partial_{j-2}-\partial_{j-1}\right)\left(-\partial_1+\tilde{\mathbb{E}}\right)\\ &=-u_{[j-2]}\left(\left(\partial_{j-2}-\partial_{j-1}\right)u_{[j-2]}-1\right)\left(-\partial_1+\tilde{\mathbb{E}}\right)\\ &=-u_{[j-2]}\left(\partial_{j-2}-\partial_{j-1}\right)u_{[j-2]}\left(-\partial_1+\tilde{\mathbb{E}}\right)+u_{[j-2]}\left(-\partial_1+\tilde{\mathbb{E}}\right).\end{aligned}$$

In line 3 we used Lemma 1. Let

$$\mathbb{L}_3^{(j)} := u_{[j-2]}\left(\partial_{j-2} - \partial_{j-1}\right)$$
$$\mathbb{L}_4^{(j)} := u_{[j-2]}\left(-\partial_1 + \tilde{\mathbb{E}}\right).$$

Both $\mathbb{L}_3^{(j)}$ and $\mathbb{L}_4^{(j)}$ can be expressed in function of the generators of $\mathcal{U}(\mathfrak{D}_{n-1})$, again because of expression (1). The operator $\widetilde{C}_{2j}$ can be expressed as follows:

$$\widetilde{C}_{2j} = -\mathbb{L}_3^{(j)}\mathbb{L}_4^{(j)} - (2\nu_j - 1)\mathbb{L}_4^{(j)} + 2\nu_2\mathbb{L}_3^{(j)} + (\nu_2 + \nu_j)(\nu_2 + \nu_j - 1).$$

This means that $\widetilde{C_{2j}}$ is also in $\mathcal{U}(\mathfrak{D}_{n-1})$.

Consider the first term of the operator $\widetilde{C_{ij}}$:

$$\begin{aligned}
&- u_{[j-1,i-2]}^2\left(\partial_{i-2} - \partial_{i-1}\right)\left(\partial_{j-2} - \partial_{j-1}\right)\\
&\quad = -u_{[j-1,i-2]}\left(\left(\partial_{i-2} - \partial_{i-1}\right)u_{[j-1,i-2]} - 1\right)\left(\partial_{j-2} - \partial_{j-1}\right)\\
&\quad = -u_{[j-1,i-2]}\left(\partial_{i-2} - \partial_{i-1}\right)u_{[j-1,i-2]}\left(\partial_{j-2} - \partial_{j-1}\right)\\
&\qquad + u_{[j-1,i-2]}\left(\partial_{j-2} - \partial_{j-1}\right).
\end{aligned}$$

In line 2 we used Lemma 1. Let

$$\mathbb{L}_5^{(ij)} := u_{[j-1,i-2]}\left(\partial_{i-2} - \partial_{i-1}\right)$$
$$\mathbb{L}_6^{(ij)} := u_{[j-1,i-2]}\left(\partial_{j-2} - \partial_{j-1}\right).$$

Both $\mathbb{L}_5^{(ij)}$ and $\mathbb{L}_6^{(ij)}$ can be expressed in function of the generators of $\mathcal{U}(\mathfrak{D}_{n-1})$. The operator $\widetilde{C_{ij}}$ can be expressed as follows:

$$\widetilde{C_{ij}} = -\mathbb{L}_5^{(ij)}\mathbb{L}_6^{(ij)} - (2\nu_i - 1)\mathbb{L}_6^{(ij)} + 2\nu_j\mathbb{L}_5^{(ij)} + (\nu_i + \nu_j)(\nu_i + \nu_j - 1).$$

This means that $\widetilde{C_{ij}}$ is also in $\mathcal{U}(\mathfrak{D}_{n-1})$.This proves Conjecture 1 for this differential realization.

5 Conclusions

In this paper we have considered the higher rank Racah algebra in one of its differential operator realizations, obtained in the recent paper [5]. We have shown that this realization can be embedded in the enveloping algebra of a differential operator realization of $\mathfrak{sl}_n$ verifying Conjecture 1 in this realization.

This embedding gives us a good idea of what should be the abstract embedding of $\mathcal{R}_n$ in $U(\mathfrak{sl}_{n-1})$. Achieving the abstract construct remains, however, nontrivial. The differential embedding simplifies the problem since, e.g. all central elements $\widetilde{C}_i$ become scalars and it is not immediate therefore how to lift the differential case to the abstract one. An alternative construction might be better suited for that purpose and we believe that the route taken in [2] for the case of the Heisenberg algebra is promising in this respect.

Acknowledgments The work of HDB is supported by the Research Foundation Flanders (FWO) under Grant EOS 30889451. HDB and WVDV are grateful for the hospitality extended to them by the Centre de Recherches Mathématiques in Montréal, where part of this research was carried out. The research of LV is funded in part by a discovery grant of the Natural Sciences and Engineering Council (NSERC) of Canada.

References

1. N. Crampé, L. Poulain d'Andecy, L. Vinet, Temperley-Lieb, Brauer and Racah algebras and other centralizers of $su(2)$. Trans. Amer. Math. Soc. **373**(7), 4907–4932 (2020)
2. N. Crampé, W. van de Vijver, L. Vinet, Racah problems for the oscillator algebra, the Lie algebra $\mathfrak{sl}_n$ and multivariate Krawtchouk polynomials (2019). arXiv:1909.12643, 28 pages
3. H. De Bie, W. van de Vijver, A discrete realization of the higher rank Racah algebra. Constr. Approx. **52**, 1–29 (2020). https://doi.org/10.1007/s00365-019-09475-0
4. H. De Bie, V.X. Genest, W. van de Vijver, L. Vinet, A higher rank Racah algebra and the $(\mathbb{Z}_2)^n$ Laplace-Dunkl operator. J. Phys. A Math. Theor. **51**, 025203 (20 pp.) (2018)
5. H. De Bie, P. Iliev, L. Vinet, Bargmann and Barut-Giradello models for the Racah algebra. J. Math. Phys. **60**, 011701 (2019)
6. J. Gaboriaud, L. Vinet, S. Vinet, A. Zhedanov, The generalized Racah algebra as a commutant. J. Phys. Conf. Ser. **1194**(1), 012034 (2019)
7. V.X. Genest, L. Vinet, A. Zhedanov, The equitable Racah algebra from three $su(1, 1)$ algebras. J. Phys. A Math. Theor. **47**(2), 025203, 12 pp. (2014)
8. V.X. Genest, L. Vinet, A. Zhedanov, The Racah algebra and superintegrable models. J. Phys. Conf. Ser. **512**, 012011 (2014)
9. P. Iliev, The generic quantum superintegrable system on the sphere and Racah operators. Lett. Math. Phys. **107**(11), 2029–2045 (2017)
10. P. Iliev, Symmetry algebra for the generic superintegrable system on the sphere. J. High Energy Phys. **2018**(2), 44 (2018), front matter+22 pp.
11. R. Koekoek, P.A. Lesky, R.F. Swarttouw, *Hypergeometric Orthogonal Polynomials and Their q-Analogues* (Springer, New York, 2010)
12. M.V. Tratnik, Some multivariable orthogonal polynomials of the Askey tableau-discrete families. J. Math. Phys. **32**, 2337–2342 (1991)
13. A.S. Zhedanov, "Hidden symmetry" of the Askey-Wilson polynomials. Theor. Math. Phys. **89**, 1146–1157 (1991)

On Reducible Verma Modules over Jacobi Algebra

V. K. Dobrev

Abstract With this paper we start the study of reducible representations of the Jacobi algebra with the ultimate goal of constructing differential operators invariant w.r.t. the Jacobi algebra. In this first paper we show examples of the low level singular vectors of Verma modules over the Jacobi algebra. According to our methodology these will produce the invariant differential operators.

Keywords Jacobi algebra · Verma modules · Singular vectors

1 Introduction

The role of nonrelativistic symmetries in theoretical physics was always important. Currently one of the most popular fields in theoretical physics—string theory, pretending to be a universal theory—encompasses together relativistic quantum field theory, classical gravity, and certainly, nonrelativistic quantum mechanics, in such a way that it is not even necessary to separate these components.

Since the cornerstone of quantum mechanics is the Schrödinger equation then it is not a surprise that the Schrödinger group—the group that is the maximal group of symmetry of the Schrödinger equation—was the first to play a prominent role in theoretical physics. The latter is natural since originally the Schrödinger group, actually the Schrödinger algebra, was introduced in [1, 2] as a nonrelativistic limit of the vector-field realization of the conformal algebra. For a review on these developments we refer to [3].

Another interesting nonrelativistic example is the Jacobi algebra [4, 5] which is the semi-direct sum of the Heisenberg algebra and the $sp(n)$ algebra. Actually the lowest case of the Jacobi algebra coincides with the lowest case of the Schrödinger algebra which makes it interesting to apply to the Jacobi algebra the methods we

V. K. Dobrev (✉)
Institute of Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, Sofia, Bulgaria

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_20

applied to the Schrödinger algebra. This is a project we start in the present short paper. Actually here we give as examples the low level singular vectors of Verma modules over the Jacobi algebra.

2 Preliminaries

The Jacobi algebra is the semi-direct sum $\mathcal{G}_n := \mathcal{H}_n \ni sp(n,\mathbb{R})_{\mathbb{C}}$ [4, 5]. The Heisenberg algebra $\mathcal{H}_n$ is generated by the boson creation (respectively, annihilation) operators a_i^+ (a_i^-), $i, j = 1, \dots, n$, which verify the canonical commutation relations

$$\left[a_i^-, a_j^+\right] = \delta_{ij}, \qquad [a_i^-, a_j^-] = \left[a_i^+, a_j^+\right] = 0. \tag{1}$$

$\mathcal{H}_n$ is an ideal in $\mathcal{G}_n$, i.e., $[\mathcal{H}_n, \mathcal{G}_n] = \mathcal{H}_n$, determined by the commutation relations (following the notation of [6]):

$$\left[a_k^+, K_{ij}^+\right] = [a_k^-, K_{ij}^-] = 0, \tag{2a}$$

$$[a_i^-, K_{kj}^+] = \tfrac{1}{2}\delta_{ik}a_j^+ + \tfrac{1}{2}\delta_{ij}a_k^+, \qquad \left[K_{kj}^-, a_i^+\right] = \tfrac{1}{2}\delta_{ik}a_j^- + \tfrac{1}{2}\delta_{ij}a_k^-, \tag{2b}$$

$$\left[K_{ij}^0, a_k^+\right] = \tfrac{1}{2}\delta_{jk}a_i^+, \qquad \left[a_k^-, K_{ij}^0\right] = \tfrac{1}{2}\delta_{ik}a_j^-. \tag{2c}$$

$K_{ij}^{\pm,0}$ are the generators of the $\mathcal{S}_n \equiv sp(n,\mathbb{R})_{\mathbb{C}}$ algebra:

$$[K_{ij}^-, K_{kl}^-] = [K_{ij}^+, K_{kl}^+] = 0, \qquad 2\left[K_{ij}^-, K_{kl}^0\right] = K_{il}^-\delta_{kj} + K_{jl}^-\delta_{ki}, \tag{3a}$$

$$2[K_{ij}^-, K_{kl}^+] = K_{kj}^0\delta_{li} + K_{lj}^0\delta_{ki} + K_{ki}^0\delta_{lj} + K_{li}^0\delta_{kj} \tag{3b}$$

$$2\left[K_{ij}^+, K_{kl}^0\right] = -K_{ik}^+\delta_{jl} - K_{jk}^+\delta_{li}, \quad 2\left[K_{ji}^0, K_{kl}^0\right] = K_{jl}^0\delta_{ki} - K_{ki}^0\delta_{lj}. \tag{3c}$$

In order to implement our approach we introduce a triangular decomposition of $\mathcal{G}_n$:

$$\mathcal{G}_n = \mathcal{G}_n^+ \oplus \mathcal{K}_n \oplus \mathcal{G}_n^- , \tag{4}$$

using the triangular decomposition $\mathcal{S}_n = \mathcal{S}_n^+ \oplus \mathcal{K}_n \oplus \mathcal{S}_n^-$, where:

$$\begin{aligned}
\mathcal{G}_n^\pm &= \mathcal{H}_n^\pm \oplus \mathcal{S}_n^\pm \\
\mathcal{H}_n^\pm &= \text{l.s.}\{a_i^\pm : i = 1, \dots, n\}, \\
\mathcal{S}_n^+ &= \text{l.s.}\{K_{ij}^+ \ : \ 1 \le i \le j \le n\} \oplus \text{l.s.}\{K_{ij}^0 \ : \ 1 \le i < j \le n\} \\
\mathcal{S}_n^- &= \text{l.s.}\{K_{ij}^- \ : \ 1 \le i \le j \le n\} \oplus \text{l.s.}\{K_{ij}^0 \ : \ 1 \le j < i \le n\} \\
\mathcal{K}_n &= \text{l.s.}\{K_{ii}^0 \ : \ 1 \le i \le n\}.
\end{aligned} \tag{5}$$

Note that the subalgebra $\mathcal{K}_n$ is abelian and is a Cartan subalgebra of $\mathcal{S}_n$. Furthermore, not only $\mathcal{S}_n^{\pm}$, but also $\mathcal{G}_n^{\pm}$ are its eigenspaces:

$$[\mathcal{K}_n, \mathcal{G}_n^{\pm}] = \mathcal{G}_n^{\pm}. \tag{6}$$

Thus, $\mathcal{K}_n$ plays for $\mathcal{G}_n$ the role that Cartan subalgebras are playing for semi-simple Lie algebras.

3 Case $\mathcal{G}_2$

Note that the algebra $\mathcal{G}_1$ is isomorphic to the (1+1)-dimensional Schrödinger algebra (without central extension). The representations of the latter are well known, cf. [3, 7–9]. Thus, we study the first new case of the $\mathcal{G}_n$ series, namely, $\mathcal{G}_2$.

For simplicity, we introduce the following notations for the basis of $\mathcal{S}_2$:

$$\mathcal{S}^+ : \quad b_i^+ \equiv K_{ii}^+, \ i = 1, 2; \quad c^+ \equiv K_{12}^+, \quad d^+ \equiv K_{12}^0 \tag{7a}$$

$$\mathcal{S}^- : \quad b_i^- \equiv K_{ii}^-, \ i = 1, 2; \quad c^- \equiv K_{12}^-, \quad d^- \equiv K_{21}^0 \tag{7b}$$

$$\mathcal{K} : \quad h_i \equiv K_{ii}^0, \ i = 1, 2. \tag{7c}$$

Next, using (2) and (3) we give the eigenvalues of the basis of $\mathcal{G}^+$ w.r.t. $\mathcal{K}$:

$$\begin{aligned} h_1 &: (b_1^+, b_2^+, c^+, d^+, a_1^+, a_2^+) : (1, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0), \\ h_2 &: (b_1^+, b_2^+, c^+, d^+, a_1^+, a_2^+) : (0, 1, \tfrac{1}{2}, -\tfrac{1}{2}, 0, \tfrac{1}{2}), \end{aligned} \tag{8}$$

(e.g., $[h_1, b_1^+] = b_1^+$, $[h_2, d^+] = -\frac{1}{2}d^+$, etc). Naturally, the eigenvalues of the basis of $\mathcal{G}^-$ w.r.t. $\mathcal{K}$ are obtained from (8) by multiplying every eigenvalue by (-1).

Next we introduce the following grading of the basis of $\mathcal{G}_2^+$:

$$(b_1^+, b_2^+, c^+, d^+, a_1^+, a_2^+) : (2\delta_1, 2\delta_2, \delta_1 + \delta_2, \delta_1 - \delta_2, \delta_1, \delta_2). \tag{9}$$

The grading of the $\mathcal{S}_2^+$ part of the basis follows from the root system of $\mathcal{S}_2^+$, while the grading of the $\mathcal{H}_2^+$ part of the basis is determined by consistency with commutation relations (2). It is consistent also with formulae (8).

Naturally, the grading of the basis of $\mathcal{G}^-$ w.r.t. are obtained from (9) by multiplying every grading by (-1).

4 Verma Modules and Singular Vectors

4.1 Definitions

We shall introduce Verma modules over the Jacobi algebra analogously to the case of semi-simple algebras. Thus, we define a lowest weight *Verma module* V^{Λ} over $\mathcal{G}_n$ as the lowest weight module over $\mathcal{G}_n$ with lowest weight $\Lambda \in \mathcal{K}_n^*$ and lowest weight vector $v_0 \in V^{\Lambda}$, induced from the one-dimensional representation $V_0 \cong \mathbb{C}v_0$ of $U(\mathcal{B}_n)$, (where $\mathcal{B}_n = \mathcal{K}_n \oplus \mathcal{G}_n^-$ is a Borel subalgebra of $\mathcal{G}_n$), such that:

$$\begin{aligned} X\, v_0 &= 0, \quad \forall\, X \in \mathcal{G}_n^- \\ H\, v_0 &= \Lambda(H)\, v_0\,, \quad \forall\, H \in \mathcal{K}_n. \end{aligned} \tag{10}$$

Pursuing the analogy with the semi-simple case and following our approach we are interested in the cases when the Verma modules are reducible. Namely, we are interested in the cases when a Verma module V^{Λ} contains an invariant submodule which is also a Verma module $V^{\Lambda'}$, where $\Lambda' \neq \Lambda$, and holds the analog of

$$X\, v_0' = 0, \quad \forall\, X \in \mathcal{G}_n^- \tag{11a}$$

$$H\, v_0' = \Lambda'(H)\, v_0', \quad \forall\, H \in \mathcal{K}_n. \tag{11b}$$

Since $V^{\Lambda'}$ is an invariant submodule then there should be a mapping such that v_0' is mapped to a singular vector $v_s \in V^{\Lambda}$ fulfilling exactly (11). Thus, as in the semi-simple case there should be a polynomial $\mathcal{P}$ of $\mathcal{G}_n^+$ elements which is eigenvector of $\mathcal{K}_n$: $[H, \mathcal{P}] = \Lambda'(H)\mathcal{P}$, $(\forall H \in \mathcal{K}_n)$, and then we would have: $v_s = \mathcal{P}v_0$.

4.2 Case $\mathcal{G}_2$

We shall consider several examples of reducible Verma modules with different weights.

Weight $2\delta_1$

As first example we try to find a singular vector of weight $\Lambda' \sim 2\delta_1$. There are six possible terms in $U(\mathcal{G}_2)$ with this weight, thus, we try:

$$\begin{aligned} v_s^{2\delta_1} = \big(&\nu_1 b_1^+ + \nu_2 c^+ d^+ + \nu_3 b_2^+ (d^+)^2 + \nu_4 (a_1^+)^2 + \nu_5 a_1^+ a_2^+ d^+ \\ &+ \nu_6 (a_2^+)^2 (d^+)^2\big) v_0, \end{aligned} \tag{12}$$

where ν_k are numerical coefficients which may be fixed when we impose (11a) on (12). (Note that (11b) is fulfilled by every term of (12).)

After we impose (11a) on (12) we find the solution:

$$\begin{aligned}
&\Lambda(H_1) = \tfrac{3}{4}\,, \quad \nu_3 \;=\; -2\nu_6,\\
&\nu_1 = -\nu_6(\Lambda(H_2) - \Lambda(H_1))(2\Lambda(H_2) - 2\Lambda(H_1) - 1),\\
&\nu_2 = 2\nu_6(2\Lambda(H_2) - 2\Lambda(H_1) - 1),\\
&\nu_4 = \nu_6(\Lambda(H_2) - \Lambda(H_1))(\Lambda(H_2) - \Lambda(H_1) - \tfrac{1}{2}),\\
&\nu_5 = -\nu_6(2\Lambda(H_2) - 2\Lambda(H_1) - 1).
\end{aligned} \tag{13}$$

Thus, the singular vector is:

$$\begin{aligned}
v_s^{2\delta_1} \;=\; \nu_6\big(\, &(\Lambda(H_2) - \tfrac{3}{4})(\Lambda(H_2) - \tfrac{5}{4})((a_1^+)^2 - 2b_1^+) \;+\\
&+ 2(\Lambda(H_2) - \tfrac{5}{4})(2c^+ - a_1^+ a_2^+)d^+ \;+\\
&+ ((a_2^+)^2 - 2b_2^+)(d^+)^2\,\big)v_0\,, \quad \Lambda(H_1) \;=\; \tfrac{3}{4}.
\end{aligned} \tag{14}$$

Weight $2\delta_2$

As next example we try to find a singular vector of weight $\Lambda' \sim 2\delta_2$. The possible singular vector is:

$$v_s^{2\delta_2} \;=\; \big(\,\mu_1 b_2^+ + \mu_2 (a_2^+)^2\,\big)v_0. \tag{15}$$

Imposing (11a) on (15) we obtain:

$$\Lambda(H_2) \;=\; \tfrac{1}{4}\,, \quad \mu_1 \;=\; -2\mu_2, \tag{16}$$

Thus, the singular vector is:

$$v_s^{2\delta_2} \;=\; \mu_2((a_2^+)^2 - 2b_2^+)v_0\,, \quad \Lambda(H_2) \;=\; \tfrac{1}{4}. \tag{17}$$

Weight $\delta_1 + \delta_2$

Next we try a singular vector of weight $\Lambda' \sim \delta_1 + \delta_2$. The possible singular vector is:

$$v_s^{\delta_1+\delta_2} \;=\; \big(\,\kappa_1 c^+ + \kappa_2 b_2^+ d^+ + \kappa_3 a_1^+ a_2^+ + \kappa_4 (a_2^+)^2 d^+\,\big)v_0. \tag{18}$$

Imposing (11a) on (18) we obtain:

$$\Lambda(H_2) \ = \ \tfrac{3}{2} - \Lambda(H_1) \, , \quad \kappa_1 = (3 - 4h(1))\kappa_4 \, ,$$
$$\kappa_2 = -2\kappa_4 \, , \quad \kappa_3 = (2h(1) - \tfrac{3}{2})\kappa_4 . \tag{19}$$

Thus, the singular vector is:

$$v_s^{\delta_1+\delta_2} \ = \ \kappa_4 \big(\, (\tfrac{3}{2} - 2h(1))(2c^+ - a_1^+ a_2^+) + ((a_2^+)^2 - 2b_2^+)d^+ \, \big) v_0 . \tag{20}$$

Weight $\delta_1 - \delta_2$

Next we try a singular vector of weight $\Lambda' \sim \delta_1 - \delta_2$. The only possible singular vector is:

$$v_s^{\delta_1-\delta_2} \ = \ \lambda d^+ v_0 . \tag{21}$$

Imposing (11a) on (21) we obtain that $v_s^{\delta_1-\delta_2}$ is a singular vector iff:

$$\Lambda(H_2) \ = \ \Lambda(H_1) . \tag{22}$$

Weight δ_1

Next we try a singular vector of weight $\Lambda' \sim \delta_1$. The possible singular vector is:

$$v_s^{\delta_1} \ = \ \big(\, \lambda_1 a_1^+ + \lambda_2 a_2^+ d^+ \, \big) v_0 . \tag{23}$$

Imposing (11a) on (23) we obtain:

$$\lambda_1 = \lambda_2 = 0 . \tag{24}$$

Thus, there is no singular vector of weight δ_1 .

Weight δ_2

Finally, we try a singular vector of weight $\Lambda' \sim \delta_2$. The only possible singular vector is:

$$v_s^{\delta_2} \ = \ \mu a_2^+ v_0 . \tag{25}$$

Imposing (11a) on (25) we obtain:

$$\mu = 0 . \tag{26}$$

Thus, there is no singular vector of weight δ_2 .

Weight $3\delta_2$

The only possible singular vector is:

$$v_s^{\delta_2} = \mu b_2^+ a_2^+ v_0 + \nu (a_2^+)^3 v_0. \tag{27}$$

Imposing (11a) on (27) we obtain:

$$\mu = \nu = 0. \tag{28}$$

Thus, there is no singular vector of weight $3\delta_2$.

Acknowledgment The author acknowledges partial support from Bulgarian NSF Grant DN-18/1.

References

1. U. Niederer, The maximal kinematical invariance group of the free Schrodinger equation. Helv. Phys. Acta **45**, 802–810 (1972). https://doi.org/10.5169/seals-114417
2. C.R. Hagen, Scale and conformal transformations in Galilean-covariant field theory. Phys. Rev. **D5**, 377–388 (1972). https://doi.org/10.1103/PhysRevD.5.377
3. V.K. Dobrev, *Invariant Differential Operators, Volume 4: AdS/CFT, (Super-)Virasoro and Affine (Super-)Algebras*. De Gruyter Studies in Mathematical Physics, vol. 53 (De Gruyter, Berlin, 2019)
4. M. Eichler, D. Zagier, *The Theory of Jacobi Forms*. Program in Mathematics, vol. 55 (Birkhäuser, Boston, 1985)
5. R. Berndt, R. Schmidt, *Elements of the Representation Theory of the Jacobi Group*. Program in Mathematics, vol. 163 (Birkhäuser, Basel, 1998)
6. S. Berceanu, A holomorphic representation of the semidirect sum of symplectic and Heisenberg Lie algebras. J. Geom. Symmetry Phys. **5**, 5–13 (2006). https://doi.org/10.7546/jgsp-5-2006-5-13
7. V.K. Dobrev, H.D. Doebner, C. Mrugalla, Lowest weight representations of the Schrödinger algebra and generalized heat equations. Rep. Math. Phys. **39**, 201–218 (1997). https://doi.org/10.1016/S0034-4877(97)88001-9
8. N. Aizawa, V.K. Dobrev, Intertwining operator realization of non-relativistic holography. Nucl. Phys. **B828**, 581–593 (2010). https://doi.org/10.1016/j.nuclphysb.2009.10.019
9. B. Dubsky, R. Lue, V. Mazorchuk, K. Zhao, Category O for the Schrödinger algebra. Linear Algebra Appl. **460**, 17–50 (2014). https://doi.org/10.1016/j.laa.2014.07.030

Howe Duality and Algebras of the Askey–Wilson Type: An Overview

Julien Gaboriaud, Luc Vinet, and Stéphane Vinet

Abstract The Askey–Wilson algebra and its relatives such as the Racah and Bannai–Ito algebras were initially introduced in connection with the eponym orthogonal polynomials. They have since proved ubiquitous. In particular they admit presentations as commutants that are related through Howe duality. This paper surveys these results.

Keywords Howe duality · Racah · Bannai–Ito and Askey–Wilson algebras · Commutants · Reductive dual pairs

1 Introduction

The quadratic algebras of Askey–Wilson type such as the Askey–Wilson algebra itself, the Racah and Bannai–Ito algebras and their specializations and contractions encode the bispectral properties of orthogonal polynomials that arise in recoupling coefficients such as the Clebsch–Gordan or Racah coefficients. It is therefore natural that these algebras be encountered in centralizers of the diagonal action of an algebra of interest $\mathfrak{g}'$ such as $\mathfrak{sl}(2)$, $\mathfrak{osp}(1|2)$, or $U_q(\mathfrak{sl}(2))$, on n-fold tensor products of representations of $\mathfrak{g}'$. Indeed, elements of these centralizers will be used as labeling operators to define bases whose overlaps will be expressed in terms of the corresponding orthogonal polynomials.

Often the algebra $\mathfrak{g}'$ forms a reductive pair with another algebra $\mathfrak{g}$ in which case the Howe duality operates in certain modules. This leads to alternative

J. Gaboriaud (✉) · L. Vinet
Centre de Recherches Mathématiques, Montréal, QC, Canada

Département de Physique, Université de Montréal, Montréal, QC, Canada
e-mail: julien.gaboriaud@umontreal.ca; vinet@crm.umontreal.ca

S. Vinet
Département de Physique, Université de Montréal, Montréal, QC, Canada
e-mail: stephane.vinet@umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_21

characterizations of the quadratic algebras that are in correspondence: on the one hand commutants in representations of the universal enveloping algebra $U(\mathfrak{g})$ and on the other hand, realizations of the type mentioned above as centralizers in recoupling problems for $\mathfrak{g}'$. This is the topic of this brief review which is organized as follows. Section 2 presents the general framework. Section 3 describes as illustration the dual commutant picture for the Racah algebra; this will involve the reductive pair $(\mathfrak{o}(6), \mathfrak{su}(1,1))$. Section 4 gives a summary of the different cases that have been analyzed and Sect. 5 provides a short outlook.

2 General Framework

We shall say following [1] that two algebras $\mathfrak{g}$ and $\mathfrak{g}'$ have dual representations on a Hilbert space $\mathcal{H}$ if (1) this space carries fully reducible representations of both $\mathfrak{g}$ and $\mathfrak{g}'$, (2) the action of $\mathfrak{g}$ and $\mathfrak{g}'$ commute, (3) the representation ρ of the direct sum $\mathfrak{g} \oplus \mathfrak{g}'$ defined by the actions of $\mathfrak{g}$ and $\mathfrak{g}'$ on $\mathcal{H}$ is multiplicity-free, and (4) each irreducible representation of $\mathfrak{g}$ occurring in the decomposition of ρ is paired with a unique irreducible representation of $\mathfrak{g}'$ and vice versa. This is the essence of Howe duality which can be proved in a number of situations. We shall consider such instances in this paper.

Consider now a setup with the representation of $\mathfrak{g}'$ in $\mathcal{H} = V^{\otimes 2n}$ given by $\bar{\sigma}^{\otimes 2n}[\Delta^{(2n-1)}(\mathfrak{g}')]$ where $\bar{\sigma} : \mathfrak{g}' \to End\ V$ is a representation of $\mathfrak{g}'$ on the vector space V, $\Delta : \mathfrak{g}' \to \mathfrak{g}' \otimes \mathfrak{g}'$ is the coproduct, and $\Delta^{(n)}$ is defined recursively by $\Delta^{(n)} = (\Delta \otimes 1^{\otimes(n-1)}) \circ \Delta^{(n-1)}$, with $\Delta^{(0)} = 1$. This symmetric situation makes it natural that there be an action of some other algebra $\mathfrak{g}$ on the carrier space $\mathcal{H}$ that commutes with the action of $\mathfrak{g}'$. Take the maximal Abelian subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ to be $\mathfrak{h} \simeq \mathfrak{X}^{\oplus n}$ with $\mathfrak{X}$ one-dimensional. The pairing under Howe duality with the representations of $\mathfrak{X}^{\oplus n}$ implies that $\bar{\sigma}^{\otimes 2n}[\Delta^{(2n-1)}(\mathfrak{g}')] = \bar{\sigma}^{\otimes 2n}[\Delta^{\otimes n} \circ \Delta^{(n-1)}(\mathfrak{g}')]$ decomposes into representations of the form $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n(\Delta^{(n-1)}(\mathfrak{g}'))$ with the σ_i's being irreducible representations arising in the decomposition of $\bar{\sigma}^{\otimes 2}$. This quotienting by $\mathfrak{h}$ is a way of posing a generalized Racah problem for the recoupling of the n representations σ_i of $\mathfrak{g}'$.

We indicated in the introduction that the quadratic algebras $\mathcal{A}$ of Askey–Wilson type can be obtained as (subalgebras of) centralizers of diagonal actions in n-fold tensor products of representations. The intermediate Casimir elements in $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n$ manifestly centralize the action of $\mathfrak{g}'$ on $\mathcal{H}$ mod $\mathfrak{h}$. They are taken to generate the quadratic algebra of interest. This provides the first presentation of $\mathcal{A}$ as a commutant. The dual one is identified as follows in the present context. We know that $\mathfrak{g}$ is the commutant of $\mathfrak{g}'$ in $\mathcal{H}$. Moreover from the application of Howe duality, the generators of the representation $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n$ of $\mathfrak{g}'$ are known to commute with those that represent the subalgebra $\mathfrak{h} \simeq \mathfrak{X}^{\oplus n}$. The non-trivial part of the centralizer of $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n$ must therefore be obtained, in the

given representation on $\mathcal{H}$ mod $\mathfrak{h}$, by those elements in the universal enveloping algebra of $\mathfrak{g}$ that commute with $\mathfrak{X}^{\oplus n}$. In other words, $\mathcal{A}$ can also be identified as the commutant of $\mathfrak{h} \subset \mathfrak{g}$ in $\mathcal{U}(\mathfrak{g})$ as represented on $\mathcal{H}$.

There is an equivalent way of looking at this. The pairing of the representations of $\mathfrak{g}$ and $\mathfrak{g}'$ through Howe duality manifests itself in the fact that the Casimir elements of $\mathfrak{g}$ and $\mathfrak{g}'$ are affinely related. Let $\mathcal{C}$ be a Casimir element of $\mathfrak{g}'$. Consider for example the intermediate Casimir element given by $\bar{\sigma}^{\otimes 4}[((\Delta \otimes \Delta) \circ \Delta)(\mathcal{C})] \otimes 1^{\otimes(2n-4)}$ corresponding to the embedding of $\mathfrak{g}'$ in the first four factors of $\mathfrak{g}'^{\otimes 2n}$. There will be a subalgebra $\mathfrak{g}_1$ of $\mathfrak{g}$ that will be dually related to $\mathfrak{g}'$ on the restriction of $\mathcal{H}$ to $V^{\otimes 4}$ so that its Casimir element will be essentially the one of $\mathfrak{g}'$. Next, looking at the intermediate Casimir element of $\mathfrak{g}'$ associated with a different embedding, for instance in the four last factors of $\mathfrak{g}'^{\otimes 2n}$, there will be a dual pairing with a different embedding in $\mathfrak{g}$ of the same subalgebra $\mathfrak{g}_1$ and again the two Casimir elements will basically coincide. These observations lead to the conclusion that the set of intermediate Casimir elements associated with the representation of $\mathfrak{g}'$ is algebraically identical to the set of Casimir elements of the subalgebras of $\mathfrak{g}$ that form dual pairs with $\mathfrak{g}'$ when intermediate representations of the latter are taken. It is not difficult to convince oneself that the set of invariants connected to the relevant subalgebras of $\mathfrak{g}$ consists in the commutant of the maximal Abelian subalgebra of $\mathfrak{g}$ as concluded differently before.

To summarize, in situations where Howe duality prevails with $(\mathfrak{g}, \mathfrak{g}')$ the pair of algebras that are dually represented on $\mathcal{H}$ and if the representation of $\mathfrak{g}'$ is of the form $\bar{\sigma}^{\otimes 2n}[\Delta^{(2n-1)}(\mathfrak{g}')]$, the quadratic algebras $\mathcal{A}$ of Askey–Wilson type can be viewed on one hand as the commutant of this action of $\mathfrak{g}'$ on $\mathcal{H}$ and thus generated by the intermediate Casimir elements of $\mathfrak{g}'$, or on the other hand as the commutant of $\mathfrak{h} \subset \mathfrak{g}$ in the intervening representation of $\mathcal{U}(\mathfrak{g})$. We shall present next an example of how this can be concretely realized.

3 The Dual Presentations of the Racah Algebra

The Racah algebra $\mathcal{R}$ has three generators K_1, K_2, K_3 that are subjected to the relations [2]:

$$\begin{aligned} [K_1, K_2] &= K_3, \qquad & [K_2, K_3] &= {K_2}^2 + \{K_1, K_2\} + dK_2 + e_1, \\ & & [K_3, K_1] &= {K_1}^2 + \{K_1, K_2\} + dK_1 + e_2, \end{aligned} \tag{1}$$

where $[A, B] = AB - BA$, $\{A, B\} = AB + BA$ and d, e_1, e_2 are central.

We shall explain how dual presentations of the algebra $\mathcal{R}$ as a commutant are obtained in the fashion described in Sect. 2. The dual pair will be $(\mathfrak{o}(6), \mathfrak{su}(1,1))$ and the representation space $\mathcal{H}$ will be that of the state space of six quantum harmonic oscillators with annihilation and creation operators $a_\mu, a_\nu^\dagger$, $\mu, \nu = 1, \ldots, 6$ verifying $[a_\mu, a_\nu^\dagger] = \delta_{\mu\nu}$. The corresponding Hamiltonian $H = a_1^\dagger a_1 + \cdots + a_6^\dagger a_6$

is manifestly invariant under the rotations in six dimensions. These are encoded in the Lie algebra $\mathfrak{o}(6)$, realized by the generators $L_{\mu\nu} = a_\mu^\dagger a_\nu - a_\mu a_\nu^\dagger$ and possessing the Casimir element $\mathcal{C} = \sum_{\mu<\nu} {L_{\mu\nu}}^2$.

The Lie algebra $\mathfrak{su}(1,1)$ has generators J_0, $J_\pm$ that obey the following commutation relations: $[J_0, J_\pm] = \pm J_\pm$, $[J_+, J_-] = -2J_0$, and its Casimir operator is given by $C = {J_0}^2 - J_+J_- - J_0$. The six harmonic oscillators also provide a realization of this algebra through the addition of six copies of the metaplectic representation of $\mathfrak{su}(1,1)$, for which the generators are mapped to: $J_0^{(\mu)} = \frac{1}{2}(a_\mu^\dagger a_\mu + \frac{1}{2})$, $J_+^{(\mu)} = \frac{1}{2}(a_\mu^\dagger)^2$, $J_-^{(\mu)} = \frac{1}{2}(a_\mu)^2$, $\mu = 1, \ldots, 6$. Note that the operators $\sum_{\mu=1}^{6} J_\bullet^{(\mu)}$ are invariant under rotations. The space of state vectors $\mathcal{H}$ thus carries commuting representations of $\mathfrak{o}(6)$ and $\mathfrak{su}(1,1)$ and Howe duality takes place.

The maximal Abelian algebra of $\mathfrak{o}(6)$ is $\mathfrak{o}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{o}(2)$ and is generated by the set $\{L_{12}, L_{34}, L_{56}\}$. The non-abelian part of its commutant in the representation of $\mathcal{U}(\mathfrak{o}(6))$ on $\mathcal{H}$ is generated by the two invariants

$$K_1 = \frac{1}{8}\big({L_{12}}^2 + {L_{34}}^2 + {L_{13}}^2 + {L_{23}}^2 + {L_{14}}^2 + {L_{24}}^2\big), \tag{2}$$

$$K_2 = \frac{1}{8}\big({L_{34}}^2 + {L_{56}}^2 + {L_{35}}^2 + {L_{36}}^2 + {L_{45}}^2 + {L_{46}}^2\big). \tag{3}$$

Define K_3 by $[K_1, K_2] = K_3$. Working out the commutation relations of K_3 with K_1 and K_2, it is found that they correspond to those (1) of the Racah algebra with the central parameters given by $d = -\frac{1}{8}\big(\mathcal{C} + {L_{12}}^2 + {L_{34}}^2 + {L_{56}}^2\big)$, $e_1 = -\frac{1}{64}\big(\mathcal{C} - {L_{12}}^2 - 4\big)\big({L_{34}}^2 - {L_{56}}^2\big)$, and $e_2 = -\frac{1}{64}\big(\mathcal{C} - {L_{56}}^2 - 4\big)\big(L_{34}^2 - L_{12}^2\big)$. For details see [3]. By abuse of notation we designate the abstract generators and their realizations by the same letter.

Regarding the $\mathfrak{su}(1,1)$ picture, let $J_\bullet^{(\mu,\nu,\rho,\lambda)} = J_\bullet^{(\mu)} + J_\bullet^{(\nu)} + J_\bullet^{(\rho)} + J_\bullet^{(\lambda)}$ denote the addition of the four metaplectic representations labeled by the variables μ, ν, ρ, λ all assumed different. The corresponding Casimir operator is $C^{(\mu,\nu,\rho,\lambda)} = (J_0^{(\mu,\nu,\rho,\lambda)})^2 - J_+^{(\mu,\nu,\rho,\lambda)} J_-^{(\mu,\nu,\rho,\lambda)} - J_0^{(\mu,\nu,\rho,\lambda)}$. Quite clearly, these actions of $\mathfrak{su}(1,1)$ restricted to state vectors of four oscillators are paired with commuting actions of the Lie algebra $\mathfrak{o}(4)$ of rotations in the four dimensions labeled by μ, ν, ρ, λ. It is hence not surprising to find, owing to Howe duality, that $C^{(1234)} = -2K_1$ and $C^{(3456)} = -2K_2$, namely that the intermediate $\mathfrak{su}(1,1)$ Casimir operators corresponding to the recouplings of the first four and last four of the six metaplectic representations are equal (up to a factor) to the Casimir elements of the two corresponding $\mathfrak{o}(4)$ subalgebras of $\mathfrak{o}(6)$ which together generate as we observed the non-trivial part of the commutant of $\mathfrak{o}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{o}(2)$ in $\mathcal{U}(\mathfrak{o}(6))$. This entails the description of the Racah algebra as the commutant in $\mathcal{U}(\mathfrak{su}(1,1)^{\otimes 3})$ of the action of $\mathfrak{su}(1,1)$ on $\mathcal{H}$. Alternatively, picking the $\mathfrak{su}(1,1)$ representations associated with those of $\mathfrak{o}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{o}(2)$ under Howe duality yields the sum of three irreducible representations of $\mathfrak{su}(1,1)$ belonging to the discrete series; these are realized as dynamical algebras of three singular oscillators.

Note that corresponding to the $\mathfrak{su}(1,1)$ representation $J_\bullet^{(\mu,\nu)} = J_\bullet^{(\mu)} + J_\bullet^{(\nu)}$ is the Casimir $C^{(\mu\nu)} = -\frac{1}{4}\left(L_{\mu\nu}{}^2 + 1\right)$. With the dependence on the polar angles "rotated out", the total Casimir element $C^{(123456)}$ becomes the Hamiltonian of the generic superintegrable system on the two-sphere; the constants of motion are the quotiented intermediate Casimir elements and the symmetry algebra that they generate is hence that of Racah.

4 More Dual Pictures: An Overview

The main algebras of Askey–Wilson type have been studied recently from the commutant and Howe duality viewpoints. We summarize in the following the main results and give in particular the dualities that are involved.

4.1 The Racah Family

The higher rank extension of the Racah algebra [4] defined as the algebra generated by all the intermediate Casimir elements of $\sigma_1 \otimes \sigma_2 \otimes \cdots \otimes \sigma_n(\Delta^{(n-1)}(\mathfrak{su}(1,1)))$ can be described in the framework of the preceding section with the help of the dual pair $(\mathfrak{o}(2n), \mathfrak{su}(1,1))$ using in this case the module formed by the state vectors of $2n$ harmonic oscillators. It is then seen to be dually the commutant of $\mathfrak{o}(2)^{\oplus n}$ in the oscillator representation of $\mathcal{U}(\mathfrak{o}(2n))$ [5].

The case $n = 2$ is special and of particular interest since it pertains to the Clebsch–Gordan problem for $\mathfrak{su}(1,1)$, that is, the recoupling of the two irreducible representations σ_1 and σ_2. There are no intermediate Casimirs here; the relevant operators associated with the direct product basis and the recoupled one are respectively $M_1 = \sigma_1(J_0) - \sigma_2(J_0)$ and the total Casimir $M_2 = (\sigma_1 \otimes \sigma_2)\Delta(C)$. These are seen to obey the commutation relations of the Hahn algebra [6]:

$$\begin{aligned} &[M_1, M_2] = M_3, \qquad [M_2, M_3] = -2\{M_1, M_2\} + \delta_1, \\ &\qquad\qquad\qquad\quad [M_3, M_1] = -2{M_1}^2 - 4M_2 + \delta_2, \end{aligned} \tag{4}$$

where $\delta_1 = 4(\sigma_1(J_0) + \sigma_2(J_0))(\sigma_1(C) - \sigma_2(C))$ and $\delta_2 = 2(\sigma_1(J_0) + \sigma_2(J_0))^2 + (\sigma_1(C) + \sigma_2(C))$ are central. The name of the algebra comes from the fact that the $3j$-coefficients involve dual Hahn polynomials. In the setup with four harmonic oscillators, with $\mathcal{H}$ carrying the product of four metaplectic representations, Howe duality will imply that the total Casimir element $C^{(1234)}$ of $\mathfrak{su}(1,1)$ coincides with the Casimir of $\mathfrak{o}(4)$—this is the same computation as the one described above. It is easily seen that $\sigma_1(J_0) - \sigma_2(J_0)$ is derived from $\frac{1}{2}(N_1 + N_2 - N_3 - N_4)$ under the quotient by $\mathfrak{o}(2) \oplus \mathfrak{o}(2)$ with $N_i = a_i^\dagger a_i$, $i = 1, \ldots, 4$. It can in fact

be checked directly, again abusing notation, that $M_1 = \frac{1}{2}(N_1 + N_2 - N_3 - N_4)$ and $M_2 = -\frac{1}{4}\big({L_{12}}^2 + {L_{34}}^2 + {L_{13}}^2 + {L_{23}}^2 + {L_{14}}^2 + {L_{24}}^2\big)$ satisfy the relations given in Eq. (4) with $\delta_1 = -\frac{1}{2}(N_1 + N_2 + N_3 + N_4 + 2)({L_{12}}^2 - {L_{34}}^2)$ and $\delta_2 = \frac{1}{2}(N_1 + N_2 + N_3 + N_4 + 2)^2 - ({L_{12}}^2 + {L_{34}}^2 + 2)$, in correspondence with the preceding expressions for δ_1 and δ_2 in the realization $J_\bullet^{(1234)}$ of $\mathfrak{su}(1,1)$. From the expressions of these last M_1 and M_2, we can claim that the Hahn algebra is the commutant of $\mathfrak{o}(2) \oplus \mathfrak{o}(2)$ in $\mathcal{U}(\mathfrak{u}(4))$ represented on $\mathcal{H}$. Let us stress that it is the universal enveloping algebra of $\mathfrak{u}(4)$ that intervenes here.

4.2 The Bannai–Ito Ensemble

The Bannai–Ito algebra [7] takes its name after the Bannai–Ito polynomials that enter in the Racah coefficients of the Lie superalgebra $\mathfrak{osp}(1|2)$. This algebra has three generators $K_i,\ i = 1, \ldots, 3$ that satisfy the relations

$$\{K_i, K_j\} = K_k + \omega_k, \qquad i \neq j \neq k \in \{1, 2, 3\} \tag{5}$$

with ω_i central and $\{X, Y\} = XY + YX$. The relevant reductive pair in this case is $(\mathfrak{o}(6), \mathfrak{osp}(1|2))$ and the representation space $\mathcal{H}$ is that of Dirac spinors in six dimensions with the Clifford algebra generated by the elements γ_μ verifying $\{\gamma_\mu, \gamma_\nu\} = -2\delta_{\mu\nu},\ \mu, \nu = 1, \ldots, 6$. That the pair $(\mathfrak{o}(6), \mathfrak{osp}(1|2))$ is dually represented on $\mathcal{H}$ is seen as follows: The spinorial representation of $\mathfrak{o}(6)$ with generators

$$J_{\mu\nu} = -iL_{\mu\nu} + \Sigma_{\mu\nu}, \qquad L_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu, \qquad \Sigma_{\mu\nu} = \frac{i}{2}\gamma_\mu\gamma_\nu \tag{6}$$

leaves invariant the following operators:

$$J_- = -i \sum_{1 \leq \mu \leq 6} \gamma_\mu \partial_\mu, \qquad J_+ = -i \sum_{1 \leq \mu \leq 6} \gamma_\mu x_\mu, \qquad J_0 = \sum_{1 \leq \mu \leq 6} x_\mu \partial_\mu, \tag{7}$$

which in turn realize the commutation relations of the Lie superalgebra $\mathfrak{osp}(1|2)$: $[J_0, J_\pm] = \pm J_\pm,\ \{J_+, J_-\} = -2J_0$ with J_0 even and $J_\pm$ odd. Howe duality thus takes place. As a matter of fact, for any subset $A \subset \{1, \ldots, 6\}$ of cardinality $|A|$ the operators $J_-^A = -i \sum_{\mu \in A} \gamma_\mu \partial_\mu,\ J_+^A = -i \sum_{\mu \in A} \gamma_\mu x_\mu$, and $J_0^A = \frac{|A|}{2} + \sum_{\mu \in A} x_\mu \partial_\mu$ realize $\mathfrak{osp}(1|2)$. The Casimir element of $\mathfrak{osp}(1|2)$ is given by $C = \frac{1}{2}([J_-, J_+] - 1)S$ with S the grade involution obeying $S^2 = 1$, $[S, J_0] = 0$, $\{S, J_\pm\} = 0$. In the realizations at hand, $S^A = i^{|A|/2} \prod_{\mu \in A} \gamma_\mu$ with $|A|$ even.

It can be checked that the operators

$$K_1 = M_1 + \tfrac{3}{2}\Sigma_{12}\Sigma_{34}, \qquad K_2 = M_2 + \tfrac{3}{2}\Sigma_{34}\Sigma_{56}, \qquad K_3 = M_3 + \tfrac{3}{2}\Sigma_{12}\Sigma_{56},$$

$$M_1 = (L_{12}\gamma_1\gamma_2 + L_{13}\gamma_1\gamma_3 + L_{14}\gamma_1\gamma_4 + L_{23}\gamma_2\gamma_3 + L_{24}\gamma_2\gamma_4 + L_{34}\gamma_3\gamma_4)\Sigma_{12}\Sigma_{34},$$

$$M_2 = (L_{34}\gamma_3\gamma_4 + L_{35}\gamma_3\gamma_5 + L_{36}\gamma_3\gamma_6 + L_{45}\gamma_4\gamma_5 + L_{46}\gamma_4\gamma_6 + L_{56}\gamma_5\gamma_6)\Sigma_{34}\Sigma_{56},$$

$$M_3 = (L_{12}\gamma_1\gamma_2 + L_{15}\gamma_1\gamma_5 + L_{16}\gamma_1\gamma_6 + L_{25}\gamma_2\gamma_5 + L_{26}\gamma_2\gamma_6 + L_{56}\gamma_5\gamma_6)\Sigma_{12}\Sigma_{56}$$

realize the relations (5) of the Bannai-Ito algebra upon taking the following: $\omega_{ij} = 2\Gamma_k\Gamma_{123} + 2\Gamma_i\Gamma_j$, where $\Gamma_1 = J_{12}$, $\Gamma_2 = J_{34}$, $\Gamma_3 = J_{56}$, and $\Gamma_{123} = \left(\frac{5}{2} - i\sum_{1\le\mu<\nu\le 6} L_{\mu\nu}\Sigma_{\mu\nu}\right)\Sigma_{12}\Sigma_{34}\Sigma_{56}$. That these arise from dual pictures is explained as follows (see [8] for details). On the one hand, K_1, K_2, K_3 are observed to belong to the commutant in $\mathcal{U}(\mathfrak{o}(6))$ of the $\mathfrak{o}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{o}(2)$ subalgebra of $\mathfrak{o}(6)$ spanned by $\{J_{12}, J_{34}, J_{56}\}$. On the other hand, considering the Casimir elements C^A of $\mathfrak{osp}(1|2)$ associated with the realization by the operators $\{J_0^A, J_\pm^A, S^A\}$, we find that $C^{(1234)} = K_1$, $C^{(3456)} = K_2$, and $C^{(1256)} = K_3$. This confirms that the Bannai–Ito algebra can be dually presented either as the commutant of $\mathfrak{o}(2) \oplus \mathfrak{o}(2) \oplus \mathfrak{o}(2)$ in the spinorial representation of $\mathcal{U}(\mathfrak{o}(6))$ or as the centralizer of the action of $\mathfrak{osp}(1|2)$ on $\mathcal{H}$. These considerations can be extended to higher dimensions [8] so as to obtain analogously dual commutant pictures for the Bannai–Ito algebras of higher ranks [9].

4.3 The Askey–Wilson Class

The Askey–Wilson algebra can be presented as follows:

$$\frac{[K_A, K_B]_q}{q^2 - q^{-2}} + K_C = \frac{\gamma}{q + q^{-1}}, \qquad \begin{aligned} \frac{[K_B, K_C]_q}{q^2 - q^{-2}} + K_A &= \frac{\alpha}{q + q^{-1}}, \\ \frac{[K_C, K_A]_q}{q^2 - q^{-2}} + K_B &= \frac{\beta}{q + q^{-1}}, \end{aligned} \tag{8}$$

with $[A, B]_q = qAB - q^{-1}BA$ and α, β, γ central.

The $U_q(\mathfrak{su}(1,1))$ algebra has three generators, $J_\pm$ and J_0, obeying $[J_0, J_\pm] = \pm J_\pm$ and $J_-J_+ - q^2 J_+ J_- = q^{2J_0}[2J_0]_q$ with $[x]_q = \frac{q^x - q^{-x}}{q - q^{-1}}$. Its coproduct is defined by $\Delta(J_0) = J_0 \otimes 1 + 1 \otimes J_0$, $\Delta(J_\pm) = J_\pm \otimes q^{2J_0} + 1 \otimes J_\pm$. The Casimir operator C of $U_q(\mathfrak{su}(1,1))$ is given by $C = J_+J_-q^{-2J_0+1} - \frac{q}{(1-q^2)^2}\left(q^{2J_0-1} + q^{-2J_0+1}\right) + \frac{1+q^2}{(1-q^2)^2}$.

The q-deformation $\mathfrak{o}_{q^{1/2}}(N)$ of $\mathfrak{o}(N)$ is defined as the algebra with generators $L_{i,i+1}$ $(i = 1, \ldots, N-1)$ obeying the relations

$$L_{i-1,i}\,L_{i,i+1}^2 - (q^{1/2}+q^{-1/2})L_{i,i+1}\,L_{i-1,i}\,L_{i,i+1} + L_{i,i+1}^2\,L_{i-1,i} = -L_{i-1,i},$$
$$L_{i,i+1}\,L_{i-1,i}^2 - (q^{1/2}+q^{-1/2})L_{i-1,i}\,L_{i,i+1}\,L_{i-1,i} + L_{i-1,i}^2\,L_{i,i+1} = -L_{i,i+1},$$
$$[L_{i,i+1}, L_{j,j+1}] = 0 \quad \text{for} \quad |i-j| > 1.$$

We shall use the notation $L_{ik}^{\pm} = [L_{ij}^{\pm}, L_{jk}^{\pm}]_{q^{\pm 1/4}}$ for any $i < j < k$, and by definition $L_{i,i+1}^{\pm} = L_{i,i+1}$.

The reductive pair $(\mathfrak{o}_{q^{1/2}}(6), U_q(\mathfrak{su}(1,1))$ is the one which is of relevance for the Askey–Wilson algebra. Let us indicate how $\mathfrak{o}_{q^{1/2}}(2n)$ and $U_q(\mathfrak{su}(1,1))$ are dually represented on the standard state space $\mathcal{H}$ of $2n$ independent q-oscillators described by operators $\{A_i^{\pm}, A_i^0\}$ such that $[A_i^0, A_i^{\pm}] = \pm A_i^{\pm}$, $[A_i^-, A_i^+] = q^{A_i^0}$, $A_i^- A_i^+ - q A_i^+ A_i^- = 1$, $i = 1, \ldots, 2n$. The algebra $U_q(\mathfrak{su}(1,1))$ is represented on $\mathcal{H}$ by using the coproduct to embed it in the tensor product of $2n$ copies of the q-deformation of the metaplectic representation, this gives

$$J_0^{(2n)} = \Delta^{(2n-1)}\Big(\tfrac{1}{2}\Big(A_i^0 + \tfrac{1}{2}\Big)\Big) = \frac{1}{2}\sum_{i=1}^{2n}\Big(A_i^0 + \frac{1}{2}\Big),$$
$$J_{\pm}^{(2n)} = \Delta^{(2n-1)}\left(\frac{1}{[2]_{q^{1/2}}}(A_i^{\pm})^2\right) = \frac{1}{[2]_{q^{1/2}}}\sum_{i=1}^{2n}\left((A_i^{\pm})^2 \prod_{j=i+1}^{2n} q^{A_j^0+\frac{1}{2}}\right). \tag{9}$$

The algebra $\mathfrak{o}_{q^{1/2}}(2n)$ can also be realized in terms of $2n$ q-oscillators. The $2n-1$ generators take the form

$$L_{i,i+1} = q^{-\frac{1}{2}(A_i^0+\frac{1}{2})}\Big(q^{\frac{1}{4}}A_i^+ A_{i+1}^- - q^{-\frac{1}{4}}A_i^- A_{i+1}^+\Big), \qquad i = 1, \ldots, 2n-1.$$

It can be checked that $[J_0^{(2n)}, L_{i,i+1}] = [J_{\pm}^{(2n)}, L_{i,i+1}] = 0$, $i = 1, \ldots, 2n-1$, in other words, that $U_q(\mathfrak{su}(1,1))$ and $\mathfrak{o}_{q^{1/2}}(2n)$ have commuting actions on the Hilbert space $\mathcal{H}$ of $2n$ q-oscillators. This sets the stage for Howe duality. In order to connect with the Askey–Wilson algebra we take $n = 3$. The expressions of the operators K_A and K_B acting on $\mathcal{H}$ that realize the relations (8) (together with the specific central elements) are rather involved and we shall refer the reader to [10] for the formulas. We shall only stress that these operators can be obtained in a dual way: They are affinely related to the generators of the commutant of $\mathfrak{o}_{q^{1/2}}(2)^{\oplus 3}$ in $\mathfrak{o}_{q^{1/2}}(6)$ as well as to the intermediate $U_q(\mathfrak{su}(1,1))$ Casimir elements $C^{(1234)} = \Delta^{(3)}(C)\otimes 1\otimes 1$ and $C^{(3456)} = 1\otimes 1\otimes \Delta^{(3)}(C)$ of the q-metaplectic representation (see (9)). This can be extended to higher ranks by letting n be arbitrary [11]. For $n = 2$ we are looking at the Clebsch–Gordan problem for $U_q(\mathfrak{su}(1,1))$. The q-Hahn algebra that arises has two dual realizations [12]: one as the commutant of $\mathfrak{o}_{q^{1/2}}(2)^{\oplus 2}$ in $U_q(\mathfrak{u}(4))$ and the other in terms of the following two $U_q(\mathfrak{su}(1,1))$ operators, $(\Delta(J_0)\otimes 1\otimes 1) - (1\otimes 1\otimes \Delta(J_0))$ and $\Delta^{(2)}(C)$ (the full Casimir element) in the q-metaplectic representation.

5 Conclusion

This paper has offered a summary of how the quadratic algebras of Racah, Hahn, Bannai–Ito, Askey–Wilson, and q-Hahn types can be given dual descriptions as commutant of Lie algebras, superalgebras, and quantum algebras. The connection between these dual pictures is rooted in Howe dualities whose various expressions have been stressed. The attentive reader will have noticed that the Clebsch–Gordan problem for $\mathfrak{osp}(1|2)$ has not been mentioned; this is because it has not been analyzed yet. We plan on adding this missing piece to complete the picture.

Acknowledgments The authors thank Luc Frappat, Eric Ragoucy, and Alexei Zhedanov for collaborations that led to the results reviewed here. JG holds an Alexander-Graham-Bell Scholarship from the Natural Science and Engineering Research Council (NSERC) of Canada. LV gratefully acknowledges his support from NSERC through a Discovery Grant.

References

1. D.J. Rowe, M.J. Carvalho, J. Repka, Dual pairing of symmetry groups and dynamical groups in physics. Rev. Mod. Phys. **84**, 711 (2012)
2. V.X. Genest, L. Vinet, A. Zhedanov, The Racah algebra and superintegrable models. J. Phys. Conf. Ser. **512**, 012011 (2014)
3. J. Gaboriaud, L. Vinet, S. Vinet, A. Zhedanov, The Racah algebra as a commutant and Howe duality. J. Phys. A Math. Theor. **51**, 50LT01 (2018)
4. H. De Bie, V.X. Genest, W. van de Vijver, L. Vinet, A higher rank Racah algebra and the Laplace-Dunkl operator. J. Phys. A Math. Theor. **51**, 025203 (2017)
5. J. Gaboriaud, L. Vinet, S. Vinet, A. Zhedanov, The generalized Racah algebra as a commutant. J. Phys. Conf. Ser. **1194**, 012034 (2019)
6. L. Frappat, J. Gaboriaud, L. Vinet, S. Vinet, A. Zhedanov, The Higgs and Hahn algebras from a Howe duality perspective. Phys. Lett. A **383**, 1531–1535 (2019)
7. H. De Bie, V.X. Genest, S. Tsujimoto, L. Vinet, A. Zhedanov, The Bannai–Ito algebra and some applications. J. Phys. Conf. Ser. **597**, 012001 (2015)
8. J. Gaboriaud, L. Vinet, S. Vinet, A. Zhedanov, The dual pair $Pin(2n) \otimes \mathfrak{osp}(1|2)$, the Dirac equation and the Bannai–Ito algebra. Nucl. Phys. B **937**, 226–239 (2018)
9. H. De Bie, V.X. Genest, L. Vinet, The $\mathbb{Z}_2^n$ Dirac-Dunkl operator and a higher rank Bannai-Ito algebra. Adv. Math. **5**, 390–414 (2016)
10. L. Frappat, J. Gaboriaud, E. Ragoucy, L. Vinet, The dual pair $\left(U_q(\mathfrak{su}(1,1)), \mathfrak{o}_{q^{1/2}}(2n)\right)$, q-oscillators and the higher rank Askey–Wilson algebra $AW(n)$. J. Math. Phys. **61**, 041701 (2020). https://doi.org/10.1063/1.5124251
11. H. De Bie, H. De Clerq, W. van de Vijver, The higher rank q-deformed Bannai–Ito and Askey–Wilson algebra. Commun. Math. Phys. **374**(1), 277 (2020)
12. L. Frappat, J. Gaboriaud, E. Ragoucy, L. Vinet, The q-Higgs and Askey–Wilson algebras. Nucl. Phys. B **944**, 114632 (2019)

Second-Order Supersymmetric Partners of the Trigonometric Rosen–Morse Potential

Rosa Reyes, D. J. Fernández, and H. Gasperín

Abstract The second-order supersymmetric partners of the trigonometric Rosen–Morse potential are studied. The stationary Schrödinger equation for this potential is solved in such a way that the general solution supplies straightforwardly the eigenstates of the Hamiltonian while the non-physical solutions turn out to be conveniently expressed for characterizing its global properties. This allows to implement in a simple and systematic way the second-order supersymmetry transformations.

Keywords Supersymmetric quantum mechanics · Trigonometric Rosen–Morse potential

1 Introduction

The trigonometric Rosen–Morse (TRM) potentials belong to the exactly solvable class of potentials, i.e., there exist explicit analytic expressions for their energy eigenstates and eigenvalues [1, 2]. These potentials are interesting in physics mainly for two reasons: the first one is their possible use for describing the quark-gluon interaction in quantum chromodynamics [3]; the second one is their intrinsic properties, making them ideal as a toy model for studying nonlinear algebras and supersymmetric quantum mechanics (SUSY QM) [4–6]. For example, they have a relatively simple dependence of the x-coordinate in a finite domain. In addition, they have an infinite discrete energy spectrum, with a nonlinear dependence of the energy levels on the index labeling them, making these potentials a clear case study for nonlinear algebras.

On the other hand, SUSY QM is a powerful tool for generating, from an exactly solvable initial Hamiltonian, new families of exactly solvable Hamiltonians whose spectra are quite similar to the initial one [7–11]. In this work we will apply the

R. Reyes (✉) · D. J. Fernández · H. Gasperín
Departamento de Física, Cinvestav, AP, Ciudad de México, Mexico
e-mail: rmreyes@fis.cinvestav.mx

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_22

second-order SUSY QM to the TRM potential for generating new families of exactly solvable potentials. Through these transformations we can design the spectra for the new potentials in several different ways, as we will exhibit in this article.

This work consists of three parts, the first one contains a brief review of the second-order SUSY QM, while the second will address the TRM potential and the solution to the corresponding Schrödinger equation. In the third part we will show results of the second-order SUSY transformation when applied to the TRM potential. At the end we will highlight the main results of this paper.

2 Supersymmetric Quantum Mechanics

The basic idea of SUSY QM is to deal with an intertwining relation which involves the operators H_i $(i = 0, 2)$ and $B^\dagger$ as follows:

$$H_2 B^\dagger = B^\dagger H_0. \tag{1}$$

We suppose that H_i $(i = 0, 2)$ are two one-dimensional Schrödinger Hamiltonians

$$H_i = -\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + V_i(x), \quad i = 0, 2,$$

where, for simplicity, we are working in dimensionless coordinates and $B^\dagger$ is the second-order differential intertwining operator

$$B^\dagger = \frac{1}{2}\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \eta(x)\frac{\mathrm{d}}{\mathrm{d}x} + \gamma(x)\right),$$

with $\eta(x)$, $\gamma(x)$ being two real unknown functions. If we plug the expressions for $B^\dagger$ and H_i $(i = 0, 2)$ into the intertwining relationship (1) we arrive to a coupled system of equations which, after some work, leads to

$$V_2 = V_0 - \eta', \qquad \gamma = \frac{\eta'}{2} + \frac{\eta^2}{2} - 2V_0 + d, \tag{2}$$

$$\frac{\eta\eta''}{2} - \frac{\eta'^2}{4} + \eta^2\eta' + \frac{\eta^2}{4} - 2V_0\eta^2 + d\eta^2 + c = 0, \tag{3}$$

where d, c are two real integration constants. It is important to note that in these expressions we have four unknown functions V_0, V_2, η, γ, but only three equations to determine them, thus we need some extra information to deal with the problem.

Suppose now that $V_0(x)$ is given, then we can determine $V_2(x)$ and $\gamma(x)$ once the solution $\eta(x)$ to the nonlinear differential equation (3) is obtained (see Eq. (2)). This is done by using the *ansatz*

$$\eta' = -\eta^2 + 2\beta\eta + 2\xi,$$

with β, ξ being two functions of x to be determined. This *ansatz* transforms Eq. (3) into the following set of equations:

$$\xi^2 = c, \qquad \epsilon = \tfrac{1}{2}(d + \xi), \qquad \beta' + \beta^2 = 2\left(V_0 - \epsilon\right).$$

The first two equations produce the solutions $\xi_{1,2} = \pm\sqrt{c}$ and $\epsilon_{1,2} = \left(d \pm \sqrt{c}\right)/2$. The third one is a first-order nonlinear differential equation known as Riccati equation, which can be transformed into a linear equation through the change $\beta_i = u'_{0i}/u_{0i}$, leading to

$$-\frac{1}{2}u''_{0i} + V_0 u_{0i} = \epsilon_i u_{0i}, \qquad i = 1, 2. \tag{4}$$

This is the initial stationary Schrödinger equation with potential V_0 for the two factorization energies $\epsilon_1,\ \epsilon_2$. The functions u_{0i}, $i = 1, 2$ are named seed solutions in the literature; depending on whether they are square integrable or not, they are called physical or non-physical solutions of the initial Hamiltonial H_0.

The second-order transformations can be classified according to the sign of the constant c involved in the factorization energies $\epsilon_{1,2}$. Thus, three different cases appear: the real case for $c > 0$, the complex case for $c < 0$, and the confluent case for $c = 0$ [10–12]. In all three cases the new potential is given by

$$V_2(x) = V_0(x) - \left[\ln\left(W(u_{01}, u_{02})\right)\right]''. \tag{5}$$

The function $W(u_{01}, u_{02})$ denotes the Wronskian of the two seed solutions u_{01} and u_{02} in the real and complex cases, while for the confluent case it is given by

$$W(u_{01}, u_{02}) = w_0 + \int_{x_0}^{x} [u_{01}(y)]^2 \, \mathrm{d}y, \tag{6}$$

where u_{01} is a seed solution satisfying Eq. (4), u_{02} fulfills $(H_0 - \epsilon_1)u_{02} = u_{01}$, and w_0 is an integration constant that can be adjusted to avoid that $W(u_{01}, u_{02})$ will have a zero in the x-domain. For the real case u_{01} and u_{02} must be taken as real solutions to Eq. (4), for the complex case u_{01} and u_{02} are complex conjugate seed solutions of (4) such that $u_{02} = u^*_{01}$. For further details on the conditions that the seed solutions must fulfill in each case, see for example [10–12].

On the other hand, the eigenfunctions $\psi_{0n}(x)$ of the initial Hamiltonian H_0 are related with those of the new Hamiltonian H_2 ($\psi_{2n}(x)$) as follows:

$$\psi_{2n}(x) = \frac{B^{\dagger}\psi_{0n}(x)}{\sqrt{(E_n - \epsilon_1)(E_n - \epsilon_2)}}.$$

Moreover, there exist solutions to the stationary Schrödinger equation for H_2 with factorization energies $\epsilon_1,\ \epsilon_2$, which are given by:

$$\psi_{2\epsilon_1}(x) \propto \frac{u_{02}(x)}{W(u_{01}, u_{02})}, \qquad \psi_{2\epsilon_2}(x) \propto \frac{u_{01}(x)}{W(u_{01}, u_{02})}.$$

The kind of modifications that can done in the spectrum of the resultant Hamiltonian H_2, as compared with the initial one, depends on the factorization energies chosen, as well as on the square-integrability of $\psi_{2\epsilon_1}$ and $\psi_{2\epsilon_2}$.

3 Trigonometric Rosen–Morse Potential

In this section we describe briefly the trigonometric Rosen–Morse potentials. They form a biparametric family of one-dimensional potentials in a finite domain, which in the dimensionless coordinate x are given by

$$V_0(x) = \frac{a(a+1)}{2}\csc^2(x) - b\cot(x), \qquad a > 0, \qquad b \in \mathbb{R}, \qquad x \in (0, \pi), \tag{7}$$

with a, b being the parameters of the potential. Since these potentials are time independent, it is required just to solve the corresponding stationary Schrödinger equation

$$\left(-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{a(a+1)}{2}\csc^2(x) - b\cot(x)\right)\psi(x) = E\psi(x). \tag{8}$$

One way to solve this equation is to transform it into the hypergeometric equation. After doing this, the general solution to the Schrödinger equation (8) is

$$\psi(x) = A\psi_L(x) + B\psi_R(x), \qquad A, B \in \mathbb{C}, \tag{9}$$

where

$$\psi_L(x) = \kappa(a, \nu, \mu)e^{-\left[\frac{\mu}{2} - i(\nu+a)\right]x}\sin^{a+1}(x)\,{}_2F_1\left(\nu+a, a+1+\tfrac{i\mu}{2}; \nu+\tfrac{i\mu}{2}; e^{2ix}\right)$$
$$+\rho(a, \nu, \mu)e^{\left[\frac{\mu}{2} + i(1-\nu-a)\right]x}\sin^{-a}(x)$$

$$\times {}_2F_1\left(1-\nu-a,\,-a-\frac{i\mu}{2};\,2-\nu-\frac{i\mu}{2};\,e^{2ix}\right), \tag{10}$$

$$\mu=\frac{2b}{\sqrt{E+\sqrt{E^2+b^2}}}, \qquad \nu=1-\sqrt{E+\sqrt{E^2+b^2}},$$

$$\kappa(a,\nu,\mu)=\frac{\Gamma(2a+2)\Gamma\left(1-\nu-i\frac{\mu}{2}\right)}{\Gamma\left(a+1-i\frac{\mu}{2}\right)\Gamma(a+2-\nu)}, \quad \rho(a,\nu,\mu)=\left(\frac{i}{2}\right)^{2a+1}\frac{\Gamma(2a+2)\Gamma\left(\nu-1+i\frac{\mu}{2}\right)}{\Gamma\left(a+1+i\frac{\mu}{2}\right)\Gamma(a+\nu)},$$

and

$$\psi_R(x)=\psi_L(\pi-x). \tag{11}$$

In this expression, in order to obtain $\psi_R(x)$, in addition of making the reflection of $\psi_L(x)$ with respect to $\pi/2$ we have to change as well in $\psi_L(\pi-x)$ the parameter b by $-b$. Note that $\{\psi_L(x),\psi_R(x)\}$ is a set of two linearly independent solutions of Eq. (8) vanishing to the left ($x=0$) and to the right ($x=\pi$), respectively. These expressions make easy to study the behavior of the solutions, particularly the non-physical ones. If the condition of square-integrability is imposed, we obtain the energy spectrum of the TRM Hamiltonian, whose energy levels are given by

$$E_n=\frac{1}{2}(n+a+1)^2-\frac{b^2}{2(n+a+1)^2}, \qquad n\in\mathbb{N}, \tag{12}$$

with bound state solutions

$$\psi_n(x)=C_n e^{-\left[\frac{b}{n+1+a}-in\right]x}\sin^{a+1}(x) \times {}_2F_1\left(-n,\,a+1-\frac{ib}{n+1+a};\,2a+2;\,2ie^{-ix}\sin(x)\right), \tag{13}$$

where C_n is a normalization constant. We have now all the information required to implement the second-order SUSY transformations for the TRM potentials.

4 SUSY Partners of the Trigonometric Rosen–Morse Potential

Once the general solution (9) to the stationary Schrödinger equation (8) has been constructed, we can study the second-order SUSY transformations that lead to a final non-singular real potential by exploring all possible combinations of factorization energies and associated seed solutions for the real, complex, and confluent cases [10–12]. Some examples of the resulting potentials for the different kinds of transformations are now discussed.

The real case has been partially analyzed in the past, for factorization energies coinciding with two consecutive energy levels of the initial Hamiltonian [13]. To illustrate this example, let us take as seed solutions the two bound states associated with the factorization energies $\epsilon_1 = E_2$ and $\epsilon_2 = E_1$, which leads to the following potential:

$$V_2(x) = \frac{(a+2)(a+3)}{2}\csc^2(x) - b\cot(x) + 4[(a+3)^2 + \tilde{b}^2] \times \frac{(a+2)^2 + \tilde{b}^2 + [(a+2)(a+3) - \tilde{b}^2]\cos(2x) - (2a+5)\tilde{b}\sin(2x)}{\{(a+3)^2 + \tilde{b}^2 + [(a+2)(a+3) - \tilde{b}^2]\cos(2x) - (2a+5)\tilde{b}\sin(2x)\}^2} \tag{14}$$

where $\tilde{b} = b/(a+1)$. Up to our knowledge, the explicit expression for this potential is new.

Note that in this work we extend these results, by using general seed solutions (9) whose factorization energies are not in the spectrum of the TRM potential. For doing this, we need to choose carefully such seed solutions, since their behavior depends on the two constants A, B involved in the linear combination (9), which will be taken real in order to guarantee that the seed solutions will be real. Moreover, if A, B have the same sign, then the number of zeros of the seed solution will be even, otherwise it will be odd. This information is enough for selecting appropriately the seed solutions, in order to implement non-singular transformations [10].

In Fig. 1 we can see several examples of how the new potential changes as the spectrum of the TRM potential is modified by a second-order transformation in the real case. In the left side we observe the potentials resulting from erasing two consecutive energy levels (dotted and dashed curves) of $V_0(x)$ (continuous curve). Let us note that the dashed curve in the graph corresponds to a potential given by Eq. (14). On the right side of Fig. 1 it is seen the potentials resulting from adding two levels in the same energy gap (dotted and dashed curves).

For the complex and confluent cases there exist some restrictions on the behavior of the seed solutions at the edges of the x-domain (see for example [10–12]). As the solutions (10) and (11) satisfy precisely such requirements, we can use them directly to implement the corresponding transformations.

For two complex conjugate seed solutions, and factorization energies, the implemented SUSY transformations generate isospectral potentials. Some examples are shown in Fig. 2.

On the other hand, the confluent case generates Hamiltonians which can be whether or not isospectral to the initial one. This depends on the w_0-parameter of Eq. (6), as well as of the factorization energy chosen. Due to the difficulty involved in evaluating the integral for general factorization energies, in this work we show only examples with a factorization energy coinciding with one of the energy levels of the initial Hamiltonian. In Fig. 3 we can see examples of the new potentials resulting from applying the confluent second-order SUSY transformation to the TRM potentials for the two factorization energies $\epsilon_1 = E_0,\ E_3$.

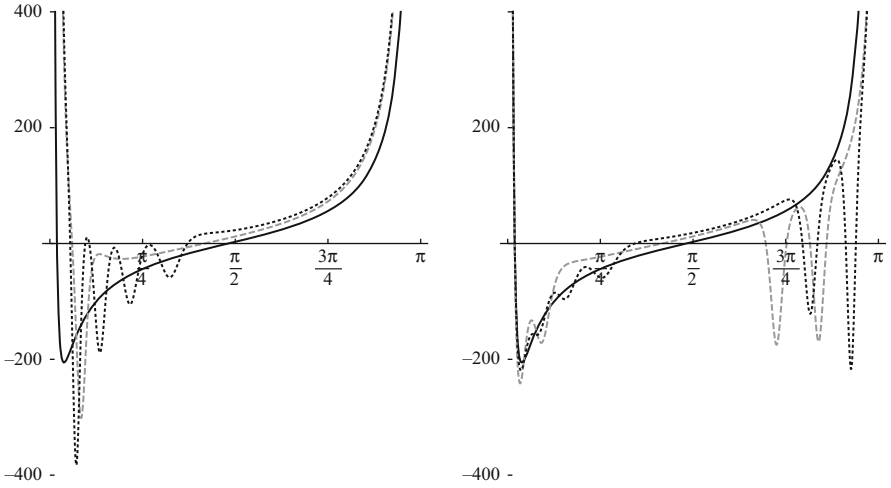


Fig. 1 TRM potential with $a = 2, b = 50$ (continuous curve) and its second-order SUSY partners in the real case (dashed and dotted curves). Left: potentials resulting from a transformation with $\epsilon_1 = E_2 = -37.5$, $\epsilon_2 = E_1 = -70.125$ (dashed curve), and with $\epsilon_1 = E_5 = 12.4688$, $\epsilon_2 = E_4 = -1.0102$ (dotted curve). Right: potentials appearing from a transformation with $\epsilon_1 = -40$, $\epsilon_2 = -60$ (dashed curve), and with $\epsilon_1 = -2$, $\epsilon_2 = -10$ (dotted curve)

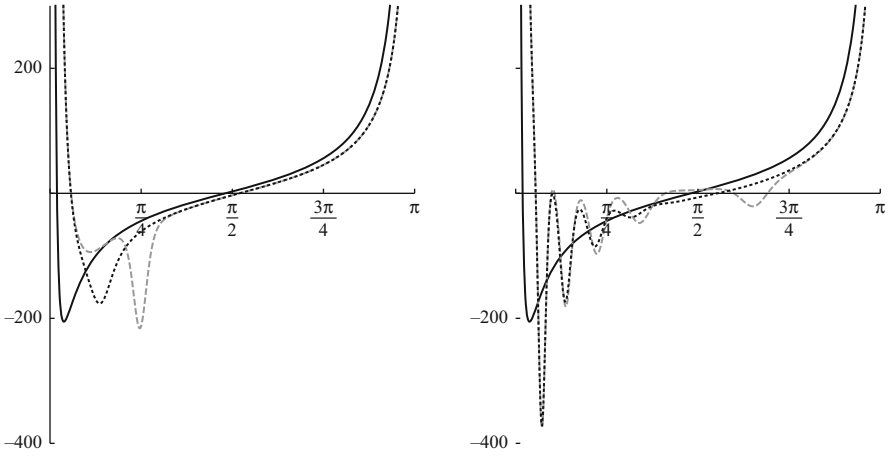


Fig. 2 Second-order SUSY partners (dashed and dotted curves) of the TRM potential with $a = 2$, $b = 50$ (continuous curve) for two complex conjugate factorization energies $\epsilon_2 = \epsilon_1^*$. Left: for $\epsilon_1 = E_1 + 0.5i$ ($E_1 = -70.125$) and for $\epsilon_1 = E_1 + 20i$ (dashed and dotted curves, respectively). Right: for $\epsilon_1 = 0.5i$ and $\epsilon_1 = 20i$ (dashed and dotted curves, respectively)

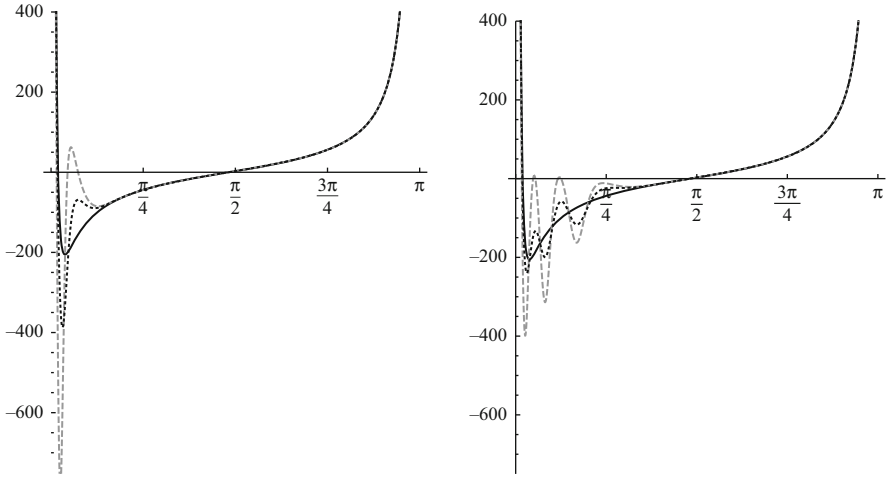


Fig. 3 Confluent second-order SUSY partner potentials (dashed and dotted curves) of the TRM potential with $a = 2$, $b = 50$ (continuous curve). Left: for $\epsilon_1 = E_0 = -134.389$, $w_0 = 0.1$ (dashed curve), and for $\epsilon_1 = E_0$, $w_0 = 0.5$ (dotted curve). Right: for $\epsilon_1 = E_2 = -37.5$, $w_0 = 0.5$ (dashed curve), and for $\epsilon_1 = E_2$, $w_0 = 0.1$ (dotted curve)

5 Conclusions

In this work we have expressed in an appropriate way the general solution to the stationary Schrödinger equation for the TRM potential. The main advantage of this construction is the possibility of characterizing simply its global properties, namely, the number of zeros and the behavior at the edges of the domain of the potential. This allows us to implement in a systematic way the second-order SUSY transformations once the seed solutions have been conveniently chosen.

We have reproduced the results reported in [13], but we have gone beyond by completing the study for the real case, also we have developed in full the complex case, and we have partially studied the confluent case, by considering only the situation when the factorization energy becomes one of the energy levels of H_0 in the last case.

Acknowledgment Rosa Reyes acknowledges the Conacyt scholarship No. 280723.

References

1. G.H. Sun, S.H. Dong, Quantum information entropies of the eigenstates for a symmetrically trigonometric Rosen–Morse potential. Phys. Scr. **87**, 045003 (2013)
2. C.L. Morrison, B. Shizgal, Pseudospectral solution of the Schrödinger equation for the Rosen-Morse and Eckart potentials. J. Math. Chem. **57**, 1035–1052 (2019)

3. C.B. Compeán, M. Kirchbach, Trigonometric quark confinement potential of QCD traits. Eur. Phys. J. A **33**, 1–4 (2007)
4. C.B. Compeán, M. Kirchbach, The trigonometric Rosen Morse potential in the supersymmetric quantum mechanics and its exact solutions. J. Phys. A Math. Gen. **39**, 547–557 (2006)
5. D.J. Fernández, V. Hussin, O. Rosas-Ortiz, Coherent states for Hamiltonians generated by supersymmetry. J. Phys. A Math. Theor. **40**, 6491–6511 (2007)
6. A. Chenaghlou, O. Faizy, Gazeau-Klauder coherent states for trigonometric Rosen-Morse potential. J. Math. Phys. **49**, 022104 (2008)
7. F. Cooper, A. Khare, U. Sukhatme, Supersymmetry and quantum mechanics. Phys. Rep. **251**, 267–385 (1995)
8. G. Junker, *Supersymmetric Methods in Quantum, Statistical and Solid State Physics* (IOP Publishing, Bristol, 2019)
9. B.K. Bagchi, *Supersymmetry in Quantum and Classical Mechanics* (Chapman & Hall, New York, 2001)
10. D.J. Fernández, N. Fernández-García, Higher-order supersymmetric quantum mechanics. AIP Conf. Proc. **744**, 236–273 (2005)
11. D.J. Fernández, Supersymmetric quantum mechanics. AIP Conf. Proc. **1287**, 3–36 (2010)
12. D.J. Fernández, Trends in supersymmetric quantum mechanics, in *Integrability, Supersymmetry and Coherent States. CRM Series in Mathematical Physics*, ed. by S. Kuru, J. Negro, L.M. Nieto (Springer, Cham, 2019), pp. 37–68
13. S. Domínguez-Hernández, D.J. Fernández, Rosen-Morse potential and its supersymmetric partners. Int. J. Theor. Phys. **50**, 1993–2001 (2011)

A Noncommutative Geometric Approach to the Batalin–Vilkovisky Construction

Roberta A. Iseppi

Abstract In this paper we argue why noncommutative geometry offers a natural geometrical framework to describe the Batalin–Vilkovisky construction for gauge theories over algebraic spaces. A key role is played by the notion of *BV-spectral triple*, which encodes all the elements of a BV-extended theory within a purely noncommutative geometrical object. An interesting aspect of this approach is that it provides all physical properties, like being a ghost field or anti-ghost field, with a geometrical interpretation. We present our results for the case of $U(2)$-matrix models. However, indications are given on how to perform the construction in the general setting of $U(n)$-theories.

Keywords Noncommutative geometry · Batalin–Vilkovisky construction · Finite spectral triple · Gauge theory · Matrix models

1 Introduction: Why Noncommative Geometry

Since its early days, noncommutative geometry [9] has shown a reciprocal and valuable interconnection with several areas of mathematics, such as motivic and foliation theory, operator algebras, and KK-theory. However, maybe even more remarkably, noncommutative geometry revealed a deep relation to quantum field theory and gauge theory in particular. A confirmation of that can be found in a series of the celebrated results, which began with the pioneering papers by Connes [8, 10], had a breakthrough in [3, 4, 11], and finally arrived to the key result obtained by Chamseddine, Connes, and Marcolli [5] of deriving the full Standard Model of particles, with neutrino mixing and minimally coupled to gravity, from a purely noncommutative geometrical object. Furthermore, recently new approaches have been suggested to go beyond the Standard Model, using the framework provided by

R. A. Iseppi (✉)
Center for Quantum Geometry of Moduli Spaces, Aarhus University, Aarhus, Denmark
e-mail: roberta.iseppi@qgm.au.dk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_23

noncommutative geometry as starting point to search for a new spectral model of gravity coupled with matter [6, 7, 13].

Therefore, as confirmed by the quoted monumental results, it is natural to investigate whether noncommutative geometry could provide a new mathematical description of other constructions developed in the context of gauge theories. In this article we focus in particular on the Batalin–Vilkovisky approach to the BRST construction. After a concise recall of the needed notions from the noncommutative geometric world (cf. Sect. 2), we will briefly outline the algebraic approach to the Batalin–Vilkovisky construction for gauge theories with an affine configuration space (cf. Sect. 3). Then, Sect. 4 is devoted to present our main result: focusing on a $U(2)$-matrix model, we present how all the elements playing a role in the BV construction can be successfully encoded in a purely noncommutative geometrical object, the so-called *BV-spectral triple*. Finally, in Sect. 5 we explain how the problem can be approached and solved in the general context of finite spectral triples on the algebra $M_n(\mathbb{C})$.

2 Finite Spectral Triples and Induced Gauge Theories

Without any doubt, the notion of spectral triple plays a central role in contemporary noncommutative geometry. In its full generality, a spectral triple can be viewed as a noncommutative version of the classical concept of Riemannian spin manifold. However, conversely to what happens in the classical setting, a spectral triple presents a very rich and interesting structure also when the underline topological space is 0-dimensional, and hence the corresponding spectral triple is finite dimensional. Even more, it is precisely a finite spectral triple that, in the description of the full Standard Model as induced by an almost commutative spectral triple [5], encodes the particle content of the theory. Therefore, we briefly recall the needed notions in this finite dimensional context, where also our construction will take place (cf. [9, 12]).

Definition 1 A *spectral triple* $(\mathcal{A}, \mathcal{H}, D)$ consists of an involutive unital algebra $\mathcal{A}$, faithfully represented as operators on a Hilbert space $\mathcal{H}$, together with a self-adjoint operator D on $\mathcal{H}$, with a compact resolvent, such that the commutators $[D, a]$ are bounded operators for each $a \in A$. A spectral triple $(\mathcal{A}, \mathcal{H}, D)$ is *finite* if the Hilbert space $\mathcal{H}$ and hence the algebra $\mathcal{A}$ are finite dimensional.

Given a spectral triple $(\mathcal{A}, \mathcal{H}, D)$, its structure can be further enriched via the introduction of a *real structure*, determining a *real spectral triple* $(\mathcal{A}, \mathcal{H}, D, J)$.

Definition 2 A *real structure of odd KO-dimension n (mod 8)* on a spectral triple $(\mathcal{A}, \mathcal{H}, D)$ is an anti-linear isometry $J : \mathcal{H} \to \mathcal{H}$ that satisfies

$$J^2 = \epsilon \quad \text{and} \quad JD = \epsilon' DJ.$$

The constants ϵ and ϵ' depend on the odd KO-dimension n (mod 8) as follows:

n	1	3	5	7
ϵ	1	−1	−1	1
ϵ'	−1	1	−1	1

Moreover, we require for all $a, b \in \mathcal{A}$ that:

- the action of $\mathcal{A}$ satisfies the *commutation rule*: $\left[a, Jb^*J^{-1}\right] = 0$;
- the operator D fulfills the *first-order condition*: $[[D, a], Jb^*J^{-1}] = 0$.

Next to its purely geometrical nature, a spectral triple is also strongly related to gauge theory: indeed, each spectral triple $(\mathcal{A}, \mathcal{H}, D)$ naturally induces a gauge theory, whose gauge-invariant action is given by the so-called *spectral action*.

Definition 3 For a finite spectral triple $(\mathcal{A}, \mathcal{H}, D)$ and a suitable real-valued function f, the *spectral action* S_0 is given by

$$S_0[D + M] := Tr\big(f(D + M)\big)$$

with domain the set of self-adjoint operators of the form $M = \sum_j a_j[D, b_j]$, for $a_j, b_j \in \mathcal{A}$.

Definition 4 Let X_0 be a vector space over $\mathbb{R}$, S_0 be a functional on X_0, $S_0 : X_0 \to \mathbb{R}$, and $\mathcal{G}$ be a group acting on X_0 through an action $F : \mathcal{G} \times X_0 \to X_0$. Then the pair (X_0, S_0) is a *gauge theory with gauge group $\mathcal{G}$ if it holds that*

$$S_0(F(g, \varphi)) = S_0(\varphi), \qquad \forall \varphi \in X_0,\ \forall g \in \mathcal{G}.$$

Concerning the terminology, X_0 is the *configuration space*, an element φ in X_0 is a *gauge field*, the functional S_0 is the *action*, and $\mathcal{G}$ is known as the *gauge group*.

Proposition 1 *Each finite spectral triple $(\mathcal{A}, \mathcal{H}, D)$ naturally induces a gauge theory (X_0, S_0), where the configuration space is the space of inner fluctuation*

$$X_0 := \Big\{\varphi = \sum_j a_j\big[D, b_j\big] : \varphi^* = \varphi,\ a_j, b_j \in \mathcal{A}\Big\},$$

for $*$ the involution on $\mathcal{A}$, and the action functional S_0 is the spectral action

$$S_0[D + \varphi] := Tr\big(f(D + \varphi)\big),$$

with f a polynomial in one real variable and Tr the classical matrix trace. Finally, the unitary elements in $\mathcal{A}$ determine the gauge group $\mathcal{G}$

$$\mathcal{G} := \mathcal{U}(\mathcal{A}) = \big\{u \in \mathcal{A} : uu^* = u^*u = 1\big\},$$

which acts on X_0 as follows:

$$\begin{aligned} \mathcal{G} \times X_0 &\longrightarrow X_0 \\ (u, \varphi) &\mapsto u\varphi u^* + u[D, u^*]. \end{aligned}$$

The proof of the above classical Proposition is a straightforward checking. We remark that a similar construction can be performed also in the infinite dimensional case. To conclude the section, we recall that there is another notion of action for spectral triples, which will play a key role in our construction: the so-called *fermionic action*.

Definition 5 For a finite spectral triple $(\mathcal{A}, \mathcal{H}, D)$ (finite real spectral triple $(\mathcal{A}, \mathcal{H}, D, J)$) the *fermionic action* on $\mathcal{H}_f \subseteq \mathcal{H}$ is given by

$$S_{\text{ferm}}[\varphi] = \frac{1}{2}\langle \varphi, D\varphi \rangle \quad \left(S_{\text{ferm}}[\varphi] = \frac{1}{2}\langle J\varphi, D\varphi \rangle \right); \qquad \text{for } \varphi \in \mathcal{H}_f.$$

3 The BV Construction in the Algebraic Context

The Batalin–Vilkovisky (BV) formalism (cf. [1, 2]) can be viewed as the end point of a long path, which had its motivation in the problem of defining the path integral (cf.[16]) for gauge theories and its origin in the introduction of the concept of *ghost field* by Faddeev and Popov in 1967 [14]. As suggested by the name, the ghost fields are non-existing particles, whose function is to compensate the presence of local symmetries and hence the appearance of divergences in the path integral. Moreover, next to the ghost fields, the BV formalism requires also the introduction of all the corresponding anti-fields/anti-ghost fields.

Definition 6 A *field/ghost field* φ is a graded variable characterized by two integers:

$$\deg(\varphi) \in \mathbb{Z} \quad \text{and} \quad \epsilon(\varphi) \in \{0, 1\}, \quad \text{with} \quad \deg(\varphi) = \epsilon(\varphi) \pmod{\mathbb{Z}/2}.$$

$\deg(\varphi)$ is the *ghost degree*, while $\epsilon(\varphi)$ is the *parity*, which distinguishes between the bosonic case, where $\epsilon(\varphi) = 0$ and φ behaves as a real variable, and the fermionic case, where $\epsilon(\varphi) = 1$ and φ behaves as a Grassmannian variable:

$$\varphi\psi = -\psi\varphi, \qquad \text{and} \qquad \varphi^2 = 0, \qquad \text{if} \quad \epsilon(\varphi) = \epsilon(\psi) = 1.$$

The *anti-field/anti-ghost field* φ^* corresponding to a field/ghost field φ satisfies

$$\deg(\varphi^*) = -\deg(\varphi) - 1, \qquad \text{and} \qquad \epsilon(\varphi^*) = \epsilon(\varphi) + 1, \pmod{\mathbb{Z}/2}.$$

Then, given an initial gauge theory (X_0, S_0), the BV construction associates with that a new pair $(\widetilde{X}, \widetilde{S})$, where the *extended configuration space* $\widetilde{X}$ and the *extended action* $\widetilde{S}$ are given by:

$$\widetilde{X} = X_0 \cup \{ \text{ ghost \& anti-ghost fields } \} \quad \widetilde{S} = S_0 + \text{ terms in ghost/anti-ghosts.}$$

However, conditions need to be imposed on how we perform this extension, on the number and type of ghost and anti-ghost fields we have to introduce and on the properties satisfied by the extended action $\widetilde{S}$ (cf. [15, 17]).

Definition 7 Let the pair (X_0, S_0) be a gauge theory. Then an *extended theory* associated with (X_0, S_0) is a pair $(\widetilde{X}, \widetilde{S})$ where $\widetilde{X} = \oplus_{i\in\mathbb{Z}}[\widetilde{X}]^i$ is a super graded vector space suitable to be decomposed as

$$\widetilde{X} \cong \mathcal{F} \oplus \mathcal{F}^*[1], \qquad \text{with} \quad [\widetilde{X}]^0 = X_0 \tag{1}$$

for $\mathcal{F} = \oplus_{i\geqslant 0}\mathcal{F}^i$ a graded locally free $\mathcal{O}_{X_0}$-module with homogeneous components of finite rank, and $\widetilde{S} \in [\mathcal{O}_{\widetilde{X}}]^0$ is a regular function on $\widetilde{X}$, with $\widetilde{S}_{|X_0} = S_0$, $\widetilde{S} \neq S_0$ and such that it solves the *classical master equation*, i.e.,

$$\{\widetilde{S}, \widetilde{S}\} = 0,$$

where $\{-,-\}$ denotes the graded Poisson structure on the algebra $\mathcal{O}_{\widetilde{X}}$.

Note The condition in (1) enforces the prescription of the BV formalism of introducing all anti-fields/anti-ghost fields corresponding to the fields/ghost fields in $\widetilde{X}$. In particular, $\mathcal{F}$ describes the fields/ghost fields in the extended theory while $\mathcal{F}^*[1]$ denotes the shifted dual module of the anti-fields/anti-ghost fields:

$$\mathcal{F}^*[1] = \oplus_{i\in\mathbb{Z}}\big[\mathcal{F}^*[1]\big]^i \qquad \text{with} \quad \big[\mathcal{F}^*[1]\big]^i = \big[\mathcal{F}^*\big]^{i+1}.$$

Moreover, the Poisson structure on $\mathcal{O}_{\widetilde{X}}$ is completely determined by requiring that, on the generators of $\widetilde{X}$, it satisfies the following conditions:

$$\big\{\beta_i, \beta_j\big\} = 0, \qquad \big\{\beta_i^*, \beta_j\big\} = \delta_{ij} \qquad \text{and} \qquad \big\{\beta_i^*, \beta_j^*\big\} = 0$$

for $\beta_i \in \mathcal{F}^p$ and $\beta_i^* \in \big[\mathcal{F}^*[1]\big]^{-p-1}$, $p \in \mathbb{Z}_{\geqslant 0}$, while its value on any other possible combination of fields/ghost fields/anti-fields and anti-ghost fields is equal to zero.

4 BV-Spectral Triple: The Notion and the Relevance

In this section we present how the BV construction for gauge theories on affine spaces can be encoded in the framework of noncommutative geometry. For simplicity, we focus on a $U(2)$-gauge theory, induced by the finite spectral triple

$$(M_2(\mathbb{C}), \mathbb{C}^2, D),$$

where D is a self-adjoint 2×2-matrix. By applying Proposition 1, we have that the above spectral triple induces a gauge theory (X_0, S_0) with

$$X_0 = \{M \in M_2(\mathbb{C}) : M^* = M\}, \quad S_0[M] = Tr(f(M)) \quad \text{and} \quad \mathcal{G} = U(2),$$

where f a polynomial in $\mathcal{O}_{X_0}$. By fixing as basis for X_0 the one given by the Pauli matrices together with the identity matrix $\{\sigma_1, \sigma_2, \sigma_3, \sigma_4 = Id\}$, we have the following identifications:

$$X_0 \simeq \mathbb{A}^4_{\mathbb{R}} = \langle M_1, M_2, M_3, M_4 \rangle_{\mathbb{R}} \qquad S_0 = \sum_{k=0}^{r} (M_1^2 + M_2^2 + M_3^2)^k g_k(M_4),$$

where M_a, $a = 1, \ldots, 4$ are real variables and $g_k(M_4)$ are suitable polynomials only in M_4. Given the gauge theory (X_0, S_0) just described, one can verify (cf. [17]) that the corresponding minimally BV-extended theory has an extended configuration space whose decomposition as $\mathbb{Z}$-graded vector space is:

$$\widetilde{X} = \langle E^* \rangle_{-3} \oplus \langle C_1^*, \cdots, C_3^* \rangle_{-2} \oplus \langle M_1^*, \ldots, M_4^* \rangle_{-1} \oplus X_0 \oplus \langle C_1, \cdots, C_3 \rangle_1 \oplus \langle E \rangle_2.$$

and an extended action

$$\widetilde{S} = S_0 + \sum_{i,j,k} \epsilon_{ijk} M_i^* M_j C_k + \sum_{i,j,k} C_i^* (M_i E + \epsilon_{ijk} C_j C_k). \tag{2}$$

for ϵ_{ijk} a totally antisymmetric tensor in the indices $i, j, k \in \{1, 2, 3\}$, with $\epsilon_{123} = 1$.

After having determined this BV-extended pair $(\widetilde{X}, \widetilde{S})$, a natural question arises: indeed, because the pair (X_0, S_0) came as the gauge theory naturally induced by a finite spectral triple, one might wonder if also the corresponding BV-extended theory can be encoded in a new BV-spectral triple. In other words, we want to determine a real spectral triple $(\mathcal{A}_{BV}, \mathcal{H}_{BV}, \mathcal{D}_{BV}, J_{BV})$ such that its fermionic action S_{ferm} coincides with the BV action $S_{BV} := \widetilde{S} - S_0$ of the model. As we explain in the theorem below, this goal can be reached. However, before constructing all the elements entering the BV-spectral triple, we remark that the reason why we have to introduce a real structure and to consider the fermionic action, instead of the spectral action, is the presence of Grassmannian variables both in $\widetilde{X}$ and $\widetilde{S}$.

We define the following data:

- $\mathcal{A}_{BV} := M_2(\mathbb{C})$;
- $\mathcal{H}_{BV} := \big[M_2(\mathbb{C})\big]_M \oplus \big[M_2(\mathbb{C})\big]_C \oplus \big[M_2(\mathbb{C})\big]_E$.
  The inner product structure is the Hilbert–Schmidt inner product on each summand. Moreover, by $\mathcal{H}_{BV,f}$, we identify the following subspace:

$$\begin{aligned} \mathcal{H}_{BV,f} &= i \cdot u(2) \oplus i \cdot su(2) \oplus u(1) \\ &\simeq \big\{[M_1, \ldots, M_4], [C_1, \ldots, C_3, 0], [0, \ldots, 0, iE]\big\} \end{aligned} \tag{3}$$

where, for now, M_a, C_j, and E are all treated as real variables;

- the self-adjoint linear operator D_{BV} acting on $\mathcal{H}_{\mathrm{BV}}$ is given by

$$D_{\mathrm{BV}} := \begin{pmatrix} 0 & R & T \\ R & S & 0 \\ T & 0 & 0 \end{pmatrix}$$

where the linear operators R, S, T are defined by

$$R : \mathcal{H}_C \to \mathcal{H}_M; \qquad S : \mathcal{H}_C \to \mathcal{H}_C; \qquad T : \mathcal{H}_M \to \mathcal{H}_M;$$
$$\varphi_C \mapsto [\beta, \varphi_C], \qquad \varphi_C \mapsto [\alpha, \varphi_C], \qquad \varphi_C \mapsto [\alpha, \varphi_C]_+.$$

Here, α and β denote Hermitian, traceless 2×2-matrices and we stress that, while R and S are defined as a commutator, T is an odd derivation, given in terms the anti-commutator. Alternatively, if we write α and β in terms of the Pauli matrices as follows:

$$\alpha = \tfrac{1}{2}\big[(-C_1^*)\sigma_1 + (-C_2^*)\sigma_2 + (-C_3^*)\sigma_3\big]$$
$$\beta = \tfrac{1}{2}\big[(-M_1^*)\sigma_1 + (-M_2^*)\sigma_2 + (-M_3^*)\sigma_3\big],$$

for C_i^* and M_i^* real variables, R, S, and T can be expressed as the following 4×4-matrices:

$$R := \begin{pmatrix} 0 & +iM_3^* & -iM_2^* & 0 \\ -iM_3^* & 0 & +iM_1^* & 0 \\ +iM_2^* & -iM_1^* & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad S := \begin{pmatrix} 0 & +iC_3^* & -iC_2^* & 0 \\ -iC_3^* & 0 & +iC_1^* & 0 \\ +iC_2^* & -iC_1^* & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$T := \begin{pmatrix} 0 & 0 & 0 & C_1^* \\ 0 & 0 & 0 & C_2^* \\ 0 & 0 & 0 & C_3^* \\ C_1^* & C_2^* & C_3^* & 0 \end{pmatrix}$$

- $J_{\mathrm{BV}} : \mathcal{H}_{\mathrm{BV}} \to \mathcal{H}_{\mathrm{BV}}$, with $J_{\mathrm{BV}}(\varphi) := i \cdot \varphi^*$, for $\varphi \in \mathcal{H}_{BV}$. Note that $(J_{BV})^2 = Id$.

We remark that the operator D_{BV} neither commutes nor anti-commutes with J_{BV}. Indeed, if we decompose D_{BV} as

$$D_{\mathrm{BV}} = D_1 + D_2 \quad \text{with} \quad D_1 = \begin{pmatrix} 0 & R & 0 \\ R^* & S & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad D_2 = \begin{pmatrix} 0 & 0 & T \\ 0 & 0 & 0 \\ T & 0 & 0 \end{pmatrix}.$$

we find that

$$J_{\mathrm{BV}} D_1 = -D_1 J_{\mathrm{BV}}, \qquad J_{\mathrm{BV}} D_2 = +D_2 J_{\mathrm{BV}}.$$

Hence, we are constructing a real spectral triple of *mixed KO-dimension*.

Theorem 1 *The data* $(\mathcal{A}_{\mathrm{BV}}, \mathcal{H}_{\mathrm{BV}}, D_{\mathrm{BV}}, J_{\mathrm{BV}})$ *define a real spectral triple (with mixed KO-dim.), whose fermionic action coincides with the BV action in* (2)*:*

$$S_{\mathrm{BV}} = \frac{1}{2}\langle J_{\mathrm{BV}}(\varphi), D_{\mathrm{BV}}\varphi\rangle, \qquad \text{with} \quad \varphi \in \mathcal{H}_{BV,f}.$$

where M_a, E *and* C_j^* *have to be treated as real variables while* M_a^* *and* C_j *behave as Grassmannian variables.*

Note It can be checked that the algebra $\mathcal{A}_{BV}$ is the largest unital algebra that complete the triple $(\mathcal{H}_{BV}, D_{BV}, J_{BV})$ defined above to a real spectral triple.

Because the theorem is proved by directly checking all the requirements appearing in the definition of a spectral triple (cf. [18]), we prefer not to present the details but to conclude with few remarks on how the physical properties of the BV-extended theory get translated in the noncommutative geometrical language. Indeed, we have that, while the anti-fields/anti-ghost fields M_a^* and C_j^* appear as entries of the operator D_{BV}, the fields/ghost fields M_a, C_j, and E determine the elements in $\mathcal{H}_{BV,f}$. Moreover, the new phenomena of a spectral triple of mixed-KO dim. accounts for the presence of bosonic and fermionic fields both $\widetilde{X}$ and $\widetilde{S}$.

5 Conclusions and Outlooks

The construction presented in the above section can be applied also to the more general case of $U(n)$-gauge theories, naturally induced by finite spectral triples on the algebra $M_n(\mathbb{C})$. What made the above construction possible was the fact that the extended action $\widetilde{S}$ was precisely linear in the anti-fields/anti-ghost fields: indeed, this property allows to rewrite the BV action as a fermionic action, having all the anti-fields/anti-ghost fields as entries of the operator D_{BV}. However, this linearity condition holds also for $U(n)$-theories, with $n > 2$. Even though in principle a BV action obtained by applying the algebraic BV construction could contain higher other terms in anti-fields/anti-ghost fields, the fact that for this class of models the algebra of gauge transformation is closed under commutations on the critical locus $X_{crit} \subset X_0$ of the action functional S_0 ensures the appearance of only linear terms in the anti-fields/anti-ghost fields. Hence the structure found for this $U(2)$-model perfectly replicates for the whole class of $U(n)$-theories.

References

1. I.A. Batalin, G.A. Vilkovisky, Gauge algebra and quantization. Phys. Lett. **B102**, 27–31 (1981)
2. I.A. Batalin, G.A. Vilkovisky, Quantization of gauge theories with linearly dependent generators. Phys. Rev. **D28**, 2567–2582 (1983). Erratum D30, 508 (1984)

3. A.H. Chamseddine, A. Connes, Universal formula for noncommutative geometry actions: unifications of gravity and the standard model. Phys. Rev. Lett. **77**, 4868–4871 (1996)
4. A.H. Chamseddine, A. Connes, The spectral action principle. Commun. Math. Phys. **186**, 731–750 (1997)
5. A.H. Chamseddine, A. Connes, M. Marcolli, Gravity and the Standard Model with neutrino mixing. Adv. Theor. Math. Phys. **11**, 991–1089 (2007)
6. A.H. Chamseddine, A. Connes, W.D. van Suijlekom, Beyond the spectral standard model: emergence of Pati-Salam unification. JHEP **1311**, 132 (2013)
7. A.H. Chamseddine, A. Connes, W.D. van Suijlekom, Grand unification in the spectral Pati-Salam model. JHEP **1511**, 011 (2015)
8. A. Connes, Essay on physics and noncommutative geometry, in *The Interfaces of Mathematics and Particle Physics (Oxford 1988)*. Inst. Math. Applu. Conf. Ser. New Ser., vol. 24 (Oxford University Press, New York, 1990), pp. 9–48
9. A. Connes, *Noncommutative Geometry* (Academic Press, San Diego, 1994)
10. A. Connes, Noncommutative geometry and reality. J. Math. Phys. **36**(11), 6194–6231 (1995)
11. A. Connes, Gravity coupled with matter and the foundation of non-commutative geometry. Commun. Math. Phys. **182**, 155–176 (1996)
12. A. Connes, M. Marcolli, *Noncommutative Geometry, Quantum Fields and Motives*. Colloquium Publications, vol. 55 (American Mathematical Society, Providence, 2008)
13. A. Devastato, F. Lizzi, P. Martinetti, Grand symmetry, spectral action, and the Higgs mass. JHEP **1401**, 042 (2014)
14. L.D. Faddeev, V.N. Popov, Feynman diagrams for the Yang-Mills field. Phys. Lett. **B25**, 29–30 (1967)
15. G. Felder, D. Kazhdan, The classical master equation, in *Perspectives in Representation Theory, Contemporary Mathematics*, ed. by P. Etingof, M. Khovanov, A. Savage (American Mathematical Society, Providence, 2014)
16. R.P. Feynman, A.R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965)
17. R.A. Iseppi, The BV formalism: theory and application to a U(2)-matrix model. Rev. Math. Phys. **31**(10), 129–141 (2019)
18. R.A. Iseppi, W.D. van Suijlekom, Noncommutative Geometry and the BV formalism: application to a matrix model. J. Geom. Phys. **120**, 129–141 (2017)

A New Method for Constructing Squeezed States for the Isotropic 2D Harmonic Oscillator

James Moran and Véronique Hussin

Abstract We introduce a new method for constructing squeezed states for the 2D isotropic harmonic oscillator. Based on the construction $SU(2)$ coherent states, we define a new set of ladder operators for the 2D system as a linear combination of the x and y ladder operators and construct the $SU(2)$ coherent states. The new ladder operators are used for generalizing the squeezing operator to 2D and the $SU(2)$ coherent states play the role of the Fock states in the expansion of the 2D squeezed states. We discuss some properties of the 2D squeezed states.

Keywords Coherent states · Squeezed states · Harmonic oscillator · $SU(2)$ coherent states · 2D coherent states · 2D squeezed states · Uncertainty principle

1 Introduction

Degeneracy in the spectrum of the Hamiltonian is one of the first problems we encounter when trying to define a new type of coherent states for the 2D oscillator. As a continuation of the work in [1] we produce a non-degenerate number basis ($SU(2)$ coherent states) for the 2D isotropic harmonic oscillator with accompanying generalized creation and annihilation operators. The squeezed states for the 2D isotropic harmonic oscillator are then defined in terms of the $SU(2)$ coherent states and generalized ladder operators.

J. Moran (✉)
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada

Département de Physique, Université de Montréal, Montréal, QC, Canada
e-mail: james.moran@umontreal.ca

V. Hussin
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada

Département de Mathématiques et de Statistique, Université de Montréal, Montréal, QC, Canada
e-mail: hussin@dms.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_24

Work on degeneracy in coherent state theory has been done, Klauder described coherent states of the hydrogen atom [2] which preserved many of the usual properties required by coherent state analysis [3]. Fox and Choi proposed the Gaussian Klauder states [4], an alternative method for producing coherent states for more general systems with degenerate spectra. An analysis of the connection between the two definitions was studied in [5].

In the first part of the paper we address the degeneracy in the energy spectrum by constructing non-degenerate states, the $SU(2)$ coherent states, and we define a generalized ladder operator formed from a linear combination of the 1D ladder operators with complex coefficients.

In the last part of the paper we use a generalized squeezing operator and Fock space expansion to define squeezed states for the 2D system. In both cases we use the same definitions as for the 1D squeezed states, but replacing the Fock states with the $SU(2)$ coherent states and the 1D ladder operators with the new generalized ladder operators. We discuss the spatial probability distributions of the 2D squeezed states, as well as their dispersions.

2 Squeezed States of the 1D Harmonic Oscillator

Squeezed states, or squeezed coherent states, are a generalization of the standard coherent states first studied by Schrödinger [6], and then formalized in the context of quantum optics by Glauber and Sudarshan [7, 8]. In terms of the displacement and squeezing operators $D(\psi) = e^{\psi a^\dagger - \bar{\psi} a}$, $S(\xi) = e^{\frac{1}{2}(\xi a^{\dagger 2} - \bar{\xi} a^2)}$ respectively, where $a, a^\dagger$ are the annihilation and creation operators, squeezed states are expressed as

$$|\psi, \xi\rangle = D(\psi) S(\xi) |0\rangle , \tag{1}$$

$\psi, \xi \in \mathbb{C}$. Writing $\xi = re^{i\theta}$, in terms of Fock states, $\{|n\rangle\}$, the squeezed states are given by

$$|z, \gamma\rangle = \frac{1}{\mathcal{N}(z,\gamma)} \sum_{n=0}^{\infty} \frac{1}{\sqrt{n!}} \left(\frac{\gamma}{2}\right)^{\frac{n}{2}} H_n\left(\frac{z}{\sqrt{2\gamma}}\right) |n\rangle , \tag{2}$$

where $\frac{1}{\mathcal{N}(z,\gamma)} = \frac{1}{\sqrt{\cosh r}} e^{-\frac{|z|^2}{2}} e^{\frac{\tanh r}{2} \mathrm{Re}(e^{i\theta} \bar{z}^2)}$. The states in Eq. (2) are solutions to the eigenvalue equation

$$(a + \gamma a^\dagger) |z, \gamma\rangle = z |z, \gamma\rangle . \tag{3}$$

Equivalence between definitions (1) and (2) is understood through the following relationships between the parameters [9]:

$$\begin{aligned} z &= \psi - \bar{\psi} e^{i\theta} \tanh r, \\ \gamma &= -e^{i\theta} \tanh r. \end{aligned} \tag{4}$$

The term "squeezing" is used because the squeezed states saturate the Robertson–Schrödinger uncertainty relation [10] but with unequal dispersions in position and momentum (unlike the standard coherent states which saturate the Heisenberg uncertainty principle with equal dispersions). The squeezed states have the following dispersions:

$$\begin{aligned} (\Delta X)^2_{|\psi,\xi\rangle} &= \langle \psi, \xi | \, X^2 - \langle X \rangle^2 \, | \psi, \xi \rangle = \frac{1}{2} + \sinh^2 r + \mathrm{Re}(e^{i\theta}) \cosh r \sinh r; \\ (\Delta P)^2_{|\psi,\xi\rangle} &= \langle \psi, \xi | \, P^2 - \langle P \rangle^2 \, | \psi, \xi \rangle = \frac{1}{2} + \sinh^2 r - \mathrm{Re}(e^{i\theta}) \cosh r \sinh r, \end{aligned} \tag{5}$$

where $(\Delta \hat{O})^2_{|\psi\rangle} \equiv \langle \psi | O \rangle^2 - \langle \hat{O} \rangle^2 \, | \psi \rangle$ is the variance of the operator $\hat{O}$ in the state $|\psi\rangle$. The position and momentum operators are expressed in the usual way $\hat{X} = \frac{1}{\sqrt{2}}(a^\dagger + a)$, $\hat{P} = \frac{1}{\sqrt{2}i}(a - a^\dagger)$. When the squeezing is purely real $\xi = r$, the dispersions become $(\Delta X)^2_{|\psi,\xi\rangle} = \frac{1}{2} e^{-2r}$, $(\Delta P)^2_{|\psi,\xi\rangle} = \frac{1}{2} e^{2r}$, in this case the squeezed states saturate the Heisenberg uncertainty relation $(\Delta X)^2_{|\psi,\xi\rangle} (\Delta P)^2_{|\psi,\xi\rangle} = \frac{1}{4}$.

Like the standard coherent states, the squeezed states are also non-orthogonal and they admit a resolution of the identity [11], therefore they represent an overcomplete basis for the Hilbert space of the 1D harmonic oscillator.

3 The 2D Oscillator

For a 2D isotropic oscillator we have the quantum Hamiltonian

$$\hat{H} = -\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}\frac{d^2}{dy^2} + \frac{1}{2}x^2 + \frac{1}{2}y^2, \tag{6}$$

where we have set $\hbar = 1$ and the mass $m = 1$ and the frequency $\omega = 1$. We solve the time independent Schrödinger equation $H \, |\Psi\rangle = E \, |\Psi\rangle$ and obtain the usual energy eigenstates (or Fock states) labeled by $|\Psi\rangle = |n, m\rangle$ with eigenvalue $E_{n,m} = n + m + 1$ and $n, m \in \mathbb{Z}^{\geq 0}$. These states may all be generated by the action of the raising and lowering operators in the following way [12]:

$$\begin{aligned} a_x^- \, |n, m\rangle &= \sqrt{n} \, |n-1, m\rangle \, , \;\; a_x^+ \, |n, m\rangle = \sqrt{n+1} \, |n+1, m\rangle \, ; \\ a_y^- \, |n, m\rangle &= \sqrt{m} \, |n, m-1\rangle \, , \;\; a_y^+ \, |n, m\rangle = \sqrt{m+1} \, |n, m+1\rangle \, . \end{aligned} \tag{7}$$

The states $|n, m\rangle$ in configuration space have the following wavefunction:

$$\langle x, y|n, m\rangle = \psi_n(x)\psi_m(y) = \frac{1}{\sqrt{2^{n+m}n!m!}}\sqrt{\frac{1}{\pi}}e^{-\frac{x^2}{2}-\frac{y^2}{2}}H_n(x)\,H_m(y)\,, \tag{8}$$

where $\psi_n(x) = \frac{1}{\sqrt{2^n n!}}\left(\frac{1}{\pi}\right)^{\frac{1}{4}} e^{-\frac{x^2}{2}} H_n(x)$ is the wavefunction of the 1D oscillator and $H_n(x)$ are the Hermite polynomials. For the physical position and momentum operators, $\hat{X}_i = \frac{1}{\sqrt{2}}(a_i^+ + a_i^-)$, $\hat{P}_i = \frac{1}{\sqrt{2}i}(a_i^- - a_i^+)$, respectively in the i direction, the states $|n, m\rangle$ have the following dispersions:

$$(\Delta\hat{X})^2_{|n,m\rangle} = (\Delta\hat{P}_x)^2_{|n,m\rangle} = \frac{1}{2} + n; \tag{9}$$

$$(\Delta\hat{Y})^2_{|n,m\rangle} = (\Delta\hat{P}_y)^2_{|n,m\rangle} = \frac{1}{2} + m. \tag{10}$$

They satisfy the Heisenberg uncertainty relation $(\Delta\hat{X})_{|n,m\rangle}(\Delta\hat{P}_x)_{|n,m\rangle} = \frac{1}{2} + n$ which grows linearly in n in the x direction. Similarly for the Y quadratures, we obtain $(\Delta\hat{Y})_{|n,m\rangle}(\Delta\hat{P}_y)_{|n,m\rangle} = \frac{1}{2} + m$.

In what follows we will construct two new ladder operators as linear combinations of the operators in (7) and proceed to define a single indexed Fock state for the 2D system which yields the $SU(2)$ coherent states. The new ladder operators and $SU(2)$ coherent states are used to extend the definitions of the 1D squeezed states in Sect. 2 to the 2D oscillator.

4 $SU(2)$ Coherent States

We use the ladder operators presented in Sect. 3 to construct a single set of creation and annihilation operators for the 2D oscillator. Introducing a set of states $\{|\nu\rangle\}$, and defining a new set of ladder operators through their action on the set,

$$A^-|\nu\rangle = \sqrt{\nu}\,|\nu-1\rangle\,,\ A^+|\nu\rangle = \sqrt{\nu+1}\,|\nu+1\rangle\,,\quad \langle\nu|\nu\rangle = 1,\quad \nu = 0, 1, 2, \ldots \tag{11}$$

These states have a linear increasing spectrum $E_\nu = \nu + 1$. We may build the states by hand starting with the only non-degenerate state, the ground state, $|0\rangle \equiv |0, 0\rangle$ and we take simple linear combinations of the 1D ladder operators

$$\begin{aligned} A^+_{\alpha,\beta} &= \alpha a_x^+ \otimes \mathbb{I}_y + \mathbb{I}_x \otimes \beta a_y^+; \\ A^-_{\alpha,\beta} &= \bar{\alpha} a_x^- \otimes \mathbb{I}_y + \mathbb{I}_x \otimes \bar{\beta} a_y^-; \\ [A^-_{\alpha,\beta}, A^+_{\alpha,\beta}] &= (|\alpha|^2 + |\beta|^2)\mathbb{I}_x \otimes \mathbb{I}_y \equiv \mathbb{I}, \end{aligned} \tag{12}$$

Table 1 Construction of the states $|\nu\rangle_{\alpha,\beta}$ using the relation $A^{+}_{\alpha,\beta}\,|\nu\rangle_{\alpha,\beta} = \sqrt{\nu+1}\,|\nu+1\rangle_{\alpha,\beta}$

$\|\nu\rangle$	$\|n, m\rangle$
$\|0\rangle$	$\|0, 0\rangle$
$\|1\rangle$	$\alpha\,\|1, 0\rangle + \beta\,\|0, 1\rangle$
$\|2\rangle$	$\alpha^2\,\|2, 0\rangle + \sqrt{2}\alpha\beta\,\|1, 1\rangle + \beta^2\,\|0, 2\rangle$
$\vdots$	$\vdots$
$\|\nu\rangle$	$\sum_{n,m}^{n+m=\nu} \alpha^n \beta^m \sqrt{\binom{\nu}{n}}\,\|n, m\rangle$

for $\alpha, \beta \in \mathbb{C}$, $\mathbb{I}_x \otimes \mathbb{I}_y = \mathbb{I}_y \otimes \mathbb{I}_x \equiv \mathbb{I}$ and normalization condition $|\alpha|^2 + |\beta|^2 = 1$. Constructing the states $\{|\nu\rangle\}$ starting with the ground state gives us Table 1.

The states, $|\nu\rangle$, in Table 1 depend on α, β and may be expressed as

$$|\nu\rangle_{\alpha,\beta} = \sum_{n=0}^{\nu} \alpha^n \beta^{\nu-n} \sqrt{\binom{\nu}{n}}\,|n, \nu - n\rangle\,. \tag{13}$$

The states $|\nu\rangle_{\alpha,\beta}$ are precisely the $SU(2)$ coherent states in the Schwinger boson representation [3]. This makes sense from our construction, the degeneracy present in the spectrum $E_{n,m}$ is an $SU(2)$ degeneracy, and so we created states which averaged out the degenerate contributions to a given ν. These states have the following orthogonality relations:

$${}_{\gamma,\delta}\langle\mu|\,|\nu\rangle_{\alpha,\beta} = (\bar{\gamma}\alpha + \bar{\delta}\beta)^{\nu}\delta_{\mu,\nu}, \tag{14}$$

which reduces to a more familiar relation when $\gamma = \alpha$ and $\delta = \beta$

$$\langle\mu|_{\alpha,\beta}\,|\nu\rangle_{\alpha,\beta} = \delta_{\mu,\nu}, \tag{15}$$

using the normalization condition $|\alpha|^2 + |\beta|^2 = 1$. The probability densities, $|\,\langle x, y|\nu\rangle_{\alpha,\beta}\,|^2$, of the quantum $SU(2)$ coherent states form ellipses when viewed as density plots, this mimics the classical 2D oscillator spatial distribution. This has been studied extensively by Chen [13] (Fig. 1).

The $SU(2)$ coherent states have the following variances for the physical position and momentum operators $\hat{X}_i = \frac{1}{\sqrt{2}}(a_i^{+} + a_i^{-})$, $\hat{P}_i = \frac{1}{\sqrt{2}i}(a_i^{-} - a_i^{+})$, respectively, in the i direction:

$$(\Delta\hat{X})^2_{|\nu\rangle_{\alpha,\beta}} = (\Delta\hat{P}_x)^2_{|\nu\rangle_{\alpha,\beta}} = \frac{1}{2} + |\alpha|^2\nu; \tag{16}$$

$$(\Delta\hat{Y})^2_{|\nu\rangle_{\alpha,\beta}} = (\Delta\hat{P}_y)^2_{|\nu\rangle_{\alpha,\beta}} = \frac{1}{2} + |\beta|^2\nu. \tag{17}$$

The results are essentially the same as those in (9) and (10), but they are tuned by the continuous parameters α, β introduced in (12).

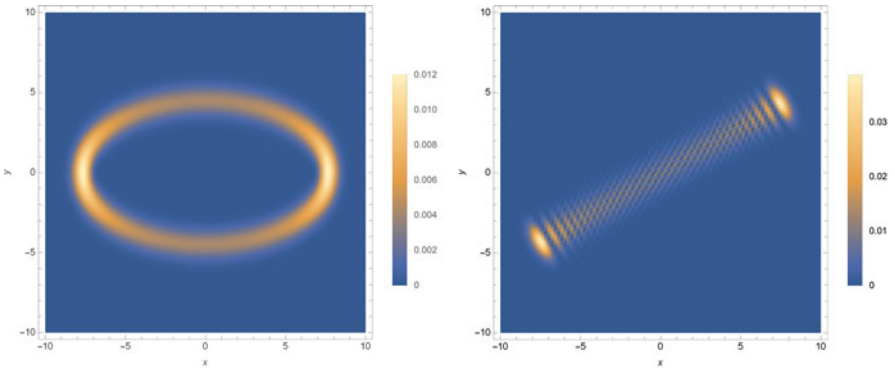


Fig. 1 Density plots of $|\langle x, y|\nu\rangle_{\alpha,\beta}|^2$ for $\alpha = \frac{\sqrt{3}}{2}e^{i\frac{\pi}{2}}$, $\beta = \frac{1}{2}$ (left) and $\alpha = \frac{\sqrt{3}}{2}$, $\beta = \frac{1}{2}$ (right) both at $\nu = 40$

5 2D Squeezed States

By analogy with the 1D case we define a 2D displacement and 2D squeezing operators

$$D(\Psi) = e^{\Psi A^+_{\alpha,\beta} - \bar{\Psi} A^-_{\alpha,\beta}}, \tag{18}$$

and

$$S(\Xi) = \exp\left(\frac{1}{2}[\Xi {A^+_{\alpha,\beta}}^2 - \bar{\Xi} {A^-_{\alpha,\beta}}^2]\right), \tag{19}$$

respectively. The generalized squeezed state is obtained through the action of the two operators on the 2D vacuum

$$|\Psi, \Xi\rangle_{\alpha,\beta} = D(\Psi)S(\Xi)\,|0\rangle_{\alpha,\beta}\,. \tag{20}$$

Using the expansion of the 1D squeezed states, we replace the basis $|n\rangle \to |\nu\rangle_{\alpha,\beta}$ and use capital lettered parameters (to indicate they are 2D states) to get the following:

$$|Z, \Gamma\rangle_{\alpha,\beta} = \frac{1}{\sqrt{\cosh R}} e^{-\frac{|Z|^2}{2}} e^{\frac{\tanh R}{2}\mathrm{Re}(e^{i\Theta}\bar{Z}^2)} \sum_{\nu=0}^{\infty} \frac{1}{\sqrt{\nu!}} \left(\frac{\Gamma}{2}\right)^{\frac{\nu}{2}} H_\nu\left(\frac{Z}{\sqrt{2\Gamma}}\right) |\nu\rangle_{\alpha,\beta}\,, \tag{21}$$

with $Z = \Psi - \bar{\Psi}e^{i\Theta}\tanh R$, $\Gamma = -e^{i\Theta}\tanh R$. In Fig. 2 we see the effect of increasing the strength of the squeezing, on the leftmost plot the squeezing is relatively small, $R = 0.1$ and the probability density is converging to a single maximum. This is in agreement with the limit $R \to 0$ which would produce a

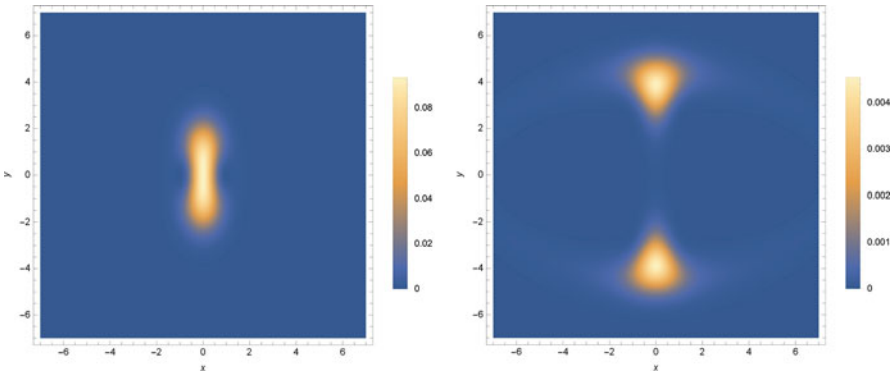


Fig. 2 Density plots of $|\langle x, y|Z, \Xi\rangle_{\alpha,\beta}|^2$ for $\alpha = \frac{\sqrt{3}}{2}e^{i\frac{\pi}{2}}$, $\beta = \frac{1}{2}$, $Z = 1$, $R = 0.1$, $\Theta = 0$ (left) and $\alpha = \frac{\sqrt{3}}{2}$, $\beta = \frac{1}{2}$, $Z = 1$, $R = 10$, $\Theta = 0$ (right) both with 20 terms kept in the expansion of Eq. (21)

Gaussian distribution with single maximum [1]. On the other hand, the rightmost plot, $R = 10$, reveals a separation of the probability density onto two distinct maxima. It is important to note that the graphs are not properly normalized as a truncated sum (20 terms) was used in the computation.

Restricting to the case of the 2D squeezed vacuum, $\Psi = 0$, the squeezing operator admits an $su(1, 1)$ decomposition [14] yielding

$$|\Xi\rangle_{\alpha,\beta} = \frac{1}{\sqrt{\cosh R}} \exp\left\{\frac{e^{i\Theta}}{2} \tanh R(\alpha^2 a_x^{+2} + \beta^2 a_y^{+2} + \alpha\beta a_x^+ a_y^+)\right\} |0, 0\rangle \tag{22}$$

in terms of the 1D ladder operators. Equation (22) does not factorize, $|\Xi\rangle \neq |\xi_x\rangle_x \otimes |\xi_y\rangle_y$; the bilinear 1D terms in the expansion of ${A_{\alpha,\beta}^+}^2$ have induced a coupling between the x and y modes of the oscillator. This represents a non-trivial generalization of the squeezed states to 2D, a two-mode-like squeezing was generated as a result of the construction, but the 2D squeezed states themselves retain most of the definitions of their 1D counterparts.

To calculate the dispersions in x and y we use the Baker–Campbell–Hausdorff identity $e^A B e^{-A} = B + [A, B] + \frac{1}{2}[A, [A, B]] + \ldots$ [15] to compute Bogoliubov transformations, for example, the x ladder operators are transformed as

$$\begin{aligned} S^\dagger(\Xi) a_x^- S(\Xi) = (|\beta|^2 + |\alpha|^2 \cosh R) a_x^- + \alpha\bar{\beta}(\cosh R - 1) a_y^- \\ + e^{i\Theta} \sinh R(\alpha^2 a_x^+ + \alpha\beta a_y^+); \end{aligned} \tag{23}$$

$$\begin{aligned} S^\dagger(\Xi) a_x^+ S(\Xi) = (|\beta|^2 + |\alpha|^2 \cosh R) a_x^+ + \bar{\alpha}\beta(\cosh R - 1) a_y^+ \\ + e^{-i\Theta} \sinh R(\bar{\alpha}^2 a_x^- + \bar{\alpha}\beta a_y^-). \end{aligned} \tag{24}$$

Using these transformations we can compute the dispersions in x

$$
\begin{aligned}
(\Delta \hat{X})^2_{|\Xi\rangle_{\alpha,\beta}} &= \frac{1}{2} + |\alpha|^2 \sinh^2 R + \mathrm{Re}(e^{i\Theta}\alpha^2) \sinh R \cosh R; \\
(\Delta \hat{P}_x)^2_{|\Xi\rangle_{\alpha,\beta}} &= \frac{1}{2} + |\alpha|^2 \sinh^2 R - \mathrm{Re}(e^{i\Theta}\alpha^2) \sinh R \cosh R,
\end{aligned}
\tag{25}
$$

and similarly for y

$$
\begin{aligned}
(\Delta \hat{Y})^2_{|\Xi\rangle_{\alpha,\beta}} &= \frac{1}{2} + |\beta|^2 \sinh^2 R + \mathrm{Re}(e^{i\Theta}\beta^2) \sinh R \cosh R; \\
(\Delta \hat{P}_y)^2_{|\Xi\rangle_{\alpha,\beta}} &= \frac{1}{2} + |\beta|^2 \sinh^2 R - \mathrm{Re}(e^{i\Theta}\beta^2) \sinh R \cosh R.
\end{aligned}
\tag{26}
$$

These results also hold for the generalized squeezed states $|\Psi, \Xi\rangle_{\alpha,\beta}$ because the action of the displacement operator has no effect on the variances. The results resemble those in Eq. (5) but are modified by α, β. We see in the limit $R \to 0$ we saturate the Heisenberg uncertainty relation in both x and y.

6 Conclusion

In this paper we have described a method for constructing squeezed states for the 2D isotropic oscillator which relies on using the minimal set of definitions used to describe the squeezed states of the 1D oscillator. Unlike the coherent states defined in a similar manner in [1], the generalized squeezed states did not factorize into the product of squeezed states on x and y independently. A coupling was induced which took the form of a two-mode like squeezing creating an entanglement between the two modes.

We found the dispersions for the 2D squeezed states and saw that they resemble the dispersions in the 1D case but modified by the parameters α, β introduced during the construction of the $SU(2)$ coherent states. As well we saw a separation of the spatial probability densities into two distinct maxima for larger values of the squeezing R.

Finally, perhaps this method can be used to construct squeezed states for more general degenerate and higher dimensional systems and oscillators. The approach presented in this paper will require modification on a case by case basis because in general a multidimensional system will admit a more complex degenerate structure, which would significantly modify the generalized ladder operators as well as the non-degenerate basis $\{|\nu\rangle\}$. If a system possesses non-algebraic degeneracies, such as the 2D particle in a box (e.g., $1^2 + 7^2 = 5^2 + 5^2$), a new method for counting states contributing to a degenerate subgroup $|\nu\rangle$ would be required.

References

1. J. Moran, V. Hussin, Coherent states for the isotropic and anisotropic 2d harmonic oscillators. Quantum Rep. **1**(2), 260–270 (2019)
2. J.R. Klauder, Coherent states for the hydrogen atom. J. Phys. A Math. Gen. **29**(12), L293–L298 (1996)
3. J.-P. Gazeau, *Coherent States in Quantum Physics* (Wiley-VCH, Berlin, 2009)
4. R.F. Fox, M.H. Choi, Generalized coherent states for systems with degenerate energy spectra. Phys. Rev. A **64**, 042104 (2001)
5. L. Dello Sbarba, V. Hussin, Degenerate discrete energy spectra and associated coherent states. J. Math. Phys. **48**(1), 012110 (2007)
6. E. Schrödinger, Der stetige Übergang von der Mikro- zur Makromechanik. Naturwissenschaften **14**, 664–666 (1926)
7. R.J. Glauber, The quantum theory of optical coherence. Phys. Rev. **130**, 2529–2539 (1963)
8. E.C.G. Sudarshan, Equivalence of semiclassical and quantum mechanical descriptions of statistical light beams. Phys. Rev. Lett. **10**, 277–279 (1963)
9. C. Gerry, P. Knight, *Introductory Quantum Optics* (Cambridge University Press, Cambridge, 2004)
10. H.P. Robertson, The uncertainty principle. Phys. Rev. **34**, 163–164 (1929)
11. S.T. Ali, J.-P. Antoine, J.-P. Gazeau, *Coherent States, Wavelets and Their Generalizations* (Springer Publishing Company, New York, 2012)
12. P.A.M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1930)
13. Y.F. Chen, K.F. Huang, Vortex structure of quantum eigenstates and classical periodic orbits in two-dimensional harmonic oscillators. J. Phys. A Math. Gen. **36**(28), 7751–7760 (2003)
14. R.A. Fisher, M.M. Nieto, V.D. Sandberg, Impossibility of naively generalizing squeezed coherent states. Phys. Rev. D **29**, 1107–1110 (1984)
15. D.J. Griffiths, D.F. Schroeter, *Introduction to Quantum Mechanics*, 3rd edn. (Cambridge University Press, Cambridge, 2018)

Electron in Bilayer Graphene with Magnetic Fields Associated with Solvable Potentials

Daniel O. Campa, Juan D. García, and David J. Fernández

Abstract The behavior of an electron in bilayer graphene with applied magnetic fields orthogonal to the surface is studied. By using second-order supersymmetric quantum mechanics some analytic solutions are found, which are expressed in terms of the eigenfunctions of two intertwined Schrödinger Hamiltonians. The case of a constant homogeneous magnetic field which leads to a pair of shifted harmonic oscillator potentials is discussed. A variant of this example produces inhomogeneous magnetic fields whose exact solutions for the electron motion in bilayer graphene are straightforwardly identified.

Keywords Bilayer graphene · SUSY QM · Solvable potentials

1 Introduction

Graphene is one of the so-called Dirac materials in the literature, which nowadays are of great interest in physics. In particular, the analysis of its electronic properties has become an extensive field of study. Here, we will focus in the electron motion in bilayer graphene. The effective Hamiltonian describing the electron behavior in either monolayer or bilayer graphene is well known [1–3]. For low energy electrons in the monolayer a tight-binding model leads to a massless Dirac equation, where the Fermi velocity replaces the speed of light. Nevertheless, to find exact analytic solutions to this equation is not always simple. The first-order supersymmetric quantum mechanics (SUSY QM) has proven useful for addressing this task [4–6]. However, for bilayer graphene a similar model does not end up with a massless Dirac equation, but with slightly more complicated equations involving second-order derivatives. This suggests to use now the second-order SUSY QM, where the intertwining operators are of second order. Thus, in this paper we will use

D. O. Campa (✉) · J. D. García · D. J. Fernández
Physics Department, Cinvestav, Mexico City, Mexico
e-mail: dortiz@fis.cinvestav.mx; dgarcia@fis.cinvestav.mx; david@fis.cinvestav.mx

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_25

the second-order SUSY QM to study the electron motion in bilayer graphene, by looking for exact solutions to the auxiliary Schrödinger problems.

2 Bilayer Graphene: A Brief Overview

2.1 Effective Hamiltonian

In the study of the electron motion in bilayer graphene, it is usual to work in the framework of the tight-binding model [1], in which the effective Hamiltonian is given by

$$H = \frac{1}{2m^*} \begin{pmatrix} 0 & \pi^2 \\ (\pi^\dagger)^2 & 0 \end{pmatrix}, \tag{1}$$

where $\pi = p_x - ip_y$, with p_j being the momentum operator in j-direction, $j = x, y$. Some important physical quantities are the effective mass of the electron $m^* = |t_\perp|/2{v_F}^2 \approx 0.054 m_e$, with m_e being the free mass electron, the Fermi velocity $v_F = \sqrt{3}a/2\hbar \approx c/300$ where c is the speed of light [2]. Several structure parameters of the graphene lattice appear as well in the Hamiltonian, e.g., the lattice constant $a = 2.46\,\text{Å}$ and the hopping parameter $t_\perp = 0.381\,\text{eV}$. In Fig. 1 it is shown the lattice structure of bilayer graphene. Each layer is divided into two sublattices A and B, with the two layers placed in positions such that the sublattices A are aligned to each other. This configuration is the most natural in graphene, and it is called *Bernal staking*. It can be noticed that the hopping parameters are labeled like γ_i; however, it is usual in the literature to denote $\gamma_0 = t$ for the in-plane hopping

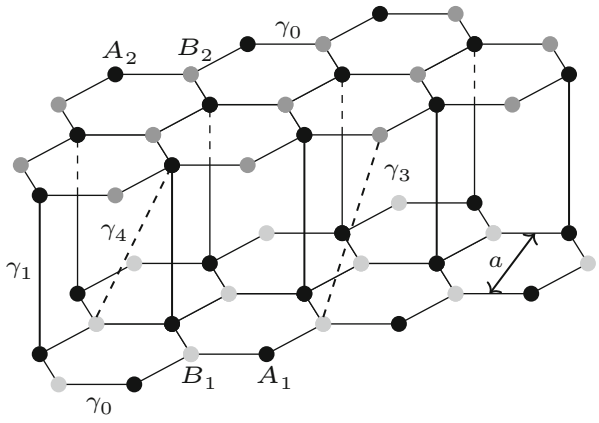


Fig. 1 Lattice structure of bilayer graphene. It is shown also the lattice constant a and the hopping parameters γ_i, $i = 0, 1, 3, 4$

parameter and $\gamma_1 = t_\perp$ for the hopping from an atom in the sublattice $A1$ to an atom in the sublattice $A2$.

2.2 Effective Hamiltonian with an Applied Magnetic Field

Let us suppose now that a magnetic field $\vec{B}$ orthogonal to the graphene surfaces is applied, which points along z-direction. For simplicity, we suppose that $\vec{B}$ just changes along a certain given direction on the surface. Working in the Landau gauge we can write $\vec{A}(x) = A(x)\hat{y}$, $\vec{B}(x) = B(x)\hat{z}$, with $B(x) = dA(x)/dx$. According with the minimal coupling rule, a term proportional to $A(x)$ must be added to the operator π [4]. Thus, we get the new operator $\Pi = p_x - ip_y - i(e/c)A(x)$, while the effective Hamiltonian which takes into account the magnetic field effects looks like

$$H = \frac{1}{2m^*}\begin{pmatrix} 0 & \Pi^2 \\ (\Pi^\dagger)^2 & 0 \end{pmatrix}. \tag{2}$$

The next step is to determine the eigenvectors and eigenvalues of H, which is a non-trivial but sometimes solvable problem.

2.3 Stationary States of H

The eigenvalue equation for H can be written as

$$H\Psi(x, y) = E\Psi(x, y). \tag{3}$$

Due to the gauge chosen, this equation is invariant under translations along y-direction, i.e., in this direction the motion is of free particle type. Thus, it is natural to propose $\Psi(x, y)$ as follows:

$$\Psi(x, y) = \frac{e^{iky}}{\sqrt{2}}\begin{pmatrix} \psi^{(2)}(x) \\ \psi^{(0)}(x) \end{pmatrix}, \tag{4}$$

where k is the wavenumber in y-direction and $1/\sqrt{2}$ is a normalization factor. Since $p_j = -i\hbar\partial_j$, $j = x, y$, after plugging equation (4) into (3) the next coupled system of equations is obtained:

$$L_2\psi^{(0)} = -\varepsilon\psi^{(2)},\ L_2^\dagger\psi^{(2)} = -\varepsilon\psi^{(0)}. \tag{5}$$

Note that $\varepsilon = 2m^* E/\hbar^2$ and L_2 is the second-order differential operator

$$L_2 = \frac{d^2}{dx^2} + \eta(x)\frac{d}{dx} + \gamma(x). \tag{6}$$

Let us remark that $\eta(x)$ is given by

$$\eta(x) = 2\left(k + \frac{e}{c\hbar}A(x)\right), \tag{7}$$

thus it is directly related with $B(x)$ as follows:

$$B(x) = \frac{c\hbar}{2e}\frac{d\eta(x)}{dx}. \tag{8}$$

For the time being, the form of the function $\gamma(x)$ is not relevant here. We will see later on that it can be expressed in terms of $\eta(x)$.

The coupled system of Eqs. (5) and the form of the operator L_2 in Eq. (6) suggest to use the second-order SUSY QM to address the problem.

3 Second-Order SUSY QM

Let us consider two Schrödinger Hamiltonians

$$H_j = -\frac{d^2}{dx^2} + V_j(x), \;\; j = 0, 2. \tag{9}$$

They are called second-order SUSY partners if there exists a second-order operator L_2 intertwining them

$$H_2 L_2 = L_2 H_0, \tag{10}$$

with L_2 being given by Eq. (6). Equation (10) produces a set of relations among $\eta(x)$, $\gamma(x)$, $V_0(x)$, $V_2(x)$, which after some work leads to

$$V_2(x) = V_0(x) + 2\eta'(x), \tag{11}$$

$$\gamma(x) = \frac{\eta^2(x)}{2} - \frac{\eta'(x)}{2} - V_0(x) + \frac{\epsilon_1 + \epsilon_2}{2}, \tag{12}$$

$$V_0(x) = \frac{\eta''(x)}{2\eta(x)} - \frac{(\eta'(x))^2}{4\eta^2(x)} - \eta'(x) + \frac{\eta^2(x)}{4} + \frac{\epsilon_1 + \epsilon_2}{2} + \left(\frac{\epsilon_1 - \epsilon_2}{2\eta(x)}\right)^2, \tag{13}$$

where $\eta'(x) \equiv d\eta(x)/dx$ and so on. The factorization energies ϵ_1 and ϵ_2 are in general complex [7]. However, in this paper we will work just with real values, specifically with two consecutive eigenvalues of H_0, i.e., $\epsilon_2 = E_j^{(0)}$, $\epsilon_1 = E_{j+1}^{(0)}$. Although the spectra of H_2 and H_0 are almost the same, H_0 has two extra energy levels at ϵ_1 and ϵ_2. Moreover, there are some relations between the eigenfunctions of H_0 and H_2, which can be derived from the next equations

$$L_2 L_2^\dagger \psi_n^{(2)} = \left(E_n^{(2)} - E_j^{(0)}\right)\left(E_n^{(2)} - E_{j+1}^{(0)}\right)\psi_n^{(2)}, \tag{14}$$

$$L_2^\dagger L_2 \psi_n^{(0)} = \left(E_n^{(0)} - E_j^{(0)}\right)\left(E_n^{(0)} - E_{j+1}^{(0)}\right)\psi_n^{(0)}, \tag{15}$$

where $\psi_n^{(l)}$ and $E_n^{(l)}$ are the corresponding eigenfunctions and eigenvalues of H_l, $l = 0, 2$, respectively. Another important fact is that the two eigenfunctions $\psi_j^{(0)}$, $\psi_{j+1}^{(0)}$ of H_0, which are also in the *kernel* of L_2, $L_2\psi_j^{(0)}(x) = L_2\psi_{j+1}^{(0)} = 0$, are required to determine the intertwining transformation of Eq. (10). In particular, the function $\eta(x)$ can be determined from $\psi_j^{(0)}$, $\psi_{j+1}^{(0)}$ through the relation

$$\eta(x) = -\frac{W'\left(\psi_j^{(0)}, \psi_{j+1}^{(0)}\right)}{W\left(\psi_j^{(0)}, \psi_{j+1}^{(0)}\right)}, \tag{16}$$

with $W(f, g) = fg' - f'g$ being the Wronskian of f and g.

Now we can solve our original problem posed in the system of Eqs. (5). First of all, according to (15) the eigenvalues ε and thus the energy levels E for electrons (positive energies) are given by

$$E_{\widetilde{n}} = \frac{\hbar^2}{2m^*}\sqrt{\left(E_n^{(0)} - E_j^{(0)}\right)\left(E_n^{(0)} - E_{j+1}^{(0)}\right)}. \tag{17}$$

The index of E is denoted $\widetilde{n}$ rather that n, since the ordering of the energy levels of H is non-standard although the orderings of the eigenvalues of the auxiliary Hamiltonians H_l are standard. In addition, the ground state zero energy has always double degeneracy, due to the choice of the factorization energies as two consecutive energy levels of H_0. According with Eq. (4), the eigenvectors $\Psi(x, y)$ can be written as

$$\Psi_{\widetilde{n}}(x, y) = \begin{cases} e^{iky}\begin{pmatrix} 0 \\ \psi_n^{(0)}(x) \end{pmatrix} & \text{for } n = j, j+1, \\ \frac{e^{iky}}{\sqrt{2}}\begin{pmatrix} \psi_n^{(2)}(x) \\ \psi_n^{(0)}(x) \end{pmatrix} & \text{for } n \neq j, j+1. \end{cases} \tag{18}$$

4 Solvable Cases

4.1 Case with j=0

Let us consider now a constant magnetic field $\vec{B} = (0, 0, B_0)$. In the Landau gauge the corresponding vector potential becomes $\vec{A} = (0, xB_0, 0)$. Equation (7) implies that $\eta(x) = 2k + \omega x$, where $\omega = 2eB_0/c\hbar$. By using Eqs. (11)–(13) the SUSY partner potentials become

$$V_0(x) = \frac{\omega^2}{4}\left(x + \frac{2k}{\omega}\right)^2 - \frac{\omega}{2}, \tag{19}$$

$$V_2(x) = \frac{\omega^2}{4}\left(x + \frac{2k}{\omega}\right)^2 + \frac{3}{2}\omega. \tag{20}$$

It is seen that these potentials are two shifted harmonic oscillators differing from each other by a constant. In the language of SUSY QM it is said that they are shape invariant potentials, in the sense that one of them can be obtained from the other by changing just some parameters and, perhaps, displacing the energy by a global quantity, see Fig. 2.

The energy levels of $V_0(x)$ and $V_2(x)$ are

$$E_0^{(0)} = 0, \quad E_1^{(0)} = \omega, \quad E_n^{(0)} = E_n^{(2)} = n\omega, \; n = 2, 3, 4, \ldots, \tag{21}$$

i.e., $Sp(H_0) = \{E_0^{(0)}, E_1^{(0)}, Sp(H_2)\}$. The factorization energies were taken in this case as $\epsilon_1 = E_1^{(0)}$ and $\epsilon_2 = E_0^{(0)}$. The eigenfunctions of H_0 and H_2 can be expressed in terms of Hermite polynomials

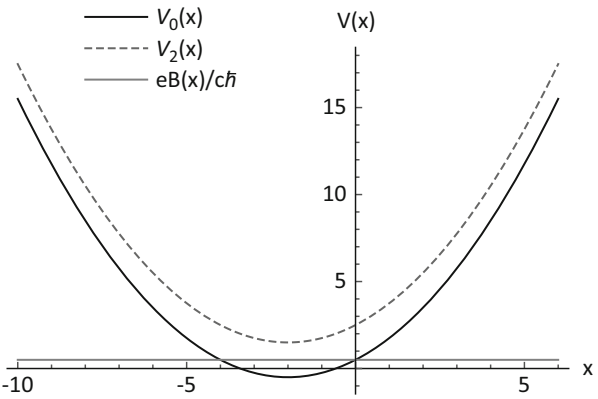


Fig. 2 Second-order SUSY partner potentials $V_0(x)$, $V_2(x)$ as functions of x and the constant magnetic field (scaled) inducing them

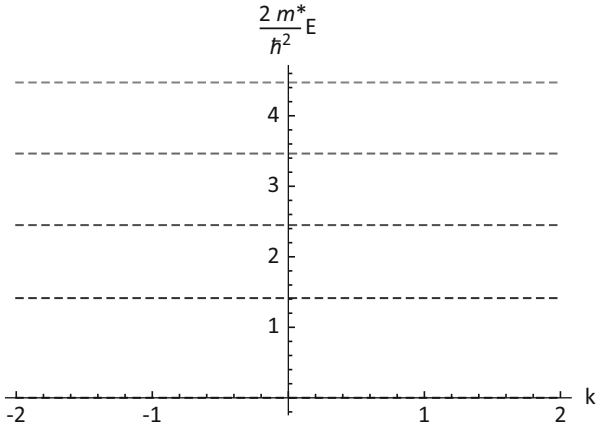


Fig. 3 The eigenvalues $E_{\tilde{n}}$ as functions of k for the constant magnetic field

$$\psi_m^{(0)}(z) = \psi_{m+2}^{(2)}(z) = e^{-\frac{1}{2}z^2} H_m(z), \ m = 0, 1, 2, \ldots, \tag{22}$$

where $z = \sqrt{\omega/2}(x - 2k/\omega)$. Thus, the eigenvectors $\Psi_{\tilde{n}}(x, y)$ looks like

$$\Psi_{\tilde{n}}(x, y) = \begin{cases} e^{iky} \begin{pmatrix} 0 \\ \psi_n^{(0)}(x) \end{pmatrix} & \text{for } n = 0, 1, \\ \frac{e^{iky}}{\sqrt{2}} \begin{pmatrix} \psi_n^{(2)}(x) \\ \psi_n^{(0)}(x) \end{pmatrix} & \text{for } n = 2, 3, 4, \ldots \end{cases} \tag{23}$$

Finally, the eigenvalues of H take the form

$$E_{\tilde{n}} = \frac{\hbar^2 \omega}{2m^*} \sqrt{n(n-1)}, \ n = 0, 1, 2, 3, \ldots \tag{24}$$

Note once again that $E_{\tilde{0}} = E_{\tilde{1}} = 0$, i.e., the ground state energy is double degenerate. Moreover, the previous eigenvalues do not depend of the wavenumber k, as it is shown in Fig. 3.

4.2 Case with j≠0

In this case we take the $V_0(x)$ of Eq. (19) as the initial potential, whose eigenvalues and eigenfunctions are given in Eqs. (21) and (22). The factorization energies are taken as $\epsilon_1 = E_{j+1}^{(0)}$ and $\epsilon_2 = E_j^{(0)}$ with $j > 0$, and the function $\eta(x)$ is calculated

from Eq. (16). The previous choice implies that the new SUSY partner potentials $\widetilde{V}_2(x)$ and $V_0(x)$ are no longer shape invariant, i.e., the new potential $\widetilde{V}_2(x)$ obtained from Eq. (11) is not just a shifted and/or displaced harmonic oscillator potential (see Fig. 4). Moreover, the new magnetic field $\widetilde{B}(x)$ generated through this technique is obtained from Eq. (8), and now it is not homogeneous as in the case with $j = 0$ (see Fig. 5). The eigenvalues $E_{\widetilde{n}}$ of H can be written as

$$E_{\widetilde{n}} = \frac{\hbar^2\omega}{2m^*}\sqrt{(n-j)(n-j-1)}. \tag{25}$$

In order that they have a standard ordering, the index $\widetilde{n}$ as a function of n should be expressed as

$$\widetilde{n} = \begin{cases} j-n & \text{for } n = 0, \ldots, j \\ n-(j+1) & \text{for } n = j+1, j+2, \ldots \end{cases} \tag{26}$$

We can see that our system has now $(j+1)$ double degenerate energy levels. The eigenfunctions $\psi_n^{(2)}(x)$ of $V_2(x)$ are obtained by using Eq. (5). Due to the double degeneracy of $\{E_{\widetilde{0}}, \ldots, E_{\widetilde{j}}\}$ the eigenvectors of H turn out to be given by Eq. (18).

Plots of the SUSY partner potentials $V_0(x)$, $\widetilde{V}_2(x)$, and the $V_2(x)$ of the previous section are shown in Fig. 4 for $\omega = k = 1$ and $j = 3$. The generated magnetic field, as compared with the constant case of the previous section, is drawn in Fig. 5.

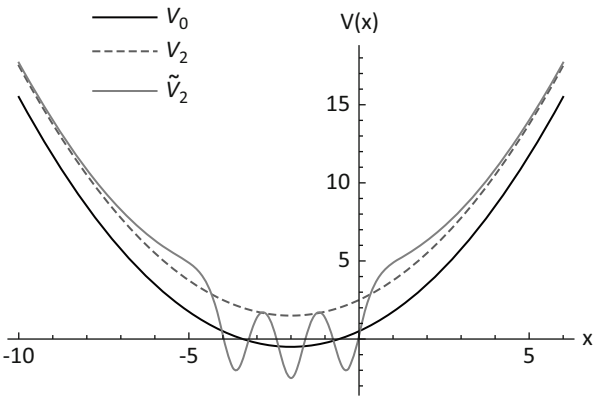


Fig. 4 Second-order SUSY partner potentials $V_0(x)$, $\widetilde{V}_2(x)$ and $V_2(x)$ as functions of x for $\omega = k = 1$ and $j = 3$

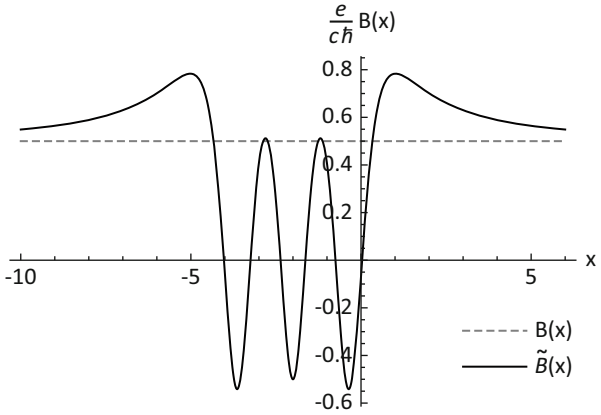


Fig. 5 Magnetic fields $\widetilde{B}(x)$ and $B(x) = B_0$ as functions of x for $\omega = k = 1$ and $j = 3$

5 Conclusions

The second-order SUSY QM is the natural technique to solve the electron motion in bilayer graphene with applied magnetic fields. When a constant magnetic field is chosen, with the factorization energies being taken as the first two energy levels of $V_0(x)$, a shape invariant SUSY partner potential $V_2(x)$ is obtained, and the effective Hamiltonian has a minimal number of eigenvalues with double degeneracy (just the ground state). In this case $V_0(x)$ and $V_2(x)$ are simple shifted harmonic oscillators and the associated magnetic field is constant. On the other hand, if the factorization energies are taken as two consecutive energy levels of $V_0(x)$, without including $E_0^{(0)}$, the new SUSY partner potential $\widetilde{V}_2(x)$ is no longer shape invariant (it leaves to be just a shifted harmonic oscillator), the number of eigenvalues $(j + 1)$ with double degeneracy grows up, and the associated magnetic field $\widetilde{B}(x)$ is not constant anymore.

Acknowledgments Daniel Ortiz Campa and Juan Domingo García acknowledge the support of CONACYT.

References

1. M.I. Katnelson, *Graphene Carbon in Two Dimensions* (Cambridge University Press, Cambridge, 2012)
2. E. McCann, M. Koshino, Rep. Prog. Phys. **76**, 056503 (2013)
3. A. Ferreira, J. Viana-Gomes, J. Nilsson, E.R. Mucciolo, N.M.R. Peres, A.H. Castro-Neto, Phys. Rev. B **83**, 165402 (2011)
4. S. Kuru, J. Negro, L.M. Nieto, J. Phys. Condens. Matter **21**, 455305 (2009)
5. B. Midya, D.J. Fernández, J. Phys. A Math. Theor. **47**, 285302 (2014)
6. M. Castillo-Celeita, D.J. Fernández, J. Phys. A Math. Theor. **53**, 035302 (2020)
7. D.J. Fernández, N. Fernández-García, AIP Conf. Proc. **744**, 236–273 (2005)

Twist Knot Invariants and Volume Conjecture

P. Ramadevi and Zodinmawia

Abstract Chern–Simons theory provides a natural framework to construct a variety of knot invariants. The calculation of colored HOMFLY-PT polynomials of knots using $SU(N)$ Chern–Simons theory requires the knowledge of $6j$-symbols for the quantum group $U_q(\mathfrak{sl}_N)$ which are not known for arbitrary representation. Interestingly, our conjectured formula for superpolynomials (categorification of colored HOMFLY-PT polynomials) of twist knots led to deducing closed form expression for these symbols for a class of multiplicity-free $U_q(\mathfrak{sl}_N)$ representation. Using the twist knot superpolynomials, we compute the classical and quantum super-A-polynomials and test the categorified version of the quantum volume conjecture.

Keywords Chern–Simons theory · Knot invariants · Colored HOMFLY-PT polynomials · Quantum $6j$-symbols · Superpolynomials · Volume conjectures · Super-A-polynomials

1 Introduction

The pioneering work of Witten [1] showed that Chern–Simons theory on a three-sphere S^3 naturally describes knots and their invariants. The Chern–Simons action $CS(S^3)$ based on $SU(N)$ gauge group is

$$CS(S^3) = k \int_{S^3} \mathrm{Tr}\left(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\right), \tag{1}$$

P. Ramadevi (✉)
Department of Physics, Indian Institute of Technology Bombay, Mumbai, India
e-mail: ramadevi@phy.iitb.ac.in

Zodinmawia
Department of Physics, Mizoram University, Aizawl, India

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_26

where k is the coupling constant and A (gauge field) are the $\mathfrak{su}(\mathfrak{n})$ Lie algebra-valued one forms: $A = \sum_{a,\mu} A^a_\mu t^a dx^\mu$ with t^a denoting the generators of the gauge group. For a knot K in S^3, the corresponding Wilson loop operator (WLO) is obtained by taking the holonomy of the gauge fields along the knot. More precisely, if T^a_R are the generators in a representation $R \in SU(N)$, then the operator colored by R is given by

$$W_R[K] = \mathrm{Tr}_R \exp\left(\oint_K A\right).$$

The expectation value of WLO, $\langle W_R[K]\rangle$, are *the knot invariants*. In fact, the $\langle W_{R=\square}[K]\rangle$ is proportional to the Jones polynomial, $J(K;q)$, for gauge group $SU(2)$ and HOMFLY-PT polynomial, $P(K;a,q)$, for $SU(N)$ gauge group. The polynomial variables are related to coupling constant and rank of the group as follows: $a = q^N$ and $q = \exp\{[2\pi i/(k+N)]\}$ [1]. For higher dimensional representation R, $\langle W_R[K]\rangle$ define the colored Jones $J_R(K;q)$ ($R \in SU(2)$) and the colored HOMFLY-PT $P_R(K;a,q)$ ($R \in SU(N)$) polynomials. Collectively, the knot invariants obtained from Chern–Simons theory are referred to as quantum invariants of knots. In this note we focus on the colored HOMFLY-PT polynomials.

An intriguing property about quantum invariants of knots is the integrality structure. We observe that the Jones polynomial for any knot has a Laurent series expansion $J(K;q) = \sum_i c_i q^i$, where $\{c_i\}$'s are integers. Other quantum invariants are also polynomials with integer coefficients. The quest to give a topological answer to such an integrality property of Jones polynomials led Khovanov to discover *knot homology* [2]. Particularly, Khovanov constructed a bigraded chain complex (naturally associated with a planar diagram of a knot K) whose homology $\mathcal{H}^{\mathfrak{sl}_2}_{i,j}(K)$ is invariant under the Reidemeister moves. Hence $\mathcal{H}^{\mathfrak{sl}_2}_{i,j}(K)$ is an invariant of the knot. More importantly, the Euler characteristics of this bigraded homology are the Jones polynomial: $J(K;q) = \sum_{i,j}(-1)^i q^j \mathrm{dim}\mathcal{H}^{\mathfrak{sl}_2}_{i,j}(K)$. Clearly, $\mathrm{dim}\mathcal{H}^{\mathfrak{sl}_2}_{i,j}(K)$ must necessarily be an integer which provides a topological meaning to the integers appearing in the Jones polynomial. Furthermore, a new, two-variable invariant polynomial called Khovanov polynomial, $Kh(K;q,t)$, can be constructed by taking the graded Poincaré polynomial,

$$Kh(K;q,t) = \sum_{i,j} t^i q^j \mathrm{dim}\mathcal{H}^{\mathfrak{sl}_2}_{i,j}(K). \tag{2}$$

Note that Jones polynomial is the $t = -1$ limit of Khovanov polynomial. In this sense, Khovanov polynomial is the lift, or categorification, of the Jones polynomial.

The categorification of Jones polynomial by Khovanov led to the study of homology theory for other quantum invariants. For the case of colored HOMFLY-PT polynomial, a triply graded colored HOMFLY homology, $(\mathcal{H}^{\mathrm{HOMFLY}}_R(K))_{i,j,k}$, was proposed in refs. [3–7] such that the graded Euler characteristic gives the colored HOMFLY polynomial:

$$P_R(K;a,q)=\sum_{i,j,k}a^i q^j(-1)^k\dim\Big(\mathcal{H}_R^{\mathrm{HOMFLY}}(K)\Big)_{i,j,k}. \tag{3}$$

$(\mathcal{H}_R^{\mathrm{HOMFLY}}(K))_{i,j,k}$ also has the physical realization as spaces of BPS states [8]. All the information about $(\mathcal{H}_R^{\mathrm{HOMFLY}}(K))_{i,j,k}$ can be encoded in its graded Poincaré polynomial $\mathcal{P}_R(K;a,q,t)$ which are called superpolynomial.

Quantum knot invariants have deep connections to the three-dimensional geometry in which they are embedded. The first of this relation is the *volume conjecture* proposed by Kashaev [9] and later reinterpreted by Murakami [10]. This conjecture relates the large color behavior of Jones polynomial to the hyperbolic volume of the complement of the knot in S^3 ($S^3\backslash K$):

$$\lim_{n\to\infty}\frac{2\pi}{n}\log\Big|J_n\Big(K;q=e^{\frac{2\pi i}{n}}\Big)\Big|=\mathrm{Vol}\big(S^3\backslash K\big)\,.$$

Here and afterward we use n to denote the $(n-1)$-th rank symmetric representation ($n\equiv\underbrace{\square\cdots\square}_{n-1}$).

The volume conjecture is further generalized by incorporating another knot invariant, known as the classical A-polynomial $A(K;x,y)$, which encodes the $SL(2,\mathbb{C})$ character variety of the fundamental group of the knot complement (S^3/K). More precisely, the generalized volume conjecture [11] states that in the double scaling limit $n\to\infty,\hbar\to 0,\quad q=e^{\hbar}\to 1,\quad x=q^n=e^{n\hbar}=\text{fixed}$,
the colored Jones polynomial has the asymptotic behavior

$$\lim_{n\to\infty,\hbar\to 0}J_n\big(K;q=e^{\hbar}\big)=\exp\left(\frac{1}{\hbar}S_0+\ldots\right),$$
$$\text{where } S_0(x)=\mathrm{Vol}\big(S^3\backslash K\big)+iCS\big(S^3\backslash K\big)+\int_1^x\frac{dx}{x}\log y. \tag{4}$$

The integral over x is done along $A(K;x,y)=0$. Differentiating the above equation, the conjecture states that

$$\log y=-x\frac{d}{dx}\left[\lim_{\substack{n\to\infty,\,\hbar\to 0\\ e^{n\hbar}=x}}\hbar\log J_n\big(K;q=e^{\hbar}\big)\right], \tag{5}$$

gives the zero locus of the classical A-polynomial of the knot K.

One can quantize the classical A-polynomial by promoting the variables (x,y) to operators $(\hat{x},\hat{y})$ such that

$$\hat{x}J_n(K;q)=q^nJ_n(K;q),\quad \hat{y}J_n(K;q)=J_{n+1}(K;q). \tag{6}$$

The quantum A-polynomial $\widehat{A}(K;\hat{x},\hat{y};q)$ is a polynomial in the operators $(\hat{x},\hat{y})$ and the variable q. In fact, $\widehat{A}(K;\hat{x},\hat{y};q)$ is defined as:

$$\widehat{A}(K;\hat{x},\hat{y};q)J_n(K;q)=0, \tag{7}$$

which is also equivalent to the q-difference equation $\sum_{m=0}^{k} a_m(\hat{x},q)J_{n+m}(K;q)$ of minimal order. We expect to recover the classical A-polynomial from the quantum A-polynomial by taking the classical limit $q=1$:

$$\widehat{A}(K;\hat{x},\hat{y};q=1)=A(K;x,y). \tag{8}$$

The above assertion is known as the quantum volume conjecture [11], or, the AJ conjecture [12, 13].

The generalized volume conjecture and the quantum volume conjecture (AJ conjecture) were further categorified for superpolynomials in [14, 15] by including the two-parameters (a,t). In this categorified version, one defines the classical super-A-polynomial, $A^{\text{super}}(K;x,y;a,t)$, to be an (a,t) deformation of $A(K;x,y)$ which can be obtained by substituting $\mathcal{P}_n(K;a,q,t)$ for $J_n(K;q)$ in (5). Likewise, the quantum super-A-polynomial, $\hat{A}^{\text{super}}(K;\hat{x},\hat{y};a,q,t)$, is also an (a,t) deformation of $\widehat{A}(K;\hat{x},\hat{y};q)$. Basically $\hat{x},\hat{y}$ are defined as

$$\hat{x}\mathcal{P}_n(K;a,q,t)=q^n\mathcal{P}_n(K;a,q,t),\quad \hat{y}\mathcal{P}_n(K;a,q,t)=\mathcal{P}_{n+1}(K;a,q,t),$$

and the corresponding quantum super-A-polynomial obeys

$$\hat{A}^{\text{super}}(K;\hat{x},\hat{y};a,q,t)\mathcal{P}_n(K;a,q,t)=0. \tag{9}$$

The categorified version of the quantum volume conjecture states that

$$\hat{A}^{\text{super}}(K;\hat{x},\hat{y};a,q=1,t))=A^{\text{super}}(K;x,y;a,t). \tag{10}$$

This note is organized as follows. In Sect. 2, we briefly review how to use the correspondence between Chern–Simons theory and Wess–Zumino–Novikov–Witten (WZNW) model given in [1] to calculate knot invariants. This method explicitly requires the knowledge of the quantum $6j$-symbols or the quantum group $U_q(\mathfrak{sl}_N)$ to write colored HOMFL-PT polynomials. In Sect. 3, we focus on a class of knots called twist knots K_p. Particularly, motivated by the structure of colored Jones polynomials $J_n(K_p,q)$ for the twist knots, we conjectured the colored superpolynomials for the twist knots. Comparing $P_n(K_p;a,q)$ with the formal Chern–Simons knot invariant, we obtained a closed form algebraic expression for the $U_q(\mathfrak{sl}_N)$ quantum $6j$-symbols for a class of multiplicity-free representation. Using our conjectured superpolynomials, we find the classical super-A-polynomial and the quantum super-A-polynomials for the $\mathbf{5_2}$ twist knot in Sect. 4 and test the categorified quantum volume conjecture.

2 Chern–Simons Theory and Knot Invariants

In order to calculate $\langle W_R[K] \rangle$ for a given knot, we slice the three-sphere S^3 containing the knot (redrawn in an appropriate way) into pieces as shown in Fig. 1. Each three-ball has one or more S^2 boundaries with punctures. *Arborescent knots* are those knots in S^3 obtained from gluing three-balls with four punctured S^2 boundaries. Other knots which cannot be viewed by gluing three-balls with four punctured S^2 boundaries are called *non-arborescent knots*. We will confine to invariants of arborescent knots carrying symmetric representation of $SU(N)$ group.

Chern–Simons functional integral on three-balls with one or more four punctured boundaries is denoted by states in the space of four-point conformal blocks of $SU(N)_k$ WZNW conformal field theory [1]. As evident from the knot 10_{152} diagram shown in Fig. 1, we require states corresponding to the fundamental building blocks in Fig. 2. The braid word $\mathcal{B}$ in v_2 are made up of concatenation of the four types of braiding between two adjacent strands shown in Fig. 3. Then, using the Chern–Simons and WZNW correspondence, the basis states of the braiding generators are the four-point conformal blocks (see Fig. 4):

$$b_1^{(\pm)} |\phi_t(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle = \lambda_t^{(\pm)}(R_1, R_2) |\phi_t(R_2, R_1, \bar{R}_3, \bar{R}_4)\rangle \,, \tag{11}$$

$$b_2^{(\pm)} |\hat{\phi}_s(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle = \lambda_s^{(\pm)}(R_2, \bar{R}_3) |\hat{\phi}_s(R_1, \bar{R}_3, R_2, \bar{R}_4)\rangle \,, \tag{12}$$

$$b_3^{(\pm)} |\phi_t(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle = \lambda_t^{(\pm)}(\bar{R}_3, \bar{R}_4) |\phi_t(R_1, R_2, \bar{R}_4, \bar{R}_3)\rangle \,. \tag{13}$$

Here b_i means braiding between i-th and the $(i+1)$-th strands and the braiding eigenvalues in the vertical framing are

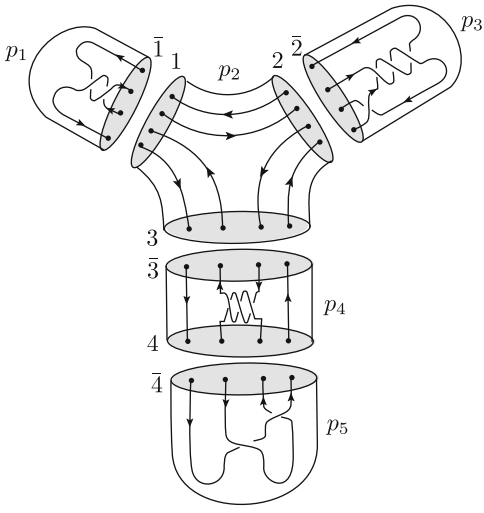


Fig. 1 Knot $\mathbf{10_{152}}$ from gluing three-balls

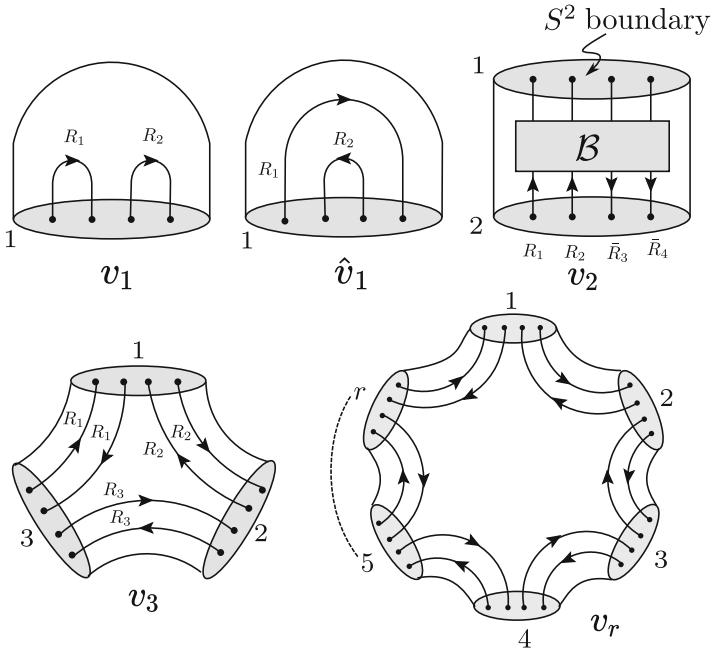


Fig. 2 Building blocks

$$\lambda_t^{(\pm)}(R_1, R_2) = \epsilon_{t;R_1,R_2}^{(\pm)} \left(q^{\frac{C_{R_1}+C_{R_2}-C_{R_t}}{2}} \right)^{\pm 1}, \tag{14}$$

where $\epsilon_{t;R_1,R_2}^{(\pm)} = \pm 1$ and C_R is the quadratic Casimir for representation R. Note that the internal representation has to satisfy the fusion rules. That is, $t \in (R_1 \otimes R_2) \cap (R_3 \otimes R_4)$ and $s \in (R_2 \otimes \bar{R}_3) \cap (\bar{R}_1 \otimes R_4)$. Since the two bases spanned the same Hilbert space, they are linearly related to each other by the fusion matrix a_{ts}

$$|\phi_t(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle = a_{ts} \begin{bmatrix} R_1 & R_2 \\ \bar{R}_3 & \bar{R}_4 \end{bmatrix} |\hat{\phi}_s(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle .$$

For $SU(N)_k$ WZNW model, the properties of the fusion matrix are same as the quantum Racah coefficients (proportional to the $6j$ symbols of the quantum groups $U_q(\mathfrak{sl}_N)$). Using these four-point conformal block basis, braiding eigenvalues and fusion matrices, the states denoting the fundamental building blocks in Fig. 2 are

$$v_1 = \sqrt{d_{R_1} d_{R_2}} |\phi_0(R_1, \bar{R}_1, R_2, R_2)\rangle^{(1)}, \quad \hat{v}_1 = \sqrt{d_{R_1} d_{R_2}} |\hat{\phi}_0(R_1, \bar{R}_2, R_2, \bar{R}_1)\rangle^{(1)},$$

$$v_2 = \sum_{l \in (R_1 \otimes R_2) \cap (R_3 \otimes R_4)} \{\mathcal{B} |\phi_l(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle\}^{(1)} |\phi_l(R_1, R_2, \bar{R}_3, \bar{R}_4)\rangle^{(2)},$$

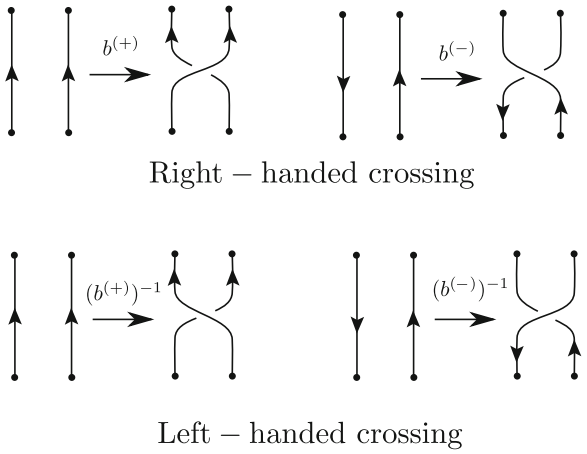


Fig. 3 Types of braiding

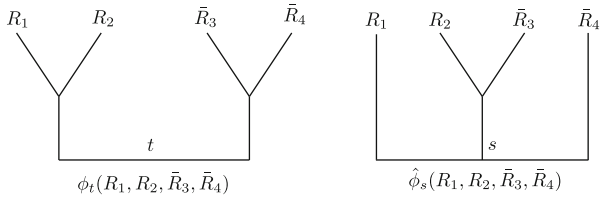


Fig. 4 Two bases for four-point conformal blocks

$$v_r = \sum_t \left(\sqrt{d_t}\right)^{2-r} \left|\phi_t\left(R_1, \bar{R}_1, R_2, R_2\right)\right\rangle^{(1)} \ldots \left|\phi_t\left(R_r, \bar{R}_r, R_1, R_1\right)\right\rangle^{(r)}, \tag{15}$$

where $t \in (R_1 \otimes \bar{R}_1) \cap \ldots . \cap (R_r \otimes \bar{R}_r)$ and d_R is the quantum dimension of a representation R. The superscripts outside the four-point conformal blocks denote the boundaries as indicated in Fig. 1. The 0 in ϕ_0 and $\hat{\phi}_0$ represents the singlet representation. Substituting the building blocks states for each three-ball, we can write $\langle W_R[10_{152}]\rangle$. Basically, gluing the three-balls along oppositely oriented S^2 boundaries involves inner product of four-point conformal bases with its dual four-point conformal blocks. The final invariant can be written in terms of braiding eigenvalues and the $U_q(\mathfrak{sl}_N)$ Racah coefficients [16]. In order to write the explicit polynomial form in variables $a = q^N, q$, we require the $U_q(\mathfrak{sl}_N)$ Racah coefficients which is not known for general $SU(N)$ representations. We will now review our work on superpolynomials for twist knots leading us to conjecture some $U_q(\mathfrak{sl}_N)$ Racah coefficients.

3 Twist Knots

The colored Jones polynomial of twist knots K_p with $2p$ half-twists [17] has a multi-sum expression. We observed that the summand consists of the polynomial of the trefoil $K_1 = \mathbf{3_1}$ for $p > 0$ ($K_{-1} = \mathbf{4_1}$ for $p < 0$) as the main body and a twisting factor. With this structure in mind, the superpolynomials for trefoil $\mathbf{3_1}$ and figure-eight $\mathbf{4_1}$ [14, 15] can be seen in a more succinct form:

$$\mathcal{P}_n(K_1;a,q,t) = (-t)^{-n+1}\sum_{k=0}^{\infty} q^k \frac{(-atq^{-1};q)_k}{(q;q)_k}\big(q^{1-n};q\big)_k\big(-at^3q^{n-1};q\big)_k$$
$$\equiv (-t)^{-n+1}\sum_{k=0}^{\infty} q^k H_{n,k} \equiv (-t)^{-n+1}\sum_{k=0}^{\infty} F_{n,k}(a,q,t), \qquad (16)$$

$$\mathcal{P}_n(K_{-1};a,q,t) = \sum_{k=0}^{\infty}\big(-at^2\big)^{-k}q^{-k(k-3)/2}H_{n,k} \equiv \sum_{k=0}^{\infty} G_{n,k}(a,q,t), \qquad (17)$$

where we use the q-Pochhammer symbol $(z;q)_k = \prod_{j=0}^{k-1}(1-zq^j)$. We expect the superpolynomial for twist knots K_p for $|p| > 1$ involving the above summand multiplied by twisting factors:

$$\mathcal{P}_n(K_{p>0};a,q,t) = (-t)^{-n+1}\sum_{s_{|p|}\geqslant\cdots\geqslant s_1\geqslant 0}^{\infty} F_{n,s_p}(a,q,t)\times\text{Twisting Factor}, \qquad (18)$$

$$\mathcal{P}_n(K_{p<0};a,q,t) = \sum_{s_{|p|}\geqslant\cdots\geqslant s_1\geqslant 0}^{\infty} G_{n,s_p}(a,q,t)\times\text{Twisting Factor}. \qquad (19)$$

The form of the twisting factor can be conjectured from the colored superpolynomials of $\mathbf{5_2} = K_2$ and $\mathbf{6_1} = K_{-2}$ calculated in refs. [4, 18] up to $n = 3$. From this data we could guess the form of the twisting factor for $|p| = 2$ and generalize to arbitrary $|p|$. We have checked our conjectured superpolynomials (See eqns.(2.18, 2.19) in ref. [19]) with known results in the literature.

In order to study asymptotic expansion of superpolynomials, it is important to reduce the multi-summation to minimum number of summation. We succeeded in converting the multi-summation into a double summation using Bailey identities (see appendix A in ref. [19]):

$$\mathcal{P}_n(K_{p>0};a,q,t) = (-t)^{-n+1}\sum_{k=0}^{\infty}\sum_{\ell=0}^{k} q^k\frac{\big(-atq^{-1};q\big)_k}{(q;q)_k}(q^{1-n};q)_k(-at^3q^{n-1};q)_k$$
$$\times(-1)^{\ell}\big(at^2\big)^{p\ell}q^{(p+1/2)\ell(\ell-1)}\frac{1-at^2q^{2\ell-1}}{(at^2q^{\ell-1};q)_{k+1}}\begin{bmatrix}k\\ \ell\end{bmatrix}_q. \qquad (20)$$

where q-binomial is given by $\begin{bmatrix} n \\ k \end{bmatrix}_q = \frac{(q;q)_n}{(q;q)_k (q;q)_{n-k}}$. $\mathcal{P}_n(K_{p<0}; a, q, t)$ has the same expression as $\mathcal{P}_n(K_{p>0}; a, q, t)$ but without the factor $(-t)^{-n+1}$. Following the methodology in section 2, the Chern–Simons field theory invariant $\langle W_R[K_p] \rangle$, in terms of braiding eigenvalues and $U_q(\mathfrak{sl}_N)$ Racah coefficients, for twist knots K_p turns out to be

$$\langle W_R[K_p] \rangle = \sum_{s,s'} \epsilon_s^{R,\bar{R}} \sqrt{d_s} \epsilon_{s'}^{R,\bar{R}} \sqrt{d_{s'}} \left(\lambda_s^{(-)}(R, \bar{R}) \right)^{-2} a_{ss'} \begin{bmatrix} R & \bar{R} \\ R & \bar{R} \end{bmatrix} \left(\lambda_{s'}^{(-)}(R, \bar{R}) \right)^{-2p},$$

where $s, s' \in R \otimes \bar{R}$. Comparing the superpolynomials at $t = -1$ with the above invariant, we obtained $U_q(\mathfrak{sl}_N)$ Racah coefficients for $SU(N)$ symmetric representations up to rank $n = 3$ [16]. Comparing with the formula of the $U_q(\mathfrak{sl}_2)$ $6j$-symbols obtained by Kirillov and Reshetikhin [20], we conjectured a closed form expression for the $U_q(\mathfrak{sl}_N)$ quantum Racah coefficients [21] for the following two types:

$$a_{ts} \begin{bmatrix} R_1 & \bar{R}_2 \\ R_3 & \bar{R}_4 \end{bmatrix} \quad ; \quad a_{ts} \begin{bmatrix} R_1 & R_2 \\ \bar{R}_3 & \bar{R}_4 \end{bmatrix},$$

where R_1, R_2, R_3, R_4 are symmetric representations with single row in the Young diagram.

4 Volume Conjectures and Super-A-Polynomial

We consider the asymptotic form of our conjectured formula of superpolynomials (20) for knot $\mathbf{5_2}$ and perform saddle point analysis to obtain classical super-A-polynomials $A^{\text{super}}(K_{p=2}; x, y; a, t)$. We introduce two variables $z = e^{\hbar k}$, $w = e^{\hbar \ell}$ and take the limits: $q = e^{\hbar} \to 1$, $a =$ fixed, $t =$ fixed, $x = q^n =$ fixed, in (20) and convert the two summation to integrals over z and w. Then using the categorified volume conjecture, we have

$$\mathcal{P}_n(K_{p>0}; a, q, t) \overset{\substack{n\to\infty \\ \hbar\to 0}}{\sim} \int dz dw \, e^{\frac{1}{\hbar}\left(\widetilde{\mathcal{W}}(K_{p>0}; z, w) + \mathcal{O}(\hbar)\right)} \sim e^{\left(\frac{1}{\hbar}\int \log y \frac{dx}{x} + \ldots\right)}, \tag{21}$$

where the integral on RHS in Eq. (21) is over the zero locus of the classical super-A-polynomial, i.e., $A^{\text{super}}(K_{p=2}; x, y; a, t) = 0$. Taking such limits replaces q-Pochhammer into di-logarithms giving the superpotential

$$\begin{aligned}
&\widetilde{\mathcal{W}}(K_{p=2};z,w)\\
&\quad=-\ln x\ln(-t)-\frac{\pi^2}{3}+i\pi\ln w+\ln w(p(\ln a+2\ln t)+\ln w^{p+\frac{1}{2}})\\
&\qquad+\mathrm{Li}_2(x^{-1})-\mathrm{Li}_2(x^{-1}z)+\mathrm{Li}_2(-at)-\mathrm{Li}_2(-atz)+\mathrm{Li}_2(-at^3x)\\
&\qquad-\mathrm{Li}_2(-at^3xz)-\mathrm{Li}_2(at^2w)+\mathrm{Li}_2(at^2wz)+\mathrm{Li}_2(w)+\mathrm{Li}_2(zw^{-1}).
\end{aligned}\tag{22}$$

The saddle point analysis of Eq. (21) in the limit $\hbar\to 0$ imposes the following conditions:

$$\left.\frac{\partial\widetilde{\mathcal{W}}(K_{p>0};z,w,x)}{\partial z}\right|_{(z,w)=(z_0,w_0)}=0=\left.\frac{\partial\widetilde{\mathcal{W}}(K_{p>0};z,w,x)}{\partial w}\right|_{(z,w)=(z_0,w_0)}.\tag{23}$$

Further, the zero locus of the classical super-A-polynomial is determined by

$$y=\exp\left(x\frac{\partial\widetilde{\mathcal{W}}(K_{p>0};z_0,w_0,x)}{\partial x}\right).\tag{24}$$

Plugging the value of $\widetilde{\mathcal{W}}(K_{p=2};z,w)$(22) into the above Eqs. (23) and (24), and eliminating the extremum z_0 and w_0, we obtained classical super-A-polynomials (see Table 6 in ref. [19]).

In order to calculate the quantum super-A-polynomials, we have to find a difference equation of minimal order for the colored superpolynomials. To find the difference equations, we used the program `iSumq.txt` written by Xinyu Sun, based on q-analogue of the algorithm developed in ref. [22] and obtained $\hat{A}^{\text{super}}(\mathbf{5_2};\hat{x},\hat{y};a,q,t)$ (see Table 7 in ref. [19]). We checked that our results obey

$$\hat{A}^{\text{super}}(\mathbf{5_2};\hat{x},\hat{y};a,q=1,t)=A^{\text{super}}(\mathbf{5_2};x,y;a,t),\tag{25}$$

showing the validity of the categorified quantum volume conjecture for knot $\mathbf{5_2}$.

Acknowledgments PR would like to acknowledge DST-RFBR grant(INT/RUS/RFBR/P-309) for support. PR would like to thank Kavli Institute for Theoretical Physics at the University of California Santa Barbara for local hospitality where initial parts of the manuscript were written. PR would like to acknowledge the support in part by the National Science Foundation under Grant No. PHY-1748958.

References

1. E. Witten, Quantum field theory and the Jones polynomial. Commun. Math. Phys. **121**, 351–399 (1989). https://doi.org/10.1007/BF01217730
2. M. Khovanov, A categorification of the Jones polynomial. Duke Math. J. **101**(02), 359–426 (2000). https://doi.org/10.1215/S0012-7094-00-10131-7

3. M. Khovanov, L. Rozansky, Matrix factorizations and link homology II. Geomet. Topol. **12**(06), 1387–1425 (2005). https://doi.org/10.2140/gt.2008.12.1387
4. N.M. Dunfield, S. Gukov, J. Rasmussen, The superpolynomial for knot homologies. Exp. Math. **15**(2), 129–159 (2006). https://doi.org/10.1080/10586458.2006.10128956
5. M. Khovanov, Triply-graded link homology and Hochschild homology of Soergel bimodules. Int. J. Math. **18**(08), 869–885 (2007). https://doi.org/10.1142/S0129167X07004400
6. M. Mackaay, M. Stošić, P. Vaz, The 1,2-coloured HOMFLY-PT link homology. Trans. Amer. Math. Soc. **363**(4), 2091–2124 (2011). https://doi.org/10.1090/S0002-9947-2010-05155-4
7. B. Webster, G. Williamson, A geometric construction of colored HOMFLY-PT homology. Geom. Topol. **21**(05), 2557–2600 (2009). https://doi.org/10.2140/gt.2017.21.2557
8. S. Gukov, A.S. Schwarz, C. Vafa, Khovanov-Rozansky homology and topological strings. Lett. Math. Phys. **74**, 53–74 (2005). https://doi.org/10.1007/s11005-005-0008-8
9. R. Kashaev, The hyperbolic volume of knots from quantum dilogarithm. Lett. Math. Phys. **39**, 269–275 (1997). https://doi.org/10.1023/A:1007364912784
10. H. Murakami, J. Murakami, The colored Jones polynomials and the simplicial volume of a knot. Acta Math. **186**(1), 85–104 (2001). https://doi.org/10.1007/BF02392716
11. S. Gukov, Three-dimensional quantum gravity, Chern-Simons theory, and the A polynomial. Commun. Math. Phys. **255**, 577–627 (2005). https://doi.org/10.1007/s00220-005-1312-y
12. S. Garoufalidis, T. Le, The colored Jones function is q-holonomic. Geom. Topol. **9**, 1253–1293 (2004). https://doi.org/10.2140/gt.2005.9.1253
13. S. Garoufalidis, On the characteristic and deformation varieties of a knot. Geom. Topol. Monogr. **7**, 291–309 (2004). https://doi.org/10.2140/gtm.2004.7.291
14. H. Fuji, S. Gukov, P. Sułkowski, H. Awata, et. al., Volume conjecture: refined and categorified. Adv. Theoret. Math. Phys. **16**(6) 1669–1777 (2012). https://doi.org/10.4310/ATMP.2012.v16.n6.a3
15. H. Fuji, S. Gukov, P. Sulkowski, Super-A-polynomial for knots and BPS states. Nucl. Phys. **B867**, 506–546 (2013). https://doi.org/10.1016/j.nuclphysb.2012.10.005
16. Zodinmawia and P. Ramadevi, SU(N) quantum Racah coefficients & non-torus links. Nucl. Phys. **B870**, 205–242 (2013). https://doi.org/10.1016/j.nuclphysb.2012.12.020
17. G. Masbaum, Skein-theoretical derivation of some formulas of Habiro. Algebr. Geom. Topol. **3**(17), 537–556 (2003). https://doi.org/10.2140/agt.2003.3.537
18. S. Gukov, M. Stosic, Homological algebra of Knots and BPS states. Proc. Symp. Pure Math. **85**, 125–172 (2012); Geom. Topol. Monographs **18**, 309 (2012). https://doi.org/10.1090/pspum/085/1377
19. S. Nawata, P. Ramadevi, Zodinmawia, X. Sun, Super-A-polynomials for Twist Knots. J. High Energy Phys. **1211**, 157 (2012). https://doi.org/10.1007/JHEP11(2012)157
20. A.N. Kirillov, N.Y. Reshetikhin, Representation algebra $U_q(sl_2)$, q-orthogonal polynomials and invariants of links, in *New Developments in the Theory of Knots*, ed. by T. Kohno (World Scientific, Singapore, 1989). https://doi.org/10.1142/1056
21. S. Nawata, P. Ramadevi, Zodinmawia, Multiplicity-free quantum $6j$-symbols for $U_q(sl_N)$. Lett. Math. Phys. **103**, 1389–1398 (2013). https://doi.org/10.1007/s11005-013-0651-4
22. S. Garoufalidis, X. Sun, A new algorithm for the recursion of hypergeometric multisums with improved universal denominator. Gems Exp. Math. Contemp. Math. **517**, 143–156 (2009). https://doi.org/10.1090/conm/517/10138

Demazure Formulas for Weight Polytopes

Mark A. Walton

Abstract The characters of simple Lie algebras are naturally decomposed into lattice-polytope sums. The Brion formula for those polytope sums is remarkably similar to the Weyl character formula. Here we start to investigate if other character formulas have analogs for lattice-polytope sums, by focusing on the Demazure character formulas. Using Demazure operators, we write expressions for the lattice sums of the weight polytopes of rank-2 simple Lie algebras, and the rank-3 algebra A_3.

Keywords Lattice polytopes · Simple Lie algebras · Characters · Demazure character formula

1 Introduction

The Brion formula [3, 4] for lattice-polytope sums is remarkably similar to the Weyl formula for characters of simple Lie algebras [6, 11, 15]. As a consequence, the expansion of Weyl characters in terms of lattice-polytope sums is natural and useful [6, 11–13, 15, 16]. This *polytope expansion* of Lie characters is highly reminiscent of the early work of Antoine and Speiser [2] and the recursive formulas found by Kass [8].

Here we explore further the relation between Lie characters and lattice-polytope sums. Other formulas exist for the characters—might corresponding formulas describe lattice-polytope sums?[1]

We focus on the Demazure character formulas [1, 5, 7], and use the Demazure operators involved to write expressions for the lattice-polytope sums. So far, we

[1] This question was already asked in [15].

M. A. Walton (✉)
Department of Physics & Astronomy, University of Lethbridge, Lethbridge, AB, Canada
e-mail: walton@uleth.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_27

have obtained results for all rank-2 simple Lie algebras and for the rank-3 algebra A_3. We hope these first formulas will help lead to general Demazure formulas for the lattice-polytope sums relevant to Lie characters, and perhaps others.

How might such a formula be useful? Starting with the Demazure character formulas, Littelmann was able to derive a generalization for all simple Lie algebras of the famous Littlewood–Richardson rule for A_r tensor-product decompositions [10]. In a similar way, formulas of the Demazure type might lead to efficient, general computational methods for lattice polytopes.

Physical applications should also be possible. An attempt to apply the Demazure character formula to Wess–Zumino–Witten conformal field theories was made in [14], and it has already been used in the study of solvable lattice models [9].

In the following section, we review the initial motivation for the present work, the similarity between the Weyl character formula and the Brion lattice-polytope sum formula, and the polytope expansion that exploits it. Section 3 is a quick review of the Demazure character formulas. Our new results are presented in Sect. 4: expressions involving Demazure operators for lattice-polytope sums for rank-2 simple Lie algebras, and A_3. The final section is a short conclusion.

2 Polytope Expansion of Lie Characters

Let X_r denote a simple Lie algebra of rank r, so that $X = A, B, C, D, E, F$, or G. The sets of fundamental weights and simple roots will be written as $F := \{ \Lambda^i \mid i \in \{1, 2, \ldots, r\} \}$ and $S := \{ \alpha_i \mid i \in \{1, 2, \ldots, r\} \}$, respectively. The weight and root lattices are $P := \mathbb{Z}\, F$ and $Q := \mathbb{Z}\, S$, respectively. The set of integrable weights is $P_+ := \mathbb{N}_0\, F$, and we write R (R_+) for the set of (positive) roots of X_r.

2.1 Weyl Character Formula

Consider an irreducible representation $L(\lambda)$ of X_r of highest weight $\lambda \in P_+$. The formal character of $L(\lambda)$ is defined to be

$$\mathrm{ch}_\lambda \;=\; \sum_{\mu \in P(\lambda)} \mathrm{mult}_\lambda(\mu)\, e^{\,\mu}\,. \tag{1}$$

Here $P(\lambda)$ is the set of weights of the representation $L(\lambda)$, and $\mathrm{mult}_\lambda(\mu)$ is the multiplicity of weight μ in $L(\lambda)$.

The formal exponentials of weights obey $e^\mu\, e^\nu \;=\; e^{\mu+\nu}$. If we write

$$e^\mu(\sigma) \;=:\; e^{\langle \mu, \sigma \rangle}\,, \tag{2}$$

where $\langle \mu, \sigma \rangle$ is the inner product of weights μ and σ, the formal exponential e^{μ} simply stands for $e^{\langle \mu, \sigma \rangle}$ before a choice of weight σ is made. A choice of σ fixes a conjugacy class of elements in the Lie group $\exp(X_r)$. The formal character then becomes a true character $\mathrm{ch}_\lambda(\sigma)$, the trace, in the representation of highest weight λ, of elements of $\exp(X_r)$ in the conjugacy class labeled by σ.

The celebrated Weyl character formula is

$$\mathrm{ch}_\lambda \;=\; \frac{\sum_{w \in W} (\det w)\, e^{w.\lambda}}{\prod_{\alpha \in R_+} (1 - e^{-\alpha})} \;. \tag{3}$$

Here W is the Weyl group of the simple Lie algebra X_r, and $w.\lambda = w(\lambda + \rho) - \rho$ denotes the shifted action of Weyl-group element $w \in W$ on the weight λ, with $\rho = \sum_{i=1}^{r} \Lambda^i$.

The Weyl invariance of the character can be made manifest:

$$\mathrm{ch}_\lambda \;=\; \sum_{w \in W} e^{w\lambda} \prod_{\alpha \in R_+} (1 - e^{-w\alpha})^{-1} \;. \tag{4}$$

Here now

$$\left(1 - e^{\beta}\right)^{-1} \;=\; \begin{cases} 1 + e^{\beta} + e^{2\beta} + \ldots \;, & \beta \in R_-; \\ -e^{-\beta} - e^{-2\beta} - \ldots \;, & \beta \in R_+. \end{cases} \tag{5}$$

The rule-of-thumb is expand in powers of e^{β}, with β a negative root.

2.2 *Brion Lattice-Polytope Sum Formula*

A polytope is the convex hull of finitely many points in $\mathbb{R}^d$. A polytope's vertices are such a set of points with minimum cardinality. A lattice polytope has all its vertices in an integral lattice in $\mathbb{R}^d$. The (formal) lattice-polytope sum is the sum of terms e^{ϕ} over the lattice points ϕ in the polytope.

Brion [3, 4] found a general formula for these lattice-polytope sums. Let the *weight polytope* Pt_λ be the polytope with vertices given by the Weyl orbit $W\lambda$. Consider the lattice-polytope sum

$$B_\lambda \;:=\; \sum_{\mu \in (\lambda + Q) \cap \mathrm{Pt}_\lambda} e^{\mu} \;, \tag{6}$$

where the relevant lattice is the λ-shifted root lattice $\lambda + Q$ of the algebra X_r. Applied to a weight polytope, the Brion formula yields

$$B_\lambda \;=\; \sum_{w\in W} e^{w\lambda} \prod_{\alpha\in S} (1-e^{-w\alpha})^{-1} \;. \tag{7}$$

Here S denotes the set of simple roots of X_r.

2.3 Polytope Expansion

The Brion formula (7) is remarkably similar to the Weyl character formula, as written in (4) [6, 15]. It is therefore natural, and fruitful, to consider the polytope expansion of Lie characters [6, 15, 16]:

$$\mathrm{ch}_\lambda \;=\; \sum_{\mu\leq\lambda} \mathrm{polyt}_\lambda(\mu)\; B_\mu \;. \tag{8}$$

The polytope multiplicities $\mathrm{polyt}_\lambda(\mu)$ are defined in analogy with weight multiplicities $\mathrm{mult}_\lambda(\mu)$.

We do not consider the polytope expansion further in this note. Instead we focus on the striking relation described above between characters and polytope sums.

3 Demazure Character Formulas

Do other character formulas point to the existence of new formulas for the lattice sums of weight polytopes? More general polytopes?

In particular, do the Demazure formulas for Lie characters indicate the existence of Demazure-type formulas for the lattice sums of weight polytopes?

Let us first sketch the Demazure character formulas. The Weyl group W is generated by the primitive reflections r_i in weight space across the hyperplanes normal to the corresponding simple roots α_i:

$$r_i\,\lambda \;=\; \lambda \;-\; \left(\lambda\cdot\alpha_i^\vee\right)\alpha_i \;; \tag{9}$$

here $\alpha_i^\vee = 2\alpha_i/(\alpha_i\cdot\alpha_i)$ is the simple co-root.

Define Demazure operators for every simple root $\alpha_i \in S$:

$$D_{\alpha_i} \;=:\; D_i \;=\; \frac{1\;-\;e^{-\alpha_i}\,r_i}{1-e^{-\alpha_i}} \;, \tag{10}$$

where $r_i(e^\lambda) \;=\; e^{r_i\lambda}$. Then

$$D_i\, e^{\lambda} \;=\; \begin{cases} e^{\lambda} + e^{\lambda-\alpha_i} + e^{\lambda-2\alpha_i} + \cdots + e^{r_i\lambda}, & \lambda\cdot\alpha_i^{\vee} \geqslant 0\,; \\ -e^{\lambda+\alpha_i} - e^{\lambda+2\alpha_i} - \cdots - e^{r_i(\lambda+\alpha_i)}, & \lambda\cdot\alpha_i^{\vee} < 0\,. \end{cases} \tag{11}$$

For every Weyl-group element $w \in W$ a Demazure operator D_w can be defined. First, identify $D_{r_i} := D_i$, and then use any reduced decomposition of w, replacing the factors r_j in the reduced decomposition with D_j. The resulting operator D_w must be independent of which reduced decomposition is used. As a result, the Demazure operators obey relations encoded in the Coxeter–Dynkin diagrams of X_r.

For example, consider the longest element w_L of the Weyl group of A_2: $w_L = r_1r_2r_1 = r_2r_1r_2$. The associated Demazure operator D_{w_L} can be written in two ways, so that $D_1D_2D_1 = D_2D_1D_2$.

If w_L is the longest element of the Weyl group W, then the Demazure character formula is

$$\mathrm{ch}_{\lambda} \;=\; D_{w_L}\, e^{\lambda}\,. \tag{12}$$

Also, define $D_i =: 1 + d_i$, and then d_w for all $w \in W$ by reduced decompositions. Then

$$\mathrm{ch}_{\lambda} \;=\; \sum_{w\in W} d_w\, e^{\lambda}\,. \tag{13}$$

Demazure operators can also be defined for every positive root $\beta \in R_+$:

$$D(\beta) \;:=\; \frac{1 - e^{-\beta}\, r(\beta)}{1 - e^{-\beta}}\,, \tag{14}$$

where

$$r(\beta)\lambda \;:=\; \lambda - \big(\lambda\cdot\beta^{\vee}\big)\beta \quad\text{and}\quad r(\beta)\big(e^{\lambda}\big) \;=\; e^{r(\beta)\lambda}\,. \tag{15}$$

Operators $d(\beta) = D(\beta) - 1$,

$$d(\beta) \;:=\; \frac{e^{-\beta}\,[1 - r(\beta)]}{1 - e^{-\beta}}\,, \tag{16}$$

are also defined for all positive roots $\beta \in R_+$.

4 Lattice-Polytope Formulas of Demazure Type

In the hopes of helping lead to a more general result, we take a direct approach here, and write formulas for low-rank weight-polytope sums that involve the Demazure

operators. We report only preliminary new results, formulas for all rank-2 cases, and for one of rank 3, related to the Lie algebra A_3.

But before treating ranks 2 and 3, let us first dispense with the unique rank-1 algebra, A_1. In this case, the character and weight-polytope lattice sum are identical,

$$\mathrm{ch}_\lambda \;=\; B_\lambda \;=\; e^\lambda + e^{\lambda-\alpha_1} + e^{\lambda-2\alpha_1} + \ldots + e^{-\lambda}\,. \tag{17}$$

The Demazure character formulas therefore apply to B_λ.

The A_1 Weyl group W has 2 elements, the identity and $r_1 = r(\alpha_1)$, where α_1 is the simple root, and only positive root. The longest element of W is therefore $w_L = r_1 = r(\alpha_1)$, with a unique reduced decomposition. Applying the Demazure formulas (12) and (13), we find

$$B_\lambda \;=\; D_1\, e^\lambda \;=\; \big[d(\alpha_1)\,+\,1\big]\, e^\lambda\,. \tag{18}$$

The last expression will turn out to be the most relevant here—see (23) and (25) below.

The 3 rank-2 algebras can be treated in a unified way. Put the $p := \dim R_+$ positive roots of your rank-2 algebra in angular order; label them γ_j. So we get, for A_2,

$$\{\gamma_1, \gamma_2, \gamma_{p=3}\} \;=\; \{\alpha_1, \alpha_1+\alpha_2, \alpha_2\}\,; \tag{19}$$

for B_2 ($\cong C_2$),

$$\{\gamma_1, \ldots, \gamma_{p=4}\} \;=\; \{\alpha_1, \alpha_1+\alpha_2, \alpha_1+2\alpha_2, \alpha_2\}\,; \tag{20}$$

and for G_2,

$$\{\gamma_1, \ldots, \gamma_{p=6}\} \;=\; \{\alpha_1, \alpha_1+\alpha_2, 2\alpha_1+3\alpha_2, \alpha_1+2\alpha_2, \alpha_1+3\alpha_2, \alpha_2\}\,. \tag{21}$$

In a generic weight diagram, these positive roots are parallel to half of the boundaries, in angular order. They specify a path from the highest weight to the lowest weight along the polytope edges labeled by γ_1 through γ_p, in that order. Correspondingly, the longest element w_L of the Weyl group can be written as a product of the reflections defined in (15):

$$w_L \;=\; r(\gamma_p)r(\gamma_{p-1})\,\cdots\, r(\gamma_2)r(\gamma_1)\,. \tag{22}$$

The weights of the first $p-1$ boundaries can be generated, and then the polytope can be filled in by γ_p-strings of weights. If $B_\lambda \;=\; \mathcal{B}\,(e^\lambda)$, then

$$\begin{aligned}\mathcal{B} = \big[d(\gamma_p)+1\big]\Big[\, & d(\gamma_{p-1})r(\gamma_{p-2})\cdots r(\gamma_1) \\ & + d(\gamma_{p-2})r(\gamma_{p-3})\cdots r(\gamma_1) \,+\, \cdots \,+\, d(\gamma_2)r(\gamma_1) \,+\, d(\gamma_1) \,+\, 1\,\Big].\end{aligned} \tag{23}$$

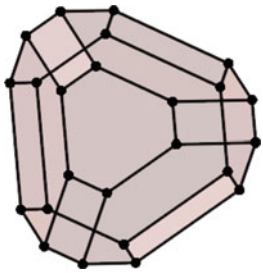


Fig. 1 Weight polytope for a regular highest weight of the Lie algebra A_3

Here we use the Demazure operators defined in (16) above. Eq. (23) is one formula of Demazure type that applies to all rank-2 cases.

Now consider a rank-3 example, the Lie algebra A_3. The weight polytope for a highest weight with all Dynkin labels non-zero, and unequal, is illustrated in Fig. 1. Notice that the facets are the weight polytopes for the rank-2 algebras whose Coxeter–Dynkin diagrams are obtained from that of A_3 by deleting a single node. Hence, the facets are hexagonal A_2 weight polytopes of 2 types, and rectangular $A_1 \oplus A_1$ polytopes.

The longest element of the A_3 Weyl group can be written as in (22), with $p = 6$ and

$$\{\gamma_1, \ldots, \gamma_6\} \;=\; \{\alpha_1, \alpha_{12}, \alpha_{123}, \alpha_2, \alpha_{23}, \alpha_3\}\,, \tag{24}$$

where $\alpha_{12} := \alpha_1 + \alpha_2$, etc. The expression corresponds to a path along the edges of the polytope from the highest to the lowest weight.

It is not difficult to see that the following expression

$$\begin{aligned}\mathcal{B} = &\left[d(\alpha_3) \,+\, 1\right]\left[d(\alpha_{23})r(\alpha_2) \,+\, d(\alpha_2) \,+\, 1\right] \\ &\times \left[d(\alpha_{123})r(\alpha_{12})r(\alpha_1) \,+\, d(\alpha_{12})r(\alpha_1) \,+\, d(\alpha_1) \,+\, 1\right]\end{aligned} \tag{25}$$

generates the A_3 weight-polytope lattice sum. Notice the order of factors follows the order of factors in the expression (22), as in the rank-2 result (23).

5 Conclusion

We have begun a search for Demazure-type formulas for the exponential sums of weight polytopes of simple Lie algebras. Our results are preliminary, mostly summarized in (23) and (25), expressions valid for all rank-2 algebras (A_2, $B_2 \cong C_2$, G_2) and A_3, respectively.

Clearly, Demazure-type formulas can be written. However, our expressions are not unique—we have obtained others. What is needed is a universal formula, one that applies to all simple Lie algebras, as the Weyl (3) and (4) and Demazure (12)

and (13) character formulas do, as well as the Brion formula (7) for a weight polytope. Similarities in the formulas given here may indicate that we are on the right track.

Acknowledgments I thank Jørgen Rasmussen for collaboration and Chad Povey for 3D-printing rank-3 weight polytopes. This research was supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

1. H.H. Andersen, Schubert varieties and Demazure's character formula. Invent. Math. **79**, 611–618 (1985)
2. J.-P. Antoine, D. Speiser, Characters of irreducible representations of the simple groups. I. General theory. J. Math. Phys. **5**, 1226–1234 (1964); Characters of irreducible representations of the simple groups. II. Application to classical groups. J. Math. Phys. **5**, 1560–1572 (1964)
3. M. Brion, Points entiers dans les polyèdres convexes. Ann. Scient. Éc. Norm. Sup., 4e série, t. **21**, 653–663 (1988)
4. M. Brion, Polyèdres et réseaux, Enseign. Math. **38**(1–2), 71–88 (1992)
5. M. Demazure, Désingularisation des variétés de Schubert généralisées. Ann. scient. Éc. Norm. sup., t. **6**, Sect. 2, 53–88 (1974); Une nouvelle formule des caractères. Bull. Sci. Math. **98**(3), 163–172 (1974)
6. G. Dhillon, A. Khare, Characters of highest weight modules and integrability (2016). arXiv:1606.09640; The Weyl-Kac weight formula. Séminaire Lotharingien de Combinatoire **78B** (2017), *Proceedings of the 29th Conference on Formal Power Series and Algebraic Combinatorics (London)*, Article #77 (2018). arXiv:1802.06974
7. A. Joseph, On the Demazure character formula. Ann. sclent. Éc. Norm. Sup., 4^{e} série, t. **18**, 389 -419 (1985)
8. S. Kass, A recursive formula for characters of simple Lie algebras. J. Alg. **137**, 126 (1991)
9. A. Kuniba, K. Misra, M. Okado, T. Takagi, J. Uchiyama, Characters of Demazure modules and solvable lattice models. Nucl. Phys. B **510**, 555–576 (1998)
10. P. Littelmann, A generalization of the Littlewood-Richardson rule. J. Alg. **130**, 328–368 (1990)
11. A. Postnikov, Permutohedra, associahedra, and beyond. Int. Math. Res. Not. **6**, 1026–1106 (2009)
12. J. Rasmussen, Layer structure of irreducible Lie algebra modules (2018). Preprint arXiv:1803.06592
13. W. Schutzer, A new character formula for Lie algebras and Lie groups. J. Lie Theory **22**(3), 817–838 (2012)
14. M.A. Walton, Demazure characters and WZW fusion rules. J. Math. Phys. **39**, 665–681 (1998)
15. M.A. Walton, Polytope sums and Lie characters, in *Symmetry in Physics*. CRM Proceedings & Lecture Notes, vol. 34 (American Mathematical Society, Providence, 2004), pp. 203–214; *Proceedings of a CRM Workshop Held in Memory of Robert T. Sharp* (2002), pp. 12–14
16. M.A. Walton, Polytope expansion of Lie characters and applications. J. Math. Phys. **54**, 121701 (2013)

Point Transformations: Exact Solutions of the Quantum Time-Dependent Mass Nonstationary Oscillator

Kevin Zelaya and Véronique Hussin

Abstract In this note, we address the exact solutions of a time-dependent Hamiltonian composed by an oscillator-like interaction with both frequency and mass terms that depend on time. The latter is achieved by constructing the appropriate point transformation that deforms the Schrödinger equation of a stationary oscillator into the one of the time-dependent model. In this form, the solutions of the latter can be seen as deformations of the well-known stationary oscillator solutions, and thus an orthogonal set of solutions can be determined straightforwardly. The latter is possible since the inner product structure is preserved by the point transformation. Moreover, any invariant operator of the stationary oscillator is transformed into an invariant of the time-dependent model. This property leads to a straightforward way to determine constants of motion without using any ansatz.

Keywords Nonstationary oscillator · Point transformations · Time-dependent Hamiltonians · Exact solutions · Quantum invariants

1 Introduction

The dynamics of non-relativistic quantum systems is determined through the solutions of the Schrödinger equation. In the latter, the information of the system under consideration is coded in the Hamiltonian operator. Most of the physical systems of interest are stationary and described by time-independent Hamiltonians. On the other hand, for open systems, time-dependent Hamiltonians are

K. Zelaya (✉)
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: zelayame@crm.umontreal.ca

V. Hussin
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
Départment de Mathématiques et de Statistique, Université de Montréal, Montréal, QC, Canada
e-mail: hussin@dms.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_28

required to provide an accurate description. Physical applications are found in electromagnetic traps of particles [1–3], in which external time-dependent electric and magnetic fields allow the confinement of particles [4]. In such a case, we describe the respective Hamiltonian through a *parametric oscillator* potential, also known as *nonstationary oscillator*, which consists of an oscillator-like interaction with a frequency that varies in time. Exact solutions were studied in detail by Lewis and Riesenfeld for the classical and quantum cases [5, 6]. Given that the Hamiltonian has an explicit dependence on time, an eigenvalue equation associated with Hamiltonian is no longer feasible and the existence of an orthonormal set of solutions cannot be taken for granted, as it is customary for the stationary quantum oscillator. Nevertheless, Lewis and Riesenfeld introduced an approach in which a nonstationary eigenvalue equation can be still found once the appropriate constant of motion is determined [6]. With the latter, solutions to the Schrödinger equation are found by adding the appropriate time-dependent complex-phase to the nonstationary eigenfunctions. The constant of motion, or *invariant operator* $\hat{I}(t)$, is usually imposed as an ansatz and determined from the condition $[i\partial/\partial t - \hat{H}(t), \hat{I}(t)] = 0$. Such approach has been applied successfully to other time-dependent models as well [7].

In this note, we address the solutions of the nonstationary oscillator with time-dependent mass. To this end, we consider the method of point transformations, which has been used in the context of quadratic time-dependent potentials [8] and nonstationary Darboux transformations [9]. In particular, we use the construction introduced in [10]. The latter allows deforming the well-known Schrödinger equation of the stationary oscillator into the one of the time-dependent model. This method leads to a straightforward way to obtain the solutions of the time-dependent model as deformations of the stationary oscillator. Remarkably, the point transformation preserves the first integrals, this means that the constants of motion and the spectral properties for the time-dependent model are inherited from the stationary oscillator, without requiring to impose any anstaz [6].

2 Nonstationary Oscillator with Time-Dependent Mass

Let us first consider the quantum harmonic oscillator, defined through the Hamiltonian

$$\hat{H}_{\text{osc}} = \frac{\hat{p}_y^2}{2} + \frac{\hat{y}^2}{2}, \tag{1}$$

where $\hat{y}$ and $\hat{p_y}$ stand for the canonical position and momentum operators, $[\hat{y}, \hat{p}_y] = i$. With the latter, the Schrödinger equation, represented in the spatial coordinate 'y', reads as

$$i\frac{\partial\Psi}{\partial\tau} = -\frac{1}{2}\frac{\partial^2\Psi}{\partial y^2} + \frac{y^2}{2}\Psi, \tag{2}$$

where τ is the time parameter, the momentum operator was represented as $\hat{p}_y = -i\frac{\partial}{\partial y}$ and $\Psi(y,\tau) = \langle y|\Psi(\tau)\rangle$ is the respective wave function. Given that $\hat{H}_{\rm osc}$ is time-independent, the solutions of (2) can be easily computed using the separation of variables $\Psi(y,\tau) = e^{-iE\tau}\Phi(y)$, where $\Phi(y) = \langle y|\Phi\rangle$ fulfills the eigenvalue equation

$$-\frac{1}{2}\frac{d^2\Phi}{dy^2} + \frac{y^2}{2}\Phi = E\Phi. \tag{3}$$

A set of physical solutions $\{\Phi_n(y)\}_{n=0}^{\infty}$ is determined with the aid of the finite-norm condition $|||\Phi_n\rangle||^2 = \langle\Phi_n|\Phi_n\rangle < \infty$, where the inner product of two eigenfunctions $\Phi_{(1)}(y)$ and $\Phi_{(2)}(y)$ is defined through

$$\langle\Phi_{(2)}|\Phi_{(1)}\rangle = \int_{-\infty}^{\infty} dy\, \Phi^*_{(2)}(y)\Phi_{(1)}(y)\,. \tag{4}$$

The spectral information of the harmonic oscillator is then given by

$$\Phi_n(y) = \sqrt{\frac{1}{2^n n!\sqrt{\pi}}}\, e^{-\frac{y^2}{2}} H_n(y)\,, \quad E_n = (n+1/2), \tag{5}$$

where $H_n(z)$ are the Hermite polynomials [11]. The set of eigenfunctions is orthonormal, $\langle\Phi_m|\Phi_n\rangle = \delta_{n,m}$, and it generates the space $\mathcal{H} = \text{span}\{|\Phi_n\rangle\}_{n=0}^{\infty}$.

Now, we introduce the nonstationary oscillator with time-dependent mass, defined in terms of the canonical position and momentum operators $\hat{x}$ and $\hat{p}_x$, respectively, together with the time parameter t through the time-dependent Hamiltonian

$$\hat{H}(t) = \frac{1}{2m(t)}\hat{p}_x^2 + \frac{1}{2}m(t)\Omega^2(t)\hat{x}^2 + F(t)\hat{x} + V_0(t), \tag{6}$$

where $m(t)$ is the time-dependent mass, $\Omega^2(t)$ the time-dependent frequency, $F(t)$ an external driving force and $V_0(t)$ a zero-point energy term. The wave functions $\psi(x,t) = \langle x|\psi(t)\rangle$ associated with the Hamiltonian (6) are thus computed from the Schrödinger equation

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m(t)}\frac{\partial^2\psi}{\partial x^2} + \frac{1}{2}m(t)\Omega^2(t)x^2\psi + F(t)x\psi + V_0(t)\psi. \tag{7}$$

In the sequel, we address the solutions of (7) by constructing the point transformation.

3 Point Transformation

In order to transform the Schrödinger equation of the stationary oscillator (2) into the one of the nonstationary oscillator with time-dependent mass (7), let us consider relationships between the elements of the set $\{y, \tau, \Psi\}$ and those of the set $\{x, t, \psi\}$ of the form [12]

$$y = y(x,t), \quad \tau = \tau(x,t), \quad \Psi = \Psi(y(x,t), \tau(x,t)) = G(x,t;\psi(x,t)). \tag{8}$$

The dependence of Ψ on x and t is implicit, so, we have introduced the function G as a reparametrization that allows to rewrite Ψ as an explicit function of x, t and ψ. In this way, we have at hand a mechanism to map any solution of (2) into a solution of (7). The explicit form of the relationships in (8) is determined through the total derivatives $\frac{d\Psi}{dx}$, $\frac{d\Psi}{dt}$ and $\frac{d^2\Psi}{dx^2}$. It allows finding relationships between the partial derivative of the initial and final models, leading to the forms

$$\frac{\partial \Psi}{\partial \tau} = G_1\left(x,t;\psi,\frac{\partial\psi}{\partial x},\frac{\partial\psi}{\partial t}\right), \quad \frac{\partial^2\Psi}{\partial y^2} = G_2\left(x,t;\psi,\frac{\partial\psi}{\partial x},\frac{\partial^2\psi}{\partial x^2},\frac{\partial\psi}{\partial t}\right). \tag{9}$$

The latter leads in general to nonlinear terms, but the conditions [10]

$$\Psi = G(x,t;\psi) = A(x,t)\psi\,, \quad \tau = \tau(t)\,, \tag{10}$$

remove such nonlinearities. With (10), and after some calculations, it can be shown that the relationships in (9) are written as

$$\begin{aligned}
\Psi_\tau &= \frac{A}{\tau_t}\left[-\frac{y_t}{y_x}\psi_x + \psi_t + \left(\frac{A_t}{A} - \frac{y_t}{y_x}\frac{A_x}{A}\right)\psi\right], \\
\Psi_{y,y} &= \frac{A}{y_x^2}\left[\psi_{x,x} + \left(2\frac{A_x}{A} - \frac{y_{xx}}{y_x}\right)\psi_x + \left(\frac{A_{xx}}{A} - \frac{y_{xx}}{y_x}\frac{A_x}{A}\right)\psi\right],
\end{aligned} \tag{11}$$

where the subindex notation denotes partial derivatives, $f_u = \frac{\partial f}{\partial u}$. The substitution of (10) and (11) into (2) leads, after some arrangements, to

$$i\psi_t = -\frac{1}{2}\frac{\tau_t}{y_x^2}\psi_{x,x} + B(x,t)\psi_x + V(x,t)\psi, \tag{12}$$

with

$$
\begin{aligned}
B(x,t) &= i\frac{y_t}{y_x} - \frac{1}{2}\frac{\tau_t}{y_x^2}\left(2\frac{A_x}{A} - \frac{y_{xx}}{y_x}\right), \\
V(x,t) &= -i\left(\frac{A_t}{A} - \frac{y_t}{y_x}\frac{A_x}{A}\right) - \frac{1}{2}\frac{\tau_t}{y_x^2}\left(\frac{A_{xx}}{A} - \frac{y_{xx}}{y_x}\frac{A_x}{A}\right) + \frac{\tau_t}{2}y^2(x,t).
\end{aligned}
\tag{13}
$$

Given that (12) must be of the form (7), we impose the conditions

$$
\frac{\tau_t}{y_x^2} = \frac{1}{m(t)}, \quad B(x,t) = 0. \tag{14}
$$

To simplify the calculations, it is convenient to introduce the real-valued functions $\mu(t)$ and $\sigma(t)$ such that $\tau_t = \sigma^{-2}(t)$ and $m(t) = \mu^2(t)$. From the first condition in (14) we get

$$
\tau(t) = \int^t \frac{dt'}{\sigma^2(t')}, \quad y(x,t) = \frac{\mu(t)x + \gamma(t)}{\sigma(t)}, \tag{15}
$$

where the real-valued function $\gamma(t)$ results from the integration with respect to x. From $B(x,t) = 0$ we obtain $A(x,t)$ as

$$
A(x,t) = \exp\left[i\frac{\mu}{\sigma}\left(\frac{\mathcal{W}_\mu}{2}x^2 + \mathcal{W}_\gamma x + \eta\right)\right], \tag{16}
$$

where $\eta(t)$ is a complex-valued function resulting from the integration with respect to x and

$$
\mathcal{W}_\mu(t) = \sigma\dot{\mu} - \dot{\sigma}\mu, \quad \mathcal{W}_\gamma(t) = \sigma\dot{\gamma} - \dot{\sigma}\gamma, \tag{17}
$$

with $\dot{f} = \frac{df}{dt}$. With (16), the new time-dependent potential $V(x,t)$ in (13) takes the form

$$
\begin{aligned}
V(x,t) &= \frac{\mu^2}{2}\left(\frac{\dot{\mathcal{W}}_\mu}{\mu\sigma} + \frac{1}{\sigma^4}\right)x^2 + \mu\left(\frac{\dot{\mathcal{W}}_\gamma}{\sigma} + \frac{\gamma}{\sigma^4}\right)x + V_0(t), \\
V_0(t) &= \frac{\mathcal{W}_\mu\xi}{\sigma^2} + \frac{\mu\dot{\xi}}{\sigma} - \frac{\mathcal{W}_\gamma^2}{2\sigma^2} + \frac{\gamma^2}{2\sigma^4} - i\frac{\mathcal{W}_\mu}{2\mu\sigma}.
\end{aligned}
\tag{18}
$$

After comparing (18) with the potential energy term in (7) we obtain a system of equations for σ, γ and η that, without loss of generality, reduces to quadratures by considering[1] $V_0(t) = 0$. We thus have

[1]For $V_0(t) \neq 0$, the solutions are just modified by adding a global complex-phase, for details see App. B of [10].

$$\ddot{\sigma}+\left(\Omega^2-\frac{\ddot{\mu}}{\mu}\right)\sigma=\frac{1}{\sigma^3},\quad \ddot{\gamma}+\left(\Omega^2-\frac{\ddot{\mu}}{\mu}\right)\gamma=\frac{F}{\mu},\quad \frac{\mu}{\sigma}\eta=\xi-\frac{i}{2}\ln\frac{\sigma}{\mu}, \tag{19}$$

where the real-valued function $\xi(t)$ is given by

$$\xi(t)=\frac{\gamma\mathcal{W}_\gamma}{2\sigma}-\frac{1}{2}\int^t dt'\frac{F(t')\gamma(t')}{\mu(t')}. \tag{20}$$

From (19) it follows that $\sigma(t)$ satisfies the Ermakov equation [13], whose solutions are well-known in the literature [13–15]. In general, for a set of nonnegative parameters $\{a, b, c\}$ we have [13]

$$\sigma(t)=\left[aq_1^2(t)+bq_1(t)q_2(t)+cq_2^2(t)\right]^{1/2},\quad b^2-4ac=-\frac{4}{W_0^2}, \tag{21}$$

where q_1 and q_2 are two linearly independent real solutions of the linear equation

$$\ddot{q}_{1,2}+\left(\Omega^2-\frac{\ddot{\mu}}{\mu}\right)q_{1,2}=0\,, \tag{22}$$

and the Wronskian $W(q_1, q_2) = W_0$ is a constant. The constraint in the constants a, b, c ensures that $\sigma > 0$ at any time [14, 15]. Thus, the transformed coordinate $y(x, t)$ and time parameter $\tau(t)$ are free of singularities at any time. Notice that (21) corresponds to the classical equation of motion of the parametric oscillator [4, 5]. On the other hand, $\gamma(t)$ is a solution to the classical parametric oscillator subjected to a driving force $F(t)$. In general, γ can be expressed as the sum of the homogeneous solution $\gamma_h = \gamma_1 q_1 + \gamma_2 q_2$ and the particular solution $\gamma_p(t)$, where the real constants $\gamma_{1,2}$ are fixed according to the initial conditions and the function $\gamma_p(t)$ is determined once the driving force $F(t)$ has been specified. Moreover, the function τ introduced in (15) can be rewritten in terms of q_1 and q_2 as well, leading to [15]

$$\tau(t)=\int^t\frac{dt'}{\sigma^2(t')}=\arctan\left[\frac{W_0}{2}\left(b+2c\frac{q_2}{q_1}\right)\right]. \tag{23}$$

From (10) and with the functions σ, γ and τ already identified, the solutions to the Schrödinger equation (7) are simply given in terms of the solutions of the stationary oscillator $\Psi(y, \tau)$ as

$$\psi(x,t)=\exp\left[-i\frac{\mu}{\sigma}\left(\frac{\mathcal{W}_\mu}{2}x^2+\mathcal{W}_\gamma x\right)-i\xi\right]\sqrt{\frac{\mu}{\sigma}}\Psi(y(x,t),\tau(t))\,. \tag{24}$$

That is, the solutions of the nonstationary oscillator with time-dependent mass $\hat{H}(t)$ can be seen as a mere deformation of the solutions of the stationary oscillator, provided by the appropriate point transformation. From the latter, it is natural to ask whether the structure of the inner product defined in terms of the new solutions

$\psi(x,t)$ is deformed as well. To this end, let us consider a pair of arbitrary solutions of the stationary oscillator, $\Psi_{(1)}(y,\tau)$ and $\Psi_{(2)}(y,\tau)$. Straightforward calculations show that

$$\begin{aligned}\langle\Psi_{(2)}(\tau)|\Psi_{(1)}(\tau)\rangle &= \int_{-\infty}^{\infty} dy\, {\Psi_{(2)}}^*(y,\tau)\Psi_{(1)}(y,\tau)\\ &= \int_{-\infty}^{\infty} dx\, {\psi_{(2)}}^*(x,t)\psi_{(1)}(x,t) = \langle\psi_{(2)}(t)|\psi_{(1)}(t)\rangle, \end{aligned} \tag{25}$$

that is, the point transformation preserves the structure of the inner product.

3.1 Orthogonal Set of Solutions and the Related Spectral Problem

With the transformation rule (24), we can find an orthogonal set of solutions for (7). From the preservation of the inner product (25), it is natural to consider the orthogonal solutions of the stationary oscillator $\Psi_n(y,\tau)$, where $\Psi_n(y,\tau) = e^{-i(n+1/2)\tau}\Phi_n(y)$ and $\Phi_n(y)$ is given in (5). Hence, the orthogonal set of solutions $\{\Psi_n(y,\tau)\}_{n=0}^{\infty}$ maps into the orthogonal set $\{\psi_n(x,t)\}_{n=0}^{\infty}$, where

$$\psi_n(x,t) = e^{-i(n+1/2)\tau(t)}\varphi_n(x,t), \tag{26}$$

with $\varphi_n(x,t) = A^{-1}(x,t)\Phi_n\left(y(x,t)\right)$ written as

$$\begin{aligned}\varphi_n(x,t) = \exp\left\{-\left(\frac{\mu^2}{\sigma^2}+i\frac{\mu\mathcal{W}_\mu}{\sigma}\right)\frac{x^2}{2}-\left(\frac{\mu\gamma}{\sigma^2}+i\frac{\mu\mathcal{W}_\gamma}{\sigma}\right)x-\left(\frac{\gamma^2}{2\sigma^2}+i\xi\right)\right\}\\ \times\sqrt{\frac{1}{2^n n!\sqrt{\pi}}}\sqrt{\frac{\mu}{\sigma}}H_n\left(\frac{\mu x+\gamma}{\sigma}\right).\end{aligned} \tag{27}$$

Notice that the orthogonality condition obtained from (25) holds provided that both solutions are evaluated at the same time, that is, $\langle\psi_n(t)|\psi_m(t)\rangle = \delta_{n,m}$. In turn, the orthogonality cannot be taken for granted at different times, $\langle\psi_m(t')|\psi_n(t)\rangle \neq \delta_{n,m}$ for $t \neq t'$. Moreover, the space of solutions generated with (26) is dynamic, $\mathcal{H}(t) = \text{Span}\{|\psi_n(t)\rangle\}_{n=0}^{\infty}$. Such a property is beyond the scope of this work and will be studied elsewhere. For information on the matter see [16].

We have shown the orthonormality of the solutions $\psi_n(x,t)$, however, it is necessary to emphasize that they are not eigenfunctions of the Hamiltonian $\hat{H}(t)$. Nevertheless, the functions $\psi_n(x,t)$ are admissible from the physical point of view. Since $\hat{H}(t)$ is not a constant of motion of the system $\frac{d}{dt}\hat{H}(t) \neq 0$, we wonder about the observable that defines the system uniquely so that it includes the set $\{\psi_n(x,t)\}_{n=0}^{\infty}$ as its eigenfunctions. Moreover, what about the related spectrum?

The latter points must be clarified in order to provide the functions (26), and any linear combination of them, with a physical meaning.

Remarkably, such information is obtained from the point transformation itself, because any conserved quantity is preserved [12]. Indeed, from (5) we see that the energy eigenvalues $E_n = (n + 1/2)$ of the stationary oscillator must be preserved since they are constant quantities. To be specific, using the relationships (11) together with $A(x, t)$=(16), the stationary eigenvalue equation (3) deforms into

$$-\frac{\sigma^2}{2\mu^2}\frac{\partial^2\varphi_n}{\partial x^2}+\frac{1}{2}\left(\mathcal{W}_\mu^2+\frac{\mu^2}{\sigma^2}\right)x^2\varphi_n-\frac{\sigma\mathcal{W}_\mu}{2\mu}i\left(2x\frac{\partial}{\partial x}+1\right)\varphi_n$$
$$-\frac{\sigma\mathcal{W}_\gamma}{\mu}i\frac{\partial\varphi_n}{\partial x}+\left(\mathcal{W}_\mu\mathcal{W}_\gamma+\frac{\mu\gamma}{\sigma^2}\right)x\varphi_n+\frac{1}{2}\left(\mathcal{W}_\gamma^2+\frac{\gamma^2}{\sigma^2}\right)\varphi_n=E_n\varphi_n, \tag{28}$$

where the eigenvalues $E_n = (n + 1/2)$ have been inherited from the stationary oscillator. It is immediate to identify the operator

$$\hat{I}(t)=\frac{\sigma^2}{2\mu^2}\hat{p}_x^2+\frac{1}{2}\left(\mathcal{W}_\mu^2+\frac{\mu^2}{\sigma^2}\right)\hat{x}^2+\frac{\sigma\mathcal{W}_\mu}{2\mu}(\hat{x}\hat{p}_x+\hat{p}_x\hat{x})+\frac{\sigma\mathcal{W}_\gamma}{\mu}\hat{p}_x$$
$$+\left(\mathcal{W}_\mu\mathcal{W}_\gamma+\frac{\mu\gamma}{\sigma^2}\right)\hat{x}+\frac{1}{2}\left(\mathcal{W}_\gamma^2+\frac{\gamma^2}{\sigma^2}\right)\mathbb{I}(t), \tag{29}$$

with $\mathbb{I}(t)=\sum_{n=0}^{\infty}|\psi_n(t)\rangle\langle\psi_n(t)|$ the representation of the identity operator in $\mathcal{H}(t)$. The invariant operator $\hat{I}(t)$ is such that we get the eigenvalue equation $\hat{I}(t)|\varphi_n(t)\rangle = (n + 1/2)|\varphi_n(t)\rangle$. Besides, straightforward calculations show that $\hat{I}(t)$ satisfies the invariant condition

$$\frac{d}{dt}\hat{I}(t)=i[\hat{H}(t),\hat{I}(t)]+\frac{\partial}{\partial t}\hat{I}(t)=0. \tag{30}$$

That is, $\hat{I}(t)$ is an integral of motion of the parametric oscillator. Remark that the invariant operator $\hat{I}(t)$ arises naturally from the point transformation, without requiring any ansatz, contrary to [6]. Moreover, with $\hat{I}(t)$ and (26) we find

$$\psi_n(x,t)=e^{-i\hat{I}(t)\tau(t)}\varphi_n(x,t)=e^{-iw(n+1/2)\tau(t)}\varphi_n(x,t). \tag{31}$$

Thus, we can conclude that the time-dependent complex-phase of the Lewis and Riesenfeld approach [6] coincides with the exponential term in (31), that is, such a phase is proportional to the deformed time parameter $\tau(t)$. Notice that, contrary to the stationary case, the operator $e^{-i\hat{I}(t)\tau(t)}$ in (31) is not the time evolution operator. It is worth to mention that, for $\gamma_1 = \gamma_2 = F(t) = 0$ and a constant mass $m(t) = \mu(t) = 1$, the operator (29) coincides with the invariant of Lewis and Riesenfeld [6].

4 Concluding Remarks

It was shown that a set of orthonormal solutions for the time-dependent mass nonstationary oscillator can be found by constructing the appropriate point transformation and deforming the solutions of the stationary oscillator. The latter is possible since the point transformation preserves the structure of the inner product. Although the Hamiltonian depends explicitly on time, a spectral problem can be identified for the appropriate constant of motion that emerges from the transformed Hamiltonian of the stationary oscillator. The procedure has been developed in general for any time-dependent mass, frequency and an external driving force. Among the examples that could be addressed we have the Caldirola–Kanai oscillator, which leads to the quantum Arnold transformation [17], and the Hermite oscillator [18]. A detailed discussion will be provided elsewhere.

Acknowledgments K.Z. acknowledges the support from the *Mathematical Physics Laboratory*, Centre de Recherches Mathématiques, through a postdoctoral fellowship. He also acknowledges the support by *Conacyt* (Mexico), grant number A1-S-24569. V.H. acknowledges the support of research grants from NSERC of Canada.

References

1. D.E. Pritchard, Phys. Rev. Lett. **51**, 1336 (1983)
2. M. Combescure, Ann. Inst. Henri Poincaré A **44**, 293 (1986)
3. B.M. Mihalcea, Phys. Scr. **2009**, 014006 (2009)
4. R.J. Glauber, in *Proceedings of the International Enrico Fermi School, Course 118, Varenna, Italy, July 1–19, 1992*, ed. by E. Arimondo, W.D. Philips, F. Sttrumia (North Holland, Amsterdam, 1992), p. 643
5. H.R. Lewis, J. Math. Phys. **9**, 1976 (1968)
6. H.R. Lewis, Jr., W.B. Riesenfeld, J. Math. Phys. **10**, 1458 (1969)
7. V.V. Dodonov, V.I. Man'ko, L. Rosa, Phys. Rev. A **57**, 2851 (1998)
8. J.R. Ray, Phys. Rev. A. **26**, 729 (1982)
9. F. Finkel, A. González-López, N. Kamran, M.A. Rodríguez, J. Math. Phys. **40**, 3268 (1999)
10. K. Zelaya, O. Rosas-Ortiz, Phys. Scr. **95**, 064004 (2020)
11. F.W.J. Olver, et al. (eds.), *NIST Handbook of Mathematical Functions* (Cambridge University Press, New York, 2010)
12. W.H. Steeb, *Invertible Point Transformations and Nonlinear Differential Equations* (World Scientific Publishing, Singapore, 1993)
13. V. Ermakov, Kiev University Izvestia, Series III **9**, 1 (1880) (in Russian). English translation by A.O. Harin, Appl. Anal. Discrete Math. **2**, 123 (2008)
14. O. Rosas-Ortiz, O. Castaños, D. Schuch, J. Phys. A: Math. Theor. **48**, 445302 (2015)
15. Z. Blanco-Garcia, O. Rosas-Ortiz, K. Zelaya, Math. Meth. Appl. Sci. **42**, 4925 (2019)
16. A. Mostafazadeh, Phys. Rev. D **98**, 046022 (2018)
17. J. Guerrero, V. Aldaya, F.F. López-Ruiz, F. Cossio, Int. J. Geom. Meth. Mod. **9**, 126011 (2012)
18. N. Ünal, J. Math. Phys. **59**, 062104 (2018)

Influence of the Electron–Phonon Interaction on the Topological Phase Transition in BiTeI

Véronique Brousseau-Couture and Michel Côté

Abstract The topological order of a material is intrinsically related to its bulk electronic structure. Topological phase transitions require that the bulk band gap vanishes. However, the gap is affected by atomic motion through electron–phonon interaction, even at $T = 0\,\mathrm{K}$. As a consequence, electron–phonon interaction can either promote or suppress topologically non-trivial phases. In this work, the temperature dependence of the pressure-induced topological phase transition in Rashba semiconductor BiTeI is investigated through first-principles methods. We first present an overview of electron–phonon interaction within the framework of density-functional perturbation theory (DFPT) and derive a qualitative argument to understand how it will affect the band gap for both typical semiconductors and topological insulators. Then, by tracking both the pressure and temperature dependence of the bulk band gap, we show how the Weyl semimetal and topological insulator phases of BiTeI evolve with temperature, thus providing a guideline for experimental detection.

Keywords Electron–phonon interaction · Density-functional theory · Topological phase transitions

1 Introduction

Since their theoretical prediction in the 1980s, the study of topological phases of matter has grown to become a very active field of research in condensed matter physics. A peculiarity of such phases resides in the fact that they cannot be explained by Landau theory, as transitioning from a topologically trivial to a non-trivial phase does not involve any symmetry breaking. From the bulk-boundary correspondence principle [1], the non-trivial phases are characterized by symmetry-

V. Brousseau-Couture (✉) · M. Côté
Université de Montréal and Regroupement Québécois sur les Matériaux de Pointe, Montréal, QC, Canada
e-mail: veronique.brousseau.couture@umontreal.ca; michel.cote@umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_29

protected boundary states, since one cannot go from a bulk insulator to a topological insulator (TI) without closing the bulk band gap.

First-principles studies have been widely used to predict and characterize new topological materials since they do not make any assumptions on the system and do not rely on external parameters besides the crystal structure. However, most of these calculations are done under the assumption of the static lattice, while experiments are inherently done at finite temperature, where atoms are necessarily in motion. Such calculations thus disregard how an increasing population of thermally activated phonons affects material properties.

1.1 Electron–Phonon Interaction in Semiconductors

When studying topological insulators from first-principles, one aims to identify topologically non-trivial electronic band structures, which manifest themselves through a local inversion of the leading orbital character of the valence and conduction bands in the vicinity of their respective extrema (*band inversion*) [1]. To accurately predict the topology of a given band structure at some finite temperature, one must, therefore, assess how the temperature affects the electronic structure. In analogy with the Lamb shift in QED, where the electronic energy levels are modified by an interaction with the vacuum fluctuations, in condensed matter, electrons interact with phonons even at $T = 0\,\mathrm{K}$, through the zero-point motion of the ions. This electron–phonon interaction (EPI) gives rise to many distinctive features in the band structure, such as a variation of the band gap energy with temperature, band broadening induced by the finite lifetime of electronic excitations, the presence of replica bands, and the formation of kinks near the Fermi level for metals [2].

1.2 Phonon-Induced Topological Insulation

From this point a view, since a topological phase transition requires the bulk band gap to close and EPI governs the temperature dependence of the electronic structure, one can naturally wonder if EPI could be strong enough to close the gap and drive a topological phase transition (TPT). This question was first addressed by Saha and Garate [3, 4] through model Hamiltonians, and later investigated with first-principles calculations for $BiTl(S_{1-\delta}Se_{\delta})_2$ [5] and Sb_2Se_3 [6]. In those last two studies, the TPT is driven by an experimentally controllable parameter, respectively, stoichiometric substitution and hydrostatic pressure. TPTs can, therefore, be detected by tracking the variation of the bulk electronic band gap, which must vanish at the critical parameter. Since temperature affects the gap energy, the critical parameter at which the TPT takes place evolves with temperature. Understanding how temperature influences the topological properties of such materials is crucial for the efficient design of technological applications

that would rely on the robust topologically protected metallic surface states, such as spintronics, quantum computing, or topological transistors.

2 Electron–Phonon Interaction from First-Principles

In this section, we briefly present how EPI can be computed from first-principles. For more details, we refer our readers to the review by Giustino [2]. We use the Hartree atomic unit system ($\hbar = m_e = e = 1$). Let us start from the Hamiltonian:

$$H = \sum_{\mathbf{k}n} \varepsilon^0_{\mathbf{k}n} c^\dagger_{\mathbf{k}n} c_{\mathbf{k}n} + \sum_{\mathbf{q}\nu} \omega_{\mathbf{q}\nu} \left(a^\dagger_{\mathbf{q}\nu} a_{\mathbf{q}\nu} + \tfrac{1}{2} \right) + \sum_{\mathbf{k}nn',\mathbf{q}\nu} g_{\mathbf{k}nn',\mathbf{q}\nu} c^\dagger_{\mathbf{k}+\mathbf{q}n'} c_{\mathbf{k}n} \left(a_{\mathbf{q}\nu} + a^\dagger_{-\mathbf{q}\nu} \right). \quad (1)$$

The first term describes Bloch electrons, with wavevector $\mathbf{k}$, band index n, and bare electronic eigenenergy $\varepsilon^0_{\mathbf{k}n}$. The second accounts for the phonons, with wavevector $\mathbf{q}$, branch index ν, and frequency $\omega_{\mathbf{q}\nu}$. The last term captures the interaction between the $\mathbf{k}n$-electron and the $\mathbf{q}\nu$-phonon, with EPI vertex $g_{\mathbf{k}nn',\mathbf{q}\nu}$. $c^\dagger_{\mathbf{k}n}$, $c_{\mathbf{k}n}$ and $a^\dagger_{\mathbf{q}\nu}$, $a_{\mathbf{q}\nu}$ are, respectively, the creation and annihilation operators for electrons and phonons.

Within many-body perturbation theory, the temperature-dependent correction to the electronic eigenenergies corresponds to the energy of the quasiparticle peak in the real part of the frequency-dependent electron–phonon self-energy, $\Sigma^{\mathrm{EPI}}_{\mathbf{k}n}(T, \omega)$. Applying the on-the-mass-shell approximation, thus evaluating the self-energy at the poles of the Green's function, namely at the bare electronic eigenvalues, the renormalized eigenenergies take the form:

$$\varepsilon_{\mathbf{k}n}(T) \approx \varepsilon^0_{\mathbf{k}n} + \mathfrak{Re}\left[\Sigma^{\mathrm{EPI}}_{\mathbf{k}n}\left(T, \omega = \varepsilon^0_{\mathbf{k}n}\right) \right]. \quad (2)$$

We work within the Migdal approximation, where the EPI vertex corrections can be neglected because of the large mass difference between electrons and ions. Expanding the third term of Eq. (1) to the lowest order, one obtains a contribution known in the literature as the Fan term (Fig. 1, middle):

$$\Sigma^{\mathrm{Fan}}_{\mathbf{k}n}\left(T, \omega = \varepsilon^0_{\mathbf{k}n}\right) = \sum_{\mathbf{q}\nu} \sum_{n'} \frac{1}{2\omega_{\mathbf{q}\nu}} \left| \langle \psi_{\mathbf{k}n} | \, \delta V^{(1)}_{\mathbf{q}\nu} \, | \psi_{\mathbf{k}+\mathbf{q}n'} \rangle \right|^2 \times \left[\frac{n_{\mathbf{q},\nu}(T) + f_{\mathbf{k}+\mathbf{q},n'}(T)}{\omega - \varepsilon^0_{\mathbf{k}+\mathbf{q}n'} + \omega_{\mathbf{q}\nu} + i\eta_{\mathbf{k}}} + \frac{n_{\mathbf{q},\nu}(T) + 1 - f_{\mathbf{k}+\mathbf{q},n'}(T)}{\omega - \varepsilon^0_{\mathbf{k}+\mathbf{q}n'} - \omega_{\mathbf{q}\nu} + i\eta_{\mathbf{k}}} \right]. \quad (3)$$

In this expression, $|\psi_{\mathbf{k}n}\rangle$ is a static lattice eigenstate and $\eta_{\mathbf{k}}$ is an infinitesimal parameter introduced to maintain causality. $\delta V^{(1)}_{\mathbf{q}\nu}$ is the first-order variation of

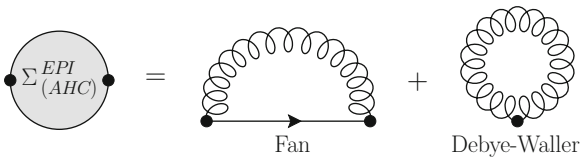


Fig. 1 Feynman diagrams for the Fan and Debye–Waller contributions to the EPI self-energy, Σ^{EPI}, within the Allen–Heine–Cardona (AHC) formalism

the ionic potential created by the $\mathbf{q}\nu$-phonon. The two terms inside the square brackets, respectively, describe phonon absorption and emission processes. The whole temperature dependence is captured by the Fermi–Dirac and Bose–Einstein occupation factors, $f_{\mathbf{k}+\mathbf{q}n'}(T)$ and $n_{\mathbf{q}\nu}(T)$.

Since the Fan term is a second-order term in first-order perturbation theory (it has two first-order vertices in δV), for consistency we must also include the contribution of a first-order term in second-order perturbation theory (one second-order vertex in δV), known as the Debye–Waller term (Fig. 1, right). Approximating the self-energy to the sum of the Fan and Debye–Waller terms is known in the literature as the Allen–Heine–Cardona (AHC) theory [7].

2.1 A Competition Between Intraband and Interband Couplings

We will now consider a simple two-band model to understand qualitatively how EPI impacts the band gap. In a typical semiconductor, the Fan contribution (Eq. (3)) usually dominates the self-energy. The sum on band index, n', can be split into contributions from the occupied and unoccupied subsets of bands. Couplings within a given subset are mapped onto intraband interactions, while couplings between different subsets are mapped to interband interactions.

One can see from Eq. (3) that the sign of each contribution to the self-energy is entirely governed by the sign of the energy difference between the coupled electronic states in the denominators. Thus, when considering intraband couplings (Fig. 2a, left), the valence band maximum (VBM) energy increases since it couples to states of lower energy, while the conduction band minimum (CBM) energy decreases since it couples to states of higher energies, thus reducing the band gap energy. Those behaviors are reversed for interband couplings, such that the band gap energy increases. The leading interaction will dictate the sign of the total band gap renormalization. In a typical semiconductor, the band gap is usually of the order of eV, such that interband couplings are strongly disfavored because of a larger energy difference in the denominators. Hence, intraband couplings dominate the EPI, leading to a band gap closing with increasing temperature, known in the literature as the Varshni effect.

For a topological insulator, the situation is more subtle, since we must also consider the leading orbital character of the valence and conduction bands, and the band inversion phenomenon that occurs in the TI phase. Let us consider the case of

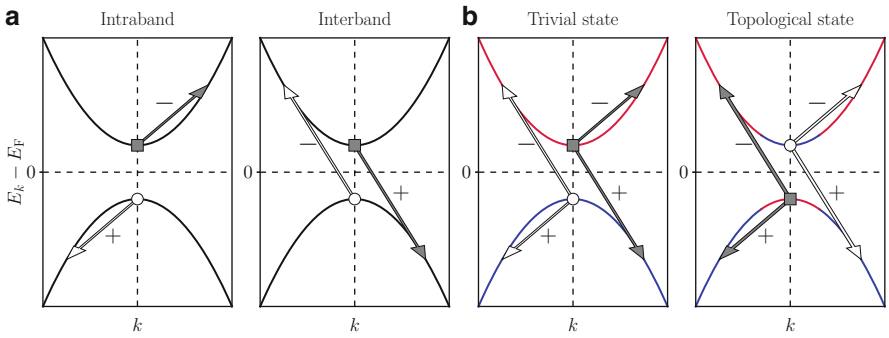


Fig. 2 Heuristic model of the band gap renormalization. (**a**) Intraband and interband couplings in a typical semiconductor. (**b**) In a topological insulator, one must take into account the leading orbital character of the band extremum, which differs between the trivial and topological states due to the band inversion phenomena. The + and − refer to the sign of each contribution to the Fan self-energy (Eq. (3))

the VBM of "blue" character in Fig. 2b (left, white circle). Suppose that, in the trivial phase, the intraband contribution dominates, such that the sum of couplings between bands of the same character ("blue-blue") and between bands of different characters ("blue-red") moves this band extremum to higher energy. This behavior remains the same in the topological phase, but this extremum is now the CBM because of the band inversion. By a similar argument, the band extremum of "red" character (gray square) will be moved to lower energies, both in the trivial phase where it is the CBM and in the TI phase, where it is the VBM. We can, therefore, deduce from this simple heuristic argument that if the EPI interaction closes the gap in the trivial phase, thus facilitation a band inversion, it will open it in the TI phase, further stabilizing an already inverted gap. On the contrary, if EPI opens the gap in the trivial phase, thus delaying the band inversion, it will close it in the TI phase, eventually reversing the band inversion. In summary, intraband contributions globally promote the topological phase, while interband contributions favor the trivial phase.

3 Topological Phase Transition in BiTeI

In this section, we specifically address the EPI contribution to the temperature dependence of the TPT in BiTeI. For a more refined analysis of our results and more details about our methodology, we refer to our forthcoming paper [8].

BiTeI is a trigonal crystal composed of atomic trilayers weakly bound by Van der Waals interaction along the normal crystallographic axis. Due to strong spin–orbit interaction and broken inversion symmetry, it exhibits one of the largest Rashba splitting known so far (∼100 meV) [9]. As a consequence, its well-defined spin-

momentum locking in-plane allows enhanced control of the spin degree of freedom, making this material a promising candidate for spintronic applications.

In 2012, a DFT calculation predicted that BiTeI would turn into a strong $\mathbb{Z}_2$ topological insulator when subjected to hydrostatic pressure [10]. Following this prediction, evidences of the TPT were documented experimentally, using x-ray diffraction and infrared spectroscopy [11], electrical resistivity [12] and Shubnikov–de Haas oscillations [13]. Liu and Vanderbilt later showed that broken inversion symmetry imposed the existence of a mandatory (albeit narrow) Weyl semimetal phase (WSM) between the trivial and topological insulator phases [14]. As a consequence, the TPT of BiTeI exhibits *two* distinct critical pressures: the first when the Weyl nodes are created in the H-A-L plane of the Brillouin zone, and the second when they annihilate each other in the M-L-A mirror planes of the Brillouin zone after migrating in the $\pm k_z$ direction (see Fig. 11 of ref. [14]).

3.1 Static Lattice

Our first step was to reproduce the TPT in the static lattice approximation. To do so, we tracked the band gap energy as a function of pressure as well as the leading orbital character of the valence and conduction bands on either side of the TPT. The TI phase was confirmed by computing the $\mathbb{Z}_2$ topological invariant using a hybrid Wannier charge center analysis [15]. We find the critical pressures $P_{C1} = 2.08$ GPa and $P_{C2} = 2.28$ GPa (gray vertical lines of Fig. 3a). We also observe a band inversion between Bi-$6p_z$ and Te/I- $5p_z$ states (Fig. 3b), as originally predicted [10]. We stress that the current implementation of our DFPT methodology assumes that the gap energy is larger than the maximal optical phonon frequency (dashed line on Fig. 3a). We thus restricted our EPI calculations inside this pressure range (red markers). For more details about the underlying assumptions of our analysis of the temperature-dependent topological phases, we refer to our forthcoming paper [8].

3.2 Phonon-Induced Gap Renormalization

Figure 4a shows the EPI induced temperature-dependent renormalization for the CBM (upper panel), VBM (center), and band gap (bottom). Three main elements can be highlighted from this figure. First, both band extrema show a positive renormalization in the trivial phase, while they shift in opposite directions in the TI phase. As a consequence, the total gap renormalization is much larger in the TI phase, since both contributions reinforce each other rather than cancel out. Second, the Bi extremum (CBM in trivial phase and VBM in TI phase) is almost unaffected by the increased pressure and the change in topology, while the Te/I extremum (VBM in trivial phase and CBM in TI phase) demonstrates a greater dependency with respect to pressure and even changes sign as the system undergoes the TPT.

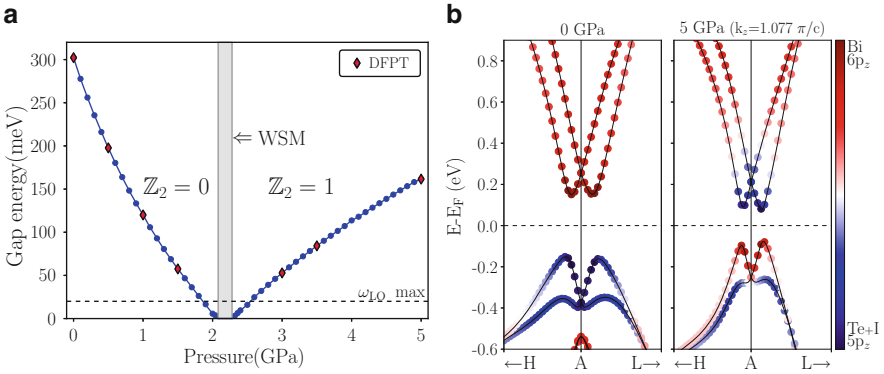


Fig. 3 TPT in the static lattice approximation. (**a**) Indirect band gap energy as a function of pressure. We find a WSM phase width of 0.2 GPa (shaded gray). The red markers locate pressures at which the EPI was explicitly calculated, for pressures where the maximal optical phonon frequency is smaller than the band gap energy. (**b**) Band inversion between Bi-6p_z (red) and Te/I-5p_z (blue) states

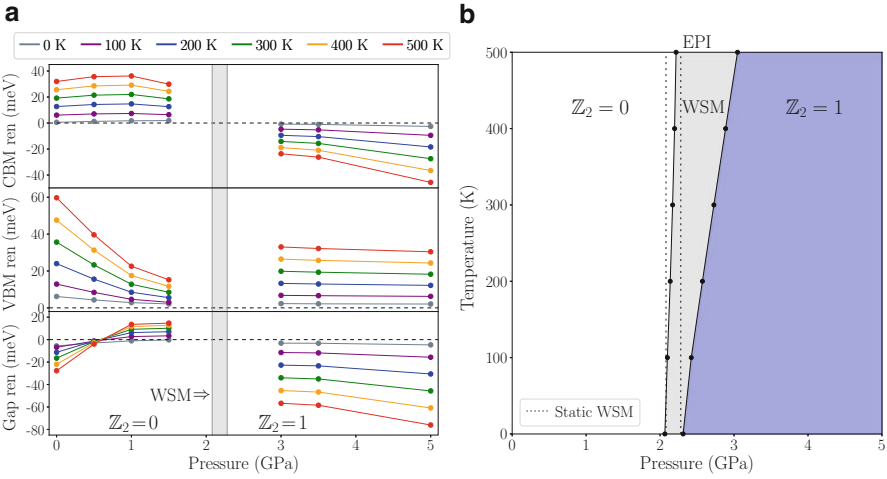


Fig. 4 Electron–phonon interaction contribution to the temperature dependence of the electronic structure. (**a**) Renormalization for CBM (upper panel), VBM (middle), and band gap (bottom). (**b**) Pressure-temperature topological phase diagram of BiTeI, induced by electron–phonon interaction

Lastly, the band gap opens with temperature in the trivial phase in the vicinity of the TPT and closes in the TI phase. Thus, in both phases, EPI is not favorable to the TI phase, as it tends to delay or reverse the band inversion phenomena. From the heuristic model of Sect. 2.1, this suggests that interband couplings between states across the gap play an important role in this material.

By extrapolating the VBM and CBM renormalization towards the TPT within each phase and applying this correction to the static band gap shown in Fig. 3a, we evaluate the temperature dependence of critical pressures P_{C1} and P_{C2}. From the above discussion, the EPI being globally unfavorable to the TI phase leads to both critical pressures being pushed towards higher values for increasing temperature, as seen in the topological phase diagram of Fig. 4b. Nevertheless, the stronger renormalization of P_{C2} compared to P_{C1} leads to a widening of the WSM phase with increasing temperature. While the WSM phase width remains small, it has increased by $\sim$50% at 100K and has more than doubled by 300K. Our findings, therefore, indicate that EPI interaction does not obstruct experimental detection of the WSM phase in BiTeI.

4 Conclusion and Outlook

We have computed the EPI contribution to the temperature dependence of the band gap energy for BiTeI using first-principles methods. We explain the different possible gap behaviors in the light of a simplified heuristic two-band model. We show that the band extrema are affected differently by EPI depending on their leading orbital character and that EPI is globally not favorable to the non-trivial topology in BiTeI. With increasing temperature, both critical pressures are pushed towards higher values, and the WSM phase is widened. A complete description of the temperature dependence of the TPT would require evaluating the thermal expansion contribution [8]. Further analysis of the individual couplings could also provide meaningful insight into how the Rashba interaction affects the EPI strength within and between spin-split bands.

References

1. M.Z. Hasan, C.L. Kane, Colloquim: topological insulators. Rev. Modern Phys. **82**(4), 3045–3067 (2010). https://doi.org/10.1103/RevModPhys.82.3045
2. F. Giustino, Electron-phonon interactions from first principles. Rev. Modern Phys. **89**(1), 015003 (2017). https://doi.org/10.1103/RevModPhys.89.015003
3. I. Garate, Phonon-induced topological transitions and crossovers in Dirac materials. Phys. Rev. Lett. **110**(4), 046402 (2013). https://doi.org/10.1103/PhysRevLett.110.046402
4. K. Saha, I. Garate, Phonon-induced topological insulation. Phys. Rev. B **89**(20), 205103 (2014). https://doi.org/10.1103/PhysRevB.89.205103
5. G. Antonius, S.G. Louie, Temperature-induced topological phase transitions: promoted versus suppressed nontrivial topology. Phys. Rev. Lett. **117**(24), 246401 (2016). https://doi.org/10.1103/PhysRevLett.117.246401
6. B. Monserrat, D. Vanderbilt, Temperature dependence of the bulk Rashba splitting in the bismuth tellurohalides. Phys. Rev. Mat. **1**(5), 054201 (2017). https://doi.org/10.1103/PhysRevMaterials.1.054201

7. P.B. Allen, V. Heine, Theory of the temperature dependence of electronic band structures. J. Phys. C Solid State Phys. **9**(12), 2305 (1976). https://doi.org/10.1088/0022-3719/9/12/013
8. V. Brousseau-Couture, G. Antonius, M. Côté, Temperature dependence of the topological phase transition of BiTeI from first principles. Phys. Rev. Research **2**(2), 023185 (2020). https://doi.org/10.1103/PhysRevResearch.2.023185
9. K. Ishikaza, M. Bahramy, S. Murakawa, M. Sakano, S. Shimojima, T. Sonobe, Giant Rashba-type spin splitting in bulk BiTeI. Nat. Mat. **10**(3051), 521–526 (2011). https://doi.org/10.1038/NMAT3051
10. M. Bahramy, B. Yang, R. Arita, N. Nagaosa, Emergence of non-centrosymmetric topological insulating phase in BiTeI under pressure. Nat. Commun. **3**, 1–7 (2012). https://doi.org/10.1038/ncomms1679
11. X. Xi, C. Ma, Z. Liu, Z. Chen, W. Ku, H. Berger, C. Martin, D.B. Tanner, G.L. Carr, Signatures of a pressure-induced topological quantum phase transition in BiTeI. Phys. Rev. Lett. **111**(15), 155701 (2013). https://doi.org/10.1103/PhysRevLett.111.155701
12. Y. Qi, W. Shi, P.G. Naumov, N. Kumar, R. Sankar, W. Schnelle, C. Shekhar, F.C. Chou, C. Felser, B. Yan, S.A. Medvedev, Topological quantum phase transition and superconductivity induced by pressure in the Bismuth tellurohalide BiTeI. Adv. Mat. **29**(18), 1605965 (2017). https://doi.org/10.1002/adma.201605965
13. J. Park, K.H. Jin, Y.J. Jo, E.S. Choi, W. Kang, E. Kampert, J.S. Rhyee, S.H. Jhi, J.S. Kim, Quantum oscillation signatures of pressure-induced topological phase transition in BiTeI. Sci. Rep. **5**, 15973 (2015). https://doi.org/10.1038/srep15973
14. J. Liu, D. Vanderbilt, Weyl semimetals from noncentrosymmetric topological insulators. Phys. Rev. B **90**(15), 155316 (2014). https://doi.org/10.1103/PhysRevB.90.155316
15. A.A. Soluyanov, D. Vanderbilt, Computing topological invariants without inversion symmetry. Phys. Rev. B **83**(23), 235401 (2011). https://doi.org/10.1103/PhysRevB.83.235401

Part III
Condensed Matter Physics

Nonlinear Coherent States for Anisotropic 2D Dirac Materials

E. Díaz-Bautista, Y. Concha-Sánchez, and A. Raya

Abstract We construct nonlinear coherent states for electrons in anisotropic 2D Dirac materials in a homogeneous magnetic field orthogonal to the sample. By solving the anisotropic Dirac equation in Landau gauge, we identify the ladder operators of the system. Nonlinear coherent states are obtained as eigenstates of a generalized annihilation operator with complex eigenvalues which depends on an arbitrary function f of the number operator. The anisotropy effects on these states are analyzed by using the probability density and the Heisenberg uncertainty relation for three different choices of f.

Keywords Coherent states · Anisotropic Dirac matter · Magnetic field

1 Introduction

2-Dimensional Dirac materials (2DDMs) like graphene [1–3], topological insulators [4, 5], and organic conductors [6, 7] stand out because their charge carriers at low-energy are described by a Dirac equation. Recently, the interest to control

E. Díaz-Bautista (✉)
Departamento de Formación Básica Disciplinaria, Unidad Profesional Interdisciplinaria de Ingeniería Campus Hidalgo, del Instituto Politécnico Nacional, Pachuca: Ciudad del Conocimiento y la Cultura, San Agustín Tlaxiaca, Hidalgo, Mexico
e-mail: ediazba@ipn.mx; ediaz@fis.cinvestav.mx

Y. Concha-Sánchez
Facultad de Ingeniería Civil, Universidad Michoacana de San Nicolás de Hidalgo, Edificio C, Ciudad Universitaria, Morelia, Michoacán, México
e-mail: yconcha@umich.mx

A. Raya
Instituto de Física y Matemáticas, Universidad Michoacana de San Nicolás de Hidalgo, Edificio C-3, Ciudad Universitaria, Morelia, Michoacán, México
e-mail: raya@ifm.umich.mx

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_30

their physical properties, e.g., their stiffness, strength, and optical conductivity by manipulation of their mechanical properties emerged, as in straintronics [8].

On the other hand, coherent states (CSs) were conceived by Schrödinger [9] as the most classical states that describe the motion of a particle in a quadratic potential. Although it is not always possible to obtain these states, they have been used in many branches of physics, as optics, atomic, nuclear, condensed matter, and particle physics (see Ref. [10] and references therein). For the harmonic oscillator, the CSs are constructed from the ladder operators a and $a^\dagger$ that together with the identity operator are generators of the Heisenberg–Weyl (HW) algebra. This algebra can be f-deformed by replacing them with [11]

$$\mathcal{A} = af(N) = f(N+1)a, \quad \mathcal{A}^\dagger = f(N)a^\dagger = a^\dagger f(N+1), \quad N = a^\dagger a, \tag{1}$$

where f is a well-behaved real function of the number operator N, verifying

$$[N, \mathcal{A}] = -\mathcal{A}, \quad [N, \mathcal{A}^\dagger] = \mathcal{A}^\dagger, \quad [\mathcal{A}, \mathcal{A}^+] = (N+1)f^2(N+1) - Nf^2(N). \tag{2}$$

Thus, nonlinear coherent states (NLCSs) are introduced as eigenstates of $\mathcal{A}|\alpha\rangle_f = \alpha|\alpha\rangle_f$ [11]. $f(N)$ is selected to guarantee that such states belong to the Hilbert space. NLCS are physically realized as stationary states of the center-of-mass motion of a trapped ion [12] or the vibrations of polyatomic molecules [13]. By generalizing the results in [14] to anisotropic 2DDMs, we aim to obtain a semi-classical description of the effects of magnetic fields and anisotropy in physical properties of these materials [10, 15–18].

For that purpose, this contribution is organized as follows. In Sect. 2 the anisotropic 2D Dirac equation is solved analytically. In Sect. 3 a generalized annihilation operator is presented and the NLCSs are constructed as its eigenstates. We obtain their probability density and the Heisenberg uncertainty relation (HUR). In Sect. 4 we present our conclusions.

2 Anisotropic 2D Dirac Hamiltonian

The isotropic 2D Dirac Hamiltonian $H = v_\mathrm{F}\,\vec{\sigma}\cdot\vec{p}$, where $\vec{\sigma} = (\sigma_x, \sigma_y)$ are the Pauli matrices and $\vec{p}$ is the canonical momentum, may be modified either because the material is inherently anisotropic or it has been mechanically deformed, yielding a Fermi velocity v_F which is no longer isotropic. The anisotropic 2D Dirac equation (see [19] and references therein) in a magnetic field is

$$H\Psi(x,y) = \vec{\sigma}\cdot\overleftrightarrow{v}\cdot\vec{\Pi}\,\Psi(x,y) = (v_{xx}\sigma_x\pi_x + v_{yy}\sigma_y\pi_y)\Psi(x,y) = E\Psi(x,y), \tag{3}$$

where $\overleftrightarrow{v}$ is the 2×2 symmetric Fermi velocity tensor with non-vanishing diagonal components v_{xx} and v_{yy} (see Fig. 1) and $\pi_{x,y} = p_{x,y} + eA_{x,y}$, with $\vec{A}$ denoting the

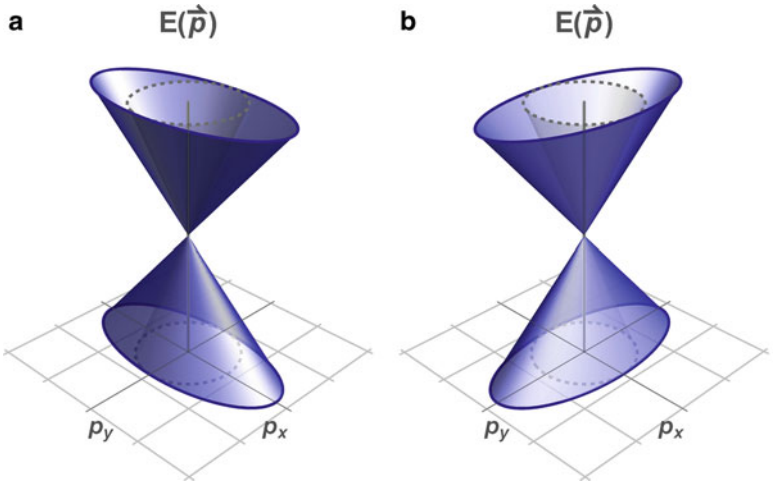


Fig. 1 Dirac cones for an isotropic (dashed gray contour) and anisotropic material (solid blue contour). For the former, their cross-sections are circles and ellipses for latter with semi-major axis along (**a**) the p_x-axis ($v_{xx} < v_{yy}$) or (**b**) the p_y-axis ($v_{xx} > v_{yy}$)

vector potential which defines a magnetic field orthogonal to the sample. In Landau gauge, such that $\vec{A}(x, y) = B_0 x\,\hat{j}$ and $\vec{B} = \nabla \times \vec{A} = B_0\hat{k}$, we can write

$$\Psi(x, y) = \mathcal{N} \exp(iky) \begin{pmatrix} \psi^+(x) \\ i\psi^-(x) \end{pmatrix}, \tag{4}$$

where $\mathcal{N}$ is for normalization. Substituting (4) into (3), and decoupling, we get

$$H_\zeta^\pm \psi^\pm(x) = \left[-\frac{d^2}{dx^2} + \frac{\omega_\zeta^2}{4} \left(x + \frac{2k}{\omega_B} \right)^2 \pm \frac{1}{2}\omega_\zeta \right] \psi^\pm(x) = (\epsilon^\pm)^2 \psi^\pm(x), \tag{5}$$

with $\epsilon^\pm = E/v_{xx}\hbar$, $\omega_\zeta = \omega_B/\zeta$, $\omega_B = 2eB_0/\hbar$ and $\zeta = v_{xx}/v_{yy}$ being the anisotropy parameter. The spectrum is

$$E_0^- = 0, \quad E_n^- = E_{n-1}^+ = \pm\hbar\sqrt{v_{xx} v_{yy}\, \omega_B\, |n|}, \quad n = 0, \pm 1, \pm 2, \ldots. \tag{6}$$

with $\pm$ standing for the conduction and valence bands. Eigenfunctions are

$$\psi_n(x) = \sqrt{\frac{1}{2^n n!} \left(\frac{\omega_\zeta}{2\pi} \right)^{1/2}} \exp\left[-\frac{\omega_\zeta}{4} \left(x + \frac{2k}{\omega_B} \right)^2 \right] H_n\left[\sqrt{\frac{\omega_\zeta}{2}} \left(x + \frac{2k}{\omega_B} \right) \right]. \tag{7}$$

Finally, we take $\mathcal{N}^2 = 2^{(\delta_{0n}-1)}$, $\psi_n^- \equiv \psi_n$ and $\psi_n^+ \equiv \psi_{n-1}$ in (4).

3 Annhilation Operator

Let us define the dimensionless differential operators

$$\theta^{\pm} = \frac{1}{\sqrt{2}}\left(\mp\frac{d}{d\xi} + \xi\right), \quad \theta^{+} = \left(\theta^{-}\right)^{\dagger}, \quad \xi = \sqrt{\frac{\omega_{\zeta}}{2}}\left(x + \frac{2k}{\omega_{\mathrm{B}}}\right), \tag{8}$$

where ξ is the guiding center and the $\theta^{\pm}$ satisfy

$$\theta^{-}\psi_n = \sqrt{n}\psi_{n-1}, \quad \theta^{+}\psi_n = \sqrt{n+1}\psi_{n+1}, \quad [\theta^{-}, \theta^{+}] = 1. \tag{9}$$

Defining $\mathbb{S}_q = \mathrm{diag}(s_q, s_q)$ and $\mathbb{S}_q^2 = \mathrm{diag}(s_q^2, s_q^2)$, $q = 0, 1$, where

$$s_q = \frac{1}{\sqrt{2}i^q}\left(\theta^{-} + (-1)^q\theta^{+}\right), \quad s_q^2 = \frac{1}{2}\left(2N + 1 + (-1)^q((\theta^{-})^2 + (\theta^{+})^2)\right), \tag{10}$$

from $\sigma_{S_q} = \left(\langle\mathbb{S}_q^2\rangle - \langle\mathbb{S}_q\rangle^2\right)^{1/2}$, we have $\sigma_{S_0} \equiv \sigma_{\xi}$ and $\sigma_{S_1} \equiv \sigma_p$ with $\sigma_{\xi}\sigma_p \geq 1/2$.

Likewise, we define a deformed annihilation operator Θ_f^{-} as:

$$\Theta_f^{-} = \begin{bmatrix} \cos(\delta)\frac{\sqrt{N+2}}{\sqrt{N+1}}f(N+2)\theta^{-} & \sin(\delta)\frac{f(N+2)}{\sqrt{N+1}}(\theta^{-})^2 \\ -\sin(\delta)f(N+1)\sqrt{N+1} & \cos(\delta)f(N+1)\theta^{-} \end{bmatrix}, \quad \Theta_f^{+} = (\Theta_f^{-})^{\dagger}, \tag{11}$$

in order to guaranty the action $\Theta_f^{-}\Psi_n = c_n\Psi_n$ on eigenstates in (4), namely:

$$\Theta_f^{-}\Psi_n(x, y) = \frac{f(n)}{\sqrt{2^{\delta_{1n}}}}\exp(i\delta)\sqrt{n}\Psi_{n-1}(x, y), \quad n = 0, 1, 2, \ldots, \tag{12}$$

where $f(N)$ is a well-behaved function of $N = \theta^{+}\theta^{-}$ and $\delta \in [0, 2\pi]$ allows us to consider either diagonal or non-diagonal representation for $\Theta_f^{\pm}$, which satisfy

$$\left[\Theta_f^{-}, \Theta_f^{+}\right] = \begin{bmatrix} \Omega(N+1) & 0 \\ 0 & \Omega(N) \end{bmatrix}, \quad \Omega(N) = (N+1)f^2(N+1) - Nf^2(N). \tag{13}$$

For the choice $f(N) = 1$, one recovers the HW algebra in order to obtain standard-like CSs.

3.1 Nonlinear Coherent States

We construct NLCSs $\Psi_{\alpha}^{f}(x, y)$ as eigenstates of the operator Θ_f^{-}:

$$\Theta_f^{-}\Psi_{\alpha}^{f}(x, y) = \alpha\Psi_{\alpha}^{f}(x, y), \quad \alpha \in \mathbb{C}, \tag{14}$$

which are also a linear combination of Landau states (4),

$$\Psi_\alpha^f(x,y) = \sum_{n=0}^{\infty} a_n \Psi_n(x,y), \tag{15}$$

with the coefficients of expansion a_n verifying the relations

$$a_1 f(1) = \sqrt{2}\tilde{\alpha} a_0, \quad a_{n+1} f(n+1)\sqrt{n+1} = \tilde{\alpha} a_n, \quad \tilde{\alpha} = \alpha \exp(-i\delta). \tag{16}$$

Hence, Θ_f^- introduces a phase δ in the eigenvalue α and a_n is a parameter depending on whether $f(n+1) \neq 0$ or $f(n+1) = 0$.

3.2 Some Examples

We now consider some examples for $f(N)$ with the aim of obtaining coherent states as similar to the standard ones, as well as to discuss the effects of strain on the NLCSs [14, 20]. It is worth to mention that the functions $f(N)$ are chosen according to the relations in Eq. (16).

Case for $f(1) \neq 0$ Choosing $f(N+1) = 1$, that satisfies $f(1) \neq 0$, the corresponding NLCSs, their probability density $\rho_\alpha(x)$ and the mean values of the operators $\mathbb{S}_q$ and $\mathbb{S}_q^2$ in such states (see Fig. 2) are

$$\Psi_\alpha^f(x,y) = \frac{1}{\sqrt{2\exp\left(|\tilde{\alpha}|^2\right) - 1}} \left[\Psi_0(x,y) + \sum_{n=1}^{\infty} \frac{\sqrt{2}\,\tilde{\alpha}^n}{\sqrt{n!}} \Psi_n(x,y) \right], \tag{17a}$$

$$\begin{aligned} \rho_\alpha(x) = \frac{1}{2\exp\left(|\tilde{\alpha}|^2\right) - 1} \Bigg[& \psi_0^2(x) + \left| \sum_{n=1}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \psi_n(x) \right|^2 \\ & + \left| \sum_{n=1}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \psi_{n-1}(x) \right|^2 + 2\,\mathrm{Re}\left(\sum_{n=1}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \psi_n(x) \psi_0(x) \right) \Bigg], \end{aligned} \tag{17b}$$

$$\langle \mathbb{S}_q \rangle_\alpha = \frac{\tilde{\alpha} + (-1)^q \tilde{\alpha}^*}{\sqrt{2} i^q \left(2\exp\left(|\tilde{\alpha}|^2\right) - 1\right)} \left[\exp\left(|\tilde{\alpha}|^2\right) + \sum_{n=1}^{\infty} \frac{|\tilde{\alpha}|^{2n}}{\sqrt{(n-1)!(n+1)!}} \right], \tag{17c}$$

$$\langle \mathbb{S}_q^2 \rangle_\alpha = \frac{1}{2\left(2\exp\left(|\tilde{\alpha}|^2\right) - 1\right)} \left[1 + 4|\tilde{\alpha}|^2 \exp\left(|\tilde{\alpha}|^2\right) + (-1)^q (\tilde{\alpha}^2 + \tilde{\alpha}^{*2}) \right.$$

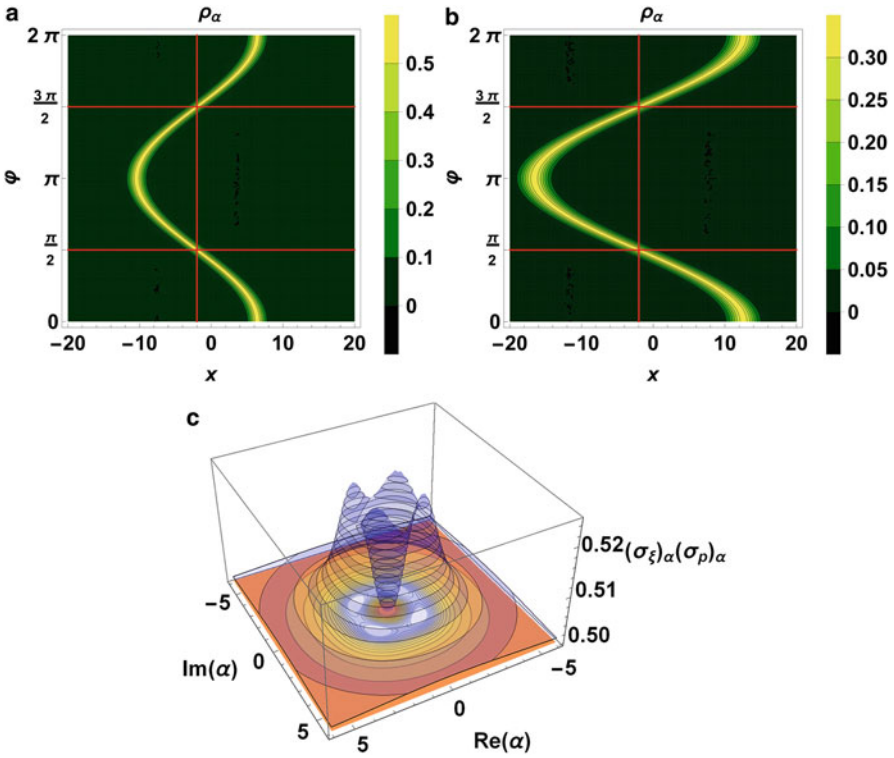


Fig. 2 (**a**, **b**) Probability density $\rho_\alpha(x)$ in Eq. (17b) for $|\alpha| = 6$ and some values of the parameter ζ. For $\varphi = \mathrm{Arg}(\alpha) = (2m+1)\pi/2$, $m = 0, 1, \ldots$, $\rho_\alpha(x)$ is located around the position x_0 (horizontal red lines). (**a, c**) $(\sigma_\xi)_\alpha(\sigma_p)_\alpha$ as function of α. As $|\alpha|$ increases, the HUR tends to the value $1/2$. In these cases, we set $B_0 = 1/2$, $k = \omega_\mathrm{B} = 1$ and $\delta = 0$

$$\times \left[\exp\left(|\tilde{\alpha}|^2\right) + \sum_{n=1}^{\infty} \frac{\sqrt{n+1}\,|\tilde{\alpha}|^{2n}}{\sqrt{(n-1)!(n+2)!}} \right] \Bigg]. \tag{17d}$$

Case for $f(1) = 0$ Assuming that $f(1) = 0$, two new cases arise.

(a) $f(2) \neq 0$. Choosing $f(N+1) = g(N) = \sqrt{N}/\sqrt{N+1}$, the NLCSs, their probability density $\rho_\alpha(x)$ and the mean values of the operators $\mathbb{S}_q$ and $\mathbb{S}_q^2$ are

$$\Psi_\alpha^f(x, y) = \exp\left(-\frac{|\tilde{\alpha}|^2}{2}\right) \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \Psi_{n+1}(x, y), \tag{18a}$$

$$\rho_\alpha(x) = \frac{\exp\left(-|\tilde{\alpha}|^2\right)}{2} \left[\left| \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \psi_{n+1}(x) \right|^2 + \left| \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!}} \psi_n(x) \right|^2 \right], \tag{18b}$$

$$\langle \mathbb{S}_q \rangle_\alpha = \frac{\tilde{\alpha} + (-1)^q \tilde{\alpha}^*}{2\sqrt{2} i^q} \left[1 + \exp\left(-|\tilde{\alpha}|^2\right) \sum_{n=0}^{\infty} \frac{\sqrt{n+2}\, |\tilde{\alpha}|^{2n}}{\sqrt{n!(n+1)!}} \right], \tag{18c}$$

$$\langle \mathbb{S}_q^2 \rangle_\alpha = 1 + |\tilde{\alpha}|^2 + (-1)^q \frac{(\tilde{\alpha}^2 + \tilde{\alpha}^{*2})}{4} \times \left[1 + \exp\left(-|\tilde{\alpha}|^2\right) \sum_{n=0}^{\infty} \frac{\sqrt{n+3}\, |\tilde{\alpha}|^{2n}}{\sqrt{n!(n+1)!}} \right]. \tag{18d}$$

(b) $f(2) = 0$. Finally, considering $f(N+2) = h(N) = \sqrt{N}\sqrt{N+1}/\sqrt{N+2}$, which satisfies $f(2) = 0$, the corresponding NLCSs, their probability density and $\langle \mathbb{S}_q \rangle_\alpha$ and $\langle \mathbb{S}_q^2 \rangle_\alpha$, are (see Fig. 4):

$$\Psi_\alpha^f(x, y) = \left(\frac{|\tilde{\alpha}|}{I_1(2|\tilde{\alpha}|)} \right)^{1/2} \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!(n+1)!}} \Psi_{n+2}(x, y), \tag{19a}$$

$$\rho_\alpha(x) = \left(\frac{|\tilde{\alpha}|}{2\, I_1(2|\tilde{\alpha}|)} \right) \left[\left| \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!(n+1)!}} \psi_{n+2}(x) \right|^2 + \left| \sum_{n=0}^{\infty} \frac{\tilde{\alpha}^n}{\sqrt{n!(n+1)!}} \psi_{n+1}(x) \right|^2 \right], \tag{19b}$$

$$\langle \mathbb{S}_q \rangle_\alpha = \frac{\tilde{\alpha} + (-1)^q \tilde{\alpha}^*}{2\sqrt{2} i^q} \left(\frac{|\tilde{\alpha}|}{I_1(2|\tilde{\alpha}|)} \right) \left[\sum_{n=0}^{\infty} \frac{|\tilde{\alpha}|^{2n}}{\sqrt{n![(n+1)!]^3}} + \sum_{n=0}^{\infty} \frac{\sqrt{n+3}\, |\tilde{\alpha}|^{2n}}{\sqrt{n!(n+2)!(n+1)!}} \right], \tag{19c}$$

$$\langle \mathbb{S}_q^2 \rangle_\alpha = 2 + |\tilde{\alpha}| \frac{I_2(2|\tilde{\alpha}|)}{I_1(2|\tilde{\alpha}|)} + (-1)^q \frac{(\tilde{\alpha}^2 + \tilde{\alpha}^{*2})}{4} \left(\frac{|\tilde{\alpha}|}{I_1(2|\tilde{\alpha}|)} \right) \times \left[\sum_{n=0}^{\infty} \frac{|\tilde{\alpha}|^{2n}}{\sqrt{n!(n+2)!(n+1)!}} + \sum_{n=0}^{\infty} \frac{\sqrt{n+4}\, |\tilde{\alpha}|^{2n}}{\sqrt{n!(n+1)!(n+2)!}} \right], \tag{19d}$$

where $I_1(x)$ denotes the Bessel function of first kind.

In a semi-classical interpretation, the eigenvalue $\alpha = |\alpha| \exp(i\varphi)$ determines the initial conditions of the motion of the electrons. For $\varphi \in [0, 2\pi]$ and a fixed value of ζ, the maximum probability performs an oscillatory motion around the equilibrium position $x_0 = 2k/\omega_B$ (vertical red lines in panels (a) and (b) in Figs. 2, 3 and 4). This fact shows that according to the phase φ, electrons can be found either close to the turning point x or the equilibrium point x_0 at a certain time t, as it happens for a classical particle moving in a closed path with center at x_0, seen from x-axis, due to the Lorentz force $\vec{F} = q\, \vec{v} \times \vec{B}$. Likewise, the distance between the points x, the position of the center of the probability density $\rho_\alpha(x)$ along the x-axis, and x_0 is also affected by ζ, depending on whether $v_{xx} < v_{yy}$ or $v_{xx} > v_{yy}$. On the other hand, panel (c) in Figs. 2, 3 and 4 shows that the HUR behaves differently for each NLCS considered due to the state of minimum energy Ψ_n that contributes to the linear combination of Ψ_α^f.

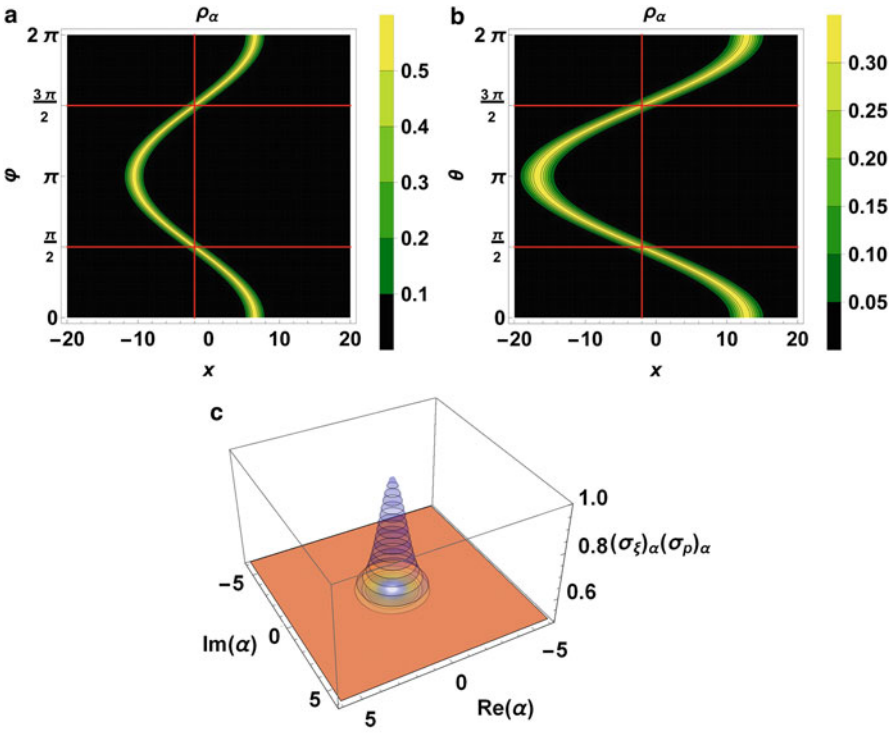


Fig. 3 (**a**, **b**) Probability density $\rho_\alpha(x)$ in Eq. (18a) for $|\alpha| = 6$ and some values of the parameter ζ. For $\varphi = (2m+1)\pi/2$, $m = 0, 1, \ldots$, $\rho_\alpha(x)$ is centered in the position x_0 (horizontal red lines). $(\sigma_\xi)_\alpha(\sigma_p)_\alpha$ as function of α. As $|\alpha|$ increases, the HUR tends to $1/2$. In these cases, we set $B_0 = 1/2$, $k = \omega_B = 1$ and $\delta = 0$

4 Conclusions

In this work, we have supplied a semi-classical description of the effects of anisotropy on the dynamics of the Dirac particles in a magnetic field. Describing the background field in Landau gauge, we construct the coherent states that depend on anisotropy through $\zeta = v_{xx}/v_{yy}$. According to these states, when the anisotropy is aligned to the p_x-axis ($\zeta < 1$), the distance between the turning points x of the semi-classical oscillatory motion of electrons and the equilibrium position $x_0 = 2k/\omega_B$ is less than when the anisotropy is aligned to the p_y-axis ($\zeta > 1$), for which the points x move away from x_0. For the first case, the NLCSs in (19a) could better describe such situation because they tend to oscillate close to the equilibrium point x_0 (Fig. 4), while the NLCSs (17a) and (18a) could be used to describe the second one (Figs. 2 and 3).

We consider that the results obtained in this work can be useful to describe phenomena on 2DDMs by using the coherent states formulation, which has been

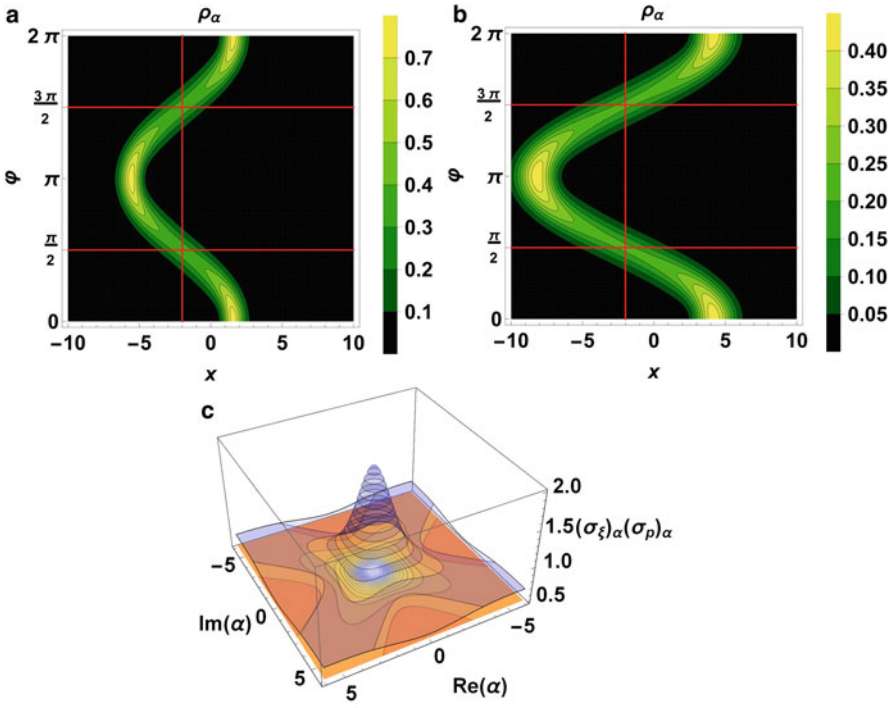


Fig. 4 (**a**, **b**) Probability density $\rho_\alpha(x)$ in Eq. (19b) for $|\alpha| = 6$ and different values of the parameter ζ. For $\varphi = (2m+1)\pi/2$, $m = 0, 1, \ldots$, $\rho_\alpha(x)$ is placed in the position x_0 (horizontal red lines). $(\sigma_\xi)_\alpha(\sigma_p)_\alpha$ as function of α. As φ changes, the HUR shows variations. In all these cases, we set $B_0 = 1/2$, $k = \omega_B = 1$ and $\delta = 0$

extended to the symmetric gauge for the background field [21], in order to describe the bidimensional effects of the anisotropy on these materials. This formalism can also be extended to the situation where the velocities v_{ij} can depend on the spatial coordinates [22].

Acknowledgments YCS acknowledges support from CIC-UMSNH under grant 3820801. AR acknowledges support from Consejo Nacional de Ciencia y Tecnología (México) under grant 256494. EDB acknowledges support from UPIIH of the National Polytechnic Institute (IPN, México) under grant 20201196.

References

1. K.S. Novoselov, et al., Nat. Phys. **2**, 177 (2006). https://doi.org/10.1038/nphys245
2. K.S. Novoselov, et al., Science **315**, 1379 (2007). https://doi.org/10.1126/science.1137201
3. A.K. Geim, K.S. Novoselov, Nat. Mater. **6**, 183 (2007). https://doi.org/10.1038/nmat1849

4. M.Z. Hasan, C.L. Kane, Rev. Mod. Phys. **82**, 3045 (2010). https://doi.org/10.1103/RevModPhys.82.3045
5. X.-L. Qi, S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011). https://doi.org/10.1103/RevModPhys.83.1057
6. S. Katayama, A. Kobayashi, Y. Suzumura, Eur. Phys. J. B **67**, 139 (2009). https://doi.org/10.1140/epjb/e2009-00020-0
7. K. Kajita, et al., J. Phys. Soc. Jpn. **83**, 072002 (2014). https://doi.org/10.7566/JPSJ.83.072002
8. V.M. Pereira, A.H. Castro Neto, N.M.R. Peres, Phys. Rev. B **80**, 045401 (2009) . https://doi.org/10.1103/PhysRevB.80.045401
9. E. Schrödinger, Naturwissenschaften **14**, 664–666 (1926). https://doi.org/10.1007/BF01507634
10. J.R. Klauder, B.S. Skagerstam, *Coherent States: Applications in Physics and Mathematical Physics* (World Scientific, Singapore, 1985). https://doi.org/10.1142/0096
11. V.I. Man'ko, et al., in ed. by N. Atakishiyev, T.H. Seligman, *Proceedings of the IV Wigner Symposium* (World Scientific, New York, 1996), pp. 421. https://doi.org/10.1142/9789814531207; Phys. Scripta **55**, 528 (1997). https://doi.org/10.1088/0031-8949/55/5/004
12. R.L. de Matos Filho, W. Vogel, Phys. Rev. A **54**, 4560 (1996). https://doi.org/10.1103/PhysRevA.54.4560
13. D. Bonatsos, C. Daskaloyannis, Phys. Rev. A **46**, 75 (1992). https://doi.org/10.1103/PhysRevA.46.75; Phys. Rev. A **48**, 3611 (1993). https://doi.org/10.1103/PhysRevA.48.3611
14. E. Díaz-Bautista, D.J. Fernández, Eur Phys. J. Plus **132**, 499 (2017). https://doi.org/10.1140/epjp/i2017-11794-y
15. A. Feldman, A.H. Kahn, Phys. Rev. B **1**, 4584 (1970). https://doi.org/10.1103/PhysRevB.1.4584
16. F.T. Arecchi, et al., Phys. Rev. A **6**, 2211 (1972). https://doi.org/10.1103/physreva.6.2211
17. Y.K. Wang, F.T. Hioe, Phys. Rev. A **7**, 831 (1973). https://doi.org/10.1103/PhysRevA.7.831
18. W.-M. Zhang, et al., Rev. Mod. Phys. **62**, 867 (1990). https://doi.org/10.1103/RevModPhys.62.867
19. M. Oliva-Leyva, C. Wang, Ann. Phys. **384**, 61 (2017). https://doi.org/10.1016/j.aop.2017.06.013
20. E. Díaz-Bautista, Y. Concha-Sánchez, A. Raya, J. Phys. Condens. Mat. **31**, 435702 (2019). https://doi.org/10.1088/1361-648X/ab2d18
21. E. Díaz-Bautista, M. Oliva-Leyva, Y. Concha-Sánchez, A. Raya, J. Phys. A: Math. Theor. **53**, 105301 (2019). https://doi.org/10.1088/1751-8121/ab7035
22. M. Oliva-Leyva, G.G. Naumis, Phys. Lett. A **379**, 2645–2651 (2015). https://doi.org/10.1016/j.physleta.2015.05.039

Monopole Operators and Their Symmetries in QED_3-Gross–Neveu Models

Éric Dupuis, M. B. Paranjape, and William Witczak-Krempa

Abstract Monopole operators are topological disorder operators in $2+1$ dimensional compact gauge field theories appearing notably in quantum magnets with fractionalized excitations. For example, their proliferation in a spin-1/2 kagome Heisenberg antiferromagnet triggers a quantum phase transition from a Dirac spin liquid phase to an antiferromagnet. The quantum critical point (QCP) for this transition is described by a conformal field theory: Compact quantum electrodynamics (QED_3) with a fermionic self-interaction, a type of QED_3-Gross–Neveu model. We obtain the scaling dimensions of monopole operators at the QCP using a state-operator correspondence and a large-N_f expansion, where $2N_f$ is the number of fermion flavors. We characterize the hierarchy of monopole operators at this $SU(2) \times SU(N_f)$ symmetric QCP.

Keywords Topological disorder operators · Gauge theories and dualities · Quantum phase transition · Quantum spin liquids · Conformal field theory

1 Confinement of a Dirac Spin Liquid

An abelian gauge theory consists of the Maxwell term for the gauge field a_μ and potentially some matter coupled through a U(1) charge

$$\mathcal{L} = \frac{1}{2e^2}\left(\epsilon_{\mu\nu\rho}\partial_\nu a_\rho\right)^2 + \text{Matter charged under U(1)}. \tag{1}$$

É. Dupuis (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: eric.dupuis.1@umontreal.ca

M. B. Paranjape · W. Witczak-Krempa
Département de physique, Université de Montréal, Montréal, QC, Canada

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: paranj@lps.umontreal.ca; w.witczak-krempa@umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_31

In 2 + 1 dimensions, the magnetic current is constructed by contracting the rank-3 antisymmetric tensor with the field strength $j^{\mu}_{\text{top}} = \frac{1}{2\pi}\epsilon^{\mu\nu\rho}\partial_\nu a_\rho$. The current conservation then expected is violated when regularizing the theory on the lattice. Indeed, the gauge field a_μ then becomes periodic, taking values in the compact U(1) gauge group. This implies the existence of monopole operators $\mathcal{M}_q^\dagger$ which create gauge field configurations A^q that may be written as[1]

$$A^q = q(1-\cos\theta)\mathrm{d}\phi\,, \tag{2}$$

where the magnetic charge q is half-quantized to respect the Dirac condition. In turn, this implies 2π quantization of the magnetic flux $\Phi = \int_{S^2} \mathrm{d}A^q = 4\pi q$.

Monopole operators may render a gauge theory unstable. In the compact pure U(1) gauge theory in 2 + 1 dimensions, monopole operators are relevant and condense. This leads to confinement and to the emergence of a mass gap [1, 2]. Adding massless matter may, however, stabilize the gauge theory. For a large number N of massless matter flavors, the monopole two-point function is

$$\left\langle \mathcal{M}_q(x)\mathcal{M}_q^\dagger(0)\right\rangle \sim |x|^{-2\Delta_{\mathcal{M}_q}} \approx |x|^{-2N(\dots)}\,, \quad N \gg 1\,, \tag{3}$$

where $\Delta_{\mathcal{M}_q}$ is the scaling dimension of the monopole operator, and the ellipses $(\dots)$ denote a number of order 1 [3, 4]. Monopoles are thus suppressed for a sufficient number of flavors N. Interestingly, a confinement–deconfinement transition can therefore be achieved by tuning an interaction which gaps the massless matter and removes its stabilizing screening effect.

The stability of compact gauge theories plays a key role in strongly correlated systems where fractionalized quasiparticles and gauge excitations emerge. In particular, certain frustrated 2D quantum magnets may be described at low energy by a Dirac spin liquid (DSL). This is a version of quantum electrodynamics in three dimensions (QED_3),

$$\mathcal{L}_{\text{QED}_3} = -\bar{\Psi}\not{D}_a\Psi + \frac{1}{2e^2}\left(\epsilon_{\mu\nu\rho}\partial_\nu a_\rho\right)^2\,, \tag{4}$$

with $2N$ flavors of two-component gapless Dirac fermions $\Psi = \left(\psi_1\,,\,\dots\,,\,\psi_{2N}\right)^{\mathsf{T}}$. The gauge covariant derivative is defined as $\not{D}_a = \gamma_\mu\left(\partial_\mu - ia_\mu\right)$ where γ_μ are the Pauli matrices. The $2N$ fermion flavors can represent the two magnetic spins and N Dirac nodes in momentum space, typically two as well. In particular, many numerical studies suggest that a DSL with $N = 2$ Dirac cones may describe the ground state of the spin-1/2 kagome Heisenberg antiferromagnet [5–10].

The stability of the DSL then hinges on the irrelevance of monopole excitations allowed by the lattice. Whether it is stable or not at $N = 2$ is still an ongoing

[1] In vector notation using spherical coordinates on Euclidean spacetime $\mathbb{R}^3$, it would be written as $A^q_\mu = q(1-\cos\theta)/(r\sin\theta)\delta^\phi_\mu$.

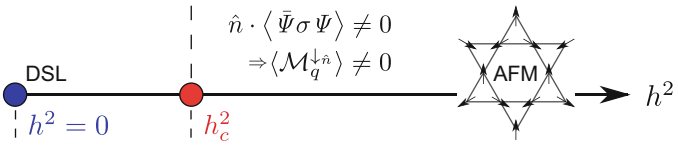


Fig. 1 Phase transition from a Dirac spin liquid to a coplanar antiferromagnet. A condensed spin-Hall mass lets a monopole with spin quantum numbers proliferate

question [11–13]. Here, we suppose a stable DSL and we focus on the possible confinement–deconfinement transition. In this context, the transition may be driven by a Gross–Neveu type interaction $\delta\mathcal{L} = -\frac{1}{2}h^2(\bar{\Psi}T^a\Psi)^2$, where T^a is a generator of the flavor symmetry group SU($2N$). For a sufficiently strong coupling, the corresponding fermion bilinear condenses $\langle \bar{\Psi}T^a\Psi \rangle \neq 0$, allowing monopoles to proliferate. Importantly, different types of monopole operators exist in QED$_3$. This is because half of the $4qN$ fermion zero modes on a monopole must be filled to minimize its scaling dimension and yield a vanishing fermion number [3, 14]. Therefore, the possible zero modes dressings define monopoles with the same scaling dimension but distinct quantum numbers. Which type of monopole proliferates is determined by the type of condensed fermion bilinear and its spontaneously chosen direction.

In particular, by tuning a chiral-Heisenberg Gross–Neveu (cHGN) interaction,

$$\mathcal{L}_{\text{QED}_3\text{-cHGN}} = \mathcal{L}_{\text{QED}_3} - \frac{h^2}{2}(\bar{\Psi}\boldsymbol{\sigma}\Psi)^2\,, \tag{5}$$

a spin-Hall mass $\hat{n} \cdot \langle \bar{\Psi}\boldsymbol{\sigma}\Psi \rangle$ is condensed, and a monopole with non-zero magnetic spin $\mathcal{M}_q^{\downarrow \hat{n}}$ proliferates. Here, the 3-vector of Pauli matrices $\boldsymbol{\sigma}$ acts on the magnetic spin. Coming back to the spin-1/2 kagome Heisenberg antiferromagnet, this model describes the transition from a DSL to a coplanar antiferromagnet [15, 16] as shown in Fig. 1. To characterize this QED$_3$-cHGN quantum critical point (QCP), we obtain the scaling dimension of monopoles.

2 Scaling Dimension of a Monopole Operator

We first warm up by examining the non-compact theory and establish the existence of the QCP. To do so, an auxiliary vectorial boson $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)$ is introduced to decouple the quartic interaction, $\mathcal{L}' = -\bar{\Psi}\left(\not{D}_a + \boldsymbol{\phi}\cdot\boldsymbol{\sigma}\right)\Psi - \frac{1}{2h^2}\boldsymbol{\phi}^2$. The fermions are then integrated out, yielding the following effective action:

$$S_{\text{eff}} = -N \ln\det\left(\not{D}_a + \boldsymbol{\phi}\cdot\boldsymbol{\sigma}\right) + \frac{1}{2h^2}\int d^3x\, \boldsymbol{\phi}^2\,. \tag{6}$$

In the $N = \infty$ saddle point solution, the gauge field vanishes due to gauge invariance $\langle a_\mu \rangle = 0$, and we take a homogeneous ansatz for the boson $\langle \boldsymbol{\phi} \rangle = M\hat{n}$. By solving the resulting gap equation $\partial S_{\text{eff}}/\partial M = 0$, we find that the order parameter $\langle \boldsymbol{\phi} \rangle$ condenses for $h^{-2} < h_c^{-2} = 0$.[2] The effective action at the critical point is then

$$S_{\text{eff}}^c = -N \ln \det \left(\not{D}_a + \boldsymbol{\phi} \cdot \boldsymbol{\sigma} \right) . \tag{7}$$

We now proceed to compute the scaling dimension of monopole operators in QED$_3$-cHGN. Following similar work done in QED$_3$ [3] and in the CP^{N-1} model [4], we employ the state-operator correspondence: An operator $\mathcal{O}$ in a conformal field theory on $\mathbb{R}^3$ can be mapped to a state $|\mathcal{O}\rangle$ in an alternate theory on $S^2 \times \mathbb{R}$. The two spacetimes aforementioned related by a conformal transformation

$$\mathrm{d}r^2 + r^2 \mathrm{d}\Omega^2 \to \mathrm{d}\tau^2 + R^2 \mathrm{d}\Omega^2 , \quad r = Re^{\tau/R} , \tag{8}$$

where the radius of the sphere is R. With this transformation, the dilatation operator on $\mathbb{R}^3$ is mapped to the Hamiltonian on $S^2 \times \mathbb{R}$. This implies that the scaling dimension of the operator $\mathcal{O}$ is equal to the energy of its related state, $\Delta_{\mathcal{O}} = E_{|\mathcal{O}\rangle}$.

For a monopole operator $\mathcal{O} = \mathcal{M}_q^\dagger$, the alternate theory is obtained by adding a monopole background gauge field A^q [3] and by putting the effective action (7) on $S^2 \times \mathbb{R}$. This selects the topological sector of operators with charge q. We restrain our study to the monopole operator with the minimum scaling dimension $\Delta_q \equiv \min(\Delta_{\mathcal{M}_q})$, which corresponds to the ground state of this alternate theory. But the ground state energy is the free energy in the zero temperature limit. The scaling dimension at leading order in $1/N$, $\Delta_q = N\Delta_q^{(0)} + \mathcal{O}(1/N^0)$, is then given by the saddle point value of the effective action

$$\Delta_q^{(0)} = \frac{1}{N} S_{\text{eff}}^c[A^q]\Big|_{\langle a_\mu \rangle = 0\,, \langle \boldsymbol{\phi} \rangle = M_q \hat{z}} = -\ln \mathrm{Det} \left(\not{D}_{A^q}^{S^2 \times \mathbb{R}} + M_q \sigma_z \right) , \tag{9}$$

where $\langle \boldsymbol{\phi} \rangle$ is again taken homogeneous and along $\hat{z}$ without loss of generality.

The Dirac operator in Eq. (9) depends on generalized angular momentum $\boldsymbol{L}_q = \boldsymbol{r} \times \left(\boldsymbol{p} + \boldsymbol{A}^q \right) - q\hat{r}$ and total spin $\boldsymbol{J}_q = \boldsymbol{L}_q + \boldsymbol{\tau}/2$ where $\boldsymbol{\tau}$ acts on particle-hole (Lorentz) space. Spinor monopole harmonics $S_{q;\ell,m}^{\pm}$ diagonalize L_q^2, J_q^z, and $J_q^2 \to j_\pm (j_\pm + 1)$ where $j_\pm = \ell \pm 1/2$. These are two-component spinors built with generalized spherical harmonics $Y_{q;\ell m}$ such that $L_q^2 \to \ell(\ell+1)$ and $L_q^z \to m$, where $\ell = |q|, |q|+1, \ldots$ [17]. Working in the $j = \ell - 1/2$ basis $(S_{q;\ell,m}^-, S_{q;\ell-1,m}^+)^\intercal$, the Dirac operator reduces to a matrix with c-number entries [3]. As for the spin-Hall mass $M_q\sigma_z$, it is already diagonal in spacetime and particle-

[2] Divergences in the gap equation are treated with a Zeta regularization.

hole space. The operator in the determinant of Eq. (9) then yields the following spectrum [18]:

$$\omega + i\sigma M_q\,, \quad \ell = q\,, \tag{10}$$

$$\pm\sqrt{\omega^2 + \varepsilon_\ell^2}\,, \quad \ell = \{q+1, q+2, \dots\}\,, \tag{11}$$

where $\varepsilon_\ell = R^{-1}\sqrt{\ell^2 - q^2 + M_q^2 R^2}$ is the energy, $\sigma = \pm 1$ is the spin projection and the magnetic charge is taken positive $q > 0$. Importantly, the zero modes of QED$_3$ corresponding to a minimum angular momentum $\ell = q$ now have a non-zero energy, positive for spin up modes and negative for spin down modes. As monopole operators are still dressed with half of those "zero" modes, the minimal scaling dimension is now obtained by filling only the spin down "zero" modes. The spectrum (10) and (11) is inserted in Eq. (9) to obtain the scaling dimension

$$\Delta_q = -N\left(d_q M_q + \sum_{\ell=q+1}^{\infty} 2 d_\ell \varepsilon_\ell\right) + \mathcal{O}(1/N^0)\,, \quad d_\ell = 2\ell\,, \tag{12}$$

where the radius R was eliminated with rescaling. This indeed corresponds to the energy obtained by filling all valence modes and spin down "zero" modes as represented in Fig. 2a.

The gap equation $\partial\Delta^{(0)}/\partial M_q = 0$ can be solved for the non-trivial expectation value of the spin-Hall mass $\langle|\boldsymbol{\phi}|\rangle = M_q$. We stress that this mass M_q defines a monopole operator at the QCP and is not an indication of spontaneous symmetry breaking in the model. The scaling dimension is computed by inserting M_q in Eq. (12).[3] The spin-Hall mass M_q and the monopole operator scaling dimension Δ_q obtained numerically are shown for a few magnetic charges in Fig. 2b. The case of QED$_3$, where there is no fermion self-interaction and $M_q = 0$, is also shown. Full lines in Fig. 2b are analytical approximations obtained for a large magnetic charge q. Note that the scaling dimension Δ_q is smaller in QED$_3$-cHGN than in QED$_3$. For the minimal magnetic charge $q = 1/2$, the scaling dimension at the QCP is

$$\Delta_{q=1/2} = 2N \times 0.195 + \mathcal{O}(1/N^0)\,. \tag{13}$$

For $2N = 2.56$, this seems to indicate a unitarity bound violation, $\Delta_{q=1/2} < 1/2$.

A Brief Aside on the Monopole Fermion Number Monopole operators are gauge invariant and must have a vanishing fermion number. In QED$_3$-cHGN, this is obtained by dressing monopoles with half of the fermion "zero" modes. While

[3] Divergences in the scaling dimension and the related gap equation are treated with a zeta function regularization. It is important to keep the same regularization scheme that was used to determine the critical effective action in the non-compact theory (7).

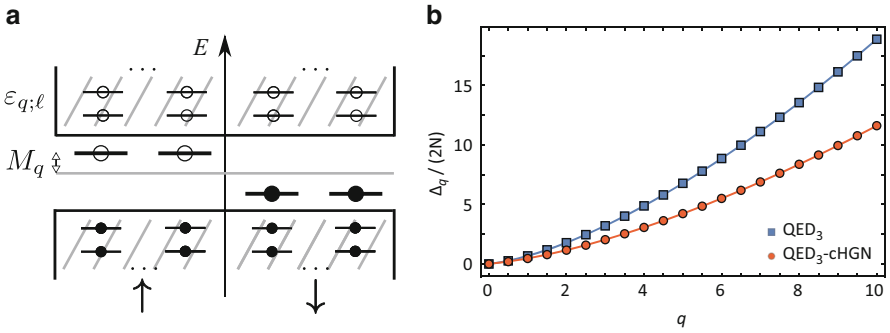


Fig. 2 (**a**) Schematic representation of the fermion occupation of the monopole operator with the lowest scaling dimension and with a spin-Hall mass $\langle \bar{\Psi} \boldsymbol{\sigma} \Psi \rangle \propto \langle \boldsymbol{\phi} \rangle = M_q \hat{z}$. (**b**) The scaling dimension as a function of the magnetic charge q in QED$_3$-cHGN and in QED$_3$ ($M_q = 0$) with the analytical approximation from the large-q expansion. This figure appears with a shorter range in q in Ref. [18]

this condition is not imposed in the formalism above, it just so happens that, in QED$_3$-cHGN, $4\pi q$-flux operators with a minimal scaling dimension have the correct fermion number. In general, the fermion number can be imposed by adding a chemical potential μ that couples to the fermion number density $\Psi^\dagger \Psi$. The chemical potential can be viewed as an imaginary gauge field whose saddle point equation is solved when the fermion number vanishes [18, 19]. In QED$_3$-cHGN, the chemical potential is independent of other parameters and vanishes [18], which explains why it can simply be ignored. As spin up and spin down zero modes are shifted by opposite energies, $\mu = 0$ remains at the center of the energy spectrum. This not the case if we instead tune a SU($2N$)-symmetric Gross–Neveu interaction $\delta\mathcal{L} \sim (\bar{\Psi}\Psi)^2$ in QED$_3$. This model, dubbed QED$_3$-GN, describes the transition to chiral spin liquid via the condensation of the symmetric fermion mass.[4] Here, monopoles at the QCP are described by a flavor independent mass parameter, which, if non-zero, shifts all fermion zero modes by the same energy $M_q = \langle \bar{\Psi}\Psi \rangle$. A chemical potential set to this energy, $\mu = \langle \bar{\Psi}\Psi \rangle$, yields a vanishing fermion number. Using this additional constraint, it is found that the monopole scaling dimension in QED$_3$-GN is minimized by a vanishing mass $M_q = 0$ [18], thus $\Delta_q^{\text{QED}_3\text{-GN}} = \Delta_q^{\text{QED}_3} + \mathcal{O}(1/N^0)$. Interestingly, this large-N result may be extrapolated to test a conjectured duality between the QED$_3$-GN model with $2N = 2$ flavors of massless fermions and the CP^1 model [20] (see Ref. [18]).

[4] Note that contrary to other transitions described in Sect. 1, here the symmetric mass generates a Chern–Simons term which prevents monopole proliferation.

3 Hierarchy Among Monopole Operators

So far, we have only computed the minimal scaling dimension of monopole operators in QED$_3$-cHGN. Monopole operators with larger scaling dimensions are expected at this QCP since the degeneracy of monopoles in QED$_3$ should be lifted [18] by the cHGN interaction which breaks the flavor symmetry

$$\mathrm{SU}(2N) \to \mathrm{SU}(2) \times \mathrm{SU}(N)\,. \tag{14}$$

We first review how monopole operators are organized in QED$_3$ with flavor symmetry SU($2N$). We focus on the simplest case with the minimal magnetic charge $q = 1/2$ where monopole operators are automatically Lorentz scalars [3]. A monopole operator can then be written as half of the $2N$ zero modes creation operators $c^\dagger_{I_i}$ multiplying a bare monopole operator $\mathcal{M}^\dagger_{\mathrm{Bare}}$ which creates all negative energy modes in a 2π magnetic flux background

$$\mathcal{M}^\dagger_{I_1 \ldots I_N} = c^\dagger_{I_1} \ldots c^\dagger_{I_N} \mathcal{M}^\dagger_{\mathrm{Bare}}\,, \quad I_i \in \{1, 2, \ldots, 2N\}\,. \tag{15}$$

Antisymmetry between these fermionic operators yields a rank-N antisymmetric tensor in flavor space. The first step in understanding the hierarchy of monopole operators at the QCP is to obtain the reduction of this SU($2N$) irreducible representation (irrep) in terms of irreps of the subgroup SU(2) $\times$ SU(N). This is a specific case of the reduction SU(MN) $\to$ SU(M) $\times$ SU(N), where M and N are integers, whose branching rules are well studied [21].

For $N = 2$, there are two valleys $v = L, R$ and monopoles form the rank-2 antisymmetric irrep of SU(4), denoted by its dimension **6**. We note that for the specific examples discussed in this section, the irreps are uniquely determined by their dimension. Monopole operators are then expressed as

$$c^\dagger A \left(c^\dagger\right)^{\mathsf{T}} \mathcal{M}^\dagger_{\mathrm{Bare}}\,, \tag{16}$$

where A acts on vectors in flavor space $c^\dagger = \left(c^\dagger_{\uparrow,L},\, c^\dagger_{\uparrow,R},\, c^\dagger_{\downarrow,L},\, c^\dagger_{\downarrow,R}\right)$. At the QCP, monopole operators in the irrep **6** of SU(4) of QED$_3$ reorganize as irreps $(\boldsymbol{m}, \boldsymbol{n})$ with dimension $m \times n$ of the remaining subgroup SU(2)$_{\mathrm{Spin}}$ $\times$ SU(2)$_{\mathrm{Nodal}}$ [21]

$$\mathbf{6} \to (\mathbf{3}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{3})\,. \tag{17}$$

Monopoles are then decomposed as spin and nodal triplets, respectively $(\mathbf{3}, \mathbf{1})$ and $(\mathbf{1}, \mathbf{3})$, which may be written as [15]

$$\vec{\mathcal{M}}^\dagger_{\mathrm{Spin}} = c^\dagger \left(\sigma_y \boldsymbol{\sigma} \otimes \mu_y\right) \left(c^\dagger\right)^{\mathsf{T}} \mathcal{M}^\dagger_{\mathrm{Bare}}\,, \quad \vec{\mathcal{M}}^\dagger_{\mathrm{Nodal}} = c^\dagger \left(\sigma_y \otimes \mu_y \boldsymbol{\mu}\right) \left(c^\dagger\right)^{\mathsf{T}} \mathcal{M}^\dagger_{\mathrm{Bare}}\,, \tag{18}$$

where $\boldsymbol{\mu}$ are Pauli matrices acting on nodal subspace. The scaling dimension of monopole operators in these spin and nodal triplets are expected to differ given that no hidden symmetry connects the multiplets. Monopole operators with the largest total spin (here, $S = 1$) have access to the largest polarization and as a result can minimize the contribution from the spin-Hall mass. These operators have the lowest scaling dimension and are the finite-N analogues of monopoles with minimal scaling dimension $\Delta_{1/2}$ studied in Sect. 2. In our large-N analysis, we gave the example of a monopole filled only with spin down "zero" modes along $\hat{z}$ such that the corresponding spin-Hall mass $\langle\bar{\Psi}\sigma_z\Psi\rangle \propto M_q > 0$ is minimized. As we perform a $\mathrm{SU}(2)_{\mathrm{Spin}}$ transformation rotating the spin down modes to spin up modes, the spin-Hall mass sign changes $\langle\bar{\Psi}\sigma_z\Psi\rangle \to -\langle\bar{\Psi}\sigma_z\Psi\rangle$. This leaves the scaling dimension unchanged as we rotate to another monopole of the triplet as schematically shown in Fig. 3. A rotation could also make the spin-Hall mass point in the $\hat{x}$ direction, in which case a combination of all the states in the spin triplet would yield the monopole operator polarized along $\hat{x}$. Note that this specific element was not well addressed in [18] as the possibility of a rotating the spin-Hall mass was not discussed.

Further insight is obtained by considering the QCP with larger values of N. For $N = 3$, the flavor symmetry is reduced as $\mathrm{SU}(6) \to \mathrm{SU}(2)_{\mathrm{Spin}} \times \mathrm{SU}(3)_{\mathrm{Nodal}}$ at the QCP, and the rank-3 antisymmetric representation **20** of SU(6) decomposes as [21]

$$\mathbf{20} \to (\mathbf{4}, \mathbf{1}) \oplus (\mathbf{2}, \mathbf{8}) \ . \tag{19}$$

The RHS has dimension $4 \times 1 + 2 \times 8 = 20$ as required. Again, monopole operators with the highest $\mathrm{SU}(2)_{\mathrm{spin}}$ spin, here a spin quadruplet $S = 3/2$, the **4** in Eq. (19), have the lowest scaling dimension. As for the spin doublet $S = 1/2$ (denoted **2**), it is obtained by the composition of three spins: Two spins form a singlet while the remaining spin is either up or down. For general N, monopole operators reorganize as various magnetic spin multiplets with total spin $S_{\mathrm{min}} \leq S \leq N/2$, with minimum spin $S_{\mathrm{min}} = 0$ for N even and $S_{\mathrm{min}} = 1/2$ for N odd. For large-N, this distinction does not affect the scaling dimension. The almost equally populated spin up and spin down "zero" modes of a spin doublet $S = 1/2$ monopole mostly cancel their contribution to the scaling dimension. The remaining unpaired "zero" mode has a contribution order $\mathcal{O}(1/N^0)$ which is neglected in our leading order

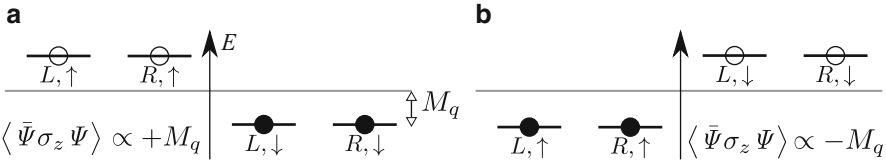


Fig. 3 Fermion zero modes occupation of monopole operators for $N = 2$ and $q = 1/2$. (**a**) The spin down monopole and (**b**) the spin up monopole are related by a $\mathrm{SU}(2)_{\mathrm{spin}}$ rotation which changes the sign of the spin-Hall mass. Here, $M_q > 0$

treatment. Therefore, the largest scaling dimension (characterizing monopoles with either $S = 0$ or $S = 1/2$) is given, at leading order in $1/N$, by the second term in (12). The gap equation for this expression is solved for $M_q = 0$, which yields the same scaling dimension as in QED$_3$. Monopole operators in QED$_3$-cHGN then have scaling dimensions $\Delta_{\mathcal{M}_q}$ which vary between the lowest monopole scaling dimension in QED$_3$-cHGN and QED$_3$,

$$\Delta_q^{\text{QED}_3\text{-cHGN}} \leq \Delta_{\mathcal{M}_q}^{\text{QED}_3\text{-cHGN}} \leq \Delta_q^{\text{QED}_3}, \quad \text{leading order in } 1/N\,. \tag{20}$$

This upper boundary was overestimated in [18] as the possibility of different masses M_q defining each monopole multiplet was not considered.

Conclusion The DSL is a parent state of many quantum phases and exotic non-Landau transitions in frustrated quantum magnets. We characterized monopole operators in QED$_3$-GN models describing various QCPs between this DSL and other ordered phases. Notably, we obtained the lowest scaling dimension of monopole operators in the QED$_3$-cHGN model. At leading order in $1/N$, we find it is lower than its counterpart in QED$_3$. Specifically, for a minimal magnetic charge, it is given by $\Delta_{q=1/2}^{\text{QED}_3\text{-cHGN}} = 2N \times 0.195 + \mathcal{O}(1/N^0)$. We discussed how monopoles are reorganized as irreps of the QED$_3$-cHGN QCP reduced symmetry group SU(2) $\times$ SU(N). Monopoles multiplets with the highest SU(2) spin have the lowest scaling dimension. A more detailed exploration of this monopole hierarchy at the QCP is reserved for future work.

Acknowledgments We thank Tarun Grover, Joseph Maciejko, David Poland, Zi Yang Meng and Chong Wang for the interesting discussions and useful observations on our work during the symposium. We also thank Jaume Gomis and Sergueï Tchoumakov for useful discussions. É.D. was funded by an Alexander Graham Bell CGS from NSERC. M.B.P. was funded by a Discovery Grant from NSERC. W.W.-K. was funded by a Discovery Grant from NSERC, a Canada Research Chair, a grant from the Fondation Courtois, and a "Établissement de nouveaux chercheurs et de nouvelles chercheuses universitaires" grant from FRQNT.

References

1. A.M. Polyakov, Compact gauge fields and the infrared catastrophe. Phys. Lett. B **59**, 82–84 (1975)
2. A.M. Polyakov, Quark confinement and topology of gauge theories. Nuclear Phys. B **120**, 429–458 (1977)
3. V. Borokhov, A. Kapustin, X. Wu, Topological disorder operators in three-dimensional conformal field theory. J. High Energy Phys. **2002**, 049 (2002)
4. M.A. Metlitski, M. Hermele, T. Senthil, M.P.A. Fisher, Monopoles in $\mathbb{C}\text{P}^{N-1}$ model via the state-operator correspondence. Phys. Rev. B **78**, 214418 (2008)
5. Y. Ran, M. Hermele, P.A. Lee, X.-G. Wen, Projected-wave-function study of the spin-1/2 heisenberg model on the kagome lattice. Phys. Rev. Lett. **98**, 117205 (2007)
6. P. Sindzingre, C. Lhuillier, Low-energy excitations of the kagome antiferromagnet and the spin-gap issue. Europhys. Lett. **88**, 27009 (2009)

7. Y. Iqbal, F. Becca, S. Sorella, D. Poilblanc, Gapless spin-liquid phase in the kagome spin-1/2 Heisenberg antiferromagnet. Phys. Rev. B **87**, 060405 (2013)
8. Y. Iqbal, D. Poilblanc, F. Becca, Spin-1/2 Heisenberg J_1-J_2 antiferromagnet on the kagome lattice. Phys. Rev. B **91**, 020402 (2015)
9. Y.-C. He, M.P. Zaletel, M. Oshikawa, F. Pollmann, Signatures of Dirac Cones in a DMRG study of the Kagome Heisenberg model. Phys. Rev. X **7**, 031020 (2017)
10. W. Zhu, X. Chen, Y.-C. He, W. Witczak-Krempa, Entanglement signatures of emergent Dirac fermions: Kagome spin liquid and quantum criticality. Sci. Adv. **4**(11), eaat5535 (2018)
11. S.S. Pufu, Anomalous dimensions of monopole operators in three-dimensional quantum electrodynamics. Phys. Rev. D **89**, 065016 (2014)
12. X.Y. Xu, Y. Qi, L. Zhang, F.F. Assaad, C. Xu, Z.Y. Meng, Monte Carlo study of lattice compact quantum electrodynamics with fermionic matter: the parent state of quantum phases. Phys. Rev. X **9**, 021022 (2019)
13. X.-Y. Song, C. Wang, A. Vishwanath, Y.-C. He, Unifying description of competing orders in two-dimensional quantum magnets. Nat. Commun. **10**(1), 4254 (2019)
14. M.F. Atiyah, I.M. Singer, The index of elliptic operators on compact manifolds. Bull. Amer. Math. Soc. **69**(3), 422–433 (1963)
15. M. Hermele, Y. Ran, P.A. Lee, X.-G. Wen, Properties of an algebraic spin liquid on the kagome lattice. Phys. Rev. B **77**, 224413 (2008)
16. Y.-M. Lu, G.Y. Cho, A. Vishwanath, Unification of bosonic and fermionic theories of spin liquids on the kagome lattice. Phys. Rev. B **96**, 205150 (2017)
17. T.T. Wu, C.N. Yang, Dirac monopole without strings: monopole harmonics. Nuclear Phys. B **107**(3), 365–380 (1976)
18. É. Dupuis, M.B. Paranjape, W. Witczak-Krempa, Transition from a Dirac spin liquid to an antiferromagnet: monopoles in a QED_3-Gross-Neveu theory. Phys. Rev. B **100**, 094443 (2019)
19. S.M. Chester, L.V. Iliesiu, M. Mezei, S.S. Pufu, Monopole operators in U(1) Chern-Simons-matter theories. J. High Energy Phys. **2018**, 157 (2018)
20. C. Wang, A. Nahum, M.A. Metlitski, C. Xu, T. Senthil, Deconfined quantum critical points: symmetries and dualities. Phys. Rev. X **7**, 031051 (2017)
21. C. Itzykson, M. Nauenberg, Unitary groups: representations and decompositions. Rev. Mod. Phys. **38**, 95 (1966)

Critical Exponents for the Valence-Bond-Solid Transition in Lattice Quantum Electrodynamics

Rufus Boyack and Joseph Maciejko

Abstract Recent sign-problem-free quantum Monte Carlo simulations of (2+1)-dimensional lattice quantum electrodynamics (QED_3) with N_f flavors of fermions on the square lattice have found evidence of continuous quantum phase transitions between a critical phase and a gapped valence-bond-solid (VBS) phase for flavor numbers $N_f = 4$, 6, and 8. We derive the critical theory for these transitions, the chiral $O(2)$ QED_3-Gross–Neveu model, and show that the latter is equivalent to the gauged Nambu–Jona-Lasinio model. Using known large-N_f results for the latter, we estimate the order parameter anomalous dimension and the correlation length exponent for the transitions mentioned above. We obtain large-N_f results for the dimensions of fermion bilinear operators, in both the gauged and ungauged chiral $O(2)$ Gross–Neveu models, which, respectively, describe the long-distance power-law decay of two-particle correlation functions at the VBS transition in lattice QED_3 and the Kekulé-VBS transition for correlated fermions on the honeycomb lattice.

Keywords Lattice gauge theory · Valence-bond solid · Quantum phase transition · Quantum electrodynamics · Gross–Neveu model · Nambu–Jona-Lasinio model · Conformal field theory

Quantum phase transitions that involve fractionalized degrees of freedom fall outside the traditional Landau paradigm and have been the focus of much interest in recent years. The classic example is deconfined quantum critical points between conventional phases of quantum antiferromagnets [1, 2], where emergent fractionalized matter fields and gauge fields appear at the critical point but are confined in the phases themselves. A class of transitions comparatively less studied, but also beyond the Landau paradigm, are transitions between phases supporting fractionalized excitations, such as different types of spin liquids, or between a fractionalized

R. Boyack · J. Maciejko (✉)
Department of Physics and Theoretical Physics Institute, University of Alberta, Edmonton, AB, Canada
e-mail: boyack@ualberta.ca; maciejko@ualberta.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_32

phase and a conventional phase. In the language of lattice gauge theories of quantum antiferromagnets [3], where spin degrees of freedom fractionalize into emergent fermions coupled to gauge fields, these correspond to transitions between distinct deconfined phases of a lattice gauge theory or between a deconfined phase and a confined phase, respectively. Besides their application to frustrated magnetism and elementary particle physics, lattice gauge theories may now be experimentally realized using ultracold atoms in optical lattices [4, 5], and thus constitute an important class of interacting many-body systems whose phases and phase transitions are of intrinsic interest.

Recently, sign-problem-free quantum Monte Carlo (QMC) simulations of (2+1)-dimensional lattice quantum electrodynamics (QED_3) with an even number N_f of flavors of fermions on the square lattice [6, 7] have found evidence for a $U(1)$ deconfined phase with power-law correlations, and for continuous quantum phase transitions from this phase to conventional confined phases. For $N_f = 2$, the putative $U(1)$ phase is adiabatically connected to the algebraic spin liquid [8] and the transition is towards a Néel antiferromagnet. This transition is described by the chiral $O(3)$ QED_3-Gross-Neveu (GN) model, whose universal critical properties were studied recently using both ϵ-expansion [9, 10] and large-N_f techniques [10]. For $N_f = 4$, 6, and 8, the confinement transition is found to be towards a gapped valence-bond-solid (VBS) phase. The corresponding critical theory was conjectured to be of the chiral $O(2)$ QED_3-GN type [6], but its critical properties have thus far not been investigated. In this paper, we establish the precise form of the critical theory, show its equivalence to the gauged Nambu–Jona-Lasinio (NJL) model [11, 12], and determine various critical exponents using the large-N_f expansion. In both the gauged and ungauged chiral $O(2)$ GN models, we obtain new results for the scaling dimensions of fermion bilinears that, respectively, describe the power-law decay of certain two-particle correlation functions at the $U(1)$-VBS transition and the semimetal-to-Kekulé-VBS transition for interacting fermions on the honeycomb lattice [13–15]. Critical exponents for the $U(1)$-VBS transition at $\mathcal{O}(1/N_f^2)$ in the large-N_f expansion and four-loop order in the ϵ-expansion will be reported in a future publication [16].

1 The $U(1)$-VBS Transition

The $U(1)$ lattice gauge theory studied in Ref. [6, 7] is a quantum rotor model with fermions on the square lattice. The Hamiltonian is

$$\mathcal{H} = \frac{1}{2} J N_f \sum_{\langle rr' \rangle} \frac{1}{4} L_{rr'}^2 - t \sum_{\alpha=1}^{N_f} \sum_{\langle rr' \rangle} \left(c_{r\alpha}^\dagger e^{i\theta_{rr'}} c_{r'\alpha} + \text{h.c.} \right) + \frac{1}{2} K N_f \sum_{\square} \cos(\boldsymbol{\Delta} \times \boldsymbol{\theta}), \tag{1}$$

where $c_{r\alpha}^{(\dagger)}$ annihilates (creates) a fermion of flavor $\alpha = 1, \ldots, N_f$ on site r, $\langle rr' \rangle$ denotes bonds between nearest-neighbor sites r and r', the angular bond variable $\theta_{rr'} \in [0, 2\pi)$ and the angular momentum $L_{rr'}$ are canonical conjugates, and $\boldsymbol{\Delta} \times \boldsymbol{\theta}$ denotes the lattice curl of θ around a plaquette $\square$. The magnetic coupling $K > 0$ favors a background flux of π in each plaquette. To begin, consider the fermionic part of the Hamiltonian, in the absence of gauge fluctuations ($J = 0$). A gauge for the background flux can be chosen such that the Hamiltonian is

$$\mathcal{H}_0 = \sum_{\alpha=1}^{N_f} \int_{\mathrm{BZ}} \frac{d^2k}{(2\pi)^2} c_{\boldsymbol{k}\alpha}^{\dagger} h(\boldsymbol{k}) c_{\boldsymbol{k}\alpha}, \tag{2}$$

with the two-component spinor $c_{\boldsymbol{k}\alpha} = (c_{\boldsymbol{k},\alpha,A}, c_{\boldsymbol{k},\alpha,B})$, where A and B denote the two sublattices arising from the choice of gauge, and

$$h(\boldsymbol{k}) = -t \begin{pmatrix} 0 & f(\boldsymbol{k}) \\ f^*(\boldsymbol{k}) & 0 \end{pmatrix}, \quad f(\boldsymbol{k}) = 1 - e^{i(k_x - k_y)} + e^{-i(k_x + k_y)} + e^{-2ik_y}. \tag{3}$$

Two Dirac nodes are found at $\pm \boldsymbol{Q} = (0, \pm\frac{\pi}{2})$. Keeping only the degrees of freedom near the Dirac nodes, the low-energy Hamiltonian becomes

$$\mathcal{H}_0 \approx v_F \sum_{\alpha=1}^{N_f} \sum_{\eta=\pm} \int \frac{d^2p}{(2\pi)^2} \chi_{\alpha\eta}^{\dagger}(\boldsymbol{p})(\mu_1 p_x + \mu_2 p_y)\chi_{\alpha\eta}(\boldsymbol{p}), \tag{4}$$

where $v_F = 2t$. The two-component Dirac fields are defined by

$$\chi_{\alpha,+}(\boldsymbol{p}) = \begin{pmatrix} c_{\boldsymbol{Q}+\boldsymbol{p},\alpha,A} \\ c_{\boldsymbol{Q}+\boldsymbol{p},\alpha,B} \end{pmatrix}, \qquad \chi_{\alpha,-}(\boldsymbol{p}) = \begin{pmatrix} c_{-\boldsymbol{Q}+\boldsymbol{p},\alpha,B} \\ -c_{-\boldsymbol{Q}+\boldsymbol{p},\alpha,A} \end{pmatrix}. \tag{5}$$

These can be combined into N_f flavors of four-component Dirac fermions $\Psi_\alpha = (\chi_{\alpha,+}, \chi_{\alpha,-})^T$. We introduce the following 4×4 (reducible) representation of the Euclidean Dirac algebra in $2 + 1$ dimensions:

$$\Gamma_\mu = \begin{pmatrix} \tilde{\gamma}_\mu & 0 \\ 0 & -\tilde{\gamma}_\mu \end{pmatrix}, \quad \mu = 0, 1, 2, \tag{6}$$

where $\tilde{\gamma}_\mu$ are 2×2 Euclidean Dirac matrices defined in terms of Pauli matrices by $(\tilde{\gamma}_0, \tilde{\gamma}_1, \tilde{\gamma}_2) = (\sigma_3, \sigma_2, -\sigma_1)$. Using the Dirac conjugate $\overline{\Psi}_\alpha = \Psi_\alpha^\dagger \Gamma_0$, the Lagrange density for the Hamiltonian (4) is $\mathcal{L}_0 = \sum_\alpha \overline{\Psi}_\alpha \Gamma_\mu \partial_\mu \Psi_\alpha$. For small but nonzero $J > 0$, a Maxwell kinetic term for gauge-field fluctuations A_μ about the π-flux background is generated and ∂_μ in $\mathcal{L}_0$ is promoted to the gauge-covariant derivative $D_\mu = \partial_\mu + iA_\mu$. This results in the QED$_3$ Lagrangian, which exhibits a conformal infrared fixed point for sufficiently large N_f [17]—in accordance with the critical phase observed numerically at small J [6, 7]. We are at present treating the $U(1)$

gauge field as noncompact; the effects of compactness due to the original lattice formulation will be discussed in Sect. 4.

For J larger than some critical value J_c, a VBS phase with unbroken global $SU(N_f)/\mathbb{Z}_{N_f}$ symmetry is found for $N_f = 4, 6, 8$ [6, 7]. Columnar VBS order doubles the unit cell of the square lattice and spontaneously breaks the latter's D_4 point-group symmetry to a D_2 subgroup; it is represented by a time-reversal-invariant vector order parameter $\boldsymbol{V} = (V_x, V_y)$ transforming in the two-dimensional E irreducible representation of D_4. Using the projective symmetry group approach, one can determine how gauge-invariant operators in the low-energy QED_3 theory transform under the microscopic lattice symmetries [10]. Defining the two 4×4 Hermitian matrices

$$\Gamma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \Gamma_5 = \Gamma_0\Gamma_1\Gamma_2\Gamma_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{7}$$

which square to the identity and anticommute with each other and with the Dirac matrices (6), one finds that the pair of time-reversal-invariant and flavor-symmetric Dirac bilinears $\left(\sum_\alpha i\overline{\Psi}_\alpha\Gamma_5\Psi_\alpha, \sum_\alpha i\overline{\Psi}_\alpha\Gamma_3\Psi_\alpha\right)$ transform precisely in the E irreducible representation of D_4. Furthermore, $\sum_\alpha i\overline{\Psi}_\alpha\Gamma_5\Psi_\alpha$ is odd under x-reflections and lattice x-translations and even under y-translations, while $\sum_\alpha i\overline{\Psi}_\alpha\Gamma_3\Psi_\alpha$ transforms oppositely. Thus one can identify $V_x \sim \sum_\alpha i\overline{\Psi}_\alpha\Gamma_5\Psi_\alpha$ and $V_y \sim \sum_\alpha i\overline{\Psi}_\alpha\Gamma_3\Psi_\alpha$. A nonzero expectation value of $\boldsymbol{V}$ corresponds to a nonzero fermion mass, in accordance with the gapped spectrum observed in the VBS phase [7]. Note that $\boldsymbol{V}$ is Lorentz invariant since Γ_3 and Γ_5 commute with the Euclidean transformations $\exp\left(-\frac{i}{2}\omega_{\mu\nu}\sigma_{\mu\nu}\right)$ where $\sigma_{\mu\nu} = \frac{i}{4}[\Gamma_\mu, \Gamma_\nu]$.

The occurrence of a VBS phase for $J > J_c$ can be understood as arising from a short-ranged four-fermion interaction term $\sim(g^2/N_f)\boldsymbol{V}^2$ generated by gauge fluctuations at the lattice scale. Such interactions are perturbatively irrelevant at the conformal QED_3 fixed point, but if sufficiently strong can give rise to dynamical fermion mass generation via a quantum critical point. Decoupling this interaction term with a pair $\boldsymbol{\phi} = (\phi_1, \phi_2)$ of scalar fields and tuning to the quantum critical point, we obtain the chiral $O(2)$ QED_3-GN model,

$$\mathcal{L}_{O(2)\text{QED}_3-\text{GN}} = \sum_{\alpha=1}^{N_f}\left[\overline{\Psi}_\alpha\Gamma_\mu\left(\partial_\mu + \frac{e}{\sqrt{N_f}}iA_\mu\right)\Psi_\alpha + \frac{g}{\sqrt{N_f}}i\boldsymbol{\phi}\cdot\overline{\Psi}_\alpha\boldsymbol{M}\Psi_\alpha\right] + \ldots, \tag{8}$$

where $\boldsymbol{M} = (\Gamma_3, \Gamma_5)$, and $\ldots$ includes Maxwell and gauge-fixing terms for the gauge field, and symmetry-allowed kinetic and self-interaction terms for the scalar field $\boldsymbol{\phi}$. At the free-field fixed point, the gauge coupling e^2 and the Yukawa coupling g^2 have units of mass and are thus relevant. However, the fields have been rescaled to make explicit the fact that e and g appear with a suppressing factor of $1/\sqrt{N_f}$. In the large-N_f limit, the physics at momenta $|q| \ll e^2, g^2$ is dominated by the coupling

between fermions and soft bosonic fluctuations, i.e., the terms in square brackets in Eq. (8), and can be computed systematically in powers of $1/N_f$. Conceptually similar applications of the $1/N_f$ expansion to the chiral Ising and $O(3)$ QED$_3$-GN models can be found in Refs. [18–23] and [10], respectively. To leading (zeroth) order in this expansion, the large-N_f scalar-field and gauge-field propagators in the infrared limit are, respectively:

$$D_{ab}(q) = \frac{4}{g^2|q|}\delta_{ab}, \qquad \Pi_{\mu\nu}(q) = \frac{8}{e^2|q|}\left(\delta_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}\right), \tag{9}$$

where $a, b = 1, 2$. The gauge-field propagator is given in the Landau gauge.

The extra terms ... in Eq. (8) contain a coupling of the form $\propto (\phi_1 + i\phi_2)^4 + \text{c.c.}$, which transforms trivially under C_4 rotations and is thus allowed by the microscopic symmetries. Such a term is relevant at the free-field fixed point. However, Eq. (9) implies that the scaling dimension of $\boldsymbol{\phi}$ at the chiral $O(2)$ QED$_3$-GN critical point is $\Delta_\phi = 1 + \mathcal{O}(1/N_f)$, thus this term is irrelevant at the $U(1)$-VBS critical point in the large-N_f limit. (Other D_4-allowed terms are already irrelevant at the free-field fixed point.) Thus the Lagrangian (8) acquires an emergent $SO(2)$ symmetry under $\Psi_\alpha \to e^{-iW\theta/2}\Psi_\alpha$, $\phi_a \to R_{ab}(\theta)\phi_b$, where $W = -i\Gamma_3\Gamma_5$ and $R(\theta)$ is the $SO(2)$ matrix for a rotation through angle θ.

2 The Gauged NJL Model and Critical Exponents

The NJL model was originally introduced as a toy model of chiral symmetry breaking and dynamical mass generation in high-energy physics [11]. Its gauged version [12] is described by the Lagrangian

$$\mathcal{L}_{\text{NJL}} = \sum_{\alpha=1}^{N_f}\left[\overline{\psi}_\alpha\gamma_\mu\left(\partial_\mu + \frac{e}{\sqrt{N_f}}iA_\mu\right)\psi_\alpha + \frac{g}{\sqrt{N_f}}\overline{\psi}_\alpha\left(\phi_1 + i\phi_2\gamma_5\right)\psi_\alpha\right] + \ldots, \tag{10}$$

where ψ_α are four-component Dirac spinors, and, as previously, ... denotes terms not involving fermions which, besides a gauge-fixing term, are irrelevant in the large-N_f limit of interest to us. We now show that the gauged NJL model is entirely equivalent to the chiral $O(2)$ QED$_3$-GN model (8). Define the gamma matrices in Eq. (10) in terms of those in Eq. (6)–(7) by $\gamma_\mu = i\Gamma_\mu\Gamma_3$, $\mu = 0, 1, 2$ and $\gamma_5 = -i\Gamma_3\Gamma_5$, and in addition define $\Psi_\alpha = \psi_\alpha$ and $\overline{\psi}_\alpha = \Psi_\alpha^\dagger\gamma_0$. The Hermitian matrices γ_μ and γ_5 obey the usual Euclidean Dirac algebra (i.e., they anticommute with each other and square to the identity). Using these gamma matrices, the gauged NJL Lagrangian (10) becomes equal to the chiral $O(2)$ QED$_3$-GN Lagrangian (8). The emergent $SO(2)$ symmetry of the latter is identified with the invariance of the

former under $U(1)$ chiral transformations $\psi_\alpha \to e^{-i\gamma_5\theta/2}\psi_\alpha$, with a concomitant rotation of the scalar field $\boldsymbol{\phi} = (\phi_1, \phi_2)$.

If the gauge field is absent, this also establishes the equivalence between the ungauged NJL model and the chiral $O(2)$ GN model. The latter describes the semimetal-to-Kekulé-VBS transition for interacting fermions on the honeycomb lattice [13–15]. The D_6 point-group symmetry of the honeycomb lattice allows for a term of the form $\propto (\phi_1 + i\phi_2)^3 + \text{c.c.}$ in the critical Lagrangian, which is marginal in the $N_f = \infty$ limit at the chiral $O(2)$ GN fixed point. However, a renormalization-group analysis in the large-N_f limit shows that the $\mathcal{O}(1/N_f)$ correction renders this term irrelevant [15]. QMC simulations of the joint probability distribution $P(\phi_1, \phi_2)$ of the two components of the VBS order parameter also support the emergent $SO(2)$ symmetry at the critical point [15].

The critical points of the gauged and ungauged NJL models are strongly coupled $(2+1)$-dimensional conformal field theories characterized by a spectrum of scaling dimensions that correspond to universal critical exponents. Some of these exponents have already been computed in the $1/N_f$ expansion in general d spacetime dimensions [24–27]. The order-parameter anomalous dimension for $d = 3$ is

$$\text{chiral } O(2) \text{ QED}_3\text{-GN}: \quad \eta_\phi = 1 + \frac{56}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big), \tag{11}$$

$$\text{chiral } O(2) \text{ GN}: \quad \eta_\phi = 1 - \frac{8}{3\pi^2 N_f} + \frac{544}{27\pi^4 N_f^2} + \mathcal{O}\big(1/N_f^3\big), \tag{12}$$

and is related to the scalar-field scaling dimension by $\Delta_\phi = \frac{1}{2}\big(1 + \eta_\phi\big)$. The inverse correlation length exponent is

$$\text{chiral } O(2) \text{ QED}_3\text{-GN}: \quad \nu^{-1} = 1 - \frac{80}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big), \tag{13}$$

$$\text{chiral } O(2) \text{ GN}: \quad \nu^{-1} = 1 - \frac{16}{3\pi^2 N_f} + \frac{8\big(364 + 27\pi^2\big)}{27\pi^4 N_f^2} + \mathcal{O}\big(1/N_f^3\big), \tag{14}$$

and is related to the scaling dimension of the $\boldsymbol{\phi}^2$ operator by $\Delta_{\phi^2} = 3 - \nu^{-1}$.

3 Fermion Bilinear Scaling Dimensions

The scaling dimension Δ_ϕ characterizes the power-law decay at long distances of the microscopic VBS correlation function $\langle \mathcal{O}_{\text{VBS}}(r)\mathcal{O}_{\text{VBS}}(r')\rangle \sim |r - r'|^{-2\Delta_\phi}$ at the $U(1)$-VBS critical point of the lattice gauge theory (1), where $\mathcal{O}_{\text{VBS}}(r)$ can be chosen as $(-1)^x \sum_A S_A(r)S_A(r + \hat{x})$ or $(-1)^y \sum_A S_A(r)S_A(r + \hat{y})$, describing

columnar VBS order in the x and y directions, respectively. Here $S_A(r) = \sum_{\alpha\beta} c^\dagger_{r\alpha} T_A^{\alpha\beta} c_{r\beta}$ denotes the $SU(N_f)$ spin operator, where T_A, $A = 1, \ldots, N_f^2 - 1$ is a Hermitian generator of the $SU(N_f)$ Lie algebra. The corresponding thermodynamic susceptibility $\chi_{\mathrm{VBS}}(\boldsymbol{q}) \sim |\boldsymbol{q}|^{2\Delta_\phi - 3} \sim |\boldsymbol{q}|^{-(2-\eta_\phi)}$, which can be computed in QMC, diverges as $\boldsymbol{q} \to 0$, signaling the onset of VBS order. In the $U(1)$ phase itself, which is a critical phase, other gauge-invariant observables such as the staggered density operator $\mathcal{O}_{\mathrm{CDW}}(r) = (-1)^{x+y} \sum_\alpha c^\dagger_{r\alpha} c_{r\alpha}$, the staggered $SU(N_f)$ spin $\mathcal{O}_A(r) = (-1)^{x+y} S_A(r)$, and a quantum anomalous Hall mass operator $\mathcal{O}_{\mathrm{QAH}}(r)$ defined in Ref. [10] also exhibit universal power-law correlations characterized by (non-diverging) susceptibilities $\chi_{\mathcal{O}}(\boldsymbol{q}) \sim |\boldsymbol{q}|^{2\Delta_{\mathcal{O}} - 3}$ [8], which are also in principle accessible in QMC. Such susceptibilities remain power law at the $U(1)$-VBS critical point, but with different exponents characterizing the conformal field theory associated with the chiral $O(2)$ QED$_3$-GN fixed point as opposed to that of the pure-QED$_3$ fixed point. Detecting a change in these exponents numerically upon approach to the critical point $J \to J_c^-$ would be a signature of the new universality class discussed in the present paper.

At the $U(1)$-VBS critical point, the microscopic observables above correspond in the long-wavelength limit to Lorentz-invariant fermion bilinears in the chiral $O(2)$ QED$_3$-GN field theory (8): the flavor-singlet, time-reversal-even bilinear $\overline{\Psi}\Psi \sim \mathcal{O}_{\mathrm{CDW}}$, the flavor-adjoint, time-reversal-even bilinear $\overline{\Psi} T_A \Psi \sim \mathcal{O}_A$, and the flavor-singlet, time-reversal-odd bilinear $i\overline{\Psi}\Gamma_3\Gamma_5\Psi \sim \mathcal{O}_{\mathrm{QAH}}$. We have computed the scaling dimensions of these bilinears at $\mathcal{O}(1/N_f)$ in the large-N_f expansion by adapting the methods used in Ref. [23] for the chiral Ising QED$_3$-GN model, accounting for the matrix structure in the Yukawa vertex and the anticommutation properties of Γ_3 and Γ_5. We obtain

$$\text{chiral } O(2) \text{ QED}_3\text{-GN}: \quad \Delta_{\overline{\Psi}\Psi} = \Delta_{\overline{\Psi} T_A \Psi} = 2 - \frac{40}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big), \tag{15}$$

$$\Delta_{i\overline{\Psi}\Gamma_3\Gamma_5\Psi} = 2 + \frac{80}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big), \tag{16}$$

$$\text{chiral } O(2) \text{ GN}: \quad \Delta_{\overline{\Psi}\Psi} = \Delta_{\overline{\Psi} T_A \Psi} = 2 - \frac{8}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big), \tag{17}$$

$$\Delta_{i\overline{\Psi}\Gamma_3\Gamma_5\Psi} = 2 + \frac{16}{3\pi^2 N_f} + \mathcal{O}\big(1/N_f^2\big). \tag{18}$$

4 Discussion

In Tables 1 and 2, we evaluate the previous expressions at values of N_f currently accessible to QMC simulations to obtain estimates of critical exponents at the $U(1)$-VBS and semimetal-to-Kekulé-VBS transitions, respectively. In Table 2 we also provide the values of η_ϕ and ν already obtained from QMC simulations [15].

Table 1 Large-N_f critical exponents for the chiral $O(2)$ QED_3-GN model

	η_ϕ	$1/\nu$	ν	$\Delta_{\overline{\Psi}\Psi}$	$\Delta_{i\overline{\Psi}\Gamma_3\Gamma_5\Psi}$
$N_f = 4$	1.473	0.3245	3.081	1.662	2.675
$N_f = 6$	1.315	0.5497	1.819	1.775	2.450
$N_f = 8$	1.236	0.6623	1.510	1.831	2.338

Table 2 Large-N_f and QMC critical exponents for the chiral $O(2)$ GN model

	η_ϕ	η_ϕ (QMC)	$1/\nu$	ν	ν (QMC)	$\Delta_{\overline{\Psi}\Psi}$	$\Delta_{i\overline{\Psi}\Gamma_3\Gamma_5\Psi}$
$N_f = 2$	0.9166	0.71(3)	1.209	0.8270	1.06(5)	1.865	2.270
$N_f = 3$	0.9329	0.78(2)	1.033	0.9681	1.07(4)	1.910	2.180
$N_f = 4$	0.9454	0.80(4)	0.9848	1.015	1.11(3)	1.932	2.135
$N_f = 5$	0.9542	0.85(4)	0.9686	1.032	1.07(2)	1.946	2.108
$N_f = 6$	0.9607	0.87(4)	0.9632	1.038	1.06(3)	1.955	2.090

For ν, a better agreement with QMC results is found than with a previously used renormalization-group approach [15], while the opposite is true for η_ϕ.

Our discussion of the $U(1)$-VBS transition has thus far ignored the compactness of the $U(1)$ gauge field, which may cause monopole (instanton) proliferation. In the large-N_f limit, the scaling dimension of the smallest symmetry-allowed monopole operator at the conformal QED_3 fixed point is $\Delta_M = 0.53N_f - 0.0383 + \mathcal{O}(1/N_f)$ [28]. This suggests that for $N_f = 6$ and 8, $\Delta_M > 3$ and monopoles are irrelevant, while for $N_f = 4$ the smallest monopole is relevant; however, at such values of N_f, subleading corrections in the $1/N_f$ expansion may be significant. As with the chiral $O(3)$ QED_3-GN model [9], at the chiral $O(2)$ QED_3-GN fixed point the scaling dimension of the smallest monopole operator is expected to grow linearly with N_f at leading order in $1/N_f$ but with a different coefficient than at the conformal QED_3 fixed point. Should this coefficient be sufficiently small, the $U(1)$-VBS critical point may be destabilized at sufficiently small N_f, resulting in a first-order transition. The QMC results [6, 7], however, suggest that a continuous $U(1)$-VBS transition persists with increasing $N_f \geq 4$ but is simply pushed to larger values of J_c. While the critical value of N_f above which monopole operators are irrelevant is not precisely known, the numerical observation of a continuous transition suggests that either monopoles are in fact irrelevant for $N_f \geq 4$ or the crossover length scale $L_* \sim a g_0^{-1/(3-\Delta_M)}$ beyond which they proliferate, where a is the lattice constant and g_0 the bare monopole fugacity, is much larger than the system sizes currently accessible in QMC.

Acknowledgments We thank the CRM and the QTS-XI committee for organizing this excellent conference. We thank J. A. Gracey, P. Marquard, and N. Zerf for collaboration on related topics, É. Dupuis, S. Giombi, I. F. Herbut, I. R. Klebanov, Z. Y. Meng, A. Penin, M. M. Scherer, and W. Witczak-Krempa for useful discussions, and NSERC, CIFAR, the University of Alberta's Theoretical Physics Institute (TPI), and the CRC program for financial support.

References

1. T. Senthil, et al., Science **303**, 1490 (2004)
2. T. Senthil, et al., Phys. Rev. B **70**, 144407 (2004)
3. X.-G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, New York, 2004)
4. E. Zohar, J.I. Cirac, B. Reznik, Rep. Prog. Phys. **79**, 014401 (2016)
5. E.A. Martinez, et al., Nature **534**, 516 (2016)
6. X.Y. Xu, et al., Phys. Rev. X **9**, 021022 (2019)
7. W. Wang, et al., Phys. Rev. B **100**, 085123 (2019)
8. M. Hermele, T. Senthil, M.P.A. Fisher, Phys. Rev. B **72**, 104404 (2005)
9. É. Dupuis, M.B. Paranjape, W. Witczak-Krempa, Phys. Rev. B **100**, 094443 (2019)
10. N. Zerf, et al., Phys. Rev. B **100**, 235130 (2019)
11. Y. Nambu, G. Jona-Lasinio, Phys. Rev. **122**, 345 (1961)
12. S.P. Klevansky, R.H. Lemmer, Phys. Rev. D **39**, 3478 (1989)
13. T.C. Lang, et al., Phys. Rev. Lett. **111**, 066401 (2013)
14. Z. Zhou, et al., Phys. Rev. B **93**, 245157 (2016)
15. Z.-X. Li, et al., Nat. Commun. **8**, 314 (2017)
16. N. Zerf, et al., Phys. Rev. D **101**, 094505 (2020)
17. T. Appelquist, D. Nash, L.C.R. Wijewardhana, Phys. Rev. Lett. **60**, 2575 (1988)
18. J.A. Gracey, J. Phys. A **25**, L109 (1992)
19. J.A. Gracey, J. Phys. A **26**, 1431 (1993); **51**, 479501 (2018)
20. J.A. Gracey, Ann. Phys. **224**, 275 (1993)
21. T. Alanne, S. Blasi, Phys. Rev. D **98**, 116004 (2018)
22. J.A. Gracey, Phys. Rev. D **98**, 085012 (2018)
23. R. Boyack, A. Rayyan, J. Maciejko, Phys. Rev. B **99**, 195135 (2019)
24. J.A. Gracey, Mod. Phys. Lett. A **8**, 2205 (1993)
25. J.A. Gracey, Phys. Lett. B **308**, 65 (1993)
26. J.A. Gracey, Phys. Rev. D **50**, 2840 (1994)
27. J.A. Gracey, Phys. Rev. D **59**, 109904 (1999)
28. S.S. Pufu, Phys. Rev. D **89**, 065016 (2014)

Emergent Geometry from Entanglement Structure

Sudipto Singha Roy, Silvia N. Santalla, Javier Rodríguez-Laguna, and Germán Sierra

Abstract We attempt to reveal the geometry, emerged from the entanglement structure of any general N-party pure quantum many-body state by representing entanglement entropies corresponding to all 2^N bipartitions of the state by means of a generalized adjacency matrix. We show this representation is often exact and may lead to a geometry very different than suggested by the Hamiltonian. Moreover, in all the cases, it yields a natural entanglement contour, similar to previous proposals. The formalism is extended for conformal invariant systems, and a more insightful interpretation of entanglement is presented as a flow among different parts of the system.

Keywords Quantum entanglement · Geometry · Entanglement entropy

1 Introduction

Study of the distribution of quantum entanglement in different parts of the low-lying states of quantum many-body Hamiltonians often unveils many interesting features related to the physical system [1–3]. For instance, bounded growth of quantum entanglement between a region and its exterior can be attributed to the fact that interactions in the quantum many-body systems are typically local [4–7]. In this work, we aim to explore the connection between the area-law for entanglement [4–

S. S. Roy · G. Sierra (✉)
Instituto de Física Teórica, UAM-CSIC, Universidad Autónoma de Madrid Cantoblanco, Madrid, Spain
e-mail: german.sierra@uam.es

S. N. Santalla
Dept. de Física and Grupo Interdisciplinar de Sistemas Complejos (GISC), Universidad Carlos III de Madrid, Madrid, Spain
e-mail: silvia.santalla@uc3m.es

J. Rodríguez-Laguna
Dept. de Física Fundamental, UNED, Madrid, Spain

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_33

7] and geometry which emerges from the distribution of quantum entanglement across all possible bipartitions of a pure quantum many-body state. Towards this aim, we first define the notion of geometry by means of a generalized adjacency matrix such that the approximate entanglement entropy of any given bipartition can be obtained as a linear sum of the weights of the links connecting it with its surroundings. We show that the representation is exact when there is a perfect area-law. In other cases, it still provides an efficient approximation with minimal error. Interestingly, we also report some other important states, e.g., the *rainbow state* [8–15], where though a strong violation of area-law is observed for the geometry defined by the local structure of the Hamiltonian, an area-law feature can indeed be recovered for a geometry which is completely different than that suggested by the Hamiltonian.

As an application of the formalism, we provide a route to compute the entanglement contour function for quantum many-body systems, which is radically different than that previously introduced in Ref. [16]. A quantitative comparison of the contour functions obtained using these two different approaches is made for the ground state of a non-interacting model. Additionally, we also study the behavior of contour function obtained for an interacting model, which surpasses the limitation of the previous approach [16–22].

Finally, we extend our analysis to conformal invariant physical systems [23–26]. As an important finding, we show that the conformal field theory (CFT) descriptions help us to interpret the elements of the generalized adjacency matrix as the two-point correlator of an entanglement current operator. This field theory realization provides a framework to consider entanglement as a flow among different parts of the system [27], similar to the flow of energy that is characterized by the stress tensor.

In the following sections, after briefly introducing the formalism, we elaborate on our main findings.

2 Emergent Geometry

We start with an N-party pure quantum state $|\psi\rangle$, and characterize its entanglement properties by computing the von Neumann entropies $S_A = -\mathrm{Tr}_A(\rho_A \log \rho_A)$ for all possible bipartitions of the state, namely (A, A^c), where $\rho_A = \mathrm{Tr}_{A^c}|\psi\rangle\langle\psi|$, and $\mathrm{Tr}_{A(A^c)}$ denotes partial trace on the subsystem $A(A^c)$. We then aim to investigate whether the set of entropies obtained in this way respond to an area-law for some geometry. As a first step, we assign a classical spin configuration $\{s_i\}^N$ to each such bipartitions using the rule

$$s_i = \begin{cases} 1, & \text{if } i \in A \\ -1, & \text{if } i \in A^c. \end{cases}$$

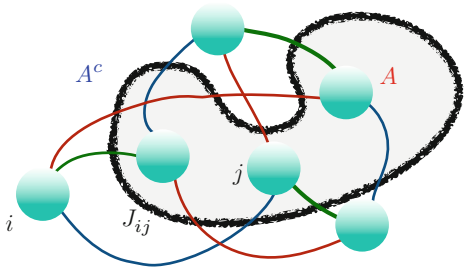


Fig. 1 Schematic of the entanglement entropy obtained for an arbitrary bipartition (A, A^c) by removing the links connecting the sites. Here the links represent the constants J_{ij}

These spins are not physical but only a convenient way to describe the different bipartitions of the system. If two spins, say i and j, belong to the same partition, A or A^c, we get $s_i s_j = 1$, while if they belong to different partitions, $s_i s_j = -1$. In the former case, there is no contribution to the entanglement entropy S_A, while in the second case, they may contribute to S_A with a certain amount that will depend on their positions. We are thus led to express the entropy S_A of the bipartition (A, A^c) as

$$S_A = \frac{1}{2} \sum_{ij} J_{ij}(1 - s_i s_j) + s_0, \tag{1}$$

where J_{ij} defines the coupling between the classical spins i and j and s_0 may constitute a topological entropy term. The entropy S_A thus can be further simplified as the sum of contributions coming from all possible pairs, J_{ij}, i.e.,

$$S_A = \sum_{i \in A, j \in A^c} J_{ij} + s_0. \tag{2}$$

A closer look at the derivation of the above entropy function reveals the fact that it is a clear manifestation of the area-law of entanglement entropy associated with the geometry revealed by the elements of J. More elaborately, the coupling function J_{ij} can be interpreted as the weight of an adjacency matrix of a generalized graph, such that the approximate entanglement entropy of the region A can be computed only by simply summing the weights (J_{ij}) associated with all the connecting edges between A and A^c. A schematic representation of the above formulation is depicted in Fig. 1. If Eq. (2) holds exactly or at least approximately, the matrix J will be termed the entanglement adjacency matrix (EAM) of the state $|\psi\rangle$. Additionally, we note that for the case when Eq. (2) is exact, J_{ij} equals to the mutual information between the sites i, j.

2.1 Graphical Representation

An equivalent way to conceptualize the entropic functions obtained for different bipartitions is through the following graphical representation we outline here. Similar to Venn diagrams, which illustrate the logical relationships between two or more sets, here we present the schematic representation of the entropy values of different bipartitions by shading different regions in the J-matrix. As an example, consider a contiguous bipartition (A, A^c), such that the sites $1 \ldots m \in A$ and $m+1 \ldots N \in A^c$. The entropy value of the block A (S_A) can be schematically represented by shading the region of the J-matrix with $i \in 1, 2, \ldots m$ and $j \in m+1 \in N$ and its conjugate part. Figure 2 depicts the single-site entropies and the mutual information of a system AA^c by shading different regions in the J-matrix.

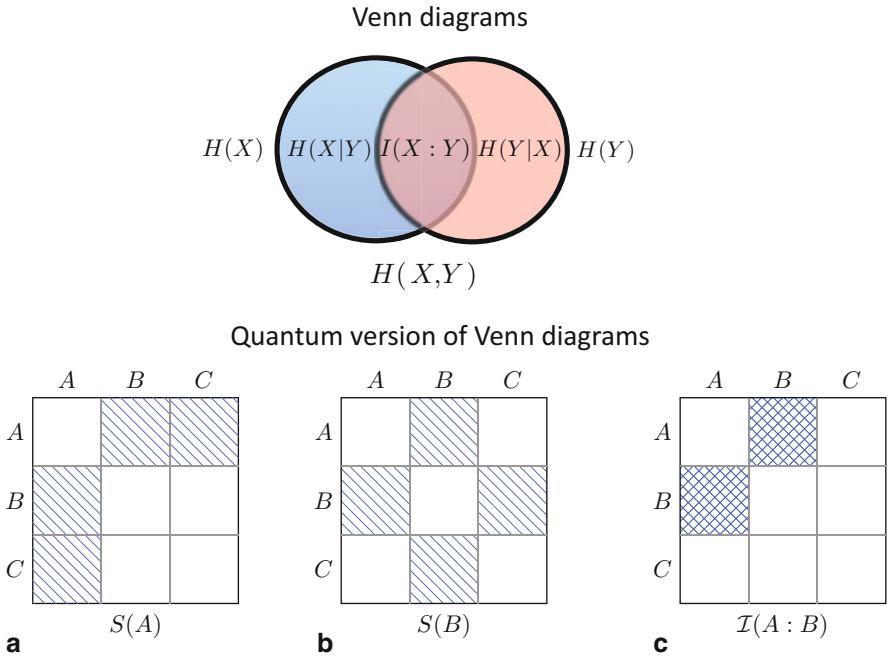


Fig. 2 Similar to the illustration of the different entropies and the mutual information between two variables X and Y using Venn diagrams, as shown in the upper panel, a graphical representation of the quantum version of different entropic relations is presented in the lower panel (**a**)–(**c**)

2.2 Exact Examples

For any general pure quantum state, computation of its entanglement adjacency matrix requires knowledge of the entanglement of all possible bipartitions of the state. The number of such bipartitions increases exponentially with the size of the system. Hence, even for moderate system size, estimation of the entanglement adjacency matrix demands lots of computational effort. However, if the quantum state possesses certain properties such that all entropies can be computed analytically, corresponding entanglement adjacency matrix can be obtained straightforwardly. Below we discuss few such cases, where the elements of the entanglement adjacency matrix can be computed exactly.

I. *Dimer model.*—We start with the *dimer state*, which can be mathematically expressed as $|\Psi\rangle = |i_1 j_1\rangle \otimes |i_2 j_2\rangle \otimes \cdots \otimes |i_k j_k\rangle$, with $k = \frac{N}{2}$ for even N and $|i_l j_l\rangle = \frac{1}{\sqrt{2}}(|\uparrow_{i_l}\downarrow_{j_l}\rangle - |\downarrow_{i_l}\uparrow_{j_l}\rangle)$. Such states belong to the family of valence bond states and appear as approximate ground states of certain strongly inhomogeneous free-fermionic model. Here, from the configuration of the state, one can observe that all the single-site entropies become $S(\rho_i) = \log(2), \forall i \in N$. On the other hand, for the two-site blocks, if i, j form a dimer $S(\rho_{ij}) = 0$, and $S(\rho_{ij}) = 2\log(2)$, otherwise. Therefore, the elements of the entanglement adjacency matrix possess non-zero values only when i, j form singlet, given by $J_{ij} = \log(2)$. In this case, one can find that the geometry revealed by the entanglement adjacency matrix is a mere restriction of the one-dimensional adjacency matrix representing the Hamiltonian.

II. *Rainbow state.*—Another important member of the family of valence bond states we consider in our work is the *rainbow state*, which is also the ground state of a local Hamiltonian [8–15]. Here, dimer are established among symmetric qubits with respect to the center: $i_k = k$, $j_k = N + 1 - k$. The state exhibits volume-law scaling of entanglement entropy with the increase of the system size. In this case also, all the single-site entropies become $S(\rho_i) = \log(2)$. Whereas, the non-zero values of the entanglement entropies can be obtained only for ρ_{ij} such that $i+j \neq N+1$, given by $S(\rho_{ij}) = 2\log(2)$. As a result, we get $J_{ij} = \log(2)$, for $i + j = N + 1$ and zero otherwise. Interestingly, one can note that in this case, the entanglement adjacency matrix is not emerging as a restriction on the adjacency matrix representing the Hamiltonian. In other words, an observer trying to determine the geometry from the distribution of the entanglement will not find the correct geometry of the Hamiltonian.

III. *GHZ state.*—A different case we consider here is the N-party GHZ state, expressed as $|GHZ\rangle = \frac{1}{\sqrt{2}}(|0\rangle^N| + |1\rangle^N)$. In this case, the entropy values of all the bipartitions, irrespective of the number of sites, become identical, given by $\log(2)$. Hence, all the J_{ij}'s become same. As a result, to represent the block entropies using our formalism, we consider $J_{ij} = 0\ \forall i, j \in N$ and put

the value of the constant term in Eq. (2), $s_0 = \log(2)$. This suggests that the GHZ state does not have a geometrical interpretation in this framework.

2.3 Numerical Computation

In this section, we describe the numerical methodology to obtain the entanglement adjacency matrix for which Eq. (2) is not exact. For any general block, the relation between parameters and entropies can be expressed through

$$\sum_{(ij)} \mathcal{D}_{I,(ij)} J_{ij} = S_I, \tag{3}$$

where $I = (x_1 \cdots x_N)$ denotes the binary expansion for the index of each block, i.e. $x_k = 1$ if site k belongs to block I (and zero otherwise), and $\mathcal{D}_{(x_1 \cdots x_N),(ij)} = 1$ if $(x_i, x_j) = (0, 1)$ or (1,0), and zero otherwise. In our case, vector J contains all the J_{ij} in order, i.e., has dimension $N(N-1)/2$, while vector S_I contains all the entropies, so it has dimension 2^N. Thus, matrix $\mathcal{D}$ has dimension $2^N \times N(N-1)/2$. In other terms: as many rows as entropies, and as many columns as couplings.

Equation (3) is a strongly overdetermined linear system which will be, in general, incompatible. Yet, it is possible to find an approximate solution in the least-squares sense, using the equation

$$\sum_{(i'j')} \mathcal{D}^{\dagger} \mathcal{D}_{(ij),(i'j')} J_{i'j'} = \sum_{I} \mathcal{D}_{I,(ij)} S_I. \tag{4}$$

Subsequently, an estimation of the relative error made in this optimization process can be made as follows. Let $\hat{S}_I$ be the estimate obtained through Eq. (4). The error will be defined as

$$\mathcal{E} = \frac{1}{2^N} \sum_{I=0}^{2^N-1} \left| S_I - \hat{S}_I \right|. \tag{5}$$

In the forthcoming section, we will use this formula to compute the error made in computation of entanglement adjacency matrix for various physical models.

3 Entanglement Contour

In this section, we discuss another important facet of our formalism, where a more refined approach to characterize entanglement entropy of any bipartition is presented in terms of the entanglement contour function introduced earlier in the

literature [16] (see also [12, 17–22]). The entanglement contour function for a given block estimates the contribution of each site to the entanglement entropy obtained for that bipartition. For a given block A, mathematically it can be expressed as

$$S_A = \sum_{i \in A} s_A(i), \qquad s_A(i) \geq 0\,. \tag{6}$$

Interestingly, from Eq. (2) one can observe that the entanglement adjacency matrix provides a natural entanglement contour which can be expressed as

$$s_A(i) \equiv \sum_{j \in A^c} J_{ij}\,. \tag{7}$$

The contour function defined above satisfies all the standard constraints listed in Ref. [16]. Here, we stress the fact that unlike the actual formulation of the entanglement contour introduced in Ref. [16], our approach aims to provide an overall description of bipartite entanglement by considering contributions of *all* bipartitions and not just the ones consisting of simply connected intervals. Moreover, the formalism includes any general quantum systems, including interacting cases.

Contour Plot for Free-Fermionic Hamiltonian For free-fermionic model, described below

$$\mathcal{H}_{\text{free-ferm}} = -\frac{1}{2} \sum_{ij} t_{ij} \left(c_i^\dagger c_j + hc\right), \tag{8}$$

where $c_i\,(c_i^\dagger)$'s is the fermionic annihilation (creation) operators at site i, and t_{ij} is the hopping matrix, a proposal for the contour is given in Ref. [16],

$$s_A(i) = \sum_{p=1}^{|A|} \left|\Phi_{p,i}^{(A)}\right|^2 H(\nu_p), \tag{9}$$

where $\Phi_{p,i}^{(A)}$ is the eigenvector with eigenvalue ν_p, of the correlation matrix block [8, 28] restricted to A and $H(x) = -\big[x \log x + (1-x)\log(1-x)\big]$. Using the above equation, in Fig. 3a, we compute the entanglement contour function for the ground state of the dimerized Hamiltonian, which can be obtained from the free-fermionic model described in Eq. (8), for $t_{ij} = (1+\delta(-1)^i)$, $|i-j| = 1$ and compare that to the contour function obtained using the elements of the entanglement adjacency matrix, J_{ij}, as described in Eq. (7). From the figure, we note that the contour functions obtained using these two different methods are very similar to each other.

Contour Plot for XXZ Hamiltonian Subsequently, we move one step further and apply the formalism to an interacting model, namely, the one-dimensional XXZ model, expressed as

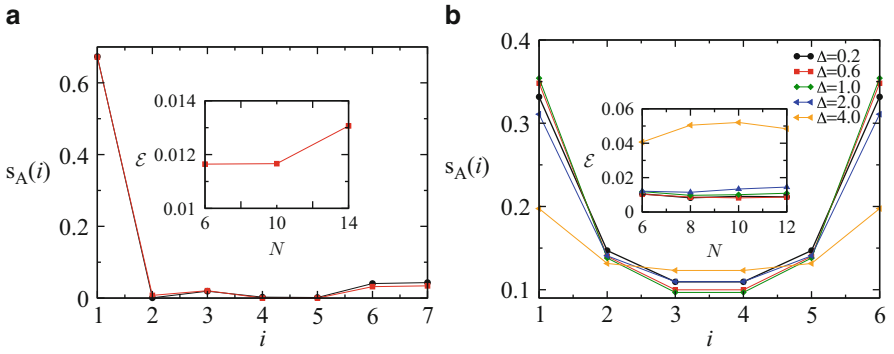


Fig. 3 In panel (**a**), we compare the contour functions for the entanglement entropy $s_A(i)$, those obtained using the Eqs. (7) and (9), for the ground state of dimerized Hamiltonian ($t_{ij} = (1 + \delta(-1)^i)$, $|i - j| = 1, \delta = 0.5, N = 14$). Whereas, in panel (**b**), we plot the contour functions for the entanglement entropy obtained using Eq. (7) for ground state of XXZ Hamiltonian for different values of the parameter Δ, and for $N = 12$. Additionally, in the inset of both the figures, the scaling of the error ($\mathcal{E}$) with the system size (N) have been shown for all the parameter values considered

$$H_{XXZ} = \sum_{i}^{N} S_i^x S_{i+1}^x + S_i^y S_{i+1}^y + \Delta S_i^z S_{i+1}^z, \tag{10}$$

where $S_l^k (k \in x, y, z)$ are the Pauli operators at site l, and Δ is the anisotropy along the z-direction. Note that in this case, to obtain the set of entropies for all possible bipartitions of the ground state of the model, we perform the exact diagonalization method. The behavior of the entanglement contour function obtained for the half-chain, for the critical ($\Delta \leq 1$) and non-critical ($\Delta > 1$) cases are depicted in Fig. 3b.

4 Entanglement Current

In this section, we extend our formalism to one-dimensional conformal invariant systems and attempt to provide an interpretation of the entanglement adjacency matrix entries, J_{ij}, as the two-point correlator of an entanglement current operator. The entanglement entropy of the ground state of a CFT for an interval $A = (u, v)$ embedded in the infinite line is given by

$$S_A = \frac{c}{3} \log \frac{v - u}{\epsilon}, \tag{11}$$

where c is the central charge and $\epsilon > 0$ a UV cut-off. One can note that Eq. (11) can be obtained from a continuous version of Eq. (2)

$$S_A = \frac{c}{6} \int_{A_\epsilon} dx \int_{A^c} dy \; J(x, y)\,, \tag{12}$$

with $A_\epsilon = (u + \epsilon, v - \epsilon)$ and $A^c = (-\infty, u) \cup (v, \infty)$, by choosing

$$J(x, y) = \frac{1}{(x - y)^2} . \tag{13}$$

This equation indicates that $J(x, y)$ is the two-point correlator, on the complex plane, of a current operator $\mathbf{J}$, whose integration along segments, as in Eq. (12), is invariant under reparametrization. $J(x, y)dxdy$ represents the amount of entanglement between the intervals $(x, x + dx)$ and $(y, y + dy)$. This field theory realization leads to think of entanglement as a flow among the parts of the system, in analogy to the flow of energy that is characterized by the stress tensor. Moreover, using the construction described in Ref. [26] for entanglement Hamiltonians in CFT, one can show that Eq. (12) reproduces the values of S_A, for the space-time geometries Σ, that are conformally equivalent to an annulus. In these cases $J(x, y)$ is given by the two-point correlator $J(x, y) = \langle \mathbf{J}(x)\, \mathbf{J}(y) \rangle_\Sigma$. Notice that in the conformal field theory systems the representation is exact only when A is an interval, but not in general.

5 Conclusion

To conclude, in this work, we aimed to unveil the geometry revealed from the entanglement properties of any pure quantum state through the elements of a generalized adjacency matrix, such that the entropy values of any bipartition of the state can be approximated as a weighted sum of all the links connecting the sites across that bipartition. We reported certain examples, where the optimal geometry emerged from the entanglement structure, turned out to be completely different from that suggested by the parent Hamiltonian of the system. Subsequently, we showed that our formalism provided a natural route to compute the entanglement contour, introduced earlier for the non-interacting models, which essentially helped us to extend the concept for interacting models as well. Finally, we showed that for conformal invariant systems, a more insightful explanation of the elements of such generalized adjacency matrices can be obtained in terms of a two-point correlator of an entanglement current operator.

Acknowledgments GS would like to thank William Witczak-Krempa for the invitation to participate in the Quantum Theory and Symmetry XI conference held in Montreal in July 2019. We acknowledge financial support from the grants PGC2018-095862-B-C21, QUITEMAD+ S2013/ICE-2801, SEV-2016-0597 of the "Centro de Excelencia Severo Ochoa" Programme and the CSIC Research Platform on Quantum Technologies PTI-001.

References

1. R. Horodecki, P. Horodecki, M. Horodecki, K. Horodecki, Quantum entanglement. Rev. Mod. Phys. **81**, 865 (2009). https://doi.org/10.1103/RevModPhys.81.865
2. A. Osterloh, L. Amico, G. Falci, R. Fazio, Scaling of entanglement close to a quantum phase transition. Nature **416**, 608 (2002). https://doi.org/10.1038/416608a
3. T.J. Osborne, M.A. Nielsen, Entanglement in a simple quantum phase transition. Phys. Rev. A **66**, 032110 (2002). https://doi.org/10.1103/PhysRevA.66.032110
4. L. Amico, R. Fazio, A. Osterloh, V. Vedral, Entanglement in many-body systems. Rev. Mod. Phys. **80**, 517 (2008). https://doi.org/10.1103/RevModPhys.80.517
5. M. Srenidcki, Entropy and area. Phys. Rev. Lett **71**, 666 (1993). https://doi.org/10.1103/PhysRevLett.71.666
6. J. Eisert, M. Cramer, M.B. Plenio, Colloquium: area laws for the entanglement entropy. Rev. Mod. Phys. **82**, 277 (2010). https://doi.org/10.1103/RevModPhys.82.277
7. M.M. Wolf, F. Verstraete, M.B. Hastings, J.I. Cirac, Area Laws in quantum systems: mutual information and correlations. Phys. Rev. Lett. **100**, 070502 (2008). https://doi.org/10.1103/PhysRevLett.100.070502
8. G. Vitagliano, A. Riera, J.I. Latorre, Volume-law scaling for the entanglement entropy in spin 1/2 chains. New J. Phys. **12**, 113049 (2010). https://doi:10.1088/1367-2630/12/11/113049
9. G. Ramírez, J. Rodríguez-Laguna, G. Sierra, From conformal to volume-law for the entanglement entropy in exponentially deformed critical spin 1/2 chains. J. Stat. Mech. **2014**, P10004 (2014). https://doi.org/10.1088/1742-5468/2014/10/P10004
10. G. Ramírez, J. Rodríguez-Laguna, G. Sierra, Entanglement over the rainbow. J. Stat. Mech. **2015**, P06002 (2015). https://doi.org/10.1088/1742-5468/2014/10/P10004
11. J. Rodríguez-Laguna, J. Dubail, G. Ramírez, P. Calabrese, G. Sierra, More on the rainbow chain: entanglement, space-time geometry and thermal states. J. Phys. A: Math. Theor. **50**, 164001 (2017). https://doi.org/10.1088/1751-8121/aa6268
12. E. Tonni, J. Rodríguez-Laguna, G. Sierra, Entanglement Hamiltonian and entanglement contour in inhomogeneous 1D critical system. J. Stat. Mech. **2018**, 043105 (2018). https://doi.org/10.1088/1742-5468/aab67d
13. V. Alba, S.N. Santalla, P. Ruggiero, J. Rodrıguez-Laguna, P. Calabrese, G. Sierra, Usual area-law violation in random inhomogeneous systems. J. Stat. Mech. **2018** 023105 (2019). https://doi.org/10.1088/1742-5468/ab02df
14. N.S.S. de Buruaga, S.N. Santalla, J. Rodríguez-Laguna, G. Sierra, Symmetry protected phases in inhomogeneous spin chains. J. Stat. Mech. **2019**, 093102 (2019). https://doi.org/10.1088/1742-5468/ab3192
15. I. MacCormack, A. Liu, M. Nozaki, S. Ryu, Holographic duals of inhomogeneous systems: the rainbow chain and the sine-square deformation model. arXiv:1812.10023
16. Y. Chen, G. Vidal, Entanglement contour. J. Stat. Mech. **2014**, P10011 (2014). https://doi.org/10.1088/1742-5468/2014/10/P10011
17. A. Botero, B. Reznik, Spatial structures and localization of vacuum entanglement in the linear harmonic chain. Phys. Rev. A **70**, 052329 (2004). https://doi.org/10.1103/PhysRevA.70.052329
18. I. Frérot, T. Roscilde, Area law and its violation: a microscopic inspection into the structure of entanglement and fluctuations. Phys. Rev. B **92**, 115129 (2015). https://doi.org/10.1103/PhysRevB.92.115129
19. A. Coser, C.D. Nobili, E. Tonni, A contour for the entanglement entropies in harmonic lattices. J. Phys. A: Math. Theor. **50**, 314001 (2017). https://doi.org/10.1088/1751-8121/aa7902
20. Q. Wen, Fine structure in holographic entanglement and entanglement contour. Phys. Rev. D 98, 106004 (2018). https://doi.org/10.1103/PhysRevD.98.106004
21. E. Tonni, Entanglement Hamiltonians and contours on a segment. Talk at the workshop It from Qubit, Centro Atómico Bariloche (2018)

22. J. Kudler-Flam, I. MacCormack, S. Ryu, Holographic entanglement contour, bit threads, and the entanglement tsunami (2019). arXiv:1902.04654. https://doi.org/10.1088/1751-8121/ab2dae
23. C. Holzhey, F. Larsen, F. Wilczek, Geometric and renormalized entropy in conformal field theory. Nucl. Phys. B **424**, 443 (1994). https://doi.org/10.1016/0550-3213(94)90402-2
24. G. Vidal, J.I. Latorre, E. Rico, A Kitaev, Entanglement in quantum critical phenomena, Phys. Rev. Lett. **90**, 227902 (2003). https://doi.org/10.1103/PhysRevLett.90.227902
25. P. Calabrese, J. Cardy,Entanglement entropy and quantum field theory. J. Stat. Mech. **2004**, P06002, (2004). https://doi.org/10.1088/1742-5468/2004/06/P06002
26. J. Cardy, E. Tonni, Entanglement Hamiltonians in two-dimensional conformal field theory. J. Stat. Mech. **2016**, 123103 (2016). https://doi.org/10.1088/1742-5468/2016/12/123103
27. T.S. Cubitt, F. Verstraete, J.I. Cirac, Entanglement flow in multipartite systems. Phys. Rev. A **71**, 052308 (2005). https://doi.org/10.1103/PhysRevA.71.052308
28. I. Peschel, Calculation of reduced density matrices from correlation functions. J. Phys. A Math. Gen. **36**, L205 (2003)

Interplay of Coulomb Repulsion and Spin–Orbit Coupling in Superconducting 3D Quadratic Band Touching Luttinger Semimetals

Serguei Tchoumakov, Louis J. Godbout, and William Witczak-Krempa

Abstract We investigate the superconductivity of 3D Luttinger semimetals, such as YPtBi, where Cooper pairs are constituted of spin-3/2 quasiparticles. Various pairing mechanisms have already been considered for these semimetals, such as from polar phonons modes, and in this work we explore pairing from the screened electron–electron Coulomb repulsion.

In these materials, the small Fermi energy and the spin–orbit coupling strongly influence how charge fluctuations can mediate pairing. We find the superconducting critical temperature as a function of doping for an $s-$wave order parameter, and determine its sensitivity to changes in the dielectric permittivity. Also, we discuss how order parameters other than $s-$wave may lead to a larger critical temperature, due to spin–orbit coupling.

Keywords Superconductivity · Luttinger semimetals · Coulomb repulsion · Critical temperature · Eliashberg equation · Spin–orbit

1 Introduction

In regular metals, Coulomb repulsion is commonly believed to compete against the superconducting pairing between electrons. For example, in the electron-phonon mechanism of superconductivity with electron-phonon coupling g, the critical temperature below which an electron gas becomes superconducting is [1]

S. Tchoumakov · L. J. Godbout
Département de Physique, Université de Montréal, Montréal, QC, Canada
e-mail: serguei.tchoumakov@umontreal.ca; louis.godbout.2@umontreal.ca

W. Witczak-Krempa (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: w.witczak-krempa@umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_34

$$T_c = \frac{\langle\omega\rangle}{1.2} \exp\left(-\frac{1.04(1+\lambda)}{\lambda - \mu^*(1+0.62\lambda)}\right), \tag{1}$$

where $\lambda = 2\int_0^\infty d\omega g^2 D(\omega)/\omega$ is the coupling constant, $\langle\omega\rangle = 2\int_0^\infty d\omega g^2 D(\omega)$ is the averaged phonon frequency, and μ^* is the Coulomb pseudo-potential. $D(\omega)$ is the phonon density of states. Equation (1) informs us that increasing the electronic Coulomb repulsion exponentially decreases the critical temperature. In a normal electron gas, the interaction strength is proportional to the Wigner–Seitz radius given in units of the effective Bohr radius, $r_s = me^2/(\alpha\epsilon^* k_F)$ where m is the band mass, e the electronic charge, $\alpha = (4/9\pi)^{1/3} \approx 0.52$, ϵ^* the background dielectric constant, and k_F is the Fermi wavevector. The superconductivity in semiconductors is often attributed to the electron-phonon coupling. However, for some materials such as $SrTiO_3$ and bismuth-based half-Heuslers, like YPtBi, the importance of the electron-phonon coupling in superconductivity is yet unresolved. In $SrTiO_3$ it was even proposed that superconductivity may come from the electron–electron repulsion [2, 3]. The qualitative explanation does not only rely on the Kohn–Luttinger mechanism [4] but also on the contribution of plasmons to screening [2]. The effective attraction between electrons is a consequence of the screening of the Coulomb potential, with a dielectric function $\epsilon(\omega, \mathbf{q})$ that is computed in the random phase approximation

$$\epsilon_{\mathrm{RPA}}(\omega, \mathbf{q}) = 1 - V_0(q)\Pi_0(\omega, \mathbf{q}), \tag{2}$$

with $V_0(q) = 4\pi e^2/(\epsilon^* q^2)$ the bare Coulomb potential and $\Pi_0(\omega, \mathbf{q})$ the bare electron polarizability. The dielectric function $\epsilon(\omega, \mathbf{q})$ depends on the system under study and has a role similar to the density of states of phonons, $D(\omega)$ that appears in Eq. (1). In Ref. [5] we use a variational approach similar to that in [6] to show how the critical temperature depends on each component $(\omega, \mathbf{q})$ of the dielectric function, as we discuss further below.

This mechanism for $SrTiO_3$, however, does not directly apply to bismuth-based half-Heusler materials, such as YPtBi, where the band structure is not well approximated by the free Hamiltonian $H_{\mathrm{N}}(\mathbf{k}) = \hbar^2 k^2/(2m) - \mu$ but also includes strong spin–orbit coupling. It is a candidate Luttinger semimetal with Hamiltonian [7]

$$\hat{H}_0(\mathbf{k}) = \frac{\hbar^2}{2m}\left[-\frac{5}{4}\mathbf{k}^2 + \left(\mathbf{k}\cdot\hat{\mathbf{J}}\right)^2\right] - \mu, \tag{3}$$

where we introduce the $j = 3/2$ total angular momentum operators $\hat{\mathbf{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)$ and the chemical potential μ. This model has inversion, rotational, and time-reversal symmetries. The spectrum consists of four bands that meet quadratically at $\mathbf{k} = 0$ with degenerate lower and upper bands with energies $\pm\hbar^2 k^2/(2m)$ as shown in Fig. 1a. In the present proceeding we outline our findings regarding screening, quasiparticles, and superconductivity in Luttinger semimetals arising from the

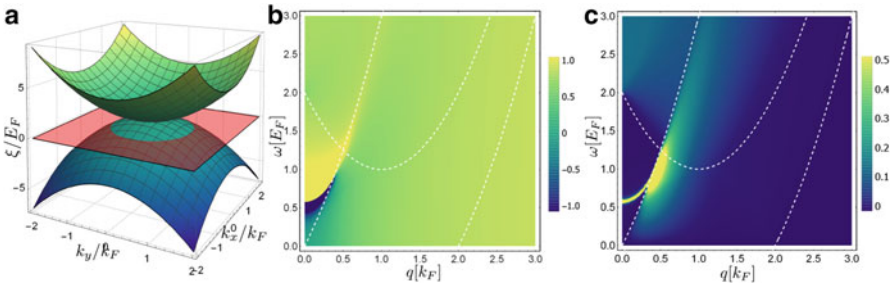


Fig. 1 (**a**) Band structure of a Luttinger semimetal, the red plane is at the Fermi level. The upper and lower bands are doubly degenerate. (**b**)–(**c**) The (**b**) real and (**c**) imaginary parts of the inverse dielectric permittivity, $1/\epsilon(\omega, q)$, for $r_s = 0.5$ as a function of wavevectors, q, and frequencies, ω. The white dashed lines are the branches of the particle-hole continuum

screened Coulomb repulsion [5, 8]. We also discuss how the $J = L = S = 1$ order parameter may have a larger critical temperature than in the $s-$wave channel, due to spin–orbit coupling.

2 Screening and Electronic Properties of Luttinger Semimetals

We perturb the bare Hamiltonian (3) with the bare Coulomb potential $V_0(q)$

$$\hat{H}_{\text{int}} = \frac{1}{2\mathcal{V}} \sum_{s_1 s_2 \mathbf{k}_1 \mathbf{k}_2, \mathbf{q}\neq 0} V_0(q) \hat{\psi}^{\dagger}_{\mathbf{k}_1+\mathbf{q}s_1} \hat{\psi}^{\dagger}_{\mathbf{k}_2-\mathbf{q}s_2} \hat{\psi}_{\mathbf{k}_2 s_2} \hat{\psi}_{\mathbf{k}_1 s_1}, \tag{4}$$

where $\mathcal{V}$ is the volume of the electron gas and introduce the annihilation operators $\hat{\psi}_{\mathbf{p}s} = \{\hat{\psi}_{\mathbf{p},3/2}, \hat{\psi}_{\mathbf{p},1/2}, \hat{\psi}_{\mathbf{p},-1/2}, \hat{\psi}_{\mathbf{p},-3/2}\}$ of the aforementioned $j = 3/2$ representation. In the following, we set $\hbar = k_B = 1$ with energies in units of the Fermi energy E_F and wavevectors in units of the Fermi wavevector k_F. The amplitude of the Coulomb potential is then given by the Wigner–Seitz radius, $r_s = me^2/(\alpha\epsilon^* k_F)$ with $\alpha \approx 0.51$. In [8] we computed the bare charge polarisability $\Pi_0(\omega, \mathbf{q})$ and the self-energy corrections $\Sigma_{\pm}(\omega, \mathbf{k})$ on the upper $(+)$ and lower $(-)$ bands. We find that, because of strong spin–orbit coupling, the plasma frequency is diminished compared to a regular quadratic band, and that screening receives important contributions from interband excitations (see Fig. 1b,c).

The difference in screening between a Luttinger semimetal and a normal electron gas affects the quasiparticle properties. We find that for Luttinger semimetals the quasiparticle residue Z_F and the first Landau coefficients, f_{0s} and f_{1s}, are less affected by the Coulomb potential [8].

3 Superconductivity in Luttinger Semimetals

We evaluate the critical temperature of a singlet $s-$wave superconductor using the linear Eliashberg equation [3, 9], with account of self-energy corrections,

$$\lambda(T)\phi_{\sigma_1}(i\omega_{n_1},k_1) = -T\sum_{\sigma_2\omega_{n_2}}\int_0^\infty dk_2 \frac{k_2}{k_1}\frac{I_{0\sigma_1\sigma_2}(i\omega_{n_1},k_1;i\omega_{n_2},k_2)\phi_{\sigma_2}(i\omega_{n_2},k_2)}{(\omega_{n_2}Z_{\sigma_2}(i\omega_{n_2},k_2))^2+(\xi_{\sigma_2}(k_2)+\chi_{\sigma_2}(i\omega_{n_2},k_2))^2}, \tag{5}$$

where ϕ_σ represents the superconducting order parameter, $\omega_n = (2n+1)\pi T$ are the Matsubara frequencies, $\sigma = \pm$ is the band index, I_0 is the angular average of the screened Coulomb potential with spin–orbit corrections, and $\Sigma_\pm(i\omega_n,k) \equiv \chi_\pm(i\omega_n,k) + i\omega_n(1-Z_\pm(i\omega_n,k))$ are the self-energy corrections. Note that we have included the pairing order parameter on the upper band (+), as it will play an important role. Equation (6) is an eigenvalue equation where the critical temperature is found for eigenvalues $\lambda(T)$ such that $\lambda(T_c)=1$.

In this approach, the absence of symmetry in Eq. (6) on parameters (σ,ω_n,k) makes its resolution complex and time consuming . We thus perform the transformation $\phi_\sigma(i\omega_n,k) \to \bar{\phi}_\sigma(i\omega_n,k) = k\phi_\sigma(i\omega_n,k)/((\omega_n Z_\sigma(i\omega_n,k))^2+(\xi_\sigma(k)+\chi_\sigma(i\omega_n,k))^2)$ to have a symmetric form of Eq. (6)

$$\rho(T)\bar{\phi} = S\bar{\phi}, \tag{6}$$

with S a symmetric operator on parameters (σ,ω_n,k) and where the critical temperature T_c is obtained for $\rho(T_c)=0$. One can show that $\rho(T>T_c)<0$, so T_c is computed from the largest eigenvalue $\rho^{\max}$ and, using the variational properties of symmetric matrices, for any test function $\bar{\phi}^{\mathrm{t}}$:

$$\rho^{\max} \geq \rho^{\mathrm{t}} = \frac{\bar{\phi}^{\mathrm{t}}\cdot S\bar{\phi}^{\mathrm{t}}}{\bar{\phi}^{\mathrm{t}}\cdot\bar{\phi}^{\mathrm{t}}} \Rightarrow T_c \geq T_c^{\mathrm{t}}, \tag{7}$$

with T_c^{t} the critical temperature obtained with the test function.

We use this equation to reproduce the critical temperature for singlet $s-$wave pairing from the screened Coulomb repulsion in a single quadratic band structure [9], and compute it for a Luttinger semimetal (see Fig. 2). For large Wigner–Seitz radii, the critical temperature of the Luttinger semimetal $T_c/T_F \approx 4.4\times10^{-4}$ is smaller than for a single quadratic band, but extends to smaller values of r_s [5]. We perform this calculation down to $r_s = 0.01$, below which we are limited by the numerical resolution of the jump in the eigenvalue $\rho(T)$ that scales likes r_s. We note that it was important to keep ϕ_+ in Eq. (6), otherwise we would not find a solution. The value we obtain is comparable to the ratio $T_c/T_F \approx (2-5)\times10^{-4}$ from measurements on the half-Heusler YPtBi [10–12]. Because we have an $s-$wave superconductor, our result stands in contradiction with a recent proposition that

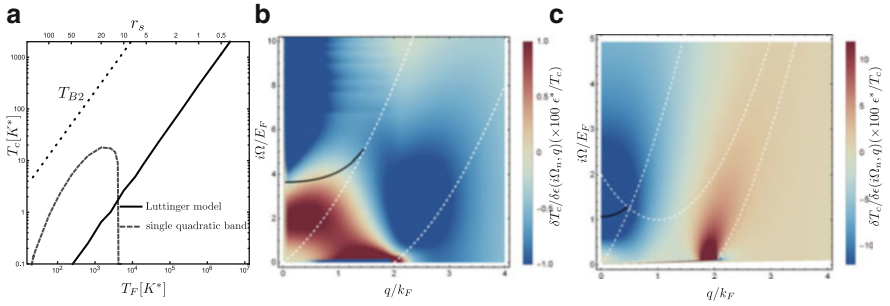


Fig. 2 (**a**) Critical temperature in units of $K^* = m/(m_e\epsilon^{*2})K$ for a single quadratic band (gray, dashed) and a Luttinger semimetal (plain, black). For comparison we superimpose the Bose–Einstein condensation temperature T_{B2} for a density $n/2$ and a mass $2m$. Reproduced with permission from [5]. (**b**)–(**c**) The functional derivative $\delta T_c/\delta\epsilon(i\Omega_n, q)$ in percent of the critical temperature for (**b**) a single quadratic band structure and (**c**) a Luttinger semimetal, for $r_s = 15$. The white dashed lines are the branches of the particle–hole excitation diagram in real frequency and the black line is the plasmon dispersion in real frequency. The critical temperature is mostly sensitive to the dielectric function in the region of plasmons and near the $q = 2k_F$ static screening

YPtBi is a line-node superconductor [13] but this interpretation, based on magnetic properties, is arguable due to the small value of the lower critical field B_{c1} in YPtBi [14], among other caveats.

Because the critical temperature depends on an integral equation involving every component $(i\Omega_n, q)$ of the dielectric function $\epsilon(i\Omega_n, q)$, it is not straightforward to understand the origin of superconductivity. If one changes $\epsilon(i\Omega_n, q)$ by $\delta\epsilon(i\Omega_n, q)$ then the critical temperature T_c changes by

$$\Delta T_c = 2\pi T \sum_{\Omega_n} \int dq \ \frac{\delta T_c}{\delta\epsilon(i\Omega_n, q)} \delta\epsilon(i\Omega_n, q). \tag{8}$$

The functional derivative $\delta T_c/\delta\epsilon(i\Omega_n, q)$ is a measure of the sensitivity of the critical temperature to screening and it can be decomposed into

$$\frac{\delta T_c}{\delta\epsilon(i\Omega_n, q)} = -\left.\frac{\delta\rho}{\delta\epsilon(i\Omega_n, q)}\right|_{T=T_c} \Bigg/ \left.\frac{\partial\rho}{\partial T}\right|_{T=T_c}. \tag{9}$$

In this equation, ρ is the maximal eigenvalue of the linear Eliashberg equation (6). We use it to evaluate numerically the derivative $\partial\rho/\partial T|_{T=T_c}$ and we use the Hellmann–Feynman theorem to compute $\delta\rho/\delta\epsilon(i\Omega_n, q)$ [5]. In Fig. 2b,c, we show the sensitivity of the critical temperature T_c to the different components of the dielectric function $\epsilon(i\Omega_n, q)$ for a quadratic band and a Luttinger semimetal [5]. We notice larger values in the area associated to plasmons and close to $2k_F$, which are, respectively, associated to plasmon and Kohn–Luttinger mechanisms of superconductivity [2, 4]. We stress that, albeit a contribution analogue to the

Kohn–Luttinger mechanism, the dynamical screening is central in the present mechanism of superconductivity and we do not expect $s-$wave pairing for a static interaction [4].

4 Pairing Beyond $s-$Wave from Spin–Orbit Coupling

In Luttinger semimetals, the quasiparticles are described with $j = 3/2$ multiplets instead of spin-1/2 as in ordinary metals. The rotational symmetry of Eqs. (3–4) allows to describe Cooper pairs by a gap function $\Delta^{J,LS}(i\omega_n, k)$ with a well-defined total angular momentum J that combines the pseudo-spin S of the Cooper pair and its orbital angular momentum L [15, 16]. At the critical temperature, these gap functions satisfy the linear Eliashberg equations

$$\begin{aligned}\lambda(T)\Delta^{J,L_1S_1}_{\sigma_1}(i\omega_{n_1}, k_1) \\ = -T \sum_{\substack{\ell\sigma_2\omega_{n_2} \\ Lambda_2S_2}} \int \frac{dk_2 k_2}{k_1} \frac{V_\ell(i(\omega_{n_1}-\omega_{n_2}), k_1, k_2) A^{J,L_1S_1L_2S_2}_{\ell,\sigma_1\sigma_2}(k_1,k_2)}{(\omega_{n_2} Z_{\sigma_2}(i\omega_{n_2}, k_2))^2 + (\xi_{\sigma_2}(k_2) + \chi_{\sigma_2}(i\omega_{n_2}, k_2))^2} \\ \times \Delta^{J,L_2S_2}_{\sigma_2}(i\omega_{n_2}, k_2),\end{aligned} \tag{10}$$

with $\lambda(T) = 1$ for $T = T_c$. Note that we have written the Eliashberg equation in its non-symmetrized form, in contrast to Eq. (6). This self-consistent relation can be complemented with off-diagonal components of the gap function [17], that we neglect in the present discussion. In Eq. (10), the electron pairing is determined by V_ℓ, the projection of the screened Coulomb potential $V_0(q)/\epsilon(i\Omega_n, \mathbf{q})$ on the Legendre polynomial P_ℓ, and by the form factor due to spin–orbit coupling:

$$\begin{aligned}A^{J,L_1S_1L_2S_2}_{\ell,\sigma_1\sigma_2} = \frac{2\ell+1}{2} \\ \times \int \frac{d^2\boldsymbol{\Omega}_1 d^2\boldsymbol{\Omega}_2}{(2\pi)^3} P_\ell\left(\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2\right) \mathrm{Tr}\left[\hat{P}_{\sigma_1}(\mathbf{k}_1)\hat{N}^{J,L_1S_1}(\mathbf{k}_1)\hat{P}_{\sigma_2}(\mathbf{k}_2)\hat{N}^{J,L_2S_2\dagger}(\mathbf{k}_2)\right],\end{aligned} \tag{11}$$

where $\boldsymbol{\Omega}_i$ is the solid angle of $\mathbf{k}_i$. Here, the matrices $\hat{N}^{J,LS}(\mathbf{k})$ correspond to the representation of the rotation symmetry on $\mathbf{J} = \mathbf{L} + \mathbf{S}$,

$$\hat{N}^{J,LS}(\mathbf{k}) = \sum_{m_L m_S} C^J_{Lm_L,Sm_S} Y_{Lm_L}(\theta_\mathbf{k}, \phi_\mathbf{k}) \hat{M}_{Sm_S}, \tag{12}$$

where $C^J_{Lm_L,Sm_S}$ are the Clebsch–Gordan coefficients, Y_{Lm_L} the spherical harmonics, and $\hat{M}_{Sm_S}$ the pairing matrices with pseudo-spin S of the Cooper pairs. Some of

these combinations, for $L = 0, 1$, are listed in [15, 16]. We introduce the projectors $\hat{P}_{\pm}$ in Eq. (12) to decompose the gap equation on the eigenstates of $\hat{H}_0$ on the upper $(+)$ and lower $(-)$ bands, that we, respectively, associate to eigenstates $\pm 3/2$ and $\pm 1/2$ of the helicity operator $\hat{\lambda} = \hat{\mathbf{k}} \cdot \hat{\mathbf{J}}$. Note that, for a given value of J, there is a finite number of components V_ℓ that contribute in the summation in Eq. (10). For example, for $s-$wave $J = L = S = 0$ only $\ell = 0$ and 2 contribute. In the following, we further simplify Eq. (10) by considering a gap function in a unique (J, L, S) sector, $\Delta^{J,LS}$, and write $A^{J,LSLS}_{\ell,\sigma_1\sigma_2} = A^{J,LS}_{\ell,\sigma_1\sigma_2}$. A more refined analysis would allow for mixing between different values of (L, S) for a fixed J.

It is expected that the amplitude of the pairing potential depends on the largest combination of the coefficients V_ℓ and $A^{J,LS}$ close the Fermi surface, where $\sigma_1 = \sigma_2 = -$ and $k_1 = k_2 = k_F$. To be more accurate, one should consider the full $k-$dependence but let us work in this simpler limit. It was shown that this amplitude is the strongest for $J = L = S = 0$ [15], which is precisely the order parameter we consider in our work (see Sect. 3). This logic of maximizing the product $V_\ell A^{J,LS}_\ell$ applies well for superconductiviy from an attractive potential, like the electron-phonon coupling, where the eigenvalue of Eq. (10) with the largest absolute value, $\lambda_1(T)$, is already positive. However, it is not straightforward to extend to superconductivity from a repulsive potential, such as the Coulomb repulsion between electrons, where in the $s-$wave channel the eigenvalue with the largest absolute value is negative, $\lambda_1(T) < 0$, because of the overall repulsive nature of the Coulomb potential. Then, the $s-$wave solution to Eq. (10) comes from the second largest-in-absolute-value eigenvalue, $\lambda_2(T) > 0$, which corresponds to the first electronic configuration where the Coulomb potential is attractive [9].

In a multi-orbital system, the coefficient $A^{J,LS}_\ell$ is a matrix that can have positive or negative eigenvalues. Then, the superconducting pairing can be favored not only by increasing the product $V_\ell A^{J,LS}_\ell$ but also by changing its sign. In the case of the repulsive Coulomb potential in Eq. (10), the eigenvalue $\lambda_1(T)$ with the largest magnitude is negative for $J = L = S = 0$ and it will not be responsible for the critical temperature defined by $\lambda(T_c) = 1 > 0$. Yet, the sign of λ_1 can be positive for other order parameters $\Delta^{J,LS}$ if A^{JLS}_ℓ has a negative eigenvalue, which we obtain for $J = L = S = 1 = \ell$, i.e. $\hat{N}^{111} = \sqrt{3}(-k_z(\hat{J}_x + i\hat{J}_y) + (k_x + ik_y)J_z)/(\sqrt{5}k)$ [16], and which decomposes as a matrix on the bands with helicity $\pm 1/2$ and $\pm 3/2$

$$A^{111}_1 = \begin{pmatrix} 2/5 & 3/10 \\ 3/10 & 0 \end{pmatrix}. \tag{13}$$

This matrix has eigenvalues $(2 \pm \sqrt{13})/10$, with one negative (≈ -0.16). For this channel, the sign of λ_1 can be positive and eventually be a solution to $\lambda_1(T_c) = 1$. This may have a larger critical temperature since we now consider the largest eigenvalue in magnitude, instead of the second largest. However, it is difficult to speculate on the resulting critical temperature and a refined study is needed to evaluate the corresponding critical temperature.

5 Conclusion

Over a wide range of doping, we find that the $s-$wave critical temperature for a Luttinger semimetal with screened Coulomb repulsion is $T_c/T_F \approx 4.4 \times 10^{-4}$. T_c/T_F is small but may be an explanation for the superconductivity of YPtBi, a candidate Luttinger semimetal, where experiments report $T_c/T_F \approx (1-8) \times 10^{-4}$. Previous theoretical works on YPtBi, with phonon-based pairing, estimate a critical temperature at least one order of magnitude smaller than in experiments [15, 18]. We quantitatively show the origins of superconductivity, in relation to the plasmon [9] and Kohn–Luttinger [4] mechanisms of superconductivity. We also analyze the Eliashberg equation of $j = 3/2$ fermions [15, 16] and propose that an unconventional order parameter, with $J = L = S = 1$, may turn the repulsive contribution of the screened Coulomb potential to attractive. This reminds a recent discussion on graphene, where the Berry curvature promotes the $\ell = 1$ component of a repulsive interaction to attractive [19]. A more involved study would be required to determine the dominant superconducting channel.

Acknowledgments This project is funded by a grant from Fondation Courtois, a Discovery Grant from NSERC, a Canada Research Chair, and a "Établissement de nouveaux chercheurs et de nouvelles chercheuses universitaires" grant from FRQNT. This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and Compute Canada (www.computecanada.ca).

Bibliography

1. W.L. McMillan, Transition temperature of strong-coupled superconductors. Phys. Rev. **167**(2), 331–334 (1968). doi:10.1103/PhysRev.167.331
2. Y. Takada, Theory of superconductivity in polar semiconductors and its application to N-type semiconducting SrTiO3. J. Phys. Soc. Jpn. **49**(4), 1267–1275 (1980). doi:10.1143/JPSJ.49.1267
3. J. Ruhman, P.A. Lee, Superconductivity at very low density: the case of strontium titanate. Phys. Rev. B **94**(22), 224515 (2016). doi:10.1103/PhysRevB.94.224515
4. W. Kohn, J.M. Luttinger, New mechanism for superconductivity. Phys. Rev. Lett. **15**(12), 524–526 (1965). doi:10.1103/PhysRevLett.15.524
5. S. Tchoumakov, L.J. Godbout, W. Wictzak-Krempa, Superconductivity from Coulomb repulsion in three-dimensional quadratic band touching Luttinger semimetals. Arxiv:1910.04189. https://arxiv.org/abs/1910.04189
6. P.B. Allen, R.C. Dynes, Transition temperature of strong-coupled superconductors reanalyzed. Phys. Rev. B **12**(3), 905–922 (1975). doi:10.1103/PhysRevB.12.905
7. J.M. Luttinger, Quantum theory of cyclotron resonance in semiconductors: general theory. Phys. Rev. **102**(4), 1030–1041 (1956). doi:10.1103/PhysRev.102.1030
8. S. Tchoumakov, W. Witczak-Krempa, Dielectric and electronic properties of three-dimensional Luttinger semimetals with a quadratic band touching. Phys. Rev. B **100**(7), 075104 (2019). doi:10.1103/PhysRevB.100.075104
9. Y. Takada, Plasmon mechanism of superconductivity in the multivalley electron gas. J. Phys. Soc. Jpn. **61**, 238–253 (1992). doi:10.1143/JPSJ.61.238

10. N.P. Butch, P. Syers, K. Kirshenbaum, A.P. Hope, J. Paglione, Superconductivity in the topological semimetal YPtBi. Phys. Rev. B **84**, 220504(R) (2011). doi:10.1103/PhysRevB.84.220504
11. T.V. Bay, T. Naka, Y.K. Huang, A. de Visser, Superconductivity in noncentrosymmetric YPtBi under pressure. Phys. Rev. B **86**, 064515 (2012). doi:10.1103/PhysRevB.86.064515
12. C. Shekhar, M. Nicklas, A.K. Nayak, S. Ouardi, W. Schnelle, G.H. Fecher, C. Felser, K. Kobayashi, Electronic structure and nonsaturating magnetoresistance of superconducting Heusler topological insulators. J. Appl. Phys. **113**, 17E142 (2013). doi:10.1063/1.4799144
13. H. Kim, K. Wang, Y. Nakajima, R. Hu, S. Ziemak, P. Syers, L. Wang, H. Hodovanets, J.D. Denlinger, P.M.R. Brydon, D.F. Agterberg, M. A. Tanatar, R. Prozorov, J. Paglione, Beyond triplet: unconventional superconductivity in a spin-3/2 topological semimetal. Sci. Adv. **4**, 4 (2018). doi:10.1126/sciadv.aao4513
14. T.V. Bay, M. Jackson, C. Paulsen, C. Baines, A. Amato, T. Orvis, M.C. Aronson, Y.K. Huang, A. de Visser, Low field magnetic response of the non-centrosymmetric superconductor YPtBi. Solid State Commun. **183**, 13–17 (2014). doi:10.1016/j.ssc.2013.12.010
15. L. Savary, J. Ruhman, J.W.F. Venderbos, L. Fu, P.A. Lee, Superconductivity in three-dimensional spin-orbit coupled semimetals. Phys. Rev. B **96**, 214514 (2017). doi:10.1103/PhysRevB.96.214514
16. B. Roy, S.A.A. Ghorashi, M.S. Foster, A.H. Nevidomskyy, Topological superconductivity of spin-3/2 carriers in a three-dimensional doped Luttinger semimetal. Phys. Rev. B **99**, 054505 (2019). doi:10.1103/PhysRevB.99.054505
17. M. Smidman, M.B. Salamon, H.Q. Yuan, D.F. Agterberg, Superconductivity and spin–orbit coupling in non-centrosymmetric materials: a review. Rep. Prog. Phys. **80**, 036501 (2017). doi:10.1088/1361-6633/80/3/036501
18. M. Meinert, Unconventional superconductivity in YPtBi and related topological semimetals. Phys. Rev. Lett. **116**, 137001 (2016). doi:10.1103/PhysRevLett.116.137001
19. T. Li, J. Ingham, H.D. Scammell, Unconventional Superconductivity in Semiconductor Artificial Graphene. arXiv:1909.07401 (2019). https://arxiv.org/abs/1909.07401
20. E. Langmann, Theory of the upper critical magnetic field without local approximation. Phys. C Supercond. **159**(5), 561–569 (1989). doi:10.1016/0921-4534(89)91286-0

Soft Degrees of Freedom, Gibbons–Hawking Contribution and Entropy from Casimir Effect

Glenn Barnich and Martin Bonte

Abstract Recent work on non-proper gauge degrees of freedom in the context of the Casimir effect is reviewed. In his original paper, Casimir starts by pointing out that, when the electromagnetic field is confined between two perfectly conducting parallel plates, there is an additional physical polarization of the electromagnetic field at zero value for the discretized longitudinal momentum besides the standard two transverse polarizations at non-zero values. In this review, the dynamics of these additional modes is obtained from first principles. At finite temperature, their contribution to the entropy is proportional to the area of the plates and corresponds to the contribution of a massless scalar field in 2+1 dimensions. When the plates are charged, there is a further contribution to the partition function from the zero mode of this additional scalar that scales with the area but does not contribute to the entropy. It reproduces the result obtained when the Gibbons–Hawking method is applied to the vacuum capacitor. For completeness, a brief discussion of the classical thermodynamics of such a capacitor is included.

Keywords Edge modes · Casimir effect · Gibbons–Hawking entropy · Black hole micro-states

1 Introduction

That seemingly unphysical polarizations of the electromagnetic field have an important role to play in the presence of charged particles is known since the work by Dirac [1], and Fock and Podolski [2], where the Coulomb force between two non-relativistic electrons is constructed in terms of creation and destruction operators associated with the scalar potential A_0.

G. Barnich (✉) · M. Bonte
Physique Théorique et Mathématique, Université libre de Bruxelles and International Solvay Institutes, Bruxelles, Belgium
e-mail: gbarnich@ulb.ac.be

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_35

When all polarizations are quantized, it is important to understand how equivalence with reduced phase space quantization is achieved. Arguably the most transparent implementation is through the quartet mechanism [3] which implies the cancellation of the contributions from unphysical polarizations and ghost variables when computing matrix elements of gauge invariant operators between gauge invariant states in the context of Hamiltonian BRST operator quantization. Furthermore, the associated path integral is simply related to the manifestly Lorentz invariant Lagrangian BRST path integral by integrating out momenta (see, e.g. [4] for a comprehensive review). More generally, as is well known in the context of topological field theories, these cancellations no longer work perfectly when there is non-trivial topology.

Whereas the quartet mechanism is relatively straightforward for the free electromagnetic field where the quartets are associated with temporal oscillators for the scalar potential and to oscillators for the longitudinal part of the vector potential on the one hand, and to oscillators for the ghost fields on the other, this is no longer the case in the presence of charged sources, where gauge invariance becomes a non-trivial issue [5]. For the simplest source representing a charged point particle at rest, it turns out that the BRST invariant vacuum state is a coherent state constructed out of unphysical null oscillators that represents the quantum Coulomb solution [6]. Some technical details and clarifications on this elementary construction are provided in Appendix "Details on Quantum Coulomb Solution".

The ultimate aim of this research is a better understanding of the degrees of freedom responsible for black hole entropy. In this context, it is intriguing to note that in one of the earliest papers on linearized quantum gravity by Bronstein [7] (see [8] for perspective), the last part of the paper follows closely the derivation by Dirac, Fock, and Podolski on the Coulomb law in order to obtain Newton's law between two test masses from the creation and destruction operators associated with the metric fluctuations h_{00}. How to extend the considerations below to the case of linearized gravity will be discussed elsewhere.

In order to avoid facing the question of the detailed interaction of the quantized electromagnetic field with charged dynamical matter, it is instructive to first consider the case where these interactions can be idealized as boundary conditions imposed on the free electromagnetic field. This naturally leads one to consider the electromagnetic field in the presence of charged conducting plates. In the absence of charge, this is precisely the set-up of the Casimir effect [9] at non-zero temperature [10] (see also [11, 12] and e.g. [13–15] for reviews).

As we will discuss in details below, that there is an additional physical polarization at zero value for the discretized longitudinal momentum, besides the two transverse ones at non-zero values, is well known in this context. We will focus on how to determine the dynamics of these "edge" modes and isolate their contribution to the partition function and the entropy, which scales with the area of the plates.

We then provide a microscopic understanding of the charged vacuum capacitor, where there is an additional contribution to the partition function that comes from the zero mode of the additional polarization and that also scales with the area but does not contribute to the entropy [16].

Before turning to these issues, we will first discuss the thermodynamics of a capacitor by standard methods. In the context of general relativity, the Euclidean approach of Gibbons and Hawking [17] consists in deriving the thermodynamics of Kerr–Newman black holes or of de Sitter space by evaluating on-shell the Euclidean action improved through suitable boundary terms. What these boundary terms are in the electromagnetic sector has been discussed, for instance, in [18, 19]. That the construction and interpretation of such boundary terms is very transparent in the first Hamiltonian formulation is discussed, for instance, in the derivation of the thermodynamics of the BTZ black hole [20]. We then review how the thermodynamics of the capacitor can easily be reproduced from the Euclidean approach [16].

2 Capacitor Thermodynamics: Textbook Approach

Consider a capacitor made of two conductors with charges $+q$ and $-q$ and area A. Its capacity $C = \frac{q}{V}$ in Lorentz–Heaviside units is

$$C_S = \frac{4\pi R_1 R_2}{R_2 - R_1} \tag{1}$$

for two concentric spheres of radii $R_1 < R_2$ and

$$C_P = \frac{A}{d} \tag{2}$$

for two parallel planes of area A separated by a distance d (see, e.g. [21, chapter 2]).

The capacitor begins with 0 charge, energy, and entropy. Charges $\pm dq$ are added on both side until one reaches $\pm q$.

In order to charge the capacitor, one may use a circuit without any resistance so that no heat would be produced in the process. One then would quickly arrive at the standard results (see, e.g., chapter 14.2 of [22] in the absence of the system and its electric polarization). A better understanding of the absence of entropy can however be gained by considering a set-up with a resistor [23].

If a potential difference V is applied on the capacitor, there will be a current $I(t) = \frac{q}{RC} e^{-\frac{t}{RC}}$. The heat lost by the system through the resistor is

$$Q = \int_0^\infty RI^2(t)dt = \frac{1}{2}CV^2. \tag{3}$$

Instead of a single step, the capacitor can be charged in N steps, each increasing the voltage by $\frac{V}{N}$. At each step, the circuit relaxes until the current vanishes. The heat lost in all N steps is then

$$Q_N = \frac{1}{2} C\,(\Delta V)^2 \times N = \frac{1}{2}\frac{CV^2}{N}. \tag{4}$$

If the ambient temperature is constant, the increase of entropy is

$$\Delta S_N = \frac{Q_N}{T} = \frac{1}{2}\frac{CV^2}{TN}. \tag{5}$$

In the limit $N \to \infty$, the charging of the capacitor becomes a quasi-static process. Since in this case, there is no increase of entropy, $dS = 0 = S(q)$, the process is reversible.

By the first law, it now follows that the increase of internal energy dU is due to the work done by the voltage source alone,

$$dU = dW = V dq. \tag{6}$$

Since $V = \frac{q}{C}$, we get

$$U(q) = \frac{1}{2}\frac{q^2}{C}, \tag{7}$$

which is the well-known energy of a charged capacitor. In this case, it is also the free energy,

$$F(T, q) = [U(q, S) - T S(q)]_{S=S(T)} = \frac{1}{2}\frac{q^2}{C}. \tag{8}$$

3 Capacitor Thermodynamics: Euclidean Approach

In the absence of gravity and of sources between the conductors, the starting point is the first-order action

$$I = \int d^4x \left[\dot{A}_i \pi^i - \mathcal{H}_0 + A_0 \partial_i \pi^i\right], \quad \mathcal{H}_0 = \frac{1}{2}\left(\pi^i \pi_i + B^i B_i\right), \tag{9}$$

where magnetic and electric fields are given, respectively, by $B^i = \epsilon^{ijk}\partial_j A_k$, $E^i = -\pi^i$. The variation of this action is

$$\begin{aligned}\delta I = &\int d^4x \left[\delta A_i \left(-\dot{\pi}^i - \epsilon^{ijk}\partial_j B_k\right) + \delta A_0 \left(\partial_i \pi^i\right) + \delta \pi^i (\dot{A}_i - \pi_i - \partial_i A_0)\right] \\ &+ \left[\int d^3x\, \delta A_i \pi^i\right]_{t_1}^{t_2} + \int_{t_1}^{t_2} dt \int d\sigma_i \left(\epsilon^{ijk} B_j \delta A_k + A_0 \delta \pi^i\right).\end{aligned} \tag{10}$$

We focus in this section on time-independent solutions for which the equations of motions reduce to $\pi_i = -\partial_i A_0$ with $\Delta A_0 = 0$ and $\Delta A_i - \partial_i \partial_j A^j = 0$. We also assume that Δ is invertible on $\partial_i \pi^i$, $\partial_j A^j$ and that the gauge condition $\partial_j A^j = 0$ may be imposed. Defining the transverse part of a vector field through $\vec{V}^T = \vec{V} - \vec{V}^L$, with the longitudinal part given by $\vec{V}^L = \vec{\nabla}(\Delta^{-1}\vec{\nabla} \cdot \vec{V})$, the gauge condition is equivalent to $\vec{A} = \vec{A}^T$, while the equations of motion determine the longitudinal part $\vec{\pi}^L$ in terms of the harmonic potential A_0 and imply $\vec{\pi}^T = 0 = \Delta \vec{A}^T$. We assume here that this implies $\vec{A}^T = \vec{v}$, with $\vec{v}$ constant.

Consider a spherical capacitor with conducting spheres at radii $R_1 < R_2$ and charges $+q$ and $-q$, respectively. Under the above assumptions, the general solution to the equations of motion is

$$A_0 = -\frac{q}{4\pi r} + c, \quad \pi^i = -\frac{q x^i}{4\pi r^3}, \quad R_1 < r < R_2, \tag{11}$$

with c a constant and 0 outside of the shell. The classical observable that captures electric charge is

$$Q = -\int_S d\sigma_i \, \pi^i, \tag{12}$$

with S a closed surface inside the shell, for instance $r = R$, $R_1 < R < R_2$ so that $Q = q$ on-shell.

In the case of planar conductors at $z = 0$ and $z = d$ with charge densities $\frac{q}{A}$ and $-\frac{q}{A}$, we have instead

$$A_0 = -\frac{q}{A} z + c, \quad \pi^i = -\delta^i_3 \frac{q}{A}, \quad 0 < z < d, \tag{13}$$

and 0 outside of the capacitor. In this case, the electric charge observable is Q in (12) with S a plane at $z = L$, $0 < L < d$.

For later purposes, note that both solutions (11) and (13) can be transformed into solutions with $A_0 = 0$ by a time dependent gauge transformation. The associated vector potential satisfies $\vec{\nabla} \cdot \vec{A} = 0$ between the conductors and is longitudinal.

When working at fixed charge, all surface terms in the second line of (10) vanish on the solutions under consideration. In the Euclidean approach, there is a contribution to the partition function from the Euclidean action evaluated at these classical solutions. It is given by

$$-\beta F(\beta, Q) = -\frac{1}{\hbar} I^E(\beta, Q)|_{\text{on-shell}}, \tag{14}$$

where

$$I_E = \int_0^{\hbar\beta} d\tau \int d^3x \left[-i \dot{A}_i \pi^i + \mathcal{H}_0 - A_0 \partial_i \pi^i \right]. \tag{15}$$

On-shell, only the longitudinal electric field in the Hamiltonian contributes and gives

$$F(\beta, q) = \frac{1}{2}\frac{q^2}{C}, \tag{16}$$

where C is the capacity given by (1) and (2) in the spherical and the flat case, respectively, in agreement with (8).

When working at fixed electric potential $A_0 = -\phi$, with $A_0|_{S_1} = -\phi_1$, $A_0|_{S_2} = -\phi_2$ constants and $\mu = \phi_1 - \phi_2$, the general solution is instead

$$A_0 = -\frac{1}{R_2 - R_1}\left(R_2\phi_2 - R_1\phi_1 + \frac{\mu R_1 R_2}{r}\right), \quad \pi^i = -\frac{\mu R_1 R_2\, x^i}{(R_2 - R_1)r^3},$$
$$Q = C_S\mu, \tag{17}$$

in the spherical and

$$A_0 = -\phi_1 + \frac{\mu}{d}z, \quad \pi^i = -\delta^i_3\frac{\mu}{d}, \quad Q = C_P\mu, \tag{18}$$

in the planar case. At fixed potential, the last surface term in (10) does no longer vanish on-shell. Instead, the improved action

$$I' = I - \int dt \int d\sigma_i A_0 \pi^i, \tag{19}$$

has a true extremum on-shell. In the Euclidean action, we have instead

$$I'_E = I_E + \int_0^{\hbar\beta} d\tau \int d\sigma_i A_0 \pi^i = I_E + (\phi_2 - \phi_1)Q. \tag{20}$$

When evaluated on-shell, this now leads to

$$F(\beta, \mu) = -\frac{1}{2}C\mu^2, \tag{21}$$

which is related to (16) through a standard Legendre transformation.

4 Boundary Conditions

In the case of the capacitor, the boundary conditions for perfect conductors are $\vec{n} \times \vec{E} = 0 = \vec{n} \cdot \vec{B}$ on the boundary defined by the conductors, with $\vec{n}$ the normal to the boundary. In the planar case, to which we limit ourselves in the following, one thus considers free electromagnetism on $\mathbf{R}^2 \times [0, d]$, with boundary conditions $E^x = 0 = E^y$ at $z = 0$ and $z = d$. It thus follows that π^a, $a = 1, 2$, satisfy Dirichlet

conditions. We then take Dirichlet conditions for A_a as well since this guarantees well-defined Poisson brackets and a standard quantization in terms of the Fourier coefficients of sine functions. The requirement that $B^3 = \partial_1 A_2 - \partial_2 A_1$ should also satisfy Dirichlet conditions then holds automatically.

There remains the boundary conditions on (A_3, π^3) and, in the case of BRST quantization, on (A_0, π^0) as well as the ghost variables $(\eta, \mathcal{P})$, $(\bar{C}, \rho)$. A natural choice is Neumann conditions for (A_3, π^3), and Dirichlet for all others in the case of BRST quantization. This choice implies that the divergence $\vec{\nabla} \cdot \vec{\pi}$ satisfies Dirichlet conditions. The constraint $\vec{\nabla} \cdot \vec{\pi} = 0$ in the space between the conductors can then be implemented by variation in the action principle (9) through a field A_0 that satisfies Dirichlet conditions as well. Proper gauge transformations are defined by gauge parameters that satisfy Dirichlet conditions, which implies the same conditions for the ghost variables. In the context of BRST quantization, this choice guarantees that the quartet mechanism for (A_0, π^0), $(\vec{A}^L, \vec{\pi}^L)$ and ghost pairs will be effective.

If $k_3 = \frac{\pi}{d} n_3$, fields with Dirichlet conditions on $\mathbf{R}^2 \times [0, d]$ are expanded as

$$\phi(x^i) = \sum_{n_3 > 0} \phi^S_{k_3}(x^a) \sin k_3 z, \quad \phi^S_{k_3}(x^a) = \frac{1}{d} \int_{-d}^{d} dz\, \phi(x^i) \sin k_3 z, \tag{22}$$

while A_3, π^3 with Neumann conditions are expanded as

$$\phi(x^i) = \sum_{n_3 \geq 0} \phi^C_{k_3}(x^a) \cos k_3 z, \quad \begin{cases} c_{k_3}(x^a) = \frac{1}{d} \int_{-d}^{d} dz\, \phi(x^i) \cos k_3 z \\ \phi^C_0(x^a) = \frac{1}{2d} \int_{-d}^{d} dz\, \phi(x^i) \end{cases}. \tag{23}$$

5 Physical Degrees of Freedom

In the Hamiltonian approach, the reduced physical phase space or rather functions thereon can be characterized through BRST cohomology in ghost number 0. This can be done independently of a choice of gauge fixation, which enters in the specification of the Hamiltonian.

In the case of free electromagnetism in Euclidean space $\mathbf{R}^3$, the Helmholtz decomposition of vector fields alluded to above allows one to show that this cohomology consists of functions of transverse vector potentials and their momenta. Alternatively, in terms of Fourier transforms, it consists of functions of transverse oscillator variables.

The analysis of the BRST cohomology in momentum space in the case of the capacitor [16] then shows that, at $k_3 \neq 0$, there are the standard two transverse polarizations, while there is in addition the mode at $n_3 = 0$ contained in (A^3, π_3). This is the additional physical polarization of the Casimir effect.

In the original paper, this additional polarization was not discussed in the context of BRST quantization. Even though not explicitly stated in [9], it is clear from the

context and from [24], that the analysis is done in radiation gauge, $A_0 = 0 = \vec{\nabla}\cdot\vec{A}$, imposed together with the constraint equations $\pi^0 = 0$, $\vec{\nabla}\cdot\vec{\pi} = 0$. When translated to momentum space with the above boundary conditions, it follows directly that the $k_3 \neq 0$ modes of $(\vec{A}, \vec{\pi})$ give rise to the two transverse polarizations, while the $k_3 = 0$ mode of $(\vec{A}, \vec{\pi})$ is also divergence-free.

The divergence-free vector fields in position space associated with the $k_3 = 0$ mode are given by

$$A_i^{\mathrm{NPG}} = \delta_i^3 A_{3,0}^C(x^a), \quad \pi^i_{\mathrm{NPG}} = \delta_3^i \pi_0^{3C}(x^a). \tag{24}$$

The argument why they are non-trivial from a position space viewpoint in equation (4.10) of [16] is incorrect. Let us focus on $\vec{\pi}_{\mathrm{NPG}}$, which has a direct interpretation in electrostatics, the argument for $\vec{A}^{\mathrm{NPG}}$ being the same. The vector field $\vec{\pi}_{\mathrm{NPG}}$ has a non-trivial longitudinal piece. The associated 1 form is co-closed without being co-exact. This follows from the Helmholtz decomposition in the presence of boundaries. Indeed, under suitable fall-off assumptions at infinity, there is a unique decomposition

$$\vec{\pi} = \vec{\nabla}\varphi + \vec{\nabla}\times\vec{\alpha}, \tag{25}$$

$$\varphi(x) = -\int d^3x' \frac{(\vec{\nabla}\cdot\vec{\pi})(x')}{4\pi|\vec{x}-\vec{x}'|} + \oint_S \frac{(\vec{n}\cdot\vec{\pi}\, d\sigma)(x')}{4\pi|\vec{x}-\vec{x}'|}, \tag{26}$$

$$\vec{\alpha}(x) = \int d^3x' \frac{(\vec{\nabla}\times\vec{\pi})(x')}{4\pi|\vec{x}-\vec{x}'|} - \oint_S \frac{(\vec{n}\times\vec{\pi}\, d\sigma)(x')}{4\pi|\vec{x}-\vec{x}'|}. \tag{27}$$

When this decomposition is applied to $\vec{\pi}^{\mathrm{NPG}}$ for the capacitor, the potential for the longitudinal part comes entirely from the boundary contribution and is explicitly given by

$$\varphi_{\mathrm{NPG}}(x) = \frac{1}{4\pi}\int dx'dy'\pi_0^{3C}(x'^a)\left(\left[\left(\rho^2+(z-d)^2\right)\right]^{-\frac{1}{2}} - \left[\rho^2+z^2\right]^{-\frac{1}{2}}\right), \tag{28}$$

while the potential for the transverse part comes entirely from the bulk contribution and is explicitly given by

$$\alpha^i_{\mathrm{NPG}}(x) = \frac{\delta^i_a}{4\pi}\int dx'dy'\epsilon^{ab}\partial'_b\pi_0^{3C}(x'^c)\ln\frac{\sqrt{\rho^2+(d-z)^2}+d-z}{\sqrt{\rho^2+z^2}-z}, \tag{29}$$

where $\rho^2 = (x-x')^2+(y-y')^2$.

One then has to decide how to deal with the transverse space $\mathbf{R}^2$. As usual, we will put the system in a finite two-dimensional box in an intermediate stage. In this box, we can adopt either perfectly conducting conditions as in [9, 24], or use periodic conditions, which is what was done in [16]. In the large area limit, where sums go to integrals, both approaches yield the same results for finite temperature

partition functions (without zero modes). In the latter, we thus consider expansions as in equation (4.4) of [25], with $d = 3$ and $p = 1$, but we explicitly keep the zero mode because we need it for the microscopic understanding of the Gibbons–Hawking contribution. This is reminiscent of the expansion of the complex scalar field in [26].

6 Dynamics and Charge

For the transverse degrees of freedom at $k_3 \neq 0$, the usual discussion in terms of two transverse polarizations applies. In addition, the canonical Hamiltonian $H_0 = \int d^3x\, \mathcal{H}_0$ induces a Hamiltonian for the non-proper gauge degrees of freedom given by

$$H_{\text{NPG}} = d \int d^2x \left[\frac{1}{2}\pi^2 + \frac{1}{2}\partial_a\phi\partial^a\phi\right], \tag{30}$$

where $\phi = A^C_{30}, \pi = \pi^{3C}_0$. When taking into account the kinetic term in the associated first-order action and after eliminating the momentum by its own equation of motions, the associated Lagrangian action is that of a massless scalar in $2+1$ dimensions with prefactor d,

$$S_{\text{NPG}} = d \int dt d^2x \left[\frac{1}{2}\dot{\phi}^2 - \frac{1}{2}\partial_a\phi\partial^a\phi\right]. \tag{31}$$

The electric charge observable can be written as a function on the phase space that includes the non-proper gauge degrees of freedom as,

$$Q = -\int d^2x\, \pi. \tag{32}$$

From this expression, it follows that charge is related to the momentum of the zero mode of the scalar field, which is a particle. For canonical commutation relations, the appropriate normalization (see e.g. [16] Appendix A for details) is

$$q = \sqrt{\frac{d}{A}} \int d^2x\, \phi, \quad p = \sqrt{\frac{d}{A}} \int d^2x\, \pi, \tag{33}$$

and the associated Hamiltonian and charge observable are given by

$$H^0_{\text{NPG}} = \frac{1}{2}p^2, \quad Q = -\sqrt{\frac{A}{d}}\, p. \tag{34}$$

7 Extra Contributions to Partition Function

When naively decomposing the additional massless scalar field into its zero mode, the particle, and the remaining bulk modes in two dimensions, their contributions to the partition function is straightforward. In the charged case, the former is given by

$$Z^0_{\mathrm{NPG}}(\beta,\mu) = \mathrm{Tr} e^{-\beta(H^0_{\mathrm{NPG}}-\mu Q)}. \tag{35}$$

This can be related to the well-known partition function of a free particle of unit mass by completing the square. The result is

$$\ln Z^0_{\mathrm{NPG}}(\beta,\mu) = \ln \Delta q - \frac{1}{2}\ln\left(2\pi\hbar^2\beta\right) + \frac{A}{2d}\beta\mu^2. \tag{36}$$

Here Δq denotes the divergent interval of integration of q, while the last term reproduces the Gibbons–Hawking contribution $-\beta F(\beta,\mu)$ to the partition function as discussed in (21).

The partition function of a massless scalar in two dimensions can be obtained as usual after putting the field in a box with periodic boundary conditions and by neglecting the zero mode. The standard result in the limit of large volume in two dimensions, which is the area of the plates in the current context, is

$$\ln Z'_{\mathrm{NPG}} = \frac{A}{2\pi}\zeta(3)(\hbar\beta)^{-2}. \tag{37}$$

A different discussion along the lines of [27, 28] gives instead

$$F^0_{\mathrm{NPG}}(\beta,0) = \beta^{-1}[\mathrm{div} + \ln\beta + \mathrm{cte}], \tag{38}$$

which differs by a factor of 2 in the $\ln\beta$ term from (36). Note however that this difference will not matter for the considerations below as long as the μ dependent part will still be given by $-\frac{A}{2d}\mu^2$.

8 Charged Black Body Partition Function

In order to discuss the full, finite result, one may follow and adapt the discussion of the finite temperature Casimir effect [10, 11] (see e.g. [13–15] for reviews).

One considers segments on the z-axis given by

$$I = [0,d], \quad II = [d,L_z], \quad III = [0,L_z/\eta], \quad IV = [L_z/\eta, L_z]. \tag{39}$$

The analog $F_C(\beta,\mu)$ of the Casimir free energy is defined as

$$F_C(\beta,\mu) = F_I(\beta,\mu) + F_{II}(\beta,0) - F_{III}(\beta,0) - F_{IV}(\beta,0). \tag{40}$$

The zero mode will then only give the Gibbons–Hawking contribution

$$F_C^0(\beta,\mu) = -\frac{A}{2d}\mu^2. \tag{41}$$

Non-zero modes, both those at $k_3 \neq 0$ and those of the additional scalar, will not contribute to the μ dependent part. As usual, one separates the zero temperature contribution from the thermal one,

$$F_C'(\beta) = F_C'(\infty) + F_C'^T(\beta). \tag{42}$$

In the limit of large plate area A and large L_z, the former is the standard zero temperature Casimir energy that may be computed from the zero point energies. Between the plates, one finds

$$F_I'(\infty) = \frac{\hbar}{2}\frac{A}{(2\pi)^2}\int d^2k \left[\sqrt{k_a k^a} + 2\sum_{n=1}^{\infty}\sqrt{k_a k^a + \frac{\pi^2 n^2}{d}}\right], \tag{43}$$

while

$$F_{II}'(\infty) - F_{III}'(\infty) - F_{IV}'(\infty) = -d\frac{\hbar}{2}\frac{A}{(2\pi)^2}\int d^2k \left[\int_{-\infty}^{+\infty}\frac{dk_z}{2\pi}\sqrt{k_a k^a + k_z^2}\right]. \tag{44}$$

After a suitable cut-off regularization and with the help of the Euler–Maclaurin formula, one then finds

$$F_C'(\infty) = -\frac{A\pi^2\hbar}{720d^3}. \tag{45}$$

Note that the first term in the square brackets of (43) is due to the additional massless scalar and gives the first term at discrete value 0 with the correct 1/2 in the Euler–Maclaurin formula. When using ζ function regularization, this divergent term is usually omitted because it does not depend on the separation distance and thus does not contribute to the Casimir force.

In the same way, the temperature dependent contribution, which needs no regularization, is given by

$$F_C'^T(\beta) = \frac{2A}{\beta}\int \frac{d^2k}{(2\pi)^2}\Bigg[\sum_{n=0}^{\infty}{}' \ln\left(1 - e^{-\hbar\beta\sqrt{k_a k^a + (\frac{n\pi}{d})^2}}\right) - d\int_{-\infty}^{+\infty}\frac{dk_z}{2\pi}\ln\left(1 - e^{-\hbar\beta\sqrt{k_a k^a + k_z^2}}\right)\Bigg], \tag{46}$$

where the prime on the sum means that the term at $n = 0$ comes with a factor $1/2$. This term is due to the non-zero modes of the additional scalar. More explicitly, if

$$b(d,\beta,n) = \frac{1}{2\beta}\int_{n^2}^{\infty} ds \ln\left(1 - e^{-\frac{\pi\hbar\beta}{d}\sqrt{s}}\right) \tag{47}$$

one finds

$$F_C'^T(\beta) = -\frac{A}{2\pi\hbar^2\beta^3}\zeta(3) + \frac{A\pi}{d^2}\sum_{n=1}^{\infty} b(d,\beta,n) + \frac{V\pi^2}{45\hbar^3\beta^4}. \tag{48}$$

The first term from the additional scalar coincides with the contribution from (37) to the free energy. It does not contribute to the Casimir force but does contribute to the entropy. The last term corresponds to the subtraction of the black body result, that is to say the contribution of the two transverse polarization in empty space. The middle term corresponds to the contribution of the two transverse polarizations at discretized non-zero values of k_3. Low and high temperature expansions are discussed in the cited literature. The full result is then

$$F_C(\beta,\mu) = F_C^0(\beta,\mu) + F_C'(\infty) + F_C'^T(\beta). \tag{49}$$

Appendix: Details on Quantum Coulomb Solution

Consider the electromagnetic field interacting with a static point particle sitting at the origin,

$$S\left[A_\mu; j^\mu\right] = \int d^4x\left[-\frac{1}{4}F^{\mu\nu}F_{\mu\nu} - j^\mu A_\mu\right], \quad j^\mu = \delta_0^\mu Q\delta^3(\vec{x}). \tag{50}$$

The modified vacuum state that is annihilated by the BRST charge in the presence of the source is given by

$$|0\rangle^Q = e^{\int d^3k\, q(\vec{k})\hat{b}^\dagger(\vec{k})}|0\rangle, \quad q(\vec{k}) = \frac{Q}{(2\pi)^{3/2}\sqrt{2}\omega(\vec{k})^{3/2}}, \tag{51}$$

if

$$A_0(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3k}{\sqrt{2\omega(\vec{k})}} \left[a_0(\vec{k}, t) e^{i\vec{k}\cdot\vec{x}} + \text{c.c.} \right], \tag{52}$$

$$A_i(x) = \frac{1}{(2\pi)^{3/2}} \int \frac{d^3k}{\sqrt{2\omega(\vec{k})}} \left[a_m(\vec{k}, t) e_i^m(\vec{k}) e^{i\vec{k}\cdot\vec{x}} + \text{c.c.} \right], \tag{53}$$

where $\omega(\vec{k}) = |\vec{k}| = k$, the polarization vectors are $e_i^3 = k_i/\omega(\vec{k})$, and $k^i e_i^a = 0, a = 1, 2$, while the unphysical null oscillators are defined by

$$a(\vec{k}) = a_3(\vec{k}) + a_0(\vec{k}), \quad b(\vec{k}) = \frac{1}{2}\left[a_3(\vec{k}) - a_0(\vec{k}) \right], \tag{54}$$

(see [4] for detailed conventions including the adapted mode expansions for the momenta, up to the correction pointed out in [6]). This state is constructed so as to be annihilated by the BRST charge in the presence of the source,

$$\hat{\Omega}^Q |0\rangle^Q = 0, \tag{55}$$

where

$$\Omega^Q = -\int d^3x \left[i\rho\pi^0 + \eta\left(\partial_i \pi^i - j^0 \right) \right], \tag{56}$$

$$\hat{\Omega}^Q = \int d^3k \left[\hat{c}^\dagger(\vec{k}) \hat{a}^Q(\vec{k}) + \hat{a}^{Q\dagger}(\vec{k}) \hat{c}(\vec{k}) \right], \quad \hat{a}^Q(\vec{k}) = \hat{a}(\vec{k}) - q(\vec{k}). \tag{57}$$

Note that it is not the only state with this property, for instance

$$|0\rangle'^Q = e^{-\int d^3k\, q(\vec{k}) \hat{a}_0^\dagger(\vec{k})} |0\rangle = e^{-\int d^3k\, \frac{q(\vec{k})}{2} \hat{a}^\dagger(\vec{k})} |0\rangle^Q, \tag{58}$$

is also annihilated by $\hat{\Omega}^Q$, and as in [1, 2], it is constructed out of temporal oscillators alone.[1]

The gauge fixed Hamiltonian

$$H_\xi = H_0 + \left\{ \Omega^Q, K_\xi \right\}, \quad H_0 = \int d^3x \frac{1}{2} \left[\pi^i \pi_i + B^i B_i \right], \quad B^i = \epsilon^{ijk} \partial_j A_k, \tag{59}$$

is constructed by using the gauge fixing fermion

$$K_\xi = -\int d^3x \left[i\bar{C} \partial_k A^k + \mathcal{P} A_0 - \xi \frac{i}{2} \bar{C} \pi^0 \right]. \tag{60}$$

[1] G.B. is grateful to M. Schmidt and S. Theisen for pointing this out and for prompting the considerations below.

Since

$$\{\Omega^Q, K_\xi\} = \int d^3x \left[\partial_k A^k \pi^0 + A_0\left(-\partial_i \pi^i + j^0\right) + i\mathcal{P}\rho + i\partial^i \bar{C}\partial_i \eta - \frac{1}{2}\xi\pi^0\pi^0\right], \tag{61}$$

the gauge fixed Hamiltonian contains in particular the correct source term. When using the decomposition $\pi^i = \pi^i_T + \frac{1}{\Delta}\partial^i\partial_j\pi^j$, it follows that

$$\frac{1}{2}\int d^3x\, \pi^i\pi_i = \frac{1}{2}\int d^3x\, \pi^i_T\pi^T_i - \frac{1}{2}\int d^3x\, \partial_j\pi^j\frac{1}{\Delta}\partial_k\pi^k. \tag{62}$$

The last term can be written as

$$-\frac{1}{2}\int d^3x\, \partial_j\pi^j\frac{1}{\Delta}\partial_k\pi^k = \left\{\Omega^Q, -\frac{1}{2}\int d^3x\, \mathcal{P}\frac{1}{\Delta}\left(\partial_i\pi^i + j^0\right)\right\} - \frac{1}{2}\int d^3x\, j^0\frac{1}{\Delta}j^0, \tag{63}$$

so that

$$H_\xi = H^{\text{ph}} + \left\{\Omega^Q, \tilde{K}^Q_\xi\right\}, \tag{64}$$

$$H^{\text{ph}} = \frac{1}{2}\int d^3x \left[\pi^i_T\pi^T_i - A^T_j\,\Delta A^j_T - j^0\frac{1}{\Delta}j^0\right], \tag{65}$$

$$\tilde{K}^Q_\xi = K_\xi - \frac{1}{2}\int d^3x\, \mathcal{P}\frac{1}{\Delta}\left(\partial_i\pi^i + j^0\right). \tag{66}$$

In Feynman gauge $\xi = 1$, when expressed in terms of modes, we have

$$\hat{H}_{\xi=1} = \hat{H}^{\text{phys}} + \left[\hat{\Omega}^Q, \hat{\tilde{K}}^Q_{\xi=1}\right], \tag{67}$$

$$\hat{H}^{\text{phys}} = \int d^3k\, \omega(\vec{k})\left[\hat{a}^\dagger_a(\vec{k})\hat{a}^a(\vec{k}) + q(\vec{k})^2\right], \tag{68}$$

$$\hat{\tilde{K}}^Q_{\xi=1} = \int d^3k\, \omega(\vec{k})\left(\hat{\bar{c}}^\dagger(\vec{k})\left[\hat{b}(\vec{k}) + \frac{\omega(\vec{k})}{2}\right] + \left[\hat{b}^\dagger(\vec{k}) + \frac{\omega(\vec{k})}{2}\right]\hat{\bar{c}}(\vec{k})\right), \tag{69}$$

and

$$\left[\hat{\Omega}^Q, \hat{\tilde{K}}^Q_{\xi=1}\right] = \int d^3k\, \omega(\vec{k})\left[\hat{a}^\dagger(\vec{k})b(\vec{k}) + \hat{b}^\dagger(\vec{k})\hat{a}(\vec{k}) + \hat{\bar{c}}^\dagger(\vec{k})\hat{c}(\vec{k}) + \hat{c}^\dagger(\vec{k})\hat{\bar{c}}(\vec{k})\right] + \int d^3k\, \omega(\vec{k})q(\vec{k})[\hat{a}_0(\vec{k}) + \hat{a}^\dagger_0(\vec{k})], \tag{70}$$

where the first line is proportional to the number operator for unphysical oscillators, while the second line contains the correct source term. Since $\hat{a}^\dagger\hat{b} + \hat{b}^\dagger\hat{a} + q\hat{a}_0 + q\hat{a}_0^\dagger = \hat{a}_3^\dagger\hat{a}_3 - (\hat{a}_0^\dagger - q)(\hat{a}_0 - q) + q^2$, it follows that $|0\rangle'^{\mathcal{Q}}$ is an eigenstate of this gauge fixed Hamiltonian

$$\hat{H}_{\xi=1}|0\rangle'^{\mathcal{Q}} = \int d^3k\,\omega(\vec{k})q(\vec{k})^2|0\rangle'^{\mathcal{Q}}. \tag{71}$$

In the context of BRST quantization, one may modify the gauge fixing fermion and remove the source dependent term therein, that is to say, one may replace $\tilde{K}_\xi^{\mathcal{Q}}$ by

$$\tilde{K}_\xi = K_\xi - \frac{1}{2}\int d^3x\,\mathcal{P}\frac{1}{\Delta}\partial_i\pi^i, \tag{72}$$

$$\hat{\tilde{K}}_{\xi=1} = \int d^3k\,\omega(\vec{k})\left[\hat{\bar{c}}^\dagger(\vec{k})\hat{b}(\vec{k}) + \hat{b}^\dagger(\vec{k})\hat{\bar{c}}(\vec{k})\right], \tag{73}$$

$$\left[\hat{\Omega}^{\mathcal{Q}}, \hat{\tilde{K}}_{\xi=1}\right] = \int d^3k\,\omega(\vec{k})\Big[\hat{a}^{\mathcal{Q}\dagger}(\vec{k})b(\vec{k}) + \hat{b}^\dagger(\vec{k})\hat{a}^{\mathcal{Q}}(\vec{k}) + \hat{\bar{c}}^\dagger(\vec{k})\hat{c}(\vec{k}) + \hat{c}^\dagger(\vec{k})\hat{\bar{c}}(\vec{k})\Big], \tag{74}$$

since this modifies the ghost number 0 part of the Hamiltonian by terms that are proportional to the constraints. It now follows that $|0\rangle^{\mathcal{Q}}$ is an eigenstate of the new gauge fixed Hamiltonian $\hat{H}'_{\xi=1}$,

$$\hat{H}'_{\xi=1} = \hat{H}^{\mathrm{ph}} + \left[\hat{\Omega}^{\mathcal{Q}}, \hat{\tilde{K}}_{\xi=1}\right], \quad \hat{H}'_{\xi=1}|0\rangle^{\mathcal{Q}} = \int d^3k\,\omega(\vec{k})q(\vec{k})^2|0\rangle^{\mathcal{Q}}, \tag{75}$$

with the same eigenvalue than $|0\rangle'^{\mathcal{Q}}$ is of $\hat{H}_{\xi=1}$.

Note also that the difference between $e^{\int d^3k\,\frac{q^2(\vec{k})}{2}}|0\rangle'^{\mathcal{Q}}$ and $|0\rangle^{\mathcal{Q}}$ is BRST exact. Indeed,

$$e^{\int d^3k\,\frac{q^2(\vec{k})}{2}}|0\rangle'^{\mathcal{Q}} - |0\rangle^{\mathcal{Q}} = \left(e^{-\int d^3k\,\frac{q(\vec{k})}{2}\hat{a}^{\mathcal{Q}\dagger}(\vec{k})} - \hat{\mathbf{1}}\right)|0\rangle^{\mathcal{Q}}. \tag{76}$$

The result follows from the fact that $-\int d^3k\,\frac{q(\vec{k})}{2}\hat{a}^{\mathcal{Q}\dagger}(\vec{k})$ is BRST exact,

$$-\int d^3k\frac{q(\vec{k})}{2}\hat{a}^{\dagger\mathcal{Q}}(\vec{k}) = [\hat{K}, \hat{\Omega}^{\mathcal{Q}}], \quad \hat{K} = -\int d^3k\,\frac{q(\vec{k})}{2}\hat{\bar{c}}^\dagger(\vec{k}), \tag{77}$$

and that the difference of the exponential of a BRST exact operator minus the unit operator is a BRST exact operator,

$$e^{[\hat{K},\hat{\Omega}^Q]} - \hat{\mathbf{1}} = \left[\hat{L}, \hat{\Omega}^Q\right], \tag{78}$$

for some operator $\hat{L}$ (see, e.g. [4], exercise 14.3 for the proof), so that

$$e^{\int d^3k \frac{q^2(\vec{k})}{2}} |0\rangle'^Q = |0\rangle^Q - \hat{\Omega}^Q \hat{L} |0\rangle^Q, \tag{79}$$

since $|0\rangle^Q$ is BRST closed.

Some additional comments on [6] are in order.

(i) In the computation (2.8), an obvious infrared regularization is understood since the Fourier transform of k^{-2} is $\frac{1}{4\pi r}$ only when using such a regulator,

$$\frac{1}{(2\pi)^3} \int d^3k \, \frac{1}{k^2 + \mu^2} e^{i\vec{k}\cdot\vec{x}} = \frac{1}{4\pi r} e^{-\mu r},$$

with the desired result obtained when $\mu \to 0^+$.

(ii) Equation (2.7) is not correct. Starting from

$$\partial_i A^i = \frac{i}{(2\pi)^{3/2}} \int d^3\vec{k} \sqrt{\frac{\omega(\vec{k})}{2}} \left[\frac{a(\vec{k}) + 2b(\vec{k})}{2} e^{i\vec{k}\cdot\vec{x}} - \text{c.c.} \right] \tag{80}$$

one finds instead of (2.7) that

$${}^Q\langle 0|\partial_i \hat{A}^i |0\rangle^Q = \frac{i}{(2\pi)^{3/2}} \int d^3\vec{k} \sqrt{\frac{\omega(\vec{k})}{2}} \left[\frac{1}{2} q(\vec{k}) e^{i\vec{k}\cdot\vec{x}} - \text{c.c.} \right] = 0. \tag{81}$$

Indeed, the two terms cancel since both $\omega(\vec{k})$ and $q(\vec{k})$ are even under $\vec{k} \to -\vec{k}$. There is no explanation needed for the difference of a factor 2 between (2.5) and (2.6) because A_μ is not a gauge invariant quantity, as opposed to $\vec{\pi}$ and $\vec{\nabla} \times \vec{A}$ whose associated expectation values are correctly given in (2.8) and (2.9). Note however that the Hamiltonian $\hat{H}'_{\xi=1}$ gives rise to the usual oscillating behavior for all oscillators in the Heisenberg picture, except for $\hat{a}(\vec{k}), \hat{a}^\dagger(\vec{k})$ which evolve according to

$$\hat{a}^Q(t, \vec{k}) \equiv \hat{a}(t, \vec{k}) - q(\vec{k}) = e^{-i\omega(\vec{k})t} \hat{a}^Q(\vec{k}), \tag{82}$$

and its complex conjugate.

(iii) In order to make contact with the original [5] and subsequent work, note that the new vacuum corresponds to the old one "dressed" by

$$e^{\int d^3x' \left[i \frac{Q}{2} \frac{x'^i}{|\vec{x}'|^3} \hat{A}_i^{(-)}(\vec{x}') - \frac{1}{2} j^0(\vec{x}')(-\Delta)^{-1/2} A_0^{(-)}(\vec{x}') \right]}, \tag{83}$$

where the subscript $(-)$ denotes the creation part.

Acknowledgments G.B. is grateful to S. Theisen for pointing out reference [7] and to M. Henneaux for suggesting to explain the thermodynamics of the charged capacitor by standard methods. The authors acknowledge useful discussions with F. Alessio, J. Crabbe, and S. Prohazka. This work is supported by the F.R.S.-FNRS Belgium through a research fellowship for Martin Bonte and also through conventions FRFC PDR T.1025.14 and IISN 4.4503.15.

Bibliography

1. P.A.M. Dirac, Proc. R. Soc. Lond. A **136**(829), 453 (1932). doi:10.1098/rspa.1932.0094
2. V. Fock, B. Podolsky, Phys. Zs. Sowjetunion **1**, 798 (1932)
3. T. Kugo, I. Ojima, Prog. Theor. Phys. Suppl. **66**, 1 (1979)
4. M. Henneaux, C. Teitelboim, *Quantization of Gauge Systems* (Princeton University, Princeton, 1992)
5. P.A.M. Dirac, Can. J. Phys. **33**, 650 (1955). doi:10.1139/p55-081
6. G. Barnich, Gen. Rel. Grav. **43**, 2527 (2011). doi:10.1007/s10714-010-0984-6
7. M. Bronstein, Gen. Rel. Grav. **44**, 267 (2012). doi:10.1007/s10714-011-1285-4
8. S. Deser, A.A. Starobinsky, Gen. Rel. Grav. **44**, 263 (2012). doi:10.1007/s10714-011-1284-5
9. H.B.G. Casimir, Indag. Math. **10**, 261 (1948). [Kon. Ned. Akad. Wetensch. Proc.100N3-4,61 (1997)]
10. J. Mehra, Physica **37**, 145 (1967). doi:10.1016/0031-8914(67)90115-2
11. M. Fierz, Helv. Phys. Acta **33**, 855 (1960)
12. L.S. Brown, G.J. Maclay, Phys. Rev. **184**, 1272 (1969). doi:10.1103/PhysRev.184.1272
13. G. Plunien, B. Muller, W. Greiner, Phys. Rept. **134**, 87 (1986). doi:10.1016/0370-1573(86)90020-7
14. B.E. Sernelius, *Surface Modes in Physics* (Wiley, New York, 2001). doi:10.1002/3527603166
15. M. Bordag, G.L. Klimchitskaya, U. Mohideen, V.M. Mostepanenko, in *Advances in the Casimir effect*. Integral Series Monographs Physics, vol. 145 (Oxford University, Oxford, 2009)
16. G. Barnich, Phys. Rev. **D99**(2), 026007 (2019). doi:10.1103/PhysRevD.99.026007
17. G.W. Gibbons, S.W. Hawking, Phys. Rev. **D15**, 2752 (1977)
18. S.W. Hawking, S.F. Ross, Phys. Rev. **D52**, 5865 (1995)
19. S. Deser, M. Henneaux, C. Teitelboim, Phys. Rev. **D55**, 826 (1997)
20. M. Banados, C. Teitelboim, J. Zanelli, Phys. Rev. Lett. **69**, 1849 (1992)
21. J. Slater, N. Frank, *Electromagnetism* (Dover Publications, New York, 1969)
22. H.B. Callen, *Thermodynamics* (Wiley, New York, 1961)
23. F. Heinrich, Am. J. Phys. **54**(8), 742 (1986)
24. H. Casimir, D. Polder, Phys. Rev. **73**(4), 360 (1948)
25. J. Ambjorn, S. Wolfram, Annals Phys. **147**, 1 (1983). doi:10.1016/0003-4916(83)90065-9
26. J.I. Kapusta, Phys.Rev. **D24**, 426 (1981). doi:10.1103/PhysRevD.24.426
27. J.S. Dowker, Phys. Rev. **D37**, 558 (1988). doi:10.1103/PhysRevD.37.558
28. J. Dowker, Class. Quant. Grav. **20**, L105 (2003). doi:10.1088/0264-9381/20/8/102

Part IV
Cosmology and Gravitation & String Theory

Probes in AdS_3 Quantum Gravity

Alejandra Castro

Abstract The Chern–Simons formulation of three-dimensional gravity is a powerful framework to explore non-perturbative aspects of quantum gravity; it allows us to describe properties of gravitational theories without resorting to a metric description. For higher spin gravity this is particularly important, where a geometric formulation can be cumbersome. Here we will review how this formalism has provided unique insights regarding the local properties of higher spin gravity in AdS_3. We will discuss various definitions of black holes in AdS_3 gravity, and how to probe them using the observables that naturally arise in Chern–Simons theory.

Keywords AdS/CFT · Black holes · Higher spin gravity

1 Introduction

Higher spin theories are gravitational theories that challenge our geometrical intuition. A higher spin theory is characterized by being somewhat crowed by symmetries: the theory introduces massless higher spin fields whose gauge symmetries and interactions spoil the standard notions of causality and curvature that we hold sacred otherwise. Very natural concepts in general relativity, such as black holes, become rather puzzling in higher spin gravity. This makes higher spin gravity an excellent arena to explore the repercussions of having violent modifications of general relativity. The aim of this review is to explore potential definitions of black holes in gravitational theories that lack such a geometrical description.

An important appeal of higher spin gravity is that it allows us to introduce non-linear and non-geometrical features classically. These are features we expect to arise in quantum gravity, but are generically difficult to quantify. Within higher spin gravity there is a rather powerful example: in three dimensions the massless higher

A. Castro (✉)
Institute for Theoretical Physics, University of Amsterdam, Amsterdam, The Netherlands
e-mail: a.castro@uva.nl

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_36

spin sector can be consistently described using Chern–Simons theory. Depending on the gauge group we assign to the theory, we will have a different spectrum of higher spin fields. For example, these include pure AdS_3 gravity [1, 2], gravity coupled to Abelian gauge fields, and a tower of massless spin-s fields coupled to a gravity, among many other examples. This can be viewed as truncations of the interacting Prokushkin–Vasiliev higher spin theory [3, 4] which includes in addition massive scalar fields.

In the context of AdS_3/CFT_2, the Chern–Simons sector captures the chiral algebra of the dual two-dimensional CFTs which have an extended symmetry algebras of $\mathcal{W}$-type [5–8]. And here is where the AdS/CFT correspondence has provided a useful framework to organize our understanding of higher spin gravity: it is rather clear how to define, e.g. correlation functions, currents, and sources on both sides of the correspondence. What this description lacks is the addition of light primary fields that are generic in CFTs. Nevertheless, the Chern–Simons sector will suffice to probe how deviations from general relativity can affect our understanding of gravity, and in particular the mechanics behind black holes. For reviews on holographic aspects of AdS_3/CFT_2 involving higher spin fields, we refer the reader to [9, 10] and references within.

As we mentioned before, our main goal here is to be capable of describing *black holes* in higher spin gravity. In the absence of a metric, which is crucial for giving a notion of causality, it is rather non-trivial to think of a definition of black hole at a non-linear and non-local level. Is it the horizon its defining property? Is it the high mass density that gives rise to a curvature singularity? Is it the thermodynamic nature the defining feature? Or the fact that it is the fastest scrambler? Or something else, such as the ring-down pattern? All of these are, in my opinion, valid starting points. The remarkable property of general relativity is that these facts are generically implied by each other. But as we modify violently the interactions, it is not clear if we should expect that these different properties are still so intimately tied: perhaps one could find important deviations from the usual lore in a two-derivative theory of gravity.

In the following I will present one general strategy our community has taken to define a black hole in higher spin gravity. This strategy was initially put forward in the original proposal of [11] and further refined in later work, as we will elaborate in the following sections. I leave it as a challenge to the reader to further explore and question this starting point: any deviations from the lore of GR would be very interesting! And as we will see, some of these deviations are already present given the modest starting point we take.

This talk covers only three topics in this field: Euclidean black holes (Sect. 3), extremal black holes (Sect. 4), and Wilson lines in higher spin gravity (Sect. 5). There are also many other topics in higher spin gravity that I will not explore. I will provide references as appropriate, but unfortunately the field is rather large so many interesting corners will be left out here.

2 AdS$_3$ Higher Spin Gravity

The simplest way to craft a higher spin theory exploits the Chern–Simons formulation of three-dimensional gravity: general relativity with a negative cosmological constant can be reformulated as a $SL(2,\mathbb{R}) \times SL(2,\mathbb{R})$ Chern–Simons theory [1, 2, 12]. A high spin theory can be crafted by simply taking instead $SL(N,\mathbb{R}) \times SL(N,\mathbb{R})$, which will produce an interacting higher spin theory for symmetric tensors of spin $s = 2, 3, \ldots, N$ [13]. There are of course other ways to build higher spin theories, but here we restrict the attention to these models. For a more complete discussion on properties of these theories, see, for example, [9, 14, 15].

The action of the $SL(N,\mathbb{R}) \times SL(N,\mathbb{R})$ Chern–Simons theory is given by

$$S = S_{CS}[A] - S_{CS}[\bar{A}]\,, \quad S_{CS}[A] = \frac{k}{4\pi}\int_{\mathcal{M}} \mathrm{tr}\Big(A \wedge dA + \frac{2}{3} A \wedge A \wedge A\Big)\,. \tag{1}$$

Here $\mathcal{M}$ is the 3-manifold that supports the $sl(N,\mathbb{R})$ algebra valued connections A and $\bar{A}$, and the trace "tr" denotes the invariant quadratic form of the Lie algebra. The equations of motion following from (1) are

$$dA + A \wedge A = 0\,, \quad d\bar{A} + \bar{A} \wedge \bar{A} = 0\,. \tag{2}$$

The conventions here follow those in [16].

The metric and higher spin fields are obtained from the Chern–Simons connection as symmetric, traceless tensors that transform in the spin s representation of $SL(2,\mathbb{R})$. For example, the metric and the spin three fields can be expressed as follows

$$g_{\mu\nu} \sim \mathrm{tr}\big(e_\mu e_\nu\big)\,, \quad \phi_{\mu\nu\rho} \sim \mathrm{tr}\big(e_{(\mu} e_\nu e_{\rho)}\big)\,, \tag{3}$$

where, in line with the pure gravity case, one defines

$$e = \frac{\ell}{2}\big(A - \bar{A}\big)\,, \quad \omega = \frac{1}{2}\big(A + \bar{A}\big) \tag{4}$$

and we introduced the AdS radius ℓ. The metric and higher spin fields can then be expressed in terms of trace invariants of the vielbein [6, 7], with the total number of inequivalent invariants being $N - 1$ for $sl(N,\mathbb{R})$. This definition for metric-like fields is appropriate for the principal embedding of $sl(2,\mathbb{R})$ in $sl(N,\mathbb{R})$.

The relation between the Chern–Simons level and the gravitational couplings is

$$k = \frac{\ell}{8 G_3 \epsilon_N}\,, \quad \epsilon_N \equiv \mathrm{tr}_f(L_0 L_0) = \frac{1}{12} N(N^2 - 1)\,, \tag{5}$$

in accordance with the pure gravity limit. The notation tr_f denotes a trace in the fundamental representation of $sl(N,\mathbb{R})$. The central charge of the asymptotic symmetry group is [5, 6]

$$c = 12k\epsilon_N = \frac{3\ell}{2G} \,. \tag{6}$$

The standard way to parametrize solutions to (2) is by gauging away the radial dependence, i.e.

$$A = b(r)^{-1}\left(a(x^+, x^-) + d\right)b(r)\,, \qquad \bar{A} = b(r)\left(\bar{a}(x^+, x^-) + d\right)b(r)^{-1}\,. \tag{7}$$

Here r is the holographic radial direction, and $x^{\pm} = t \pm \phi$ are the boundary coordinates. In Lorentzian signature we will consider solutions with $\mathbb{R} \times D_2$ topology; the compact direction on D_2 is described by $\phi \sim \phi + 2\pi$. In Euclidean signature we will analytically continue $x^{\pm}$ to complex coordinates $(z, \bar{z})$, and the topology of the bulk is now a solid torus with $z \sim z + 2\pi \sim z + 2\pi i\tau$. Here τ is the modular parameter of the boundary torus. $b(r)$ is a radial function that is normally taken to be e^{rL_0}.[1]

The connections $a(x^+, x^-)$ and $\bar{a}(x^+, x^-)$ contain the information that characterizes the state in the dual CFT. In the absence of sources there is systematic procedure to label them: a suitable set of boundary conditions on the connections results in $\mathcal{W}$-algebras as asymptotic symmetries [5–8, 19]. These are commonly known as Drinfeld–Sokolov boundary conditions. To be concrete, for $sl(N) \times sl(N)$ the connections take the form

$$a_z = L_1 - \sum_{s=2}^{N} J_{(s)}(z) W^{(s)}_{-s+1}\,, \quad \bar{a}_{\bar{z}} = -\left(L_{-1} - \sum_{s=2}^{N} \bar{J}_{(s)}(\bar{z}) W^{(s)}_{s-1}\right)\,, \tag{8}$$

while $a_{\bar{z}} = \bar{a}_z = 0$. Here $\{L_0, L_{\pm 1}\}$ are the generators of the $sl(2, \mathbb{R})$ subalgebra in $sl(N)$, and $W^{(s)}_j$ are the spin-s generators with $j = -(s-1), \ldots (s-1)$; note that $W^{(2)}_j = L_j$. And $J_{(s)}(z)$ are dimension-s currents whose algebra is $\mathcal{W}_N$, and same for the barred sector.

Our general arguments and results will not be very sensitive to the choice of gauge group, but for the sake of simplicity our explicit computations will involve connections valued in either the Lie algebra $sl(2)$ (standard spin-2 gravity on AdS_3) or $sl(3)$ (a graviton coupled to a single spin-3 field).

3 Euclidean Black Holes

We will start the discussion with the most successful (and elegant) definition: Euclidean black holes in AdS_3. We will review our current understanding of the

[1] How to choose $b(r)$ is very important when considering Lorentzian properties of the solutions, and it is usually overlooked. See [17, 18] for a recent discussion on this topic.

solutions and its properties in the Chern–Simons formulation of higher spin gravity. This section is a collection of results in [11, 15, 20–23].

Any definition of black holes should include at least two inputs:

1. A quantitative definition of physical observables; in particular, a definition of conserved charges (such as mass and angular momentum) and its counterparts potentials (such as temperature and angular velocity).
2. A notion of regularity and smoothness. The aim here would be to find a notion of horizon. But even more broadly, we need to clearly argue if a solution, at least in Euclidean signature, lacks singularities.

Let us elaborate first on how to obtain conserved charges. The properties and values of these observables are intimately tied to the boundary conditions we use. For instance, in AdS spacetimes we are mostly accustomed to Dirichlet boundary conditions and to implement a notion of asymptotically AdS spaces (AAdS). But let me emphasize: there is more than one choice! This occurs even in AdS_3 gravity, where some non-trivial examples are shown in [24, 25] and more recently a broad analysis was presented in [26] which are important deviations from the standard Brown–Henneaux boundary conditions [27].

In higher spin gravity we of course have similar choices, but in addition there are further complexities as we turn on sources. More concretely, consider the Chern–Simons connections in (7) and (8), and that we impose AAdS boundary conditions. From the CFT perspective, it is natural to capture the currents in a_z and the sources in $a_{\bar{z}}$, and vice versa for $\bar{a}$ [11]. From the gravitational perspective, the canonical prescription is to encode in $(a_\phi, \bar{a}_\phi)$ the currents [28–31]. These two choices, a_z versus a_ϕ, amount for different partition functions as shown in [23]: the a_z prescription, denoted *holomorphic* black hole, corresponds to a Lagrangian deformation of the theory; the a_ϕ prescription, denoted *canonical* black hole, corresponds to a Hamiltonian deformation. It is important to make a distinction between these two, since the Legendre transformation that connects these two prescriptions is non-trivial.

In the remainder of these lectures we will mostly use the canonical description. Moreover, we are interested in stationary black hole solutions, hence $(a, \bar{a})$ are constant flat connections that contain both charges and sources. This in particular implies that the ϕ-component will be always written as

$$a_\phi = L_1 - \sum_{s=2}^{N} Q_{(s)} W^{(s)}_{-s+1} \,, \quad \bar{a}_\phi = L_{-1} - \sum_{s=2}^{N} \bar{Q}_{(s)} W^{(s)}_{s-1} \,, \tag{9}$$

where $(Q_{(s)}, \bar{Q}_{(s)})$ are constants (not functions) and they represent the conserved charges associated with the zero modes of each higher spin current $(J_{(s)}, \bar{J}_{(s)})$. The a_t component will contain the information about the potentials, which we will

usually denote as μ_s.[2] Hence a solution that contains both charges and potentials will be interpreted in the CFT as being part of a canonical ensemble, where

$$Z_{\text{can}}[\tau, \alpha_s, \bar{\alpha}_s] = \text{Tr}_{\mathcal{H}} \exp 2\pi i \left[\sum_{s=2}^{N} \left(\alpha_s J_0^{(s)} - \bar{\alpha}_s \bar{J}_0^{(s)} \right) \right] . \tag{10}$$

Here, $J_0^{(s)}$ and $\bar{J}_0^{(s)}$ denote the zero modes of the corresponding currents; $Q_{(s)}$ and $\bar{Q}_{(s)}$ would be the eigenvalues of these operators. For $s > 3$ we have

$$\mu_s = \frac{i\alpha_s}{\text{Im}(\tau)} , \qquad \bar{\mu}_s = -\frac{i\bar{\alpha}_s}{\text{Im}(\tau)} , \tag{11}$$

which are the chemical potential associated with each operator; recall that τ is the complex structure of the torus. For $s = 2$ we have

$$J_0^{(2)} = L_0 - \frac{c}{24} , \quad \bar{J}_0^{(2)} = \bar{L}_0 - \frac{c}{24} , \tag{12}$$

and the CFT Hamiltonian and angular momentum are $H = L_0 + \bar{L}_0 - \frac{c}{12}$ and $J = L_0 - \bar{L}_0$, respectively. For the potentials the relation with the complex structure of the torus is

$$\alpha_2 = \tau = \frac{i\beta}{2\pi}\big(1 + \Omega\big) , \quad \bar{\alpha}_2 = \bar{\tau} = \frac{i\beta}{2\pi}\big(-1 + \Omega\big) , \tag{13}$$

with β the inverse temperature and Ω the angular velocity.

The feature that distinguishes black holes from other solutions is a smoothness condition, and this brings us to the second bullet point mentioned above. In a metric formulation of gravity, the Euclidean section of a black hole has the property that the compact Euclidean time direction smoothly shrinks to zero size at the horizon of the black hole, resulting in a smooth cigar-like geometry as in Fig. 1. In the Chern–Simons formulation of gravity, this property is normally thought to generalize to the idea that a black hole is a flat gauge connection defined on a solid torus, where the holonomy along the thermal cycle of the torus belongs to the center of the group, i.e.

$$\mathcal{P} \exp \left(\oint_{\mathcal{C}_E} a \right) \cong e^{2\pi(\tau a_z + \bar{\tau} a_{\bar{z}})} \cong e^{2\pi i L_0} , \tag{14}$$

[2] For a quantitative and general definition of μ_s in terms of (a_t, a_ϕ) see, for example, [23]. Here we will just define them via examples.

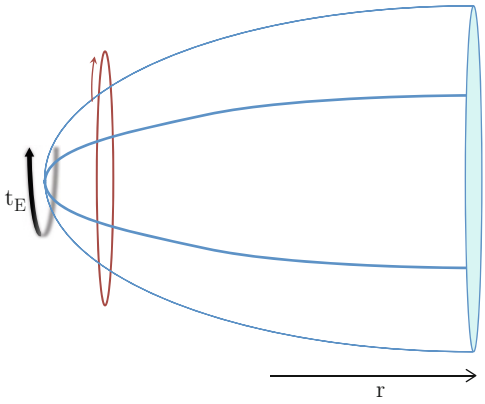


Fig. 1 Topology of the Euclidean higher spin black hole for a static solution, where the compact direction is Euclidean time $t = it_E$. The red curve depicts the cycle along which the smoothness condition (14) is imposed, and it is independent of the radial position. In Euclidean signature, the geometry ends at a finite value of r: in a metric-like formulation of gravity this end point would be the horizon

and similarly in the barred sector; here L_0 denotes the Cartan element of $sl(2)$,[3] and $\mathcal{C}_E$ is the thermal cycle $z \sim z + 2\pi i\tau$ which is contractible in the bulk.

The smoothness condition (14) is a robust and successful definition of Euclidean black holes. It reproduces in an elegant manner many properties that we expect from a thermal state in the dual CFT$_2$. This definition has also unveiled novel properties of systems in the grand canonical ensemble of $\mathcal{W}_N$, such as microscopic features of the entropy [30, 35, 36], ensemble properties [21, 23], and novel phase diagrams [37], and it inspires new observables related to entanglement entropy [33, 38, 39].

It is perhaps worth emphasizing that there exist several ways to compute the entropy of higher spin black holes, all giving the same result. In the original proposal of [11], the entropy was inferred by demanding integrability of the thermodynamic laws. For a Hamiltonian derivation of the entropy, see, e.g. [22, 28, 40]. The entropy can also be understood as the on-shell value of the appropriate action functional in a microcanonical ensemble, where the charges at infinity are held fixed [21]. The punchline is that the entropy of a higher spin black hole reads

$$S = -2\pi i k \mathrm{Tr}\Big[(a_z + a_{\bar{z}})\,(\tau a_z + \bar{\tau} a_{\bar{z}}) - (\bar{a}_z + \bar{a}_{\bar{z}})\,(\tau \bar{a}_z + \bar{\tau}\bar{a}_{\bar{z}})\Big]. \tag{15}$$

More interestingly, one can exploit the holonomy conditions to cast the entropy directly as function of the charges only. Using the smoothness conditions (14) one

[3] Depending on the gauge group, the choice of center in the rhs of (14) is not unique [32]. The choice used here has the feature that it is smoothly connected to the BTZ solution. The interpretations of other choices are discussed in [33, 34].

finds that (15) can be written equivalently as [21]

$$S = 2\pi k \mathrm{Tr}\Big[\big(\lambda_\phi - \bar{\lambda}_\phi\big)\, L_0\Big]\,, \tag{16}$$

where λ_ϕ and $\bar{\lambda}_\phi$ are diagonal matrices containing the eigenvalues of the angular component of the connection (9), which carries the values of the charges $(Q_{(s)}, \bar{Q}_{(s)})$.

3.1 Example

To illustrate the discussion in this section, we will consider black holes in $SL(3) \times SL(3)$ Chern–Simons theory. In this case we define:[4]

$$\begin{aligned}
a_+ &= L_1 - Q_{(2)} L_{-1} - \frac{Q_{(3)}}{4} W_{-2}\,,\\
a_- &= \mu_3 \left(W_2 + 2Q_{(3)} L_{-1} + Q_{(2)}^2 W_{-2} - 2Q_{(2)} W_0\right)\,,\\
\bar{a}_- &= -\left(L_{-1} - Q_{(2)} L_1 + \frac{Q_{(3)}}{4} W_2\right)\,,\\
\bar{a}_+ &= \mu_3 \left(W_{-2} - 2Q_{(3)} L_1 + Q_{(2)}^2 W_2 - 2Q_{(2)} W_0\right).
\end{aligned} \tag{17}$$

For simplicity we have turned off rotation, i.e. $Q_{(2)} = \bar{Q}_{(2)}$ and $Q_{(3)} = -\bar{Q}_{(3)}$; this as well implies that τ is purely imaginary ($\tau = i\beta$) and $\bar{\mu} = -\mu$. The interpretation of these connections as thermal states depends on the boundary conditions used to define the classical phase space. The holomorphic black hole is given by the following connections:

$$a_h = a_+ dz + a_- d\bar{z}\,, \qquad \bar{a}_h = \bar{a}_+ dz + \bar{a}_- d\bar{z}\,. \tag{18}$$

In this notation the components $(a_z, \bar{a}_{\bar{z}})$ contain the information of the charges of the system: $(Q_{(2)}, Q_{(3)})$ are the zero modes of the stress tensor and dimension-3 current of the $\mathcal{W}_3$ asymptotic symmetry group that organizes the states in this theory. (β, μ) are their respective sources which are fixed by the smoothness condition (14). The second prescription, i.e. the canonical black hole, is given by

$$a_c = a_+ d\phi + (a_+ + a_-)dt\,, \qquad \bar{a}_c = -\bar{a}_- d\phi + (\bar{a}_+ + \bar{a}_-)dt\,. \tag{19}$$

[4] Note that the equations of motion, flatness condition, simply impose that $[a_+, a_-] = 0 = [\bar{a}_-, \bar{a}_+]$ as can be checked explicit for (17).

For this prescription, again $(Q_{(2)}, Q_{(3)})$ are the zero modes of the currents in $\mathcal{W}_3$. The quantitative difference between the holomorphic and canonical definitions lies in the spatial components of the connection; both a_c and a_h have the same time component.

The smoothness condition (14) enforces relations between the parameters $Q_{(2)}$, $Q_{(3)}$, μ_3, and β. Following [11, 20], these constraints can be solved in terms of dimensionless parameter $C \geq 3$:

$$\begin{aligned} Q_{(3)} &= \frac{4(C-1)Q_{(2)}}{C^{3/2}}\sqrt{Q_{(2)}}\,, \\ \mu_3 &= \frac{3\sqrt{C}}{4(2C-3)}\sqrt{\frac{1}{Q_{(2)}}}\,, \\ \frac{\mu_3}{\beta} &= \frac{3}{4\pi}\frac{(C-3)\sqrt{4C-3}}{(3-2C)^2}\,. \end{aligned} \tag{20}$$

The limit $C \to \infty$ makes the higher spin charges vanish, and we recover the BTZ case. $C = 3$ and μ_3 fixed corresponds to a zero temperature solution which defines an extremal higher spin black hole [11, 41] which is the subject of the next section.

Applying (15) to the canonical black hole (19) we get

$$S = 8k\left(2\beta Q_{(2)} + 3\alpha_3 Q_{(3)}\right), \tag{21}$$

where the thermal spin-3 source α_3 is related to the spin-3 chemical potential μ_3 as in (11). This expression is clearly compatible with a first law of thermodynamics. It is simple to generalize this expression to restore the barred variables; this gives

$$S = -8\pi ik\left(2\tau Q_{(2)} + 3\alpha_3 Q_{(3)}\right) + 8\pi ik\left(2\bar{\tau}\,\bar{Q}_{(2)} + 3\bar{\alpha}_3\bar{Q}_{(3)}\right). \tag{22}$$

The entropy as function of the charges can be achieved via (16), and for this purpose it is convenient to trade the charges $(Q_{(2)}, Q_{(3)})$ for the eigenvalues of a_ϕ. More concretely, we parametrize

$$\text{Eigen}(a_\phi) = (\lambda_1, \lambda_2, -\lambda_1 - \lambda_2)\,, \tag{23}$$

so that

$$Q_{(2)} = \frac{1}{4}\left(\lambda_1^2 + \lambda_1\lambda_2 + \lambda_2^2\right)\,, \qquad Q_{(3)} = \frac{1}{2}\lambda_1\lambda_2\left(\lambda_1 + \lambda_2\right)\,, \tag{24}$$

with analogous expressions in the barred sector. In Lorentzian signature the eigenvalues $(\lambda_i, \bar{\lambda}_i)$ are independent and real when one chooses the connection to be valued in $sl(3;\mathbb{R})$. In Euclidean signature, we have $\lambda_i^* = -\bar{\lambda}_i$, which implies that ${Q_{(2)}}^* = \bar{Q}_{(2)}$ and ${Q_{(3)}}^* = -\bar{Q}_{(3)}$. Equation (16) then gives us immediately the entropy as a function of the charges

$$S = 2\pi k\big(\lambda_1 - \lambda_3\big) + \text{other sector}$$
$$= 2\pi k\big(2\lambda_1 + \lambda_2\big) + \text{other sector}\,, \tag{25}$$

with λ_1 and λ_2 obtained by inverting (24) and choosing the branch of the solution that connects smoothly to the BTZ black hole as one turns off the $Q_{(3)}$ charge.

4 Extremal Black Holes

In general relativity there is a wide variety of black holes which are not necessarily Euclidean. For example, there are Lorentzian black holes that do not have a real Euclidean continuation (such as five-dimensional black rings), there are eternal black holes (which are the maximal extension of the Euclidean geometry in Lorentzian signature), and black holes that arise from gravitational collapse. And there are as well *extremal black holes*. Extremal black holes have undoubtedly played a crucial role in string theory: due to their enhanced symmetries and their capacity to preserve supersymmetry, they have become a landpost for microstate counting and precursors to many aspects of holography. As such, it is very natural to wonder what is the definition of extremality in higher spin gravity. This is the question we will address in this section. The discussion here is a summary of the results presented in [41]. See also [42] for a discussion on related properties.

4.1 A Practical Definition of Extremality

In conventional gravitational theories, the notion of extremality is tied to the confluence of two horizons. This feature generically implies that the Hawking temperature of the black hole is zero. We could declare that extremality in higher spin theories is simply defined as a solution at zero temperature. However, our aim is to propose a definition that is along the lines of confluence (degeneration) of the parameters of the solution and that relies only on the topological formulation of the theory, yielding in particular the zero temperature condition as a consequence.

In this spirit, in [41] we proposed that a $3d$ extremal higher spin black hole is a solution of Chern–Simons theory corresponding to flat boundary connections a and $\bar{a}$ satisfying the following conditions:

1. They obey AAdS boundary conditions,[5]
2. Their components are constant, and therefore correspond to stationary solutions,

[5] In the literature these boundary conditions are commonly known as Drinfeld–Sokolov boundary conditions.

3. They carry charges and chemical potentials, which are manifestly real in the Lorentzian section,
4. The angular component of at least one of a and $\bar{a}$, say a_ϕ , is non-diagonalizable.

Naturally, the key point of the definition is the non-diagonalizability of the a_ϕ component. The rationale behind this requirement is as follows. Suppose both the a_ϕ and $\bar{a}_\phi$ components were diagonalizable. Since the boundary connections are assumed to be constant, by the equations of motion the (Euclidean) time components of the connection commute with the angular components, and can be diagonalized simultaneously with them. It is then possible to solve (14) and find a non-zero and well-defined temperature and chemical potentials as function of the charges. On the other hand, if at least one of a_ϕ and $\bar{a}_\phi$ is non-diagonalizable, then a_{contract} will be non-diagonalizable as well. If we insist upon (14), then both features are compatible if we take a zero temperature limit, because the smoothness condition becomes degenerate as well. This is consistent with the usual notion that the solid torus topology of the finite-temperature black hole should change at extremality.

The role of boundary conditions is crucial for our definition. For a general connection the degeneration of eigenvalues does not imply non-diagonalizability. However, the special form of the flat connections dictated by the AAdS boundary conditions will guarantee that if two eigenvalues of a_ϕ are degenerate, then the connection is non-diagonalizable. From this perspective, we could interpret that equating eigenvalues of a_ϕ is in a sense analogous to the confluence of horizons for extremal black holes in general relativity.

4.2 Example: Extremal $sl(3)$ Black Holes

In [41] several supersymmetric and non-supersymmetric cases were studied. For brevity, here we will only look at the extremal cousin of the Euclidean black hole we considered in 3.1.

Let us write again (17) but will focus on the unbarred sector for concreteness; recall that it is sufficient to impose our definition of extremality on one sector to obtain the desired features. Using canonical boundary conditions, the connections are given by

$$a_\phi = L_1 - Q_{(2)} L_{-1} - \frac{Q_{(3)}}{4} W_{-2} \,, \tag{26}$$

$$i a_{t_E} + a_\phi = 2a_- = 2\mu_3 \left(W_2 + 2Q_{(3)} L_{-1} + Q_{(2)}^2 W_{-2} - 2Q_{(2)} W_0 \right) \,. \tag{27}$$

It is also instructive to re-write the solutions to (14) for the general rotating case, i.e. the generalization of (20). This gives

$$\tau = i\frac{2\lambda_1^2 + 2\lambda_1\lambda_2 - \lambda_2^2}{(\lambda_1 - \lambda_2)(2\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)}, \tag{28}$$

$$\alpha_3 = -6i\frac{\lambda_2}{(\lambda_1 - \lambda_2)(2\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)}, \tag{29}$$

and

$$\mu_3 = 6\big(1 + \Omega\big)\left(\frac{\lambda_2}{2\lambda_1^2 + 2\lambda_1\lambda_2 - \lambda_2^2}\right), \tag{30}$$

$$\bar{\mu}_3 = -6\big(1 - \Omega\big)\left(\frac{\bar{\lambda}_2}{2\bar{\lambda}_1^2 + 2\bar{\lambda}_1\bar{\lambda}_2 - \bar{\lambda}_2^2}\right). \tag{31}$$

In the above expression we traded $Q_{(2)}$ and $Q_{(3)}$ by its eigenvalues λ_1 and λ_2 as defined in (23). With these explicit relations we can now implement our definition of extremality. Requiring that a_ϕ should be non-diagonalizable gives us a necessary condition

$$\lambda_1 = \lambda_2 \equiv \lambda \qquad \Rightarrow \qquad Q_{(2)} = \frac{3}{4}\lambda^2, \quad Q_{(3)} = \lambda^3. \tag{32}$$

As a consequence, while the finite-temperature angular holonomy is diagonalizable, in the extremal limit we obtain

$$\mathrm{Hol}_\phi(a) \sim \begin{pmatrix} e^{-4\pi\lambda} & 0 & 0 \\ 0 & e^{2\pi\lambda} & 1 \\ 0 & 0 & e^{2\pi\lambda} \end{pmatrix}. \tag{33}$$

Turning now our attention to the potentials, from (28)–(31) we see in particular that in this limit

$$\text{extremal potentials:} \qquad \beta \to \infty, \quad \mu \to 4\frac{\gamma}{\lambda}, \quad \Omega \to 1, \quad \bar{\mu} \to 0, \tag{34}$$

so the temperature is zero as expected. The spin-3 chemical potential μ remains finite and becomes a simple homogeneous function of the charges, whereas the corresponding thermal source α scales with the inverse temperature and blows up. On the other hand, the barred sector spin-3 potential $\bar{\mu}$ goes to zero because the thermal source $\bar{\alpha}$ remains unconstrained and in particular finite, as no condition is imposed on the barred charges.

Several comments are now in order.

1. *Jordan decomposition versus zero temperature:* A valid concern is to wonder if our definition of extremality implies zero temperature and vice versa. From (28) it is clear that there are three combinations of λ_1 and λ_2 that achieve $\beta \to \infty$. The additional other branches also give non-trivial Jordan forms, since they

just correspond to different pairings of eigenvalues that are degenerate. For this reason, all these cases are captured by (32): any pairing $\lambda_i = \lambda_j$ with $i \neq j$ implies the extremality bound $Q_{(2)}^3 = 27/64 Q_{(3)}^2$.[6] At least for $N = 2, 3$, a non-trivial Jordan decomposition implies zero temperature and vice versa. And from the heuristic argument in Sect. 4.1, we expect this to always be the case.

2. *Other Jordan classes:* For $\lambda \equiv \lambda_1 = \lambda_2 \neq 0$, a_ϕ has only two linearly independently eigenvectors. If take first $\lambda_2 = 0$ and then $\lambda_1 = 0$, the holonomy of a_ϕ belongs to a different Jordan class where there is only one eigenvector; this case corresponds to extremal BTZ within $sl(3) \oplus sl(3)$ Chern–Simons theory.
3. *Finite entropy:* We have a continuous family of extremal $\mathcal{W}_3$ black holes parametrized by λ, and from (25) the contribution of the extremal (unbarred) sector to the total entropy is

$$S_{\text{ext}} = 2\pi k_{cs}\,\lambda = \frac{\pi k}{3}\sqrt{48 Q_{(2)}} = 2\pi k \left(Q_{(3)}\right)^{1/3} . \tag{35}$$

The answer is clearly finite. This should be contrasted with extremal BTZ, where the contribution of the extremal sector vanishes. It would be interesting to derive such bound and residual entropy in a CFT with $\mathcal{W}_3$ symmetry.

4. *Extremality vs. unitarity:* The extremality condition we have discussed can be thought of as a bound

$$Q_{(2)}^3 \geq \frac{27}{64} Q_{(3)}^2 \tag{36}$$

on the charges of a spin-3 black hole. On the other hand, in a theory with $\mathcal{W}_3$ symmetry, the unitary bound in the semiclassical limit is [43][7]

$$\frac{64}{5c}\left(h^3 - \frac{c}{32}h^2\right) \geq 9q_3^2 , \tag{37}$$

where the map between the CFT variables (h, q_3) and the gravitational charges is

$$h - \frac{c}{24} = 4k Q_{(2)} , \quad q_3 = k Q_{(3)} . \tag{38}$$

[6] Different pairings of eigenvalues conflict with the ordering of eigenvalues used in (25), but this is easily fixed by reordering the eigenvalues appropriately.

[7] The quantum (finite-c) unitarity bound reported in [43] is

$$\frac{64}{22 + 5c} h^2 \left(h - \frac{1}{16} - \frac{c}{32}\right) - 9q_3^2 \geq 0 .$$

It is clear that (36) and (37) do not agree. However, the $\mathcal{W}_3$ unitarity bound (37) encloses the bulk extremality bound (36), indicating that all $sl(3)$ black holes are dual to states allowed by unitarity in the dual CFT.

5. *Conformal invariance:* In two-derivative theories of gravity in $D = 4, 5$ all extremal black holes contain an AdS_2 factor in its near horizon geometry [44, 45]. The enhancement of time translations to conformal transformations is non-trivial and unexpected a priori; moreover, it is key to build microscopic models of extremal black holes. Here we have not investigated this feature explicitly, but we do expect that the connection at the extremal point is invariant a larger set of gauge transformations relative to the non-extremal connection. Some evidence was reported in [18].
6. *Entropy bounds:* The extremal limit of the spin-3 higher spin black hole was first discussed in [11]. Their bound was found as the maximal value of $Q_{(3)}$ for a given $Q_{(2)}$ such that the entropy is real, and it agrees with (32). Using the reality of entropy as a bound which enhanced symmetries of the solution was also used in [42]. It is not clear if this approach is always compatible with ours, and it will be interesting to explore potential discrepancies.
7. *Supersymmetry and Extremality:* As we mentioned above, extremality can be understood as the saturation of certain inequalities involving conserved charges, and it is natural to contrast these inequalities with BPS bounds that appear in supersymmetric setups. It is well known that in two-derivative theories of supergravity these two types of conditions are intimately related: supersymmetry always implies zero temperature and therefore extremality in the context of BPS black holes. In supersymmetric theories of higher spin gravity this seems to not be true!! In [41] we showed that there exist non-extremal solutions in the class of diagonalizable connections that possess 4 independent Killing spinors. This is, within the $sl(3|2)$ theory, we managed to construct a smooth higher spin black hole that is both at finite temperature and BPS. Understanding why higher spin theories allow for this peculiar behavior is an open question that needs urgent attention.

5 Wilson Lines

As we have mentioned throughout, higher spin gravity does not admit a conventional geometric understanding. However, they do admit interesting higher spin invariant probes. In this section we will consider the Wilson line operator constructed in [39, 46]. This object will allow us to address two important questions:

1. How to capture casual properties in higher spin gravity?
2. How to probe local bulk physics using the Chern–Simons formulation of gravity?

In the following we will only provide the basic definitions, features, and main results obtained in this area. For a detailed discussion we refer to [16, 18, 39, 46].

5.1 Wilson Lines in AdS$_3$/CFT$_2$

The Wilson loop in our 3D higher spin gravity is given by

$$W_{\mathcal{R}}(C) = \mathrm{Tr}_{\mathcal{R}} \left(\mathcal{P} \exp \left(\int_C A \right) \mathcal{P} \exp \left(\int_C \bar{A} \right) \right) . \tag{39}$$

Here A and $\bar{A}$ are the connections representing a higher spin background in $SL(N, \mathbb{R})$ Chern–Simons theory. The representation $\mathcal{R}$ is the infinite-dimensional highest-weight representation of $sl(N, \mathbb{R})$, and C is a loop in the bulk. We may also consider an open-ended Wilson *line* operator. To define this object we specify the locations of its endpoints (x_i, x_f). We must also specify boundary data in the form of two specific states $|U_i\rangle, |U_f\rangle \in \mathcal{R}$ at these endpoints. The Wilson line operator is then

$$W_{\mathcal{R}}(x_i, x_f) = \langle U_f | \mathcal{P} \exp \left(- \int_\gamma A \right) \mathcal{P} \exp \left(- \int_\gamma \bar{A} \right) |U_i\rangle , \tag{40}$$

where now $\gamma(s)$ is a curve with bulk endpoints (x_i, x_f) parametrized by s. $W_{\mathcal{R}}(x_i, x_f)$ is no longer fully gauge-invariant; clearly it depends in a gauge-covariant manner on the choice of boundary data $|U_i\rangle, |U_f\rangle$. Nevertheless, for flat connections, $W_{\mathcal{R}}(x_i, x_f)$ only depends on the topology of γ, but not on the shape of the curve.

In AdS$_3$ the presence of a boundary allows the formulation of suitably diffeomorphism invariant observables—the correlation functions of the dual CFT$_2$—and thus one would expect that it would be possible to compute such objects in the Chern–Simons formulation. Some progress in this direction was made in [16, 34, 38, 39, 47], motivated largely by the computation of entanglement entropy of field theories dual to 3d bulk higher spin gravity. These developments have shown that a Wilson line in an infinite-dimensional highest-weight representation $\mathcal{R}$ under the bulk $SL(N) \times SL(N)$ gauge group could be used to compute boundary theory correlators, i.e.

$$W_{\mathcal{R}}(x_i, x_f) \underset{r \to \infty}{=} \langle \Psi | \mathcal{O}(y_i) \mathcal{O}(y_f) | \Psi \rangle , \tag{41}$$

where we have picked coordinates $x^\mu = (r, y^i)$ with r an AdS holographic coordinate and y^i a CFT coordinate. See Fig. 2. Here the Wilson line $W_{\mathcal{R}}$ ends on the boundary at $r \to \infty$, and Ψ denotes the CFT$_2$ state dual to a particular configuration of Chern–Simons gauge fields that constitute the gravitational background in the interior. The representation space $\mathcal{R}$ was generated from the Hilbert space of an auxiliary $SL(N)$-valued quantum mechanical degree of freedom $U(s)$ that lives on the Wilson line. The quadratic Casimirs of the representation $\mathcal{R}$ mapped in the usual manner to the conformal dimensions $(h, \bar{h})$ of the dual CFT operator $\mathcal{O}(y)$.

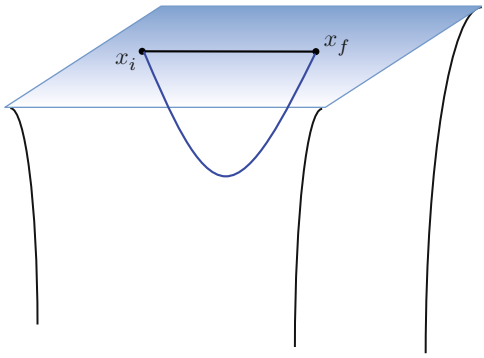


Fig. 2 Wilson line with endpoints at the boundary of AdS_3. This object computes a correlation function, or more precisely a conformal block, in the dual CFT_2

The application of Wilson lines to Lorentzian properties of higher spin gravity is extremely useful. In [18] we studied how to describe an eternal black hole in the Chern–Simons sector of AdS_3 higher spin gravity. We probed such black holes using bulk Wilson lines and motivated new regularity conditions that must be obeyed by the bulk connections in order for the geometry to be consistent with an interpretation as a thermofield state in the dual CFT_2. We demonstrated that any higher spin black hole may be placed in a gauge that satisfies these conditions: this is the Chern–Simons analogue of the construction of Kruskal coordinates that permit passage through the black hole horizon. We also argued that the Wilson line provides a higher spin notion of causality in higher spin gravity that can be used to associate a Penrose diagram with the black hole. See Fig. 3.

These Wilson lines then provide us with a sensitive probe of bulk higher spin geometries. Interestingly, we found that the study of Wilson lines on the eternal black hole background required a refined understanding of regularity properties on the bulk gauge connections. One of our main results was the description of a particular bulk gauge choice—which we call *Kruskal gauge*—that is in many ways the Chern–Simons analogue of the Kruskal choice of coordinates that permit passage through the event horizon to the full maximally extended spacetime. This gauge choice simply amounts to demanding that the connections be *smooth* when evaluated at the Euclidean origin: while this may sound like a very benign condition, it involves an interplay between the bulk radial coordinate and Euclidean time, and so is novel from the point of view of Chern–Simons theory. In particular, it is stronger than the familiar "holonomy conditions" of Euclidean regularity that are normally used to define black hole connections: however, given a black hole that satisfies the holonomy condition, there is an algorithm that can be followed to place it into Kruskal gauge. Some recent work that also implements this stronger notion of regularity is in [17].

With an understanding of this bulk gauge choice one can further study the properties of eternal higher spin black holes. We presented in [18] computations in several gauges to illustrate potential pitfalls, and verify that in Kruskal gauge, all correlators behave as expected for a thermofield state. We also studied some of the resulting physics: in particular, we demonstrated that the interior of a two-

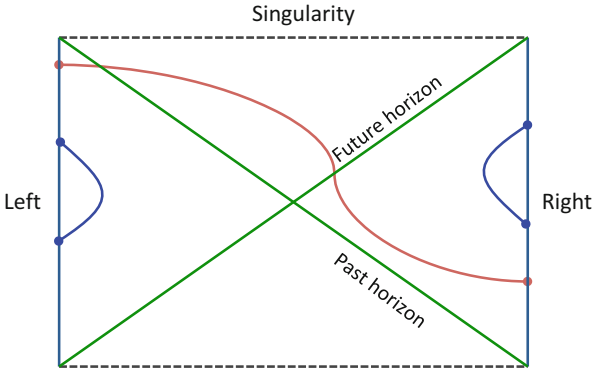


Fig. 3 Wilson line with endpoints in different boundaries of the eternal black hole background. Here $W_{\mathcal{R}}(x_i, x_f)$ allows us to test properties that rely on the presence of the left and right boundary

sided eternal black hole "grows" with time (as measured by the action of a bulk Wilson line). We also highlighted some interesting features of purely one-sided correlators, in particular the behavior of the extremal limit and provided evidence for the emergence of an infrared AdS$_2$.

Finally, another interesting development is the results we obtained in [46]. By selecting $|U\rangle$ to be singlet states of $so(2, 2)$, we showed that

$$W_{\mathcal{R}}(x_i, x_f) = \frac{e^{-2hD(x_i,x_f)}}{1 - e^{-2D(x_i,x_f)}} \tag{42}$$

where $D(x_i, x_f)$ is the geodesic length of an effective metric given by

$$g_{\mu\nu} = \frac{1}{2}\mathrm{Tr}(A_\mu - U^{-1}\bar{A}_\mu U)(A_\nu - U^{-1}\bar{A}_\nu U)\ . \tag{43}$$

Equation (42) is the familiar bulk-to-bulk propagator of a minimally coupled scalar field in a locally AdS$_3$ background [48, 49]. An expression such as (42) makes rather evident that the Wilson line is a propagator, and hence its ties to geometry. It should be viewed as an operator that satisfies

$$\left(\frac{1}{2}\Box_{x_f} - 2h(h-1)\right) W_{\mathcal{R}}(x_f, x_i) = \frac{1}{8\pi}\frac{\delta(x_f, x_i)}{\sqrt{-g}}, \tag{44}$$

where now $\Box_{x_f}$ is the Laplacian on the bulk AdS$_3$ spacetime.

An important issue that we have not addressed is quantum corrections due to fluctuations of the background connections. This would capture $1/c$ corrections, i.e. corrections controlled by the AdS radius in Planck units, or equivalently subleading terms controlled by the level of the Chern–Simons theory. Work in this direction has been done for $SL(2)$ Chern–Simons theory, where Virasoro conformal blocks are known to be tied to appropriate Wilson line in Chern–Simons [50–52]. Recent developments for this holomorphic theory include [53–59]. It would be interesting to evaluate $1/c$ corrections of our worldline quantum mechanics; in this case we

expect that the intertwining of the two copies of $sl(2)$ will produce interesting features. For example, we should be able to probe if the global conditions on the $|U\rangle$ states are enhanced to the Virasoro conditions on the Ishibashi state [60, 61], or something completely different, such as the conditions proposed in [62]. We leave these questions for future work.

Another natural direction forward is to use our construction to build probes in $SL(N) \times SL(N)$ Chern–Simons theory, and hence generalize (42) beyond pure AdS_3 gravity. This would provide a unique way to build local probes in higher spin gravity. A discussion of Ishibashi states for $\mathcal{W}_3$ algebra was done in [63], which is a natural starting point for future investigations.

Bibliography

1. A. Achucarro, P.K. Townsend, Extended supergravities in $d = (2+1)$ as Chern–Simons theories. Phys. Lett. **B229**, 383–387 (1989)
2. E. Witten, (2+1)-Dimensional gravity as an exactly soluble system. Nucl. Phys. **B311**, 46 (1988)
3. S. Prokushkin, M.A. Vasiliev, 3-D higher spin gauge theories with matter, in Proceedings of the 31st International Symposium Ahrenshoop Theory of Elementary Particles, Buckow, Germany, 2–6 Sept 1997 (1998)
4. S. Prokushkin, M.A. Vasiliev, Higher spin gauge interactions for massive matter fields in 3-D AdS space-time. Nucl. Phys. **B545**, 385 (1999)
5. M. Henneaux, S.J. Rey, Nonlinear $W_{infinity}$ as asymptotic symmetry of three-dimensional higher spin anti-de sitter gravity. JHEP **1012**, 007 (2010)
6. A. Campoleoni, S. Fredenhagen, S. Pfenninger, S. Theisen, Asymptotic symmetries of three-dimensional gravity coupled to higher-spin fields. JHEP **1011**, 007 (2010)
7. A. Campoleoni, S. Fredenhagen, S. Pfenninger, Asymptotic W-symmetries in three-dimensional higher-spin gauge theories. JHEP **1109**, 113 (2011)
8. M.R. Gaberdiel, T. Hartman, Symmetries of holographic minimal models. JHEP **1105**, 031 (2011)
9. M.R. Gaberdiel, R. Gopakumar, Minimal model holography. J.Phys. **A46**, 214002 (2013)
10. M.R. Gaberdiel, R. Gopakumar, String theory as a higher spin theory. JHEP **09**, 085 (2016)
11. M. Gutperle, P. Kraus, Higher spin black holes. JHEP **1105**, 022 (2011)
12. A. Achúcarro, P. Townsend, A Chern–Simons action for three-dimensional anti-de Sitter supergravity theories. Phys. Lett. B **180**(1–2), 89–92 (1986)
13. M. Blencowe, A consistent interacting massless higher spin field theory in D = (2+1). Class. Quant. Grav. **6**, 443 (1989)
14. X. Bekaert, S. Cnockaert, C. Iazeolla, M.A. Vasiliev, Nonlinear higher spin theories in various dimensions, in Proceedings of the 1st Solvay Workshop, Higher Spin Gauge Theories, Brussels, Belgium, 12–14 May, 2004 (2004), pp. 132–197
15. M. Ammon, M. Gutperle, P. Kraus, E. Perlmutter, Black holes in three dimensional higher spin gravity: a review. J. Phys. **A46**, 214001 (2013)
16. A. Castro, E. Llabrés, Unravelling holographic entanglement entropy in higher spin theories. JHEP **1503**, 124 (2015)
17. M. Bañados, R. Canto, S. Theisen, Higher spin black holes in three dimensions: remarks on asymptotics and regularity. Phys. Rev. **D94**(2), 024002 (2016)
18. A. Castro, N. Iqbal, E. Llabrés: Eternal higher spin black holes: a thermofield interpretation. JHEP **08**, 022 (2016)

19. J. de Boer, Six-dimensional supergravity on S**3 x AdS(3) and 2-D conformal field theory. Nucl. Phys. **B548**, 139–166 (1999)
20. M. Ammon, M. Gutperle, P. Kraus, E. Perlmutter, Spacetime geometry in higher spin gravity. JHEP **1110**, 053 (2011)
21. J. de Boer, J.I. Jottar, Thermodynamics of higher spin black holes in AdS_3. JHEP **1401**, 023 (2014)
22. C. Bunster, M. Henneaux, A. Perez, D. Tempo, R. Troncoso, Generalized black holes in three-dimensional spacetime. JHEP **05**, 031 (2014)
23. J. de Boer, J.I. Jottar, Boundary conditions and partition functions in higher spin AdS_3/CFT_2. JHEP **04**, 107 (2016)
24. G. Compère, W. Song, A. Strominger, New boundary conditions for AdS3. JHEP **05**, 152 (2013)
25. C. Troessaert, Enhanced asymptotic symmetry algebra of AdS_3. JHEP **08**, 044 (2013)
26. D. Grumiller, M. Riegler, Most general AdS_3 boundary conditions. JHEP **10**, 023 (2016)
27. J.D. Brown, M. Henneaux, On the Poisson brackets of differentiable generators in classical field theory. J. Math. Phys. **27**, 489–491 (1986)
28. M. Banados, R. Canto, S. Theisen, The action for higher spin black holes in three dimensions. JHEP **1207**, 147 (2012)
29. A. Perez, D. Tempo, R. Troncoso, Higher spin gravity in 3D: Black holes, global charges and thermodynamics. Phys. Lett. **B726**, 444–449 (2013)
30. G. Compère, J.I. Jottar, W. Song, Observables and Microscopic Entropy of Higher Spin Black Holes. JHEP **1311**, 054 (2013)
31. M. Henneaux, A. Perez, D. Tempo, R. Troncoso, Chemical potentials in three-dimensional higher spin anti-de Sitter gravity. JHEP **1312**, 048 (2013)
32. A. Castro, E. Hijano, A. Lepage-Jutier, A. Maloney, Black Holes and singularity resolution in higher spin gravity. JHEP **1201**, 031 (2012)
33. E. Hijano, P. Kraus, A new spin on entanglement entropy. JHEP **12**, 041 (2014)
34. J. de Boer, A. Castro, E. Hijano, J.I. Jottar, P. Kraus, Higher spin entanglement and $\mathcal{W}_N$ conformal blocks. JHEP **07**, 168 (2015)
35. P. Kraus, E. Perlmutter, Partition functions of higher spin black holes and their CFT duals. JHEP **1111**, 061 (2011)
36. M.R. Gaberdiel, T. Hartman, K. Jin, Higher spin black holes from CFT. JHEP **1204**, 103 (2012)
37. J.R. David, M. Ferlaino, S.P. Kumar, Thermodynamics of higher spin black holes in 3D. JHEP **1211**, 135 (2012)
38. J. de Boer, J.I. Jottar, Entanglement entropy and higher spin holography in AdS_3. JHEP **04**, 089 (2014)
39. M. Ammon, A. Castro, N. Iqbal, Wilson lines and entanglement entropy in higher spin gravity. JHEP **1310**, 110 (2013)
40. A. Perez, D. Tempo, R. Troncoso, Higher spin black hole entropy in three dimensions. JHEP **1304**, 143 (2013)
41. M. Bañados, A. Castro, A. Faraggi, J.I. Jottar, Extremal higher spin black holes. JHEP **04**, 077 (2016)
42. M. Henneaux, A. Perez, D. Tempo, R. Troncoso, Hypersymmetry bounds and three-dimensional higher-spin black holes. JHEP **08**, 021 (2015)
43. S. Mizoguchi, Determinant formula and unitarity for the $W(3)$ algebra. Phys. Lett. **B222**, 226 (1989)
44. H.K. Kunduri, J. Lucietti, H.S. Reall, Near-horizon symmetries of extremal black holes. Class. Quant. Grav. **24**, 4169–4190 (2007)
45. H.K. Kunduri, J. Lucietti, Classification of near-horizon geometries of extremal black holes. Living Rev. Rel. **16**, 8 (2013)
46. A. Castro, N. Iqbal, E. Llabrés, Wilson lines and Ishibashi states in AdS_3/CFT_2. JHEP **09**, 066 (2018)
47. A. Hegde, P. Kraus, E. Perlmutter, General results for higher spin Wilson lines and entanglement in Vasiliev theory. JHEP **01**, 176 (2016)

48. D.E. Berenstein, R. Corrado, W. Fischler, J.M. Maldacena, The Operator product expansion for Wilson loops and surfaces in the large N limit. Phys. Rev. **D59**, 105023 (1999)
49. U.H. Danielsson, E. Keski-Vakkuri, M. Kruczenski, Vacua, propagators, and holographic probes in AdS / CFT. JHEP **01**, 002 (1999)
50. H.L. Verlinde, Conformal field theory, 2-*D* quantum gravity and quantization of Teichmuller space. Nucl. Phys. **B337**, 652–680 (1990)
51. S. Elitzur, G.W. Moore, A. Schwimmer, N. Seiberg, Remarks on the canonical quantization of the Chern–Simons–Witten theory. Nucl. Phys. **B326**, 108–134 (1989)
52. E. Witten, Quantum field theory and the Jones polynomial. Commun. Math. Phys. **121**, 351 (1989)
53. K.B. Alkalaev, V.A. Belavin, Classical conformal blocks via AdS/CFT correspondence. JHEP **08**, 049 (2015)
54. E. Hijano, P. Kraus, E. Perlmutter, R. Snively, Semiclassical Virasoro blocks from AdS_3 gravity. JHEP **12**, 077 (2015)
55. A. Bhatta, P. Raman, N.V. Suryanarayana, Holographic conformal partial waves as gravitational open Wilson networks. JHEP **06**, 119 (2016)
56. K.B. Alkalaev, V.A. Belavin, Monodromic vs geodesic computation of Virasoro classical conformal blocks. Nucl. Phys. **B904**, 367–385 (2016)
57. M. Besken, A. Hegde, E. Hijano, P. Kraus, Holographic conformal blocks from interacting Wilson lines. JHEP **08**, 099 (2016)
58. A.L. Fitzpatrick, J. Kaplan, D. Li, J. Wang, Exact virasoro blocks from Wilson lines and background-independent operators. JHEP **07**, 092 (2017)
59. M. Besken, E. D'Hoker, A. Hegde, P. Kraus, Renormalization of gravitational Wilson lines. JHEP **06**, 020 (2019)
60. Y. Nakayama, H. Ooguri, Bulk local states and crosscaps in holographic CFT. JHEP **10**, 085 (2016)
61. A. Lewkowycz, G.J. Turiaci, H. Verlinde, A CFT perspective on gravitational dressing and bulk locality. JHEP **01**, 004 (2017)
62. N. Anand, H. Chen, A.L. Fitzpatrick, J. Kaplan, D. Li, An exact operator that knows its location. JHEP **02**, 012 (2018)
63. R. Nakayama, T. Suzuki, A bulk localized state and new holographic renormalization group flow in 3D Spin-3 gravity. Int. J. Mod. Phys. **A33**(12), 1850061 (2018)

Fundamental Physics, the Swampland of Effective Field Theory and Early Universe Cosmology

Robert Brandenberger

Abstract Cosmological inflation is not the only early universe scenario consistent with current observational data. I will discuss the criteria for a successful early universe cosmology, compare a couple of the proposed scenarios (inflation, bouncing cosmologies, and the *emergent* scenario), focusing on how future observational data will be able to distinguish between them. I will argue that we need to go beyond effective field theory in order to understand the early universe, and that principles of superstring theory will yield a non-singular cosmology.

Keywords Early universe cosmology · String theory

1 Introduction

In this talk I would like to convey three main messages. The first is that the *inflationary scenario* is not the only early universe scenario which is consistent with current observational data. The second message is that the inflationary scenario does not appear to naturally emerge from superstring theory. On a positive note (and this is the third message), there are arguments based on fundamental principles of superstring theory which indicate that the cosmology which emerges from string theory will be non-singular.

The two past decades have provided us with a wealth of data about the structure of the universe on large scales. From the point of view of Standard Big Bang cosmology most of the data cannot be explained. Why is the universe close to homogeneous and isotropic on scales which at the time of recombination had never been in causal contact? Why is the universe so close to being spatially flat? These are the famous *horizon* and *flatness* problems of Standard Big Bang cosmology. We now have detailed measurements of the small amplitude inhomogeneities in

R. Brandenberger (✉)
Physics Department, McGill University, Montreal, QC, Canada
e-mail: rhb@physics.mcgill.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_37

the distribution of matter and radiation, most spectacularly the high precision all sky maps of the cosmic microwave background (CMB) radiation [1]. The angular power spectrum of these anisotropies shows that the fluctuations are scale-invariant on large scales and are characterized by *acoustic oscillations* on smaller scales. What is the origin of these fluctuations?

The physics which yields the abovementioned acoustic oscillations in the angular power spectrum of CMB fluctuations was discussed in two pioneering papers [2, 3]. These authors assumed the existence of a roughly scale-invariant spectrum of curvature fluctuations on super-Hubble scales (the Hubble radius is $H^{-1}(t)$, where H is the Hubble expansion rate) at a time before recombination. These fluctuations are standing waves which are frozen in until the time when the Hubble radius becomes larger than the length scale of the fluctuations (in the linear regime fluctuations have constant wavelength in comoving coordinates; hence, in the matter-dominated epoch their physical wavelength grows as $t^{2/3}$ while the Hubble radius grows at the faster rate $\sim t$). After they enter they begin to oscillate. Modes which have performed an even (odd) number of half oscillations by the time of recombination yield maxima (local minima) in the power spectrum. The papers [2, 3] date back to 10 years before the development of inflationary cosmology. Both the CMB acoustic oscillations and the *baryon acoustic oscillations* in the power spectrum of matter fluctuations were predicted already then.

The question which was not addressed in [2, 3] is the origin of the super-Hubble fluctuations at early times. In Standard Big Bang cosmology the Hubble radius equals the horizon, and hence having super-Hubble fluctuations appears to be acausal. Inflationary cosmology [4] was the first scenario to propose an origin [5] for these fluctuations, but now we know that it is not the only one. In the following I will develop necessary criteria for an early universe scenario to be able to explain the near homogeneity of the universe and the origin of the observed cosmological perturbations. I will then introduce a couple of early universe scenarios which satisfy the criteria. In Sect. 3 I will turn to the question of which early universe scenario might emerge from superstring theory.

2 Early Universe Scenarios

The first criterion which a successful early universe scenario must satisfy is that the horizon (the radius of the forward light cone of a point on the initial value surface) is much larger than the Hubble radius $H^{-1}(t)$ at late times. This is necessary to be able to address the horizon problem of Standard Big Bang cosmology. In order to admit the possibility of a causal mechanism to generate the primordial fluctuations, comoving scales which are probed with current cosmological observations must originate inside the Hubble radius at early times. This is the second criterion. If the fluctuations emerge as quantum vacuum perturbations (as they are postulated to in inflationary cosmology), then scales we observe today must evolve for a long time on super-Hubble scales in order to obtain the squeezing of the fluctuations

which is necessary to obtain classical perturbations at late times (third criterion). Finally (fourth criterion), the structure formation scenario must produce a nearly scale-invariant spectrum of primordial perturbations (see e.g. [6] for a more detailed discussion).

Inflationary cosmology [4] is the first scenario which satisfies the four above criteria. During the time interval $t_i < t < t_R$ during which the universe undergoes nearly exponential expansion, the horizon expands exponentially while the Hubble radius remains almost unchanged. Since the physical length of a fixed comoving scale also expands nearly exponentially during the period of inflation, scales which we observe today originate inside the Hubble radius as long as the period of inflation is sufficiently long. Fluctuations are squeezed on super-Hubble scales for a long time, and the approximate time-translation symmetry of the inflationary phase ensures that the spectrum of primordial fluctuations is nearly scale-invariant [5, 7].

Bouncing cosmologies provide a second scenario in which the four criteria for a successful early universe scenario can be satisfied. In a bouncing scenario the horizon is infinite. The Hubble radius decreases during the period of contraction and then increases during the period of expansion. As long as the period of contraction is comparable in length to the period of Standard Big Bang expansion, scales which we observed today emerge from inside the Hubble radius, thus allowing a possible causal structure formation scenario. As in inflationary cosmology, there is a long period during which scales propagate with super-Hubble length, thus enabling the squeezing of the fluctuations. There are (at least) three classes of bouncing cosmologies. First, the *matter bounce* [8] in which there is a long phase of matter-dominated contraction. Second, there is the *Pre-Big-Bang* scenario [9] in which contraction is driven by a field with an equation of state $w = 1$, where w is the ratio of pressure to energy density. Finally, there is the *Ekpyrotic* scenario [10] in which contraction is obtained by means of a scalar field with equation of state $w \gg 1$. There is a duality between the evolution of curvature fluctuations in a matter-dominated phase of contraction and in an exponentially expanding background [11]. Hence, the matter bounce automatically leads to a roughly scale-invariant spectrum of fluctuations. There is a duality in the evolution of scalar field fluctuations between a cosmology with Ekpyrotic contraction and one of exponential expansion [12]. Hence, it is also possible to obtain a scale-invariant spectrum of fluctuations. In the case of the Pre-Big-Bang scenario it is possible to obtain a scale-invariant spectrum making use of axion fields [13]. See [14] for a detailed review of bouncing cosmologies. Ekpyrotic and Pre-Big-Bang cosmologies produce a steep blue spectrum of primordial gravitational waves. Hence, on cosmological scales the spectrum of primordial gravitational waves is predicted to be negligible. This contrasts with the predictions of inflationary models which forecast a roughly scale-invariant spectrum.

A third scenario for early universe cosmology is the *emergent scenario* which is based on the assumption that the universe emerged from an initial high density state in which matter was in global thermal equilibrium. One toy model for this is *String Gas Cosmology* [15] in which it is assumed that the universe loiters for a long time in a Hagedorn phase of a gas of fundamental strings, and there is a phase transition

to the expanding phase of Standard Big Bang cosmology (see e.g. [16] for a review). In the emergent scenario, the horizon is infinite, and scales which are observed today are trivially sub-Hubble in the emergent phase (since the Hubble radius is infinite in the limit that the emergent phase is static). As discovered in [17], the spectrum of cosmological perturbations originating from thermal fluctuations of the string gas is nearly scale-invariant. A prediction with which String Gas Cosmology can be distinguished from simple inflationary models is the tilt of the spectrum of primordial gravitational waves. Whereas inflationary models based on a matter content which satisfies the usual energy conditions predict a slight red tilt of the spectrum, String Gas Cosmology predicts a blue tilt n_t satisfying a consistency relation $n_t = n_s - 1$, where $n_s - 1$ is the tilt of the spectrum of curvature fluctuations [18].

None of the early universe scenarios discussed above is without problems. In the case of inflationary cosmology we can point to the *trans-Planckian problem* for fluctuations: if the period of inflation is much longer than the minimal period which inflation has to last in order to enable a causal generation mechanism of fluctuations, the length scale of all modes which are currently observed today was smaller than the Planck length at the beginning of inflation [19]. Thus, new physics must enter to give the initial conditions for the fluctuations.

As discussed in [20], the matter bounce scenario is not a local attractor in initial condition space: initial anisotropies blow up during the contracting phase. The Ekpyrotic scenario does better in this respect: initial anisotropies decay and the homogeneous Ekpyrotic contracting trajectory is a local attractor in initial condition space [21]. Note that in the case of *large field inflation*, the inflationary slow-roll trajectory is also a local attractor [22]. A key challenge for bouncing scenarios is that new physics is required to yield the cosmological bounce. An important problem for the emergent scenario is to obtain dynamical equations which describe the emergent phase.

Both inflationary and Ekpyrotic models are obtained in the context of Einstein gravity by taking the dominant component of matter to be given by a *scalar field* φ with a potential $V(\varphi)$. To obtain slow-roll inflation the potential has to be very flat

$$\frac{V'}{V} \lambda m_{pl}^{-1} , \tag{1}$$

where the prime indicates a derivative with respect to φ, and where m_{pl} is the Planck mass. For models free of an initial condition fine tuning problem the field φ must roll over a field range $|\Delta\varphi| > m_{pl}$ during inflation. In contrast, the Ekpyrotic scenario is based on scalar field matter with a negative and steep exponential potential. V'/V is large in Planck units, and the scalar field rolls a distance smaller than m_{pl}.

I highlight this point in connection with the constraints on effective field theories involving scalar fields which emerge from the considerations based on fundamental physics to be discussed in the following section.

3 Constraints from Fundamental Physics

The evolution of the very early universe should be described by the best available theory which describes physics at the highest energies. There is evidence that all forces of nature might unify at high energies. They must be described quantum mechanically. The best candidate for such a quantum theory is *superstring theory*. Superstring theory is based on the assumption that the basic building blocks of nature are not point particle, but rather elementary strings.

The quantum theory of point particles is quantum field theory. There is a huge *landscape* of quantum field theories: any number of space-time dimensions and fields is allowed, and any shape of field potentials. Superstring theory is very restrictive. The number of space-time dimensions is fixed, and the string interactions are constrained. At low energies, the physics emerging from superstring theory should be describable by an *effective field theory*.

However, the set of effective field theories compatible with string theory is constrained by what are known as the *swampland criteria*. Only theories consistent with these criteria are admissible. The vast number of field theories are not—they are said to lie in the *swampland* (see [23] for a review). Note that at the moment these criteria are not proven—they are educated guesses.

The first swampland criterion [24] is that the field range over which a given effective field theory is valid is constrained by $\Delta\varphi < \mathcal{O}(1)m_{pl}$. The second condition [25] is that, for a scalar field which is rolling and which dominates the energy density of the universe, the potential cannot be too flat:

$$\frac{V'}{V} > c_1 m_{pl}^{-1}, \tag{2}$$

where c_1 is a constant of order one. This condition is opposite to what is required for simple slow-roll inflation models. Hence, it appears that cosmological inflation is in tension with superstring theory. A corollary of the second swampland condition is that a cosmological phase dominated by a positive cosmological constant is not possible. Hence, Dark Energy cannot be a cosmological constant [26]. Scalar field models of Dark Energy are, however, consistent with (but constrained by) the swampland conditions [26, 27].

In light of the tension between inflationary cosmology and the principles of string theory it appears that we may need a new paradigm of early universe cosmology. Such a paradigm should be based on the key new degrees of freedom and symmetries which differentiate string theory from point particle theories. New degrees of freedom include the string oscillatory and winding modes. Let us for simplicity consider the background space to be toroidal. Strings on this space have momentum modes whose energies are quantized in units of $1/R$, where R is the radius of the torus, winding modes whose energies are quantized in units of R, and an tower of oscillatory modes whose energies are independent of R. Point particles only have momentum modes. If we consider a box of strings in thermal equilibrium

and compress the radius, then the temperature of the gas will initially increase since the energy of the momentum modes (which are the light modes for large values of R) increases. Eventually it becomes thermodynamically preferable to excite higher and higher energy oscillatory modes. The increase in temperature will level off: there is a maximal temperature of a gas of strings, the *Hagedorn temperature* T_H [28]. When R decreases below the string scale, the energy will flow into the winding modes (which are now the light modes), and the temperature will decrease. Hence [15], thermodynamic reasoning indicates that there is no temperature singularity in a stringy early universe cosmology.

String theory also features a new symmetry, *T-duality* symmetry. For a toroidal space, this implies that there is a symmetry between a space of radius R and a dual space of radius $1/R$ (in string units) obtained by interchanging the momentum and winding quantum numbers. As already argued in [15], the number of position operators in a quantum theory of strings must be doubled compared to a theory of point particles: there is one position operator which is the Fourier transform of the momentum

$$|x> = \sum_n |p>_n \,, \tag{3}$$

where $|p>_n$ is the eigenstate of momentum which quantum number n (n ranging over the integers), and a dual operator $|\tilde{x}>$ which is dual to the winding number eigenstates. Physical length l_p is measured in terms of $|x>$ if R is large, but in terms of $|\tilde{x}>$ if R is small. Hence, as R decreases from some large value towards zero, l_p remains finite (it is an even function of $\ln(R)$). This is another way to see the non-singularity of a stringy early universe cosmology.

The challenge for string cosmology remains to find consistent equations for the time-dependent cosmological background. Einstein gravity is not applicable since it is not consistent with the T-duality symmetry of string theory. In String Gas Cosmology [15] it was postulated that the universe emerges from a quasi-static initial Hagedorn phase. Such a phase could emerge from a better understanding of non-perturbative string theory. If we want to model the dynamics using an effective field theory, this effective field theory must live in double the number of spatial dimensions as the topological background contains in order to take into account both the $|x>$ and $|\tilde{x}>$ coordinates. A candidate for such a theory is *Double Field Theory* [29], a theory which is given by the action for a generalized metric in doubled space. The cosmology which results if we couple the Double Field Theory action for the background to "string gas matter" (matter which has an equation of state of radiative modes for large volumes of the $|x>$ space, and that of winding modes for a small volume) was recently analyzed in [30]. In this context it can be shown that the solutions in the string frame are non-singular.

4 Conclusions and Discussion

In the context of effective field theories of matter coupled to Einstein gravity, a number of early universe scenarios have been proposed which can explain current observational data. Inflationary cosmology is one of them, but not the only one. However, general considerations based on superstring theory indicate a tension between fundamental physics and inflation. In fact, they indicate that any approach based on effective field theory of matter coupled to Einstein gravity will break down in the early universe, and that we need a radically different approach which takes into account the new degrees of freedom and new symmetries which distinguish string theories from point particles theories. I presented a toy model which takes these aspects into account which indicates that the cosmology emerging from string theory will be non-singular, and that it may not include any phase of inflation.

References

1. N. Aghanim, et al. [Planck Collaboration], Planck 2018 results. VI. Cosmological parameters. arXiv:1807.06209 [astro-ph.CO]
2. R.A. Sunyaev, Y.B. Zeldovich, Small scale fluctuations of relic radiation. Astrophys. Space Sci. **7**, 3 (1970)
3. P.J.E. Peebles, J.T.Yu, Primeval adiabatic perturbation in an expanding universe. Astrophys. J. **162**, 815 (1970). https://doi.org/10.1086/150713
4. A. Guth, The inflationary universe: a possible solution to the horizon and flatness problems. Phys. Rev. D **23**, 347 (1981); R. Brout, F. Englert, E. Gunzig, The creation of the universe as a quantum phenomenon. Ann. Phys. **115**, 78 (1978); A.A. Starobinsky, A new type of isotropic cosmological models without singularity. Phys. Lett. B **91**, 99 (1980); K. Sato, First order phase transition of a vacuum and expansion of the universe. Mon. Not. Roy. Astron. Soc. **195**, 467 (1981)
5. V. Mukhanov, G. Chibisov, Quantum fluctuation and nonsingular universe (in Russian). JETP Lett. **33**, 532 (1981). [Pisma Zh. Eksp. Teor. Fiz. **33**, 549 (1981)]
6. R.H. Brandenberger, Cosmology of the very early universe. AIP Conf. Proc. **1268**, 3 (2010). https://doi.org/10.1063/1.3483879. [arXiv:1003.1745 [hep-th]]
7. W.H. Press, Spontaneous production of the Zel'dovich spectrum of cosmological fluctuations. Phys. Scr. **21**, 702 (1980). https://doi.org/10.1088/0031-8949/21/5/021
8. F. Finelli, R. Brandenberger, On the generation of a scale-invariant spectrum of adiabatic fluctuations in cosmological models with a contracting phase. Phys. Rev. D **65**, 103522 (2002). [arXiv:hep-th/0112249]
9. M. Gasperini, G. Veneziano, Pre-big bang in string cosmology. Astropart. Phys. **1**, 317 (1993). https://doi.org/10.1016/0927-6505(93)90017-8. [hep-th/9211021]
10. J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, The ekpyrotic universe: colliding branes and the origin of the hot big bang. Phys. Rev. D **64**, 123522 (2001). [hep-th/0103239]
11. D. Wands, Duality invariance of cosmological perturbation spectra. Phys. Rev. D **60**, 023507 (1999). [arXiv:gr-qc/9809062]
12. J. Khoury, B.A. Ovrut, P.J. Steinhardt, N. Turok, Density perturbations in the ekpyrotic scenario. Phys. Rev. D **66**, 046005 (2002). https://doi.org/10.1103/PhysRevD.66.046005. [hep-th/0109050]
13. E.J. Copeland, R. Easther, D. Wands, Vacuum fluctuations in axion - dilaton cosmologies. Phys. Rev. D **56**, 874 (1997). https://doi.org/10.1103/PhysRevD.56.874. [hep-th/9701082]

14. R. Brandenberger, P. Peter, Bouncing cosmologies: progress and problems. Found. Phys. **47**(6), 797 (2017). https://doi.org/10.1007/s10701-016-0057-0. [arXiv:1603.05834 [hep-th]]
15. R.H. Brandenberger, C. Vafa, Superstrings in the early universe. Nucl. Phys. B **316**, 391 (1989)
16. R.H. Brandenberger, String gas cosmology, in *String Cosmology*, ed. by J. Erdmenger (Wiley, Hoboken, 2009), pp. 193–230. [arXiv:0808.0746 [hep-th]]; T. Battefeld, S. Watson, String gas cosmology. Rev. Mod. Phys. **78**, 435 (2006). [arXiv:hep-th/0510022]
17. A. Nayeri, R.H. Brandenberger, C. Vafa, Producing a scale-invariant spectrum of perturbations in a Hagedorn phase of string cosmology. Phys. Rev. Lett. **97**, 021302 (2006). [arXiv:hep-th/0511140]
18. R.H. Brandenberger, A. Nayeri, S.P. Patil, C. Vafa, Tensor modes from a primordial hagedorn phase of string cosmology. Phys. Rev. Lett. **98**, 231302 (2007). https://doi.org/10.1103/PhysRevLett.98.231302. [hep-th/0604126]
19. J. Martin, R.H. Brandenberger, The TransPlanckian problem of inflationary cosmology. Phys. Rev. D **63**, 123501 (2001). https://doi.org/10.1103/PhysRevD.63.123501. [hep-th/0005209]
20. Y.F. Cai, R. Brandenberger, P. Peter, Anisotropy in a nonsingular bounce. Class. Quant. Grav. **30**, 075019 (2013). https://doi.org/10.1088/0264-9381/30/7/075019. [arXiv:1301.4703 [gr-qc]]
21. J.K. Erickson, D.H. Wesley, P.J. Steinhardt, N. Turok, Kasner and mixmaster behavior in universes with equation of state $w \gg 1$. Phys. Rev. **D69**, 063514 (2004). https://doi.org/10.1103/PhysRevD.69.063514. [arXiv:hep-th/0312009]
22. R.H. Brandenberger, J.H. Kung, Chaotic inflation as an attractor in initial condition space. Phys. Rev. D **42**, 1008 (1990). https://doi.org/10.1103/PhysRevD.42.1008
23. T.D. Brennan, F. Carta, C. Vafa, The string landscape, the swampland, and the missing corner. PoS TASI **2017**, 015 (2017). https://doi.org/10.22323/1.305.0015. [arXiv:1711.00864 [hep-th]]; E. Palti, The swampland: introduction and review (2019). arXiv:1903.06239 [hep-th]
24. H. Ooguri, C. Vafa, On the geometry of the string landscape and the swampland. Nucl. Phys. B **766**, 21 (2007). https://doi.org/10.1016/j.nuclphysb.2006.10.033. [hep-th/0605264]
25. G. Obied, H. Ooguri, L. Spodyneiko, C. Vafa, De Sitter space and the swampland (2018). arXiv:1806.08362 [hep-th]; H. Ooguri, E. Palti, G. Shiu, C. Vafa, Distance and de Sitter conjectures on the swampland (2018). arXiv:1810.05506 [hep-th]
26. P. Agrawal, G. Obied, P.J. Steinhardt, C. Vafa, On the cosmological implications of the string swampland. Phys. Lett. B **784**, 271 (2018). https://doi.org/10.1016/j.physletb.2018.07.040. [arXiv:1806.09718 [hep-th]]
27. L. Heisenberg, M. Bartelmann, R. Brandenberger, A. Refregier, Dark energy in the swampland. Phys. Rev. D **98**(12), 123502 (2018). https://doi.org/10.1103/PhysRevD.98.123502. [arXiv:1808.02877 [astro-ph.CO]]
28. R. Hagedorn, Statistical thermodynamics of strong interactions at high-energies. Nuovo Cim. Suppl. **3**, 147 (1965)
29. W. Siegel, Superspace duality in low-energy superstrings. Phys. Rev. D **48**, 2826 (1993). https://doi.org/10.1103/PhysRevD.48.2826. [hep-th/9305073]; C. Hull, B. Zwiebach, Double field theory. J. High Energy Phys. **0909**, 099 (2009). https://doi.org/10.1088/1126-6708/2009/09/099. [arXiv:0904.4664 [hep-th]]; G. Aldazabal, D. Marques, C. Nunez, Double field theory: a pedagogical review. Class. Quant. Grav. **30**, 163001 (2013). https://doi.org/10.1088/0264-9381/30/16/163001. [arXiv:1305.1907 [hep-th]]
30. R. Brandenberger, R. Costa, G. Franzmann, A. Weltman, Dual spacetime and nonsingular string cosmology. Phys. Rev. D **98**(6), 063521 (2018). https://doi.org/10.1103/PhysRevD.98.063521. [arXiv:1805.06321 [hep-th]]; R. Brandenberger, R. Costa, G. Franzmann, A. Weltman, T-dual cosmological solutions in double field theory. Phys. Rev. D **99**(2), 023531 (2019). https://doi.org/10.1103/PhysRevD.99.023531. [arXiv:1809.03482 [hep-th]]; H. Bernardo, R. Brandenberger, G. Franzmann, *T*-dual cosmological solutions in double field theory II. Phys. Rev. D **99**(6), 063521 (2019). https://doi.org/10.1103/PhysRevD.99.063521. [arXiv:1901.01209 [hep-th]]

Scale-Invariant Scalar Field Dark Matter Through the Higgs-Portal

Catarina Cosme

Abstract We introduce an oscillating scalar field coupled to the Higgs that can account for all dark matter in the Universe. Due to an underlying scale invariance of this model, the dark scalar only acquires mass after the electroweak phase transition. We discuss the dynamics of this dark matter candidate, showing that it behaves like dark radiation until the electroweak phase transition and like non-relativistic matter afterwards. In the case of a negative coupling to the Higgs field, the scalar gets a vacuum expectation value after the electroweak phase transition and may decay into photons although being sufficiently long-lived to account for dark matter. We show that, within this scenario, for a mass of 7 keV, the model can explain the observed galactic and extra-galactic 3.5 keV X-ray line. Nevertheless, it will be very difficult to probe this model in the laboratory in the near future. This proceedings paper is based on Refs. Cosme et al. (J High Energy Phys 1805, 129, 2018; Phys Lett B 781, 639, 2018).

Keywords Dark matter · Scalar field · Higgs boson

1 Introduction

Dark matter (DM) is one of the greatest unsolved questions in Physics. This invisible form of matter constitutes almost 27% of the Universe's energy density content and is required to explain its structure on large scales, the anisotropies in the Cosmic Microwave Background (CMB), and the galaxy rotation curves. Despite a large number of candidates that arise in theories beyond the Standard Model of Particle Physics (SM), the origin and the constitution of DM remain unknown. Although Weakly Interacting Massive Particles (WIMPs) are among the best-motivated thermally produced DM candidates, they have not been detected so

C. Cosme (✉)
Ottawa-Carleton Institute for Physics, Carleton University, Ottawa, ON, Canada
e-mail: ccosme@physics.carleton.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_38

far and the absence of new particles at the LHC motivates looking for alternatives to the WIMP paradigm.

In this work, we introduce an oscillating scalar field coupled to the Higgs as a dark matter candidate. Even though the Higgs-portal for dark matter has been explored in the context of thermal dark matter candidates (WIMPs), there are few proposals in the literature that investigate the case of a scalar field which is oscillating in the minimum of its quadratic potential, behaving like non-relativistic matter. Thus, we focus on a model where the oscillating scalar field dark matter obtains its mass only through the Higgs mechanism, i.e., through scale-invariant Higgs-portal interactions. We assume an underlying scale invariance of the theory, spontaneously broken by some mechanism that generates the Planck and the electroweak scales in the Lagrangian, but which forbids a bare mass term for the dark scalar. The scale invariance is maintained in the dark sector and, therefore, the dark scalar only gets mass after the electroweak phase transition (EWPT). Additionally, the model has a U(1) gauge symmetry which ensures the dark matter candidate stability if unbroken. The relevant interaction Lagrangian density is the following:

$$-\mathcal{L}_{int} = \pm g^2 |\Phi|^2 |\mathcal{H}|^2 + \lambda_\phi |\Phi|^4 + V(\mathcal{H}) + \xi R |\Phi|^2 \ , \tag{1}$$

where the Higgs potential, $V(\mathcal{H})$, has the usual "Mexican hat" shape, g is the coupling between the Higgs and the dark scalar, λ_ϕ is the dark scalar's self-coupling, and the last term in Eq. (1) corresponds to a non-minimal coupling of the dark matter field to curvature, where R is the Ricci scalar and ξ is a constant.

In this paper, we will focus on the case where the Higgs-dark scalar interaction has a negative sign. Hence, the U(1) symmetry may be spontaneously broken, which can lead to interesting astrophysical signatures, as we will see later.

This proceedings paper is structured as follows: in Sect. 2 , we describe the dynamics of the field from the inflationary period up to the EWPT. In Sect. 3 we discuss the behavior of the field after the EWPT, computing the present dark matter abundance. The phenomenology of this scenario is explored in Sect. 4 and the conclusions are summarized in Sect. 5. For more details and a complete list of references, see Refs. [1, 2].

2 Dynamics Before Electroweak Symmetry Breaking

In this section, we describe the evolution of the dark matter candidate before the EWPT, where the Higgs-portal coupling term has a negligible role. First, we discuss the dynamics of the dark scalar during the inflationary period, where the non-minimal coupling term dominates its behavior. Then, we examine the behavior of the field in the radiation era until the EWPT, where the self-interactions term drives the dark scalar dynamics.

2.1 Inflation

During inflation, in the regime where $\xi \gg g,\ \lambda_\phi$, the dynamics of the field is mainly driven by the non-minimal coupling to gravity in Eq. (1). This term provides an effective mass to the dark scalar, m_ϕ:

$$m_\phi \simeq \sqrt{12\,\xi}\, H_{inf}\ , \tag{2}$$

where we have used the fact that the Ricci scalar during inflation is $R \simeq 12\, H_{inf}^2$ and the Hubble parameter, written in terms of the tensor-to-scalar ratio r, reads:

$$H_{inf}\,(r) \simeq 2.5 \times 10^{13} \left(\frac{r}{0.01}\right)^{1/2} \text{GeV}\ . \tag{3}$$

Note that $m_\phi > H_{inf}$ for $\xi > 1/12$. Thus, although the classical field is driven towards the origin during inflation, its average value never vanishes due to de Sitter quantum fluctuations on super-horizon scales. Any massive field during inflation exhibits quantum fluctuations that get stretched and amplified by the Universe's expansion and, in particular, for $m_\phi/H_{inf} > 3/2$ ($\xi > 3/16$) the amplitude of each Fourier mode with comoving momentum k is given by Riotto [3]:

$$|\delta\phi_k|^2 \simeq \left(\frac{H_{inf}}{2\pi}\right)^2 \left(\frac{H_{inf}}{m_\phi}\right) \frac{2\pi^2}{\left(a\, H_{inf}\right)^3}\ , \tag{4}$$

where $a(t)$ is the scale factor. Integrating over the super-horizon comoving momentum $0 < k < a H_{inf}$, at the end of inflation, the homogeneous field variance reads:

$$\left\langle \phi^2 \right\rangle \simeq \frac{1}{3} \left(\frac{H_{inf}}{2\pi}\right)^2 \frac{1}{\sqrt{12\xi}}\ , \tag{5}$$

which sets the initial amplitude for field oscillations in the post inflationary era:

$$\phi_{inf} = \sqrt{\left\langle \phi^2 \right\rangle} \simeq \alpha\, H_{inf} \qquad \alpha \simeq 0.05\, \xi^{-1/4}\ . \tag{6}$$

After inflation, when $m_\phi \gg H$ is satisfied, the field oscillates about the minimum of its potential. Moreover, since $R = 0$ in a radiation-dominated era and $R \sim \mathcal{O}(H^2)$ in the following eras, we may neglect the effects of the non-minimal coupling term in the evolution of the field after inflation. Hence, we may conclude that the role of the non-minimal coupling to gravity is to make the field sufficiently heavy during inflation so to suppress potential isocurvature modes in the CMB anisotropy spectrum.

2.2 Radiation Era

After inflation and the reheating period (which we will assume to be instantaneous, for simplicity), the Universe becomes radiation-dominated and $R = 0$. Above the EWPT, the dominant term in the potential of the dark scalar is the quartic one (see Eq. (1)), since the thermal effects can keep the Higgs field localized about its origin. The dark scalar acquires an effective field mass $m_\phi = \sqrt{3\,\lambda_\phi}\,\phi$ and, when the condition $m_\phi \gg H$ is satisfied, it starts to oscillate about the origin with an amplitude ϕ_{rad} given by:

$$\phi_{rad}\,(T) = \frac{\phi_{inf}}{T_{rad}}\,T = \left(\frac{\pi^2\,g_*}{270}\right)^{1/4}\,\left(\frac{\phi_{inf}}{M_{Pl}}\right)^{1/2}\,\frac{T}{\lambda_\phi^{1/4}}\,. \tag{7}$$

where the temperature at the onset of fields oscillations, T_{rad}, reads

$$T_{rad} = \lambda_\phi^{1/4}\,\sqrt{\phi_{inf}\,M_{Pl}}\,\left(\frac{270}{\pi^2\,g_*}\right)^{1/4}\,, \tag{8}$$

g_* is the number of relativistic degrees of freedom and M_{Pl} is the reduced Planck mass. Since the dark scalar's amplitude decays as $a^{-1} \propto T$ and $\rho_\phi \sim a^{-4}$, we conclude that the field behaves like dark radiation during this period.

As soon as the temperature of the Universe drops below the electroweak scale, both the Higgs and the dark scalar fields acquire a vacuum expectation value (vev) and, consequently, the Higgs generate a mass for the dark scalar, as we will see in the next section. The electroweak phase transition is completed when the leading thermal contributions to the Higgs potential become Boltzmann suppressed, at approximately $T_{EW} \sim m_W$, where m_W is the W boson mass.

3 Dynamics After the Electroweak Symmetry Breaking

At the EWPT, the relevant interaction potential is

$$V(\phi, h) = -\,\frac{g^2}{4}\,\phi^2\,h^2 + \frac{\lambda_\phi}{4}\,\phi^4 + \frac{\lambda_h}{4}\,\left(h^2 - \tilde{v}^2\right)^2\,, \tag{9}$$

where the Higgs self-coupling is $\lambda_h \simeq 0.13$.

At this point, the Higgs and the dark scalar acquire a non-vanishing vev, respectively:

$$h_0 = \left(1 - \frac{g^4}{4\,\lambda_\phi\,\lambda_h}\right)^{-1/2}\,\tilde{v} \equiv \mathrm{v}, \qquad \phi_0 = \frac{g\,\mathrm{v}}{\sqrt{2\lambda_\phi}}\,, \tag{10}$$

where $\mathrm{v} = 246\,\mathrm{GeV}$. Notice that a non-vanishing vev for the dark scalar implies $g^4 < 4\lambda_\phi\lambda_h$, which we assume to hold. The mass of the dark scalar, which is generated only by the Higgs, is then:

$$m_\phi = g\,\mathrm{v}\,. \tag{11}$$

As pointed out in Refs. [1, 2], the dark scalar starts to oscillate about ϕ_0, with an amplitude $\phi_{DM} \equiv x_{DM}\,\phi_0$ with $x_{DM} \lesssim 1$ once the leading contributions to the Higgs potential become Boltzmann suppressed, below $T_{EW} \sim m_W$. This x_{DM} is not an extra parameter of the model, it is just a theoretical uncertainty that takes into account the evolution of the dark scalar during the electroweak crossover. Although a numerical simulation of the dynamics of the field during the electroweak crossover would be required, we can estimate the value of x_{DM}. Since $T_{EW} \lesssim T_{CO}$ by an $\mathcal{O}(1)$ factor, where T_{CO} corresponds to the electroweak crossover temperature, and given that $\phi \sim T$ while behaving as radiation and $\phi \sim T^{3/2}$ while behaving as non-relativistic matter, the field's amplitude might decrease by at most an $\mathcal{O}(1)$ factor as well. For more details, see Refs. [1, 2]. Hence, we may conclude that the field smoothly changes from dark radiation to a cold dark matter behavior at the EWPT, as its potential becomes quadratic about the minimum.

As soon as the dark scalar starts to behave like cold dark matter, its amplitude evolves with the temperature as $\phi(T) = \phi_{DM}(T/T_{EW})^{3/2}$ and the number of particles in a comoving volume, $\frac{n_\phi}{s}$, becomes constant:

$$\frac{n_\phi}{s} = \frac{45}{4\pi^2 g_{*S}}\frac{m_\phi \phi_{DM}^2}{T_{EW}^3}\,, \tag{12}$$

where $g_{*S} \simeq 86.25$ is the number of relativistic degrees of freedom contributing to the entropy at T_{EW}, $s = \frac{2\pi^2}{45}\,g_{*S}\,T^3$ is the entropy density of radiation, and $n_\phi \equiv \frac{\rho_\phi}{m_\phi}$ is the dark matter number density. We can use this to compute the present DM abundance, $\Omega_{\phi,0} \simeq 0.26$, obtaining the following relation for the field's mass:

$$m_\phi = \left(6\,\Omega_{\phi,0}\right)^{1/2}\left(\frac{g_{*S}}{g_{*S0}}\right)^{1/2}\left(\frac{T_{EW}}{T_0}\right)^{3/2}\frac{H_0 M_{Pl}}{\phi_{DM}}, \tag{13}$$

where g_{*S0}, T_0, and H_0 are the present values of the number of relativistic degrees of freedom, CMB temperature, and Hubble parameter, respectively. Then, plugging Eq. (11) into Eq. (13), we find a relation between g and λ_ϕ:

$$g \simeq 2\times 10^{-3}\left(\frac{x_{DM}}{0.5}\right)^{-1/2}\lambda_\phi^{1/4}\,. \tag{14}$$

This relation is a key point of our model: essentially, it has only a single free parameter, which we take to be the mass of the field. We will come back to this when discussing the phenomenology of the model.

The idea of this work is to introduce a dark matter candidate which is never in thermal equilibrium with the cosmic plasma. However, there are two main processes that can lead to the evaporation of the condensate. One of them is the Higgs annihilation into higher-momentum ϕ particles, which is prevented if [1, 2]

$$g \lesssim 8 \times 10^{-4} \left(\frac{g_*}{100}\right)^{1/8} . \tag{15}$$

The other process is the production of ϕ particles from the coherent oscillations of the background condensate in a quartic potential, which is not efficient if [1, 2]

$$\lambda_\phi < 6 \times 10^{-10} \left(\frac{g_*}{100}\right)^{1/5} \left(\frac{r}{0.01}\right)^{-1/5} \xi^{1/10} . \tag{16}$$

If the constraints of Eqs. (15) and (16) are satisfied, the dark scalar is never in thermal equilibrium with the cosmic plasma, behaving like an oscillating condensate of zero-momentum particles throughout its cosmic history. Equation (16) yields the most stringent constraint on the model, limiting the viable dark matter mass to be $m_\phi \lesssim 1\,\text{MeV}$ [1, 2].

4 Phenomenology

In this section, we will discuss two possible ways of probing the proposed model. For more examples and a complete and detailed discussion, see Refs. [1, 2].

4.1 Dark Matter Decay

Since the dark scalar and the Higgs field are coupled, they exhibit a small mass mixing, $\epsilon = \frac{g^2 \phi_0 \text{ v}}{m_h^2}$ [1]. This means that the dark scalar can decay into the same decay channels as the Higgs, provided that they are kinematically accessible. Due to the mass restriction coming from Eq. (16), which translates into $m_\phi \lesssim 1\,\text{MeV}$, the only kinematically accessible decay channel is the decay into photons. It is possible to show that the decay width of the dark matter candidate into photons is suppressed by a factor ϵ^2 with respect to the decay width of a virtual Higgs boson into photons, yielding for the dark scalar's lifetime [1, 2]:

$$\tau_\phi \simeq 7 \times 10^{27} \left(\frac{7\,\text{keV}}{m_\phi}\right)^5 \left(\frac{x_{DM}}{0.5}\right)^2 \text{s}. \tag{17}$$

Hence, although the lifetime is much larger than the age of the Universe, it can lead to an observable monochromatic line in the spectrum of galaxies and galaxy clusters.

Recently, the XMM-Newton X-ray observatory detected a line at 3.5 keV in the Galactic Center, Andromeda, and Perseus cluster [4–7]. The nature of this line has arisen some interest in the scientific community, leading to several interesting proposals in the literature, in particular, the possibility of it resulting from DM decay or annihilation [7–12]. In fact, the analysis in Refs. [6, 13] has shown that the intensity of the line observed in the astrophysical systems mentioned above could be explained by the decay of a DM particle with a mass of $\simeq 7$ keV and a lifetime in the range $\tau_\phi \sim (6-9) \times 10^{27}$ s. In the case of our dark scalar field model, fixing the field mass to this value, we predict a DM lifetime exactly in this range, up to some uncertainty in the value of the field amplitude after the EWPT parametrized by $x_{DM} \lesssim 1$. This is illustrated in Fig. 1. Notice that, for this mass, $g \simeq 3 \times 10^{-8}$ and $\lambda_\phi \simeq 4 \times 10^{-20}$, satisfying the constraints in Eqs. (15) and (16).

The uniqueness of this result should be emphasized: our model predicts that the decay of the dark scalar ϕ into photons produces a 3.5 keV line compatible with the observational data, with effectively only one free parameter: either g or λ_ϕ. Recall that, originally, the model involves four parameters—the couplings g and λ_ϕ, the non-minimal coupling ξ, and the scale of inflation r. The role of ξ is simply to suppress the potential cold dark matter isocurvature perturbations, while r only sets the initial amplitude of the field at the beginning of the radiation era. At the EWPT, the field starts to oscillate around ϕ_0, with an initial amplitude of this order—which does not depend on ξ nor r. So, when the dark scalar starts to behave effectively as cold dark matter, only g and λ_ϕ affect its dynamics. Therefore, we have three observables that rely on just two parameters (g and λ_ϕ)—the present dark matter

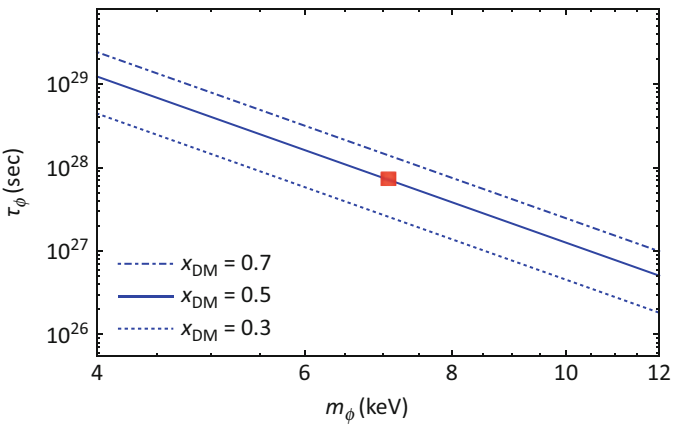


Fig. 1 Lifetime of the scalar field dark matter as a function of its mass, for different values of $x_{DM} \lesssim 1$. The horizontal red band corresponds to the values of τ_ϕ that can account for the 3.5 keV X-ray line detected by XMM-Newton for a mass around 7 keV. From [2]

abundance, the dark scalar's mass, and its lifetime. Fixing the present dark matter abundance, we get a relation between g and λ_ϕ (Eq. (14)), implying that m_ϕ and τ_ϕ depend exclusively on the Higgs-portal coupling. Hence, the prediction for the magnitude of the 3.5 keV line in different astrophysical objects is quite remarkable and, as far as we are aware, it has not been achieved by other scenarios, where the dark matter's mass and lifetime can be tuned by different free parameters.

4.2 Invisible Higgs Decays into Dark Scalars

One way to probe the Higgs-portal scalar field dark matter is to look for invisible Higgs decays into dark scalar pairs. The corresponding decay width is

$$\Gamma_{h\to\phi\phi} = \frac{1}{8\pi}\,\frac{g^4 \mathrm{v}^2}{4\,m_h}\,\sqrt{1-\frac{4m_\phi^2}{m_h^2}}\,, \tag{18}$$

where m_h is the Higgs mass. Assuming the upper limit for the dark matter mass, $m_\phi = 1\,\mathrm{MeV}$, the bound on the branching ratio is

$$Br\,(\Gamma_{h\to inv}) < 10^{-19}\,. \tag{19}$$

Considering that the current experimental limit is

$$Br\,(\Gamma_{h\to inv}) = \frac{\Gamma_{h\to inv}}{\Gamma_h + \Gamma_{h\to inv}} \lesssim 0.23\,, \tag{20}$$

where we assume that $\Gamma_{h\to inv} = \Gamma_{h\to\phi\phi}$, we conclude that this process is too small to be measured with current technology. However, it may serve as motivation for extremely precise measurements of the Higgs boson's width in future collider experiments, given any other experimental or observational hints for light Higgs-portal scalar field dark matter, such as, for instance, the 3.5 keV line that we have discussed earlier.

5 Conclusions

In this proceedings paper, we summarize the results of Refs. [1, 2], where we have shown that an oscillating scalar field coupled to the Higgs boson is a viable DM candidate that can explain the observed 3.5 keV X-ray line. This is a simple model, based on the assumed scale invariance of DM interactions, and, at the same time, extremely predictive, with effectively only a single free parameter upon fixing the present DM abundance. Hence, our scenario predicts a 3.5 keV X-ray line with the

observed properties for the corresponding value of the DM mass although it will be very difficult to probe it in the laboratory in the near future.

References

1. C. Cosme, J.G. Rosa, O. Bertolami, J. High Energy Phys. **1805**, 129 (2018). https://doi.org/10.1007/JHEP05(2018)129. [arXiv:1802.09434 [hep-ph]]
2. C. Cosme, J.G. Rosa, O. Bertolami, Phys. Lett. B **781**, 639 (2018). https://doi.org/10.1016/j.physletb.2018.04.062. [arXiv:1709.09674 [hep-ph]]
3. A. Riotto, ICTP Lect. Notes Ser. **14**, 317 (2003). [hep-ph/0210162]
4. E. Bulbul, M. Markevitch, A. Foster, R.K. Smith, M. Loewenstein, S.W. Randall, Astrophys. J. **789**, 13 (2014). https://doi.org/10.1088/0004-637X/789/1/13. [arXiv:1402.2301 [astro-ph.CO]]
5. A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, J. Franse, Phys. Rev. Lett. **113**, 251301 (2014). https://doi.org/10.1103/PhysRevLett.113.251301. [arXiv:1402.4119 [astro-ph.CO]]
6. A. Boyarsky, J. Franse, D. Iakubovskyi, O. Ruchayskiy, Phys. Rev. Lett. **115**, 161301 (2015). https://doi.org/10.1103/PhysRevLett.115.161301. [arXiv:1408.2503 [astro-ph.CO]]
7. N. Cappelluti, et al., (2017). arXiv:1701.07932 [astro-ph.CO]
8. T. Higaki, K.S. Jeong, F. Takahashi, Phys. Lett. B **733**, 25 (2014). https://doi.org/10.1016/j.physletb.2014.04.007. [arXiv:1402.6965 [hep-ph]]
9. J. Jaeckel, J. Redondo, A. Ringwald, Phys. Rev. D **89**, 103511 (2014). https://doi.org/10.1103/PhysRevD.89.103511. [arXiv:1402.7335 [hep-ph]]
10. E. Dudas, L. Heurtier, Y. Mambrini, Phys. Rev. D **90**, 035002 (2014). https://doi.org/10.1103/PhysRevD.90.035002. [arXiv:1404.1927 [hep-ph]]
11. F.S. Queiroz, K. Sinha, Phys. Lett. B **735**, 69 (2014). https://doi.org/10.1016/j.physletb.2014.06.016. [arXiv:1404.1400 [hep-ph]]
12. J. Heeck, D. Teresi, (2017). arXiv:1706.09909 [hep-ph]
13. O. Ruchayskiy, et al., Mon. Not. Roy. Astron. Soc. **460**(2), 1390 (2016). https://doi.org/10.1093/mnras/stw1026. [arXiv:1512.07217 [astro-ph.HE]]

The Moduli Portal to Dark Matter Particles

Maíra Dutra

Abstract The out-of-equilibrium production of dark matter (DM) from standard model (SM) species in the early universe (freeze-in mechanism) is expected in many scenarios in which very heavy beyond the SM fields act as mediators. In this conference, I have talked about the freeze-in of scalar, fermionic, and vector DM through the exchange of moduli fields (Chowdhury et al. Phys Rev D 99(9):095028, 2019), which are in the low-energy spectrum of many extra-dimensions and string theory frameworks. We have shown that the high temperature dependencies of the production rate densities in this model, as well as the possibility of having moduli masses at the post-inflationary reheating scale, make it crucial to consider the contribution of the freeze-in prior the start of the standard radiation era for a correct prediction of the DM relic density.

Keywords Dark matter · Moduli fields · Freeze-in production

1 Introduction

The close relationship between the couplings dark matter (DM) particles might have to particles belonging to the Standard Model of Particle Physics (SM) and their evolution through the early universe makes the DM puzzle an open problem in the interface between particle physics and cosmology.

Direct detection searches seek to detect nuclear or electronic recoils from DM scatterings, and the current status of no positive signals but more and more sensitive detectors mean that the SM-DM couplings need to be weaker and weaker. However, very weak couplings might imply that DM and SM particles were never in thermal equilibrium in the early universe. The out-of-equilibrium production of DM from SM species in the early universe, which is the so-called *freeze-in mechanism* [1–3], is expected in many scenarios in which very heavy beyond the SM (BSM) fields act

M. Dutra (✉)
Carleton University, Ottawa, ON, Canada
e-mail: mdutra@physics.carleton.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_39

as mediators. If the masses of these heavy mediators are close to the scale of the post-inflationary reheating, it is important to take into account the freeze-in during a reheating period in which entropy is being injected into the thermal bath. Here we shed light on this matter, presenting handy formulae that can be useful for any scenario.

Moduli fields are scalars which would be present in the effective limit of many string theory frameworks. Since they would need to be very feebly coupled to SM fields, it is interesting to investigate whether their feeble interactions with dark and visible matter would be enough to produce dark matter via freeze-in. In this conference, I have presented a recent study on the moduli portal to dark matter [4].

2 The Model

We have considered a complex modulus field, $\mathcal{T} = t + ia$, whose presence at temperatures below some cut-off scale Λ would appear as corrections to the free Lagrangians.[1] We consider BSM scalar, fermionic, and vector fields as feebly interacting massive particle (FIMP) candidates, which are DM candidates produced via freeze-in. Our effective Lagrangian connecting the modulus field to dark and standard scalars, fermions, and vectors, here generically denoted by Φ, Ψ, and X_μ, reads

$$\mathcal{L}_{eff} \supset \begin{cases} \left(1 + \frac{\alpha_i}{\Lambda} t\right) \left(|D_\mu \Phi|^2 - \mu_\Phi^2\right) & \text{(scalars)} \\ \frac{1}{2}\left(1 + \frac{\alpha^i_{L,R}}{\Lambda} t + i \frac{\beta^i_{L,R}}{\Lambda} a\right) \bar{\Psi}_{L,R} i \not{D} \Psi_{L,R} & \text{(fermions)} \\ -\frac{1}{4}\left(1 + \frac{\alpha_i}{\Lambda} t\right) X_{\mu\nu} X^{\mu\nu} - \frac{\beta_i}{\Lambda} a\, X_{\mu\nu} \tilde{X}^{\mu\nu} & \text{(vectors).} \end{cases} \tag{1}$$

Couplings to the real and imaginary components of the modulus are denoted, respectively, by α and β, and the tensors $X_{\mu\nu}$ and $\tilde{X}_{\mu\nu}$ are, respectively, the field strength and dual field strength of X_μ.

In order to avoid imaginary contributions to the mass and kinetic terms of the scalars, we assume that only the real component of the modulus interacts with the SM Higgs and the scalar FIMP candidate. In the case of the scalar FIMP, though, we do not assume that the modulus changes the mass term. The interactions with fermions are in principle chiral and a chirality flip would give us explicit dependence on the fermion mass in the amplitudes. For this reason, SM fermions cannot produce the FIMPs above EWSB, since they are massless. For the interactions of moduli with vectors, we have a Higgs-like operator for the real component and a Peccei–Quinn operator for the axial component.

As we are going to emphasize in the next section, the squared amplitudes of the freeze-in processes give us valuable information about the freeze-in temperature. In

[1] We here consider corrections up to the first order in the cut-off scale Λ.

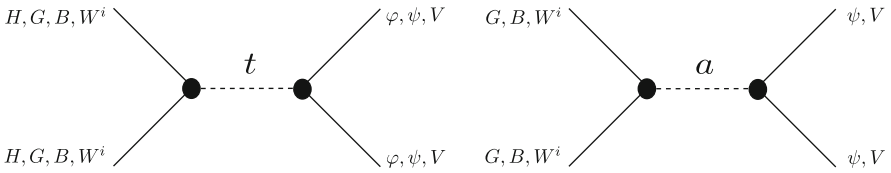


Fig. 1 Schematic Feynman diagrams leading to the freeze-in production of our FIMP candidates, scalar (ϕ), fermionic (ψ) and vector (V_μ) fields, out of annihilations of all the SM bosons (H, G^a_μ, B_μ, W^i_μ) via exchange of the real (t) and axial (a) components of a modulus field

our case, the FIMP candidates are produced from s-channel annihilations of Higgs bosons and SM gauge bosons, having t and a as mediators, as depicted in Fig. 1. The squared amplitudes are given by:

$$|\mathcal{M}|^2_0 = \frac{\alpha^2_{\rm DM}\lambda(s,\alpha_i)}{\Lambda^4}\frac{s^4\left(1-\frac{2m^2_{\rm DM}}{s}\right)^2}{(s-m_t^2)^2+m_t^2\Gamma_t^2} \tag{2}$$

$$|\mathcal{M}|^2_{1/2} = \frac{\alpha^2_{\rm DM}\lambda(s,\alpha_i)}{\Lambda^4}\frac{m^2_{\rm DM}s^3\left(1-\frac{4m^2_{\rm DM}}{s}\right)}{(s-m_t^2)^2+m_t^2\Gamma_t^2} + \frac{\beta^2_{\rm DM}\lambda(s,\beta_i)}{\Lambda^4}\frac{m^2_{\rm DM}s^3}{(s-m_a^2)^2+m_a^2\Gamma_a^2} \tag{3}$$

$$|\mathcal{M}|^2_{1} = \frac{\alpha^2_{\rm DM}\lambda(s,\alpha_i)}{\Lambda^4}\frac{s^4\left(1-\frac{4m^2_{\rm DM}}{s}+\frac{6m^4_{\rm DM}}{s^2}\right)}{(s-m_t^2)^2+m_t^2\Gamma_t^2} + \frac{\beta^2_{\rm DM}\lambda(s,\beta_i)}{\Lambda^4}\frac{s^4\left(1-\frac{4m^2_{\rm DM}}{s}\right)}{(s-m_a^2)^2+m_a^2\Gamma_a^2}\,. \tag{4}$$

Above, $\lambda(s,\alpha_i)$, $\lambda(s,\beta_i)$ are the sums over Higgs and SM gauge bosons contributions, which can be a function of the Mandelstam variable s. For further details, see [5].

3 Results

3.1 *Production Rate Densities and Evolution of FIMP Relic Density*

The freeze-in temperature is determined by the interactions between FIMPs and the SM species involved. In particular, the temperature dependence of the production rate densities tells us if the freeze-in happens at the lowest or highest scale of a given cosmological period. This is what I want to emphasize in this section.

The rate at which the number of DM particles change in a comoving volume a^3, with a the scale factor, is given by the Boltzmann fluid equation $\dot{N}_{\rm DM} = R_{\rm DM}(t)a^3$, with $N_{\rm DM} = n_{\rm DM}a^3$ the total number of DM particles and $R_{\rm DM}(t)$ the time/temperature-dependent interaction rate density, which in the case of the freeze-in only account for production and not for loss of FIMPs. On the other hand, the Hubble rate $H(t)$ determines how the scale factor varies with time, $H(t) = \dot{a}/a$, and since this quantity is proportional to the total energy density of the universe, different species dominating the expansion lead to different final total number of DM particles, as well as different time-temperature relations.

For a $12 \to 34$ process, the production rate density of species 3, in the limiting case where species 1 and 2 have Maxwell–Boltzmann distributions, is given by Dutra [5]

$$R_3^{12\to 34} \equiv n_1^{eq} n_2^{eq} \langle \sigma v \rangle = \frac{\mathcal{S}_{12}\mathcal{S}_{34}}{32(2\pi)^6} \int ds \frac{\sqrt{\lambda(s, m_3^2, m_4^2)}}{s} \int d\Omega_{13} |\mathcal{M}|^2 \times \frac{2T}{\sqrt{s}} \sqrt{\lambda(s, m_1^2, m_2^2)} K_1\left(\frac{\sqrt{s}}{T}\right), \tag{5}$$

In Fig. 2, I show the production rate densities of a scalar, fermionic, and vector FIMPs (blue, green, and red curves, respectively), as functions of the inverse of temperature. Generic features of the production rates are that, the higher the temperature, the more DM is produced. Also, notice that the rates start to vanish when the temperature of the thermal bath becomes smaller than the DM mass,

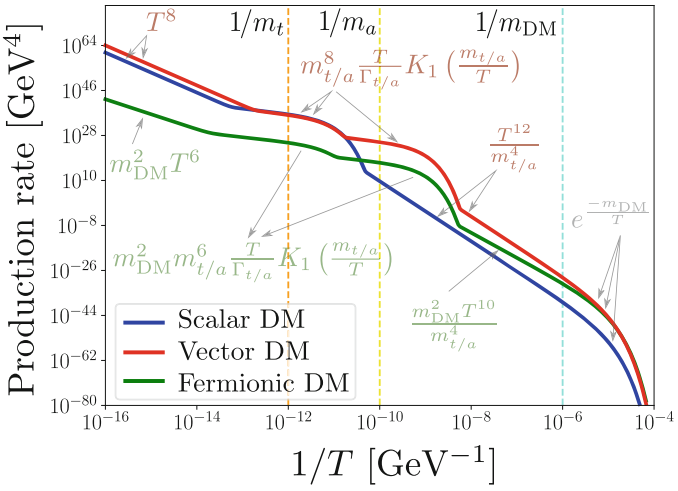


Fig. 2 Production rate densities of our FIMP candidates as functions of the inverse of the SM bath temperature. We indicate by dashed lines where T equals m_t, m_a and $m_{\rm DM}$, making it possible to identify when the on-shell production of mediators and the Boltzmann suppression occurs. Also pointed out are the approximate temperature dependencies of the rates

which is the famous Boltzmann suppression. We can recognize the presence of the poles of both components of the modulus, when the temperature of the thermal bath equals their masses. We can also notice the weaker temperature dependence of the production rate of a fermionic FIMP, due to the chirality flip. For completeness, I point out in Fig. 2 the analytic approximations of the production rates in the limiting cases where the mediators are much lighter, of the same order and much heavier than the temperature of the SM thermal bath.

As previously stated, the relic density of DM depends on which species dominates the cosmic expansion. While it is usual to assume that DM production happened during a radiation-dominated era, in which $H(T) \propto T^2$, this might not be the case in general. In inflationary theories, the universe had undergone a period of entropy production called *reheating* in which it cools down slower and $H(T) \propto T^4$. Such a period would happen from a moment when the temperature of the SM bath reaches a maximal value $T_{\rm MAX}$ up to the moment in which there is no more entropy production, the so-defined reheat temperature $T_{\rm RH}$. We do not know the scale of $T_{\rm RH}$, which could be as low as 4×10^{-3} GeV [6] and as high as 7×10^{15} GeV [7]. A general study of freeze-in through heavy portals should therefore take into account the possibility that the masses of mediators are at the reheating scale. So, as long as we have a thermal bath of SM radiation, it starts producing DM. In this context, the relic density of dark matter today receives a contribution from the reheating period and from the radiation era [5]:

$$\Omega^0_{\rm DM} h^2 = \frac{m_{\rm DM}}{2.16 \times 10^{-28}} \Big(\int_{T_0}^{T_{\rm RH}} dT \, \frac{g_s^*}{g_s \sqrt{g_e}} \frac{R_{\rm DM}(T)}{T^6} + 1.6\, c\, B_\gamma \, g_{\rm RH}^{-3/2} \, T_{\rm RH}^7 \int_{T_{\rm RH}}^{T_{\rm MAX}} dT \, g_e^* \frac{R_{\rm DM}(T)}{T^{13}} \Big) . \tag{6}$$

It is easy to see from the equation above that if $R_{\rm DM} \propto T^n$ for $n < 5$, the production during radiation era is infrared (happens at the lightest scale available) and if $n > 5$, it is ultraviolet (happens at the highest scale available, $T > T_{\rm RH}$). The same is true for the production during reheating, where the power of temperature is $n = 12$. This analysis is valid for an inflation which behaves as a matter content, with $w = 0$. For the generic case where the inflation's potential lead to $w \neq 0$, we refer the reader to Ref. [8]. Another quantity which depends on the specific inflationary reheating model is the ratio $T_{\rm MAX}/T_{\rm RH}$, here set to 100. The larger this ratio, the stronger the dilution due to inflation decay, which means the larger the production rate for a given final value of relic density.

From the temperature dependence of the production rates pointed out in Fig. 2, we can understand that the freeze-in of all the FIMP candidates happens at the highest scale of radiation era, the reheating temperature ($n > 5$), except for the case where the moduli are produced on-shell, which can make the relic density raises again after levelling-off. Neglecting the production during reheating in such a model leads to an underestimation in the relic density of many orders of magnitude, as it was explicitly shown in the case of an on-shell exchange of spin-2 fields [9].

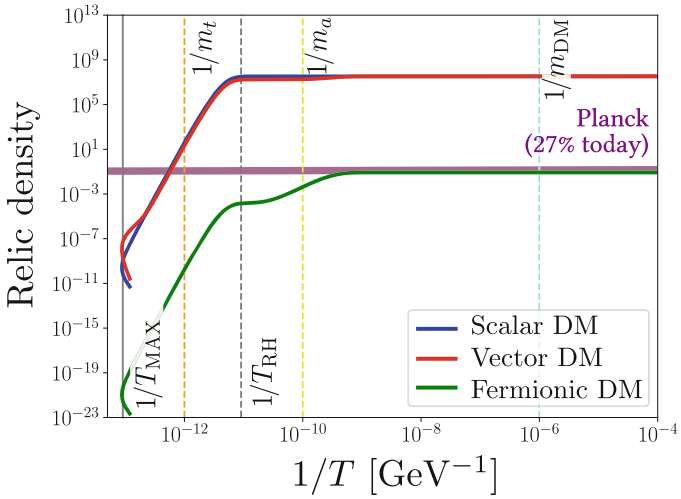


Fig. 3 Evolution of the relic density of our FIMP candidates for the same parameters of Fig. 2. This comes from the solution of Eq. (7) coupled to the evolution of the inflation and radiation contents

Such an underestimation is strongly dependent on the masses of the mediators as well as on the ratio $T_{\rm MAX}/T_{\rm RH}$.

In Fig. 3, we see the resulting evolution of the relic density, for the same set of free parameters of Fig. 2. These are the solutions of the coupled set of Boltzmann fluid equations for the evolution of DM, SM radiation, and inflation (driving the reheat period). More specifically, the DM yield $Y_{\rm DM} = N_{\rm DM}/S$ evolves as

$$\frac{dY_{\rm DM}}{dA} = \frac{R_{\rm DM}(A)}{As(A)H(A)} - \frac{Y_{\rm DM}}{S}\frac{dS}{dA}\,, \tag{7}$$

with $A \equiv aT_{\rm RH}$ the dimensionless scale factor and s the entropy density. It is the second term which couples this equation to the evolution of the energy density of the radiation and inflation contents [5].

We have fixed $T_{\rm RH} = 10^{11}$ GeV and $T_{\rm MAX} = 10^{13}$ GeV. In the presence of an on-shell production of a mediator, the production of DM is enhanced and that is why we see the relic density of vector and fermionic DM getting enhanced close to the pole of the axial modulus. Of course, the final relic density needs to agree to the Planck results, as I am going to show in the next section.

3.2 Agreement with Planck Results

We can now see the values of the new physics scale Λ and FIMP mass providing a good relic density of DM today, as inferred by the Planck satellite [10]. In Fig. 4, the reheating and maximal temperatures are set to 10^{10} and 10^{12} GeV. Since the relic

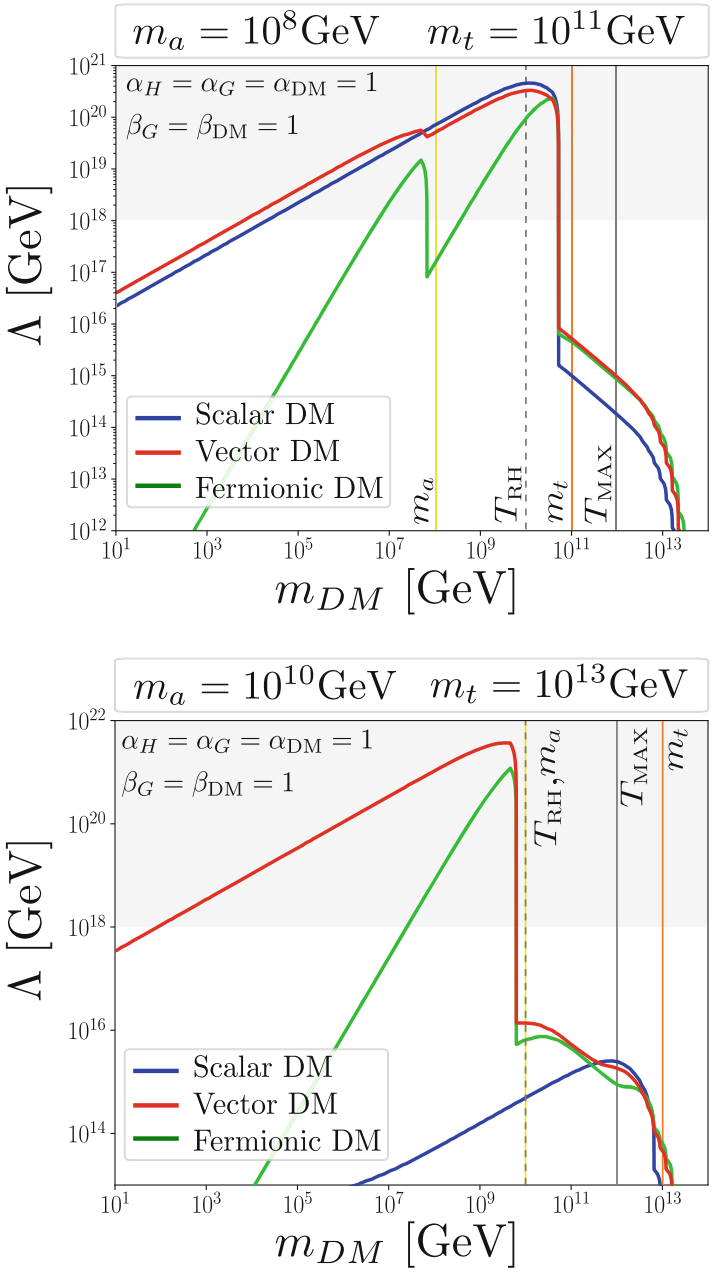


Fig. 4 Contours of good relic density in our free parameter slice comprised by the DM mass ($m_{\rm DM}$) and the cut-off scale (Λ), which needs to be the highest scale available in our framework. The enhancements in the production rates due to on-shell exchange of mediators need to be compensated as to provide the same relic density value. In this slice, the rates can be lowered by higher values of Λ. The gray region, corresponding to Λ above the Planck scale, is of course no meaningful but shown for an overview of the contours' behavior. We can see that the freeze-in is able to produce the right amount of dark matter for moduli mediator and reheating scale at intermediate scales

density increases with DM mass, we need smaller overall couplings in order to not overproduce the FIMPs and therefore Λ needs to be raised for a given relic density value. Also, due to the exponential Boltzmann suppression, the thermal bath cannot produce dark matter much before the time of maximal temperature and Λ is sharply lowered as to compensate the suppressed rates. In the upper panel, the pole of the axial modulus is reached inside the radiation era. Since it enhances the relic densities of vector and fermionic DM candidates, we need higher values of Λ. Notice that the curves for the fermionic DM depend strongly on the DM mass, due to the chirality flip. In the lower panel, the exchange of a heavier real modulus suppresses more the relic densities, so that we can have lower values of Λ, but still at intermediate scales.

4 Concluding Remarks

In this conference, I have discussed the case in which heavy moduli fields exchange between visible and dark matter are the underlying physics of the feeble couplings necessary for the freeze-in to happen. Such fields appear in many structural extensions of the SM, and our results are expected to be embedded in more realistic realizations. We have seen that if the temperature dependencies of the production rates of FIMPs are strong enough, which can be achieved in effective models with derivative couplings, FIMPs would have already been produced at the start of the radiation era. As an interesting outcome, in a wide range of our parameter space a good relic density is "naturally" achieved for scalar, fermionic, and vector FIMP candidates with the moduli masses at intermediate scales, and for reasonable scales of new physics.

Acknowledgments I want to thank the co-authors of the work presented in this conference, Debtosh Chowdhury, Emilian Dudas, and Yann Mambrini. I also acknowledge the support from the Brazilian PhD program "Ciências sem Fronteiras"-CNPQ Process No. 202055/2015-9 during the development of this work and the current support of the Arthur B. McDonald Canadian Astroparticle Physics Research Institute.

References

1. D.J.H. Chung, E.W. Kolb, A. Riotto, Production of massive particles during reheating. Phys. Rev. D **60**, 063504 (1999). https://doi.org/10.1103/PhysRevD.60.063504. [hep-ph/9809453]
2. L.J. Hall, K. Jedamzik, J. March-Russell, S.M. West, Freeze-in production of FIMP dark matter. J. High Energy Phys. **1003**, 080 (2010). https://doi.org/10.1007/JHEP03(2010)080. [arXiv:0911.1120 [hep-ph]]
3. N. Bernal, M. Heikinheimo, T. Tenkanen, K. Tuominen, V. Vaskonen, The dawn of FIMP dark matter: a review of models and constraints. Int. J. Mod. Phys. A **32**(27), 1730023 (2017). https://doi.org/10.1142/S0217751X1730023X. [arXiv:1706.07442 [hep-ph]]

4. D. Chowdhury, E. Dudas, M. Dutra, Y. Mambrini, Moduli portal dark matter. Phys. Rev. D **99**(9), 095028 (2019). https://doi.org/10.1103/PhysRevD.99.095028. [arXiv:1811.01947 [hep-ph]]
5. M. Dutra, Origins for dark matter particles: from the 'WIMP miracle' to the 'FIMP wonder'. tel-02100637, 2019SACLS059
6. S. Hannestad, What is the lowest possible reheating temperature? Phys. Rev. D **70**, 043506 (2004). https://doi.org/10.1103/PhysRevD.70.043506. [astro-ph/0403291]
7. T. Rehagen, G.B. Gelmini, Low reheating temperatures in monomial and binomial inflationary potentials. J. Cosmol. Astropart. Phys. **1506**, 039 (2015). https://doi.org/10.1088/1475-7516/2015/06/039. [arXiv:1504.03768 [hep-ph]]
8. N. Bernal, F. Elahi, C. Maldonado, J. Unwin, Ultraviolet freeze-in and non-standard cosmologies. https://doi.org/10.1088/1475-7516/2019/11/026. arXiv:1909.07992 [hep-ph]
9. N. Bernal, M. Dutra, Y. Mambrini, K. Olive, M. Peloso, M. Pierre, Spin-2 portal dark matter. Phys. Rev. D **97**(11), 115020 (2018). https://doi.org/10.1103/PhysRevD.97.115020. [arXiv:1803.01866 [hep-ph]]
10. N. Aghanim, et al. [Planck Collaboration], Planck 2018 results. VI. Cosmological parameters. arXiv:1807.06209 [astro-ph.CO]

Unified Superfluid Dark Sector

Elisa G. M. Ferreira

Abstract In this talk I present a new model of a unified dark sector, where late-time cosmic acceleration emerges from the dark matter (DM) superfluid framework. We will start by reviewing the dark matter superfluid model and show how it describes the dynamics of DM in large and small scales. Then we will show that if the superfluid consists of a mixture of two distinguishable states with a small energy gap, such as the ground state and an excited state of DM, interacting through a contact interaction a new dynamics of late-time accelerated expansion emerges in this system, without the need of dark energy, coming from a universe containing only this two-state DM superfluid. I will show the expansion history and growth of linear perturbations, and show that the difference in the predicted growth rate in comparison to ΛCDM is significant at late times.

Keywords Superfluid dark matter · Unification

1 Introduction

Our concordance model, the ΛCDM model, exhibits an outstanding agreement with current large scale cosmological observations [1–3]. In this model the present accelerated expansion is described by a cosmological constant, and the dark matter (DM) is described in the hydrodynamical limit as a fluid with negligible pressure and sound speed, with, at most, very weakly interaction with baryonic matter, the Cold Dark Matter (CDM). However, this simple coarse grained description of those components presents some challenges. The cosmological constant is

Invited talk at the "XI Symposium Quantum Theory and Symmetry" (QTS-XI), Montreal, QC, Canada, July 2019.

E. G. M. Ferreira (✉)
Max-Planck-Institute for Astrophysics, Garching, Germany
e-mail: elisagmf@mpa-garching.mpg.de

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_40

problematic since its smallness is vexing given its radiative instability under quantum corrections [4]. On small scales, a number of challenges have emerged for this hydrodynamical description of the CDM [5], with the most striking being the scaling relations like the mass discrepancy acceleration relation (MDAR), which correlates the dynamical gravitational acceleration inferred from rotation curves and the gravitational acceleration due to baryons only [6, 7].

There is a debate about the explanation for these curious relations on small scales. Within ΛCDM model it is claimed that it can be solved by the inclusion of baryonic feedback effects in simulations (see [8] for a review). An alternative is to modify the behavior of DM on small scales while maintaining the successes of CDM on large scales. Ultra-light fields have emerged as an alternative DM scenario with a different mechanism to explain the dynamics on small scales where DM forms a Bose–Einstein condensate (BEC) or a superfluid in galaxies (for a review [8] of this class of models). One model that accomplishes that is the DM superfluid [9, 10], where sub-eV mass particles with sufficiently strong self-interactions thermalize and condense in galaxies. On top of that for a certain superfluid equation of state and in the presence of coupling of the DM phonons to baryons, this theory's effective Lagrangian is similar to the MOND scalar field theory [11] that leads to a modified dynamics inside galaxies similar to Milgrom's empirical law[1] [12] known to explain and predict these scaling relations.

An interesting question is if the late-time cosmic acceleration can also emerge in the DM superfluid framework, as yet another manifestation of the same underlying substance. We show here that it is indeed possible if we consider that the DM is composed by a mixture of two superfluids, which can be in two different states of the same superfluid, that are in contact and interacting through a contact Josephson-like interaction [13], converting one species into the other. For the phonons that describe the superfluid this interaction appears as an oscillatory potential that drives the late-time acceleration. The unified vision of the dark sector is attractive for its simplicity, given that in this model needs DM in the form of a superfluid alone to describe both the DM behavior on large and small scales, and the late-time acceleration.

The DM superfluid model and the unified framework present a series of observational consequences [9, 10, 14] that successfully explain some observational challenges in galactic dynamics, cosmological evolution or present new interesting phenomenological consequences.

[1] This empirical law states that the total gravitational acceleration a is approximately the Newtonian acceleration $a_{\rm N}$ due to baryonic matter alone, in the regime $a_{\rm N} \gg a_0$, and approaches the geometric mean $\sqrt{a_{\rm N} a_0}$ whenever $a_{\rm N} \lambda a_0$.

2 Review of Dark Matter Superfluid

Superfluidity is one of the most striking quantum mechanical phenomena on macroscopic scales. It appears in fluids that when brought to very low temperatures, form a Bose–Einstein condensate, now described by a single wave-function of systems coming from the superposition of the de Broglie wavelength of these bosons. The emergent degree of freedom of this collective system has an emergent new dynamics: it flows without friction.

We want to use the physics of superfluidity to build a model of DM that on central region of galaxies DM condenses forming a Bose–Einstein condensation with a superfluid phase. The necessary conditions for condensation, assuming weakly coupling, are that de Broglie wavelength $\lambda_{\rm dB} \sim \frac{1}{mv}$ must be larger than the mean inter-particle separation $\ell \sim (m/\rho)^{1/3}$, and that particle should interact enough to thermalize. The first condition translates to an upper bound on the mass, $m \lesssim (\rho/v^3)^{1/4}$, which for a MW-like galaxy ($M = 10^{12} M_\odot$) results in $m \lesssim 4.3$ eV. The second condition requires that the particles interact strongly. An axion-like particle that obeys these conditions condenses on the central regions of galaxies, forming a core, which is enveloped by DM particles that are not condensed and behave like CDM having the usual NFW profile.

After we guaranteed the conditions for DM to condense on galactic scales, we need to describe the evolution of the superfluid. A superfluid is described by a weakly self-interacting field theory of a complex field $\Psi \propto \rho e^{{\rm i}\Theta}$ with global U(1) symmetry. This symmetry is spontaneously broken by the superfluid ground state of a system at chemical potential μ, so that $\Theta = mt + \theta$. At low energy the relevant degrees of freedom are *phonons*, which are excitations of the Goldstone boson θ for the broken symmetry. The effective theory of phonons must be invariant under the shift symmetry, $\theta \to \theta + c$, and Galilean symmetry, appropriate for a non-relativistic superfluid. Therefore, its most general form at leading order in derivatives and zero temperature is given by:

$$\mathcal{L}_{\rm phonons} = P(X)\,; \qquad X = \dot{\theta} - m\Phi - (\vec{\nabla}\theta)^2/2m\,, \tag{1}$$

where Φ is the gravitational potential. The equation of state of the superfluid is encoded in the form of $P(X)$, and the phonon sound speed is given by $c_s^2 = \frac{P_{,X}}{\rho_{,X}} = \frac{1}{m}\frac{P_{,X}}{P_{,XX}}$. Superfluids are often described by a polytropic equation of state, $P(X) \sim X^n$, corresponding to $P(\rho) \sim \rho^{\frac{n}{n-1}}$. Written in this form, we can describe a standard weakly coupled superfluid (BEC DM), for $n = 2$; for $n = 5/2$ this effective theory describes the Unitary Fermi Gas, a gas of ultra-cold fermionic atoms tuned at unitary.

In the case of the DM superfluid, since we want to reproduce MOND on galactic scales, this corresponds to $n = 3/2$, which gives the expected equation of state for MOND, $P \sim n^3$. One extra ingredient is necessary in order to mediate the MOND force, is that the phonons couple to the baryon mas density. The action for the DM superfluid is given by $\mathcal{L}_{\rm DM} = P_{\rm DM}(X) + \mathcal{L}_{int}$ where

$$P_{\rm DM}(X) = \frac{2\Lambda(2m)^{3/2}}{3} X\sqrt{|X|}\,, \qquad \mathcal{L}_{\rm int} = \alpha\Lambda\frac{\theta}{M_{\rm Pl}}\rho_{\rm b} \tag{2}$$

where α is dimensionless coupling constant. The square-root form also ensures that the Hamiltonian is bounded from below. This phenomenological interaction term breaks shift symmetry softly.

With this description in hand, we can obtain the halo profile. In the center regions of the halo, we have the superfluid region, where phonon gradients dominate,[2] the phonon-mediated acceleration matches the deep-MOND expression $a_{\rm phonon} = \sqrt{a_0 a_{\rm b}}$, where $a_{\rm b}$ is the Newtonian gravitational acceleration due to baryons only. The critical acceleration a_0 is related to the theory parameters as $a_0 = (\alpha^3\Lambda^2)/M_{\rm Pl}$. The total force experienced by baryons is the sum of the phonon-mediated force, and the Newtonian gravitational acceleration due to baryons and the DM condensate itself.

3 Unified Dark Superfluid

In this section, we are going to generalize the above model to two non-relativistic superfluid species, described in terms of two distinct phonon excitations, each given by an effective Lagrangian like (1). For instance, these could represent two distinguishable states of DM with slightly different energies, $\Delta E \lambda m$, such as a ground state (represented with subscript 1) and an excited state (2). The theory of the mixture of these two states has a $U(1) \times U(1)$ global symmetry, describing particle number conservation of each species separately. We assume that these species have a contact interaction, the simplest and ubiquitous possible interaction, of the form $\mathcal{L}_{\rm int} \propto -\left(\Psi_1^*\Psi_2 + \Psi_2^*\Psi_1\right)/|\Psi_1|\,|\Psi_2|$. At low energies, this translates into a potential for the phonons:

$$V(\theta_2 - \theta_1 + \Delta E\, t) = M^4\left[1 + \cos\left(\theta_2 - \theta_1 + \Delta E\, t\right)\right]. \tag{3}$$

This is the known and well studied in condensed matter systems Josephson or Rabi coupling. Now number density is not conserved alone anymore, $n \simeq P_{1,X_1} + P_{2,X_2}$, but there is the possibility of conversion between species. Consistent with the non-relativistic approximation we assume that $\Delta E \lambda m_i$, and that the mass splinting is large in comparison to $\dot\theta_2 - \dot\theta_1$, so $V(\theta_2 - \theta_1 + \Delta E\, t) \sim V(\Delta E\, t)$.

In the non-relativistic approximation, the pressure is given by $\mathcal{P} = P_1(X_1) + P_2(X_2) - V(\Delta E\, t)$ and the energy density of the superfluids is

[2] The phonon effective field theory breaks down for large phonon gradients, like in the vicinity of stars (e.g. in our solar system). For a more detailed, see Sect. 5 of [9].

$$\rho = \underbrace{\frac{1}{2}(m_1 + m_2)n}_{\rho_+} + \underbrace{\frac{1}{2}\Delta E\left(P_{1,X_1} - P_{2,X_2}\right)}_{\rho_-} + V(\Delta E\, t)\,. \tag{4}$$

The adiabatic sound speed of each species, governing the growth of perturbations, is $c_{s\,i}^2 = P_{i,X_i}/(m_i P_{i,X_i X_i})$, where $P_{i,X_i X_i} \geq 0$ to ensure the $c_{s\,i}^2 > 0$.

With that, the Friedmann equations for a spatially flat universe can be written, in the non-relativistic approximation as:

$$3H^2 M_{\rm Pl}^2 = \rho_+ + \rho_- + V(\Delta E\, t)\,, \qquad \dot{H} M_{\rm Pl}^2 \simeq -(1/2)\,(\rho_+ + \rho_-)\,. \tag{5}$$

The energy density ρ_+ redshifts like matter and represents the conservation of number density of DM particles, ρ_- evolves under the influence of the potential and the potential term evolves as dark energy. In the case $n = 2$, the BEC DM, $P(X_i) = \Lambda_i^4 \frac{X_i^2}{m_i^2}$, the "+" can be thought as the energy density for the sum of the phases and "−" for the difference. This can be recast in the canonical variables[3] representing the two states of the superfluid: $\xi = (1/N)(N_1^2\theta_1 + N_2^2\theta_2)$ and $\chi = (N_1 N_2/N)(\theta_1 - \theta_2)$, with $N_i = \Lambda_i^2/m_i$ and $N = \sqrt{N_1^2 + N_2^2}$.

The Friedmann equations can be combined leading to a universal equation for the Hubble parameter:

$$2\dot{H} + 3H^2 = V(\Delta E\, t)/M_{\rm Pl}^2\,. \tag{6}$$

From this equation, we can see that during matter domination the potential is not important and the density redshifts as matter; for late times, when the potential becomes dominant, so to ensure that the slow-roll approximation holds for the acceleration we need $\Delta E/2H_0 \lambda 1$, which implies, for $n = 2$, that $\Lambda \gg \sqrt{M_{pl}(m_1 + m_2)/2}$ and that the decay constant, the scale of the spontaneous symmetry breaking f_χ, is super-Planckian as in the case of pNGB models.

We can see in Fig. 1 the evolution of the unified model in comparison to the concordance model, choosing $M^4 = 2M_{pl}^2 H_0^2 \sim \text{meV}^4$, in order to have acceleration today; and $\Delta E/2H_0 = 0.2$. Our model evolves like ΛCDM until times close to today, describing the matter era and the late-time acceleration, evolving in a distinct way in the future.

4 Growth of Density Inhomogeneities

A viable alternative to the ΛCDM model must not only reproduce the evolution of the background, but it should be able to describe the growth of density perturbations

[3] Coming from the diagonalization of the Lagrangian at leading order in $\Delta E/m_i \lambda 1$.

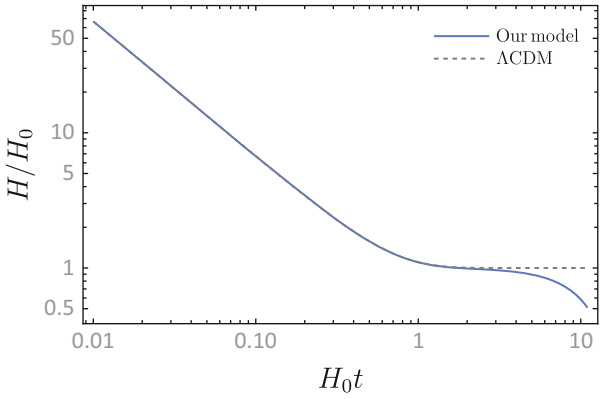


Fig. 1 Hubble parameter H as a function of time for our model (blue solid curve) for our model in comparison to the ΛCDM model (black dashed curve)

that leads to the structures we observe in our universe. In this section we turn to the analysis of density perturbations.

For simplicity we will focus on the BEC DM superfluids. Since our theory describes two interacting superfluids, it is instructive to write down their equations of motion in terms of fluid variables. The continuity and Euler's equations are first-order equations, hence to derive them we must work in the Hamiltonian description. The density is given by $\rho_\xi = \Lambda^2 \Pi_\xi$ and $\rho_\chi = (\Lambda_1^2 \Lambda_2^2/\Lambda^2)(\Delta E/m)\Pi_\chi$, with $\Pi_i = \partial \mathcal{L}/\partial \dot{i}$ where $i = \xi, \chi$. Because these were derived in the weak-field approximation, they can be applied to the cosmological context in the free-falling coordinate system (valid for $H\ell\lambda 1$) of the Friedmann–Robertson–Walker (FRW) metric: $\mathrm{d}s^2 = -(1+2\Phi)\mathrm{d}t^2 + (1-2\Psi)\mathrm{d}\vec{\ell}^2$, where ℓ is the proper distance related to the coming distance x by the scale factor, $\vec{\ell} = a(t)\vec{x}$.

Each fluid density can be decomposed into a background piece and an inhomogeneous term: $\rho_\xi = \bar{\rho}_\xi(t) + \delta\rho_\xi(\vec{x}, t)$ and $\rho_\chi = \bar{\rho}_\chi(t) + \delta\rho_\chi(\vec{x}, t)$. Note that $\delta\rho_\xi$ and $\delta\rho_\chi$ are *not* assumed small at this stage. In this expanding coordinate system, the background densities obey the equations:

$$\begin{aligned} &\dot{\bar{\rho}}_\xi + 3H\bar{\rho}_\xi = 0 \\ &\dot{\bar{\rho}}_\chi + 3H\bar{\rho}_\chi = -\Delta E\, V'(\Delta E\, t) \end{aligned} \qquad \Longrightarrow \qquad \dot{\bar{\rho}} + 3H\bar{\rho} = -\Delta E\, V'(\Delta E\, t)\ , \tag{7}$$

where $\bar{\rho} = \bar{\rho}_\xi + \bar{\rho}_\chi$. This confirms, in particular, that $\bar{\rho}_\xi$ describes dust and redshifts as $1/a^3$. Meanwhile, the evolution of ρ_χ is influenced by the potential. To study the evolution of the background energy density, we solve these equations starting at matter-radiation equality. We will set $m = (m_1 + m_2)/2 = 1\mathrm{eV}$ and $\Lambda_1 = \Lambda_2 = 500\,\mathrm{eV}$. The initial condition for $\bar{\rho}_\xi$ and $\bar{\rho}_\chi$, or for the ground or excited state of the superfluid depends on how the energy gap ΔE compares to the DM temperature at matter-radiation equality, which depends on the production mechanism of our DM particles. For $\Delta E = 5 \times 10^{-11}\mathrm{eV}$, $T_{eq} \sim 10^{-26}\,\mathrm{eV}\lambda\Delta E$, all the matter density is in

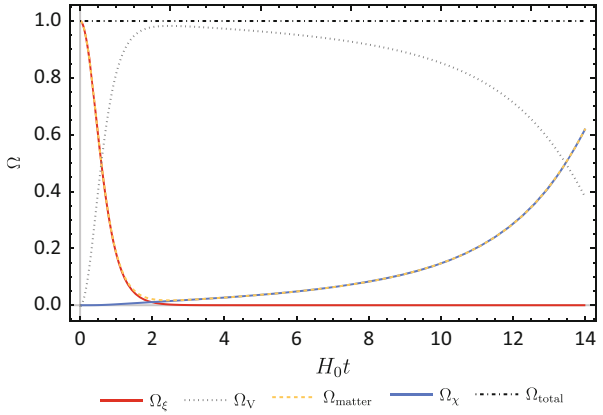


Fig. 2 Evolution of the fraction density parameters $\Omega_x = \frac{\rho_x}{3M_{\rm Pl}^2 H^2}$ if the components were separated: ξ (red), χ (blue), the total matter density given by their sum (yellow), and the potential energy (dotted gray)

the ground state θ_1: $\bar{\rho}_\xi^{eq} \simeq \rho_{\rm eq} = 0.4\,{\rm eV}^4$. In Fig. 2 we separate the energy densities of each degree of freedom of the superfluid mixture and the energy density of the potential. The sum of the two superfluid species gives the total DM density. This is only for illustrative purposes, since all of these quantities represent the same fluid. We can see that close to today we have the transition from a matter dominated period to an accelerated expansion, as matter redshifts away and the potential dominates. However, since the potential oscillates and induces conversion of species, eventually the other species of the superfluid dominates in the future and we will have a new matter domination period. Given the oscillation of the potential, this change might occur many times in the future.

Now, we are interested in analyzing the perturbations relative to the total background density, defined as $\delta_i \equiv \delta\rho_i/\bar{\rho}$ with $i = \xi, \chi$. The fully non-linear equations for the perturbations that describe the Newtonian hydrodynamical equations in an expanding universe can be written as:

$$
\begin{aligned}
&\dot{\delta}_\xi + \frac{1}{a}\vec{\nabla}\cdot\left(\left(\frac{\bar{\rho}_\xi}{\bar{\rho}} + \delta_\xi\right)\vec{v}_\xi\right) - \frac{1}{a}\vec{\nabla}\cdot\left(\left(\frac{\bar{\rho}_\chi}{\bar{\rho}} + \delta_\chi\right)\vec{v}\right) = \frac{\Delta E\, V'}{\bar{\rho}_\chi}\,\delta_\xi\,;\\
&\dot{\vec{v}}_\xi + H\vec{v}_\xi + \frac{1}{a}\left(\vec{v}_\xi\cdot\vec{\nabla}\right)\vec{v}_\xi = -\frac{\bar{\rho}}{\Lambda^4}\frac{\vec{\nabla}\delta_\xi}{a} - \frac{\vec{\nabla}\phi}{a}\,;\\
&\dot{\delta}_\chi + \frac{1}{a}\vec{\nabla}\cdot\left(\left(\frac{\bar{\rho}_\chi}{\bar{\rho}} + \delta_\chi\right)\vec{v}_\xi\right) = \frac{\Delta E\, V'}{\bar{\rho}_\chi}\,\delta_\chi\,;\\
&\dot{\vec{v}}_\chi + H\vec{v}_\chi + \frac{1}{a}\left(\vec{v}_\chi\cdot\vec{\nabla}\right)\vec{v}_\chi - \frac{1}{a}\left(\vec{v}\cdot\vec{\nabla}\right)\vec{v} = -\frac{\Lambda^4\bar{\rho}}{\Lambda_1^4\Lambda_2^4}\left(\frac{m}{\Delta E}\right)^2\frac{\vec{\nabla}\delta_\chi}{a} - \frac{\vec{\nabla}\phi}{a}\,,
\end{aligned}
\tag{8}
$$

where the velocities are given by $\vec{v}_i(\vec{l}/a(t),\ t) = \vec{u}_i - H\vec{l}$, with $u_\xi = -\vec{\nabla}\xi/\Lambda^2$ and $v_\chi = -(\Lambda^2/\Lambda_1^2\Lambda_2^2)(m/\Delta E)\vec{\nabla}\chi$, plus Poisson's equation, $\vec{\nabla}^2\Phi = \frac{a^2}{2M_{\rm Pl}^2}\bar{\rho}\,\delta$.

We can simplify these equations by taking the linear regime, where δ_i and v_i are small. We can also ignore the spatial gradients, since both $c_{s,i}\lambda 1$. We can then combine the above equations into the equation for the density perturbations:

$$\ddot{\delta} + \left(2H - \frac{\Delta E\, V'}{\bar{\rho}}\right)\dot{\delta} = \frac{1}{2M_{\rm Pl}^2}\bar{\rho}\,\delta + \frac{\Delta E\, V'}{\bar{\rho}}\left(5H + \frac{\Delta E\, V'}{\bar{\rho}}\right)\delta, \tag{9}$$

The total velocity evolves as $\dot{\vec{v}} + H\vec{v} \simeq 0$, which redshifts as $1/a$.

5 Observational Signatures

Although our model has an evolution that is very close way to ΛCDM, it predicts distinct observational implications. We cite some of those in this section.

Growth of Structures The potential has a distinct evolution than the one of a cosmological constant, affecting also the evolution equations for the density perturbations. This change is explicit in the *growth rate*, $f(z) \equiv -d\ln\delta(z)/(d\ln(1+z))$, a quantity that is interesting since various probes of structure formation are sensitive, which is shown in Fig. 3. We can see that the unified model has a smaller growth rate today than ΛCDM, which is caused by the potential which is increasing with time, suppressing more structure formation than in ΛCDM. This difference is around

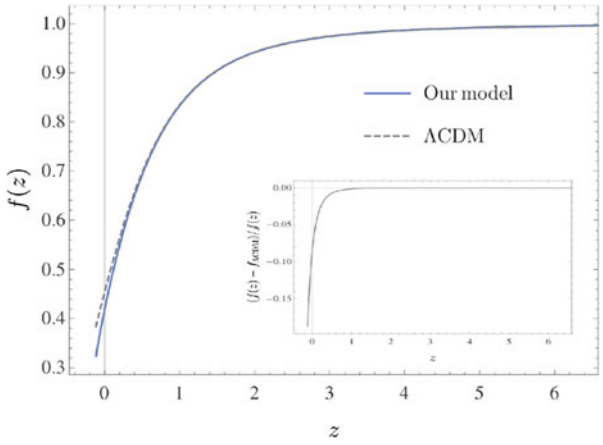


Fig. 3 Growth rate with respect to the redshift for our model (solid blue line), and the prediction for ΛCDM (dashed gray line), with initial condition at equality $\delta_{eq} = 10^{-5}$. The fractional difference between our model and ΛCDM can be seen in the small box

10% today, for the parameters chosen for the model, which is a difference that can be probed by future experiments.

Vortices Quantum vortices are a prediction of superfluids rotating faster than the critical angular velocity and its measurements would be a smoking gun for the superfluid model. To calculate the abundance and properties of those vortices, it is necessary to have a full microscopic description of the superfluid. This topic is worth further investigation since the detection of such effect would be an important evidence for the presence of superfluids in galaxies.

For a review of other effects of the DM superfluid in galaxies and clusters, one can see [8, 9, 14]

Acknowledgments I would like to thank Justin Khoury, Robert Brandenberger, and Guilherme Franzmann for the most enjoyable and exciting collaboration. I would also like to warmly thank the organizer of the QTS-XI for the stimulating conference and the invitation to speak in the cosmology session.

References

1. P.A.R. Ade, et al. [Planck Collaboration], Astron. Astrophys. **594**, A13 (2016). [arXiv:1502.01589 [astro-ph.CO]]
2. L. Anderson, et al. [BOSS Collaboration], Mon. Not. Roy. Astron. Soc. **441**(1), 24 (2014). [arXiv:1312.4877 [astro-ph.CO]]
3. T.M.C. Abbott, et al. [DES Collaboration], Phys. Rev. D **98**(4), 043526 (2018). [arXiv:1708.01530 [astro-ph.CO]]
4. A. Padilla, arXiv:1502.05296 [hep-th]
5. J.S. Bullock, M. Boylan-Kolchin, Ann. Rev. Astron. Astrophys. **55**, 343 (2017). [arXiv:1707.04256 [astro-ph.CO]]
6. S. McGaugh, F. Lelli, J. Schombert, Phys. Rev. Lett. **117**(20), 201101 (2016). [arXiv:1609.05917 [astro-ph.GA]]
7. F. Lelli, S.S. McGaugh, J.M. Schombert, M.S. Pawlowski, Astrophys. J. **836**(2), 152 (2017). [arXiv:1610.08981 [astro-ph.GA]]
8. E.G.M. Ferreira, Alternative models of dark matter: ultra-light fields. to appear (invited review for The Astronomy and Astrophysics Review)
9. L. Berezhiani, J. Khoury, Phys. Rev. D **92**, 103510 (2015). [arXiv:1507.01019 [astro-ph.CO]]
10. L. Berezhiani, J. Khoury, Phys. Lett. B **753**, 639 (2016). [arXiv:1506.07877 [astro-ph.CO]]
11. J. Bekenstein, M. Milgrom, Astrophys. J. **286**, 7 (1984)
12. M. Milgrom, Astrophys. J. **270**, 365 (1983)
13. E.G.M. Ferreira, G. Franzmann, J. Khoury, R. Brandenberger, J. Cosmol. Astropart. Phys. **1908**, 027 (2019). https://doi.org/10.1088/1475-7516/2019/08/027. [arXiv:1810.09474 [astro-ph.CO]]
14. L. Berezhiani, B. Elder, J. Khoury, arXiv:1905.09297 [hep-ph]

de Sitter Vacua in the String Landscape: La Petite Version

Keshav Dasgupta, Maxim Emelin, Mir Mehedi Faruk, and Radu Tatar

Abstract In this review we argue that four-dimensional effective field theory descriptions with de Sitter isometries are allowed in the presence of time-dependent internal degrees of freedom in type IIB string landscape. Both moduli stabilizations and time-independent Newton constants are possible in such backgrounds. However once the time-dependences are switched off, there appear no possibilities of effective field theory descriptions and these backgrounds are in the swampland.

Keywords String theory · M-theory · de Sitter vacua · Flux compactifications

1 Introduction

The late time (quasi-)de Sitter vacuum is a desired consequence of any UV complete theory of quantum gravity, therefore one would expect its appearance in string theory. Earlier attempts to reproduce such a vacuum in the context of type II theories [1] have had some successes although more recently various questions have been raised (see, for example, [2–4]) regarding the validity of such computations. The answers to some of these objections are attempted in [5] and [6], mostly by clarifying the role of anti D3-branes as one of the key ingredients in realizing a four-dimensional de Sitter vacuum. A more serious objection, not just against the computations of [1], but against the very existence of de Sitter vacua in the string landscape, has been recently raised in [7] claiming that string theory prefers quintessence instead of an inflationary evolution towards a late time de Sitter vacuum. The situation is rather paradoxical because objections against models of quintessence were around from early on, and more recently [8] showed

K. Dasgupta (✉) · M. Emelin · M. M. Faruk
Department of Physics, McGill University, Montreal, QC, Canada
e-mail: keshav@hep.physics.mcgill.ca; maxim.emelin; mir-faruk@mail.mcgill.ca

R. Tatar
Department of Mathematics, University of Liverpool, Liverpool, UK
e-mail: Radu.Tatar@Liverpool.ac.uk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_41

that the swampland criteria of [7] are in fact *inconsistent* with simple models of quintessence. More general arguments against swampland criteria themselves have also been raised mostly concerning their implications on cosmology (see, for example, [9]), or their implications on general effective field theories [10]. The latter paper in fact questioned the ad hoc nature of the criteria themselves. As a response to these objections, in [11], the criteria were given a slightly formal derivation using the trans-Planckian problems first raised for inflationary cosmology in [12]. The original swampland criteria of [7] are then elevated to a Trans-Planckian Censorship Conjecture (TCC), meaning that the validity of the criteria [7] rely on censoring the trans-Planckian modes in a theory of quantum gravity because these modes have non-unitary evolutions. Whether this is really a problem in string theory, where we expect a well-defined behavior at short distances, remains to be seen.

In this review, which basically summarizes some parts of [13] and [14], we will argue that if the internal degrees of freedom in a type IIB compactification with four-dimensional de Sitter isometries remain time-independent, then there is no four-dimensional effective field theory description possible. Such backgrounds may truly be in the swampland, thus providing some credence to [7]. However once the internal degrees of freedom become time-dependent (but preserving the four-dimensional Newton constant), then an effective field theory description becomes possible.

2 How Hard Is to Get a de Sitter Vacuum in String Theory?

Let us consider two sets of backgrounds: one, in which we allow a four-dimensional de Sitter space with a flat slicing and an internal six-dimensional compact space with time-independent warp-factors; and two, a similar compactification but now with an internal six-dimensional space with time-dependent warp-factors. Additionally, for the first case we allow all internal degrees of freedom to be time-independent (the internal degrees of freedom being the three and five-form fluxes; as well as the axio-dilaton), whereas for the second case we allow all internal degrees of freedom to be time-dependent. In both cases however we keep vanishing axion but constant dilaton. Question is which of these two backgrounds would solve the equations of motion in type IIB theory?

In the absence of quantum corrections, even if we allow fluxes, branes, anti-branes, and/or orientifold planes, neither of the two backgrounds can solve the sugra equations of motion as was shown in [15] and [16]. This is of course a consequence of the no-go theorem first proposed in [17]. The situation becomes more interesting when quantum corrections are added in. The naive expectation would be that once a sufficient number of quantum terms are added in, either of the two backgrounds should solve the equations of motion. Unfortunately this expectation did not quite turn out to be correct as was first demonstrated in [16] and [13]. In the absence of time-dependent internal degrees of freedom, the equations of motion can *never* be solved because the quantum corrections do not lead to an effective field theory

description in four-dimensions with de Sitter isometries [13, 14, 16]. Thus these backgrounds are truly in the *swampland* [7]. On the other hand, once the internal degrees of freedom become time-dependent, miraculously the effective field theory description reappears [13, 14].

Why is this the case? The answer can be ascertained from many angles, but we shall adopt the strategy of uplifting the IIB background to M-theory and then studying the dynamics from there. The reason for choosing M-theory as opposed to IIB is not just for the sheer compactness of the representation of the degrees of freedom (the *number* of degrees of freedom of course does not change from IIB to M-theory), but also for the fact that M-theory allows a Lagrangian formulation (even with non-local counter-terms, as we shall discuss a bit later) as opposed to IIB, where a Lagrangian formulation is harder to come by. Existence of a Lagrangian formulation then allows us to express the most generic form of the quantum corrections as a series in polynomial powers of four-form flux components and Riemann curvature tensors as well as with spatial and temporal derivatives as shown in eq. (2.6) of [14]. The spatial derivatives are with respect to the internal eight-manifold which is an orbifolded torus fibered over the six-dimensional base that we have in the IIB side. The existence of an effective field theory in lower dimensions then depends on whether the quantum series allows a hierarchy or not. This hierarchy should of course be with respect to M_p, but also with respect to any other relevant expansion parameter. The only other allowed expansion parameter is the type IIA string coupling g_s, so the question of hierarchy boils down to finding a *finite* number of operators at any order in $g_s^{|a|}/M_p^b$, where $(a, b) \in \left(\frac{\mathbb{Z}}{3}, \mathbb{Z}\right)$, and the $1/3$ moding has been explained in [13, 14]. Note that while b can have any sign, we take only positive powers of g_s. This is because all the negative powers of g_s can be resummed into powers of $\exp\left(-\frac{1}{g_s^{1/3}}\right)$, taking the expected non-perturbative form and expressing the full set of corrections as a resurgent trans-series.

All of these would make sense when $g_s \ll 1$. As a happy coincidence, the type IIA coupling g_s turns out to be related to the dimensionless temporal coordinate in the IIB side (the dimension being determined by the cosmological constant Λ). This means at late time, where we expect $\Lambda|t|^2$ to vanish (recall that we are in the flat slicing), the type IIA string coupling g_s also vanishes, implying that the weak coupling limit is also the late time limit in our case. This is good, but the fact that g_s now becomes time-dependent implies that one needs to deal with the quantum series (eq (2.6) in [14]) much more carefully. In fact one would also have to assign certain temporal dependences of the background G-flux components in M-theory. A simple ansatz for the G-flux is given in eq. (2.5) of [14] where, note that, we have traded the temporal dependences with g_s dependences. This means generically all degrees of freedom may be expanded into a series of perturbative as well as non-perturbative g_s-dependent terms. At late time, the non-perturbative terms decouple, so we are left with a perturbative series in g_s.

A careful study of the quantum series then reveals the following interesting fact. One can easily isolate the dominant g_s coefficient of the quantum series, which we

henceforth express as $g_s^{\theta_k}$, with θ_k is as given by eq. (2.10) in [14]. This is expressed in terms of polynomial powers of the curvature tensor and G-flux components (which are denoted in eq (2.10) in [14] by $l_i \in \mathbb{Z}$); parameters $k_i \in \frac{\mathbb{Z}}{2}$ that denote how the g_s dependences of the various G-flux components are expressed in eq. (2.5) of [14]; as well as on n_0 and $n_{1,2,3}$, the former being related to the number of temporal derivatives and the latter to the number of internal spatial derivatives in the quantum series given as eq. (2.6) of [14]. If the G-flux components are time-independent (i.e. when $k_i = 0$), then θ_k, which we write as θ_0, is as in eq. (2.11) of [14].

The important thing to note now is the difference between eq. (2.10) and eq. (2.11) in [14]. If $k_1 \geqslant 0$, $k_2 \geqslant \frac{3}{2}$ and $n_3 = 0$, then eq. (2.10) has no relative minus signs, whereas eq. (2.11) does have relative minus signs. Interestingly switching on n_3, i.e. derivatives along the toroidal directions, both the expressions have relative minus signs. For the time being let us assume that we have switched off n_3 (we will come back to non-zero n_3 just a bit later). In that case, the g_s scalings of the energy-momentum tensors along the various directions in M-theory (i.e. along the space-time, internal six-manifold, and the internal toroidal directions) now become eq. (2.12) of [14]. What does that imply?

Let us take the $2+1$ dimensional space-time directions to clarify the implications of the above g_s scalings. The energy-momentum tensors of the quantum terms will appear on the RHS of the Einstein's equation. The g_s scalings of the Einstein as well as of the classical G-flux terms can be shown to scale as g_s^0, i.e. they are at zeroth order in g_s (there is some subtlety associated with the scaling as shown in section 4.1 of [13]). Now comparing the g_s scalings of the quantum terms along the space-time directions, i.e. eq. (2.12) of [14], we see that $\theta_k = \frac{8}{3}$, implying that the quantum terms are constructed out of eighth order polynomials in curvature and G-flux components. This seems like a marvellous thing, until we notice that for the time-independent case the equations $\theta_0 = \frac{8}{3}$ do not have a *finite* number of integer solutions for $(l_i, n_{1,2}, n_0)$. The loss of finiteness is precisely because of the presence of relative *minus* signs in the expression for θ_0 in eq. (2.12) of [14]. On the other hand, with $\theta_k = \frac{8}{3}$ in eq. (2.11), there are no relative minus signs (we have made $n_3 = 0$), so there is only a *finite* number of operators to zeroth order in g_s.

Such a conclusion would appear to immediately rule out the time-independent compactifications, because allowing an infinite number of operators means that we have lost g_s hierarchy. However the scenario is a bit more subtle than what appears at face-value. In fact a careful analysis reveals a few caveats. First is of course the case with $n_3 > 0$. Second, and this may be more important, all the quantum terms (finite or infinite) at zeroth order in g_s have *different* M_p suppressions. Thus although we have lost our g_s hierarchy for the time-independent case, we seem to still retain an M_p hierarchy in the system. Could we then just make $M_p \to \infty$ and get rid of operators suppressed by higher powers of M_p and retain a finite set of operators for both time-dependent and time-independent cases, respectively? What goes wrong with such a line of reasoning?

The answer lies in the careful mapping between the degrees of freedom of IIB in M-theory. For example not all G-flux components can be allowed without breaking the four-dimensional de Sitter isometries in the IIB side. In particular, if we allow the internal G-flux components with two components along the toroidal directions, they cannot be global fluxes. In fact they can only be localized fluxes so that they appear as gauge fluxes on seven branes in the IIB side (the seven branes being related to the orbifold points in M-theory). In the simplest case, the localized pieces of the G-flux components can be viewed as a Gaussian over the toroidal space. The spread of such Gaussian pieces is measured in terms of M_p, at least for the time-independent cases, implying that the localized fluxes could in principle have hidden M_p scalings implicit in their definitions themselves. Such M_p scalings are in general harmless because they do not interfere or change the M_p suppressions of the quantum terms. However, the quantum terms, i.e. eq. (2.6) of [14], do also have derivatives along the toroidal directions which we called n_3 earlier. If we switch on these n_3 derivatives, we see that now they are able to influence the M_p scalings of the quantum terms!

As shown in [14], the n_3 derivatives change the M_p scalings of the quantum terms in an interesting way: they introduce a relative minus sign there. The g_s scaling θ_0 already had their share of minus signs, and now putting everything together we can ask, for the time-independent case, how many operators are allowed at zeroth order in g_s and M_p (we can even ask at zeroth order in g_s but second order in M_p). The answer is as shown in eq. (2.23), defined in terms of certain variables that may be expressed in terms of (l_i, n_i, n_0) in eq. (2.16), both equations being from [14]. The worrisome thing is the relative *minus* signs in eq. (2.23) whose RHS is 8, related to the eighth order polynomials in curvature and G-flux components. There are an *infinite* number of integer solutions to this equation, all to zeroth order in g_s and M_p. Now we have clearly lost both the g_s and M_p hierarchies, implying that there *cannot* be an effective field theory description in the IIB side with de Sitter isometries and with time-independent internal degrees of freedom.

The above conclusion is definitely intriguing, but couldn't we say the same thing for the time-dependent cases too when we switch on the n_3 derivatives? Why should the time-dependent cases be any different? Answering this will take us to the second level of subtleties in our construction, that we hitherto kept under the rug.

In the time-dependent case, most of the arguments related to the localized fluxes go through in a similar way to the time-independent case. However there are now a few subtle differences. The first difference lies in the scale of the Gaussian spread of the localized function. As shown in eq. (2.25) in [14], the Gaussian spread now involves *both* g_s and M_p, simply because the distances are measured in terms of *warped* metric components (that involve g_s dependences). This means the n_3 derivatives in the quantum terms of eq. (2.6) will contribute to both g_s and M_p scalings. We already discussed how the M_p scalings change, so the question now is whether the change in the g_s scalings can alter our earlier conclusion. The answer turns out to be a bit more subtle. If the G-flux ansatz for the localized flux, as given by eq. (2.15) and eq (2.25) of [14], has g_s independent coefficients (i.e. $\mathcal{A}_{n0}$ in eq. (2.25)), then even the extra g_s contributions from n_3 derivative cannot change the conclusion. The reason is simple to state: with time-independent

coefficients we are as though allowing g_s independent pieces in our original G-flux ansatz eq. (2.5) of [14]. This would be like having $k_2 = 0$ in eq (2.6) of [14], leading to the above-mentioned problems. Thus the coefficients of the localized G-flux components should also have g_s dependent factors. Such factors easily arise by simply normalizing the localized functions! Putting everything together immediately reproduces eq. (2.28) of [14], implying *finite* number of operators at zeroth orders in g_s and M_p with time-dependent internal degrees of freedom.

The story does not end here because M-theory could also have non-local quantum terms, these terms are sometime christened as non-local counter-terms. A simple way to infer their presence is to go to the infinite derivative limits of eq. (2.6) in [14]. More generic form of non-localities are possible in the limit where (n_0, n_i) themselves become *negative*. Such negative-derivative actions may be rewritten as nested integrals as shown in [13] and [14], thus allowing a Lagrangian formulation of the scenario. One however needs to be careful in introducing such terms because we do not expect the non-localities to show up in the low energy limit of M-theory. The validity of the non-local quantum terms has been discussed in great details in [13], and we urge the readers to look up the discussions therein. What we want to question here is what happens to the two cases once we switch on these non-local quantum terms.

The answer as meticulously shown in section 3 of [14], to any given order in $g_s^{|a|}/M_p^b$, there are only finite number of operators possible when all the internal degrees of freedom become time-dependent, whereas the number of such operators tend to become infinite once time-dependences are switched off. These have been demonstrated in figure 1–4 in [14], where moving vertically down in any figure counts the number of operators. The vertical axis in each figure represents the level of non-locality (i.e. how many nested integrals are taken into account), and the horizontal axis represents the M_p scalings. Thus it appears that non-localities do not seem to change the generic idea that four-dimensional effective field theory description with de Sitter isometries remains valid whenever the internal degrees of freedom become time-dependent, but fails when the time-dependences are switched off. Interestingly all of these happen while still keeping the four-dimensional Newton constant time-independent.

There are many other questions that may be asked at this stage, for example: How are the moduli stabilized? How are the equations of motion satisfied? How does the flux quantization, and anomaly cancellation work out in the time-dependent cases? All of these have been answered, and to preserve the brevity of the note we will refer the interested readers to our longer works [13] and [14]. We will however make two comments, one, related to the quantum potential as it appears in eq. (2.7) of [14] and two, related to the possibility of realizing inflation in our construction.

The quantum potential, with contributions from both local and non-local quantum terms, gives rise to an exact expression for the renormalized cosmological constant Λ as given in eq. (4.1) of [14]. This contribution appears from the zeroth order in g_s and second order in M_p (i.e. to order g_s^0/M_p^2 once we assume the non-local contributions are suppressed by the toroidal volume in M-theory). Why don't

the higher order quantum terms contribute to the cosmological constant? The answer as detailed in [13] and [14] is easy to see: at higher orders in g_s we also switch on other G-flux components that appear at higher orders in k in eq. (2.5) of [14]. These components back-react on the geometry creating *negative* gravitational potentials. These negative gravitational potentials are then exactly cancelled by the positive contributions from the quantum potential at higher orders in g_s in such a way that the zeroth order cosmological constant remains unaltered.

The final comment is on the possibility of realizing inflation on our set-up. It has been claimed in [18] that warm inflation may indeed be outside the swampland. Here however we want to comment on the possibility of using the time-dependences to allow for some variant of the D3-D7 inflationary model of [19], the D7-branes in IIB appearing from the orbifold points in M-theory and the D3-branes appearing from the M2-branes that we require the cancel anomalies in the system. This of course calls for a F-theory uplift of our IIB picture, so a natural question is to see whether the M-theory and F-theory connection could be made more precise with time-dependent internal degrees of freedom. It would also be interesting to compare our answers with those of [20].

3 Discussion and Conclusion

In this short note we have summarized some of the contents of [13] and [14] related to the possibility of the existence of four-dimensional effective field theory description with de Sitter isometries and time-dependent internal degrees of freedom in IIB. However there are still questions related to the exact meaning of Wilsonian effective actions in such spaces, for example, how the time-dependent degrees of freedom are integrated out, and whether the de Sitter solutions that we have could be interpreted as *vacua* or coherent states over some supersymmetric solitonic minima. Additionally, the precise connection to TCC needs to be spelled out, maybe along the lines of the recent work [21]. All these and other advances show that this is indeed an exciting time for cosmology.

References

1. S. Kachru, R. Kallosh, A.D. Linde, S.P. Trivedi, De Sitter vacua in string theory. Phys. Rev. D **68**, 046005 (2003). [hep-th/0301240]; C.P. Burgess, R. Kallosh, F. Quevedo, De Sitter string vacua from supersymmetric D terms. J. High Energy Phys. **0310**, 056 (2003). [hep-th/0309187]; E. Silverstein, Simple de Sitter solutions. Phys. Rev. D **77**, 106006 (2008). [arXiv:0712.1196 [hep-th]]
2. I. Bena, M. Grana, S. Kuperstein, S. Massai, Anti-D3 branes: singular to the bitter end. Phys. Rev. D **87**(10), 106010 (2013). [arXiv:1206.6369 [hep-th]]; Polchinski-Strassler does not uplift Klebanov-Strassler. J. High Energy Phys. **1309**, 142 (2013) [arXiv:1212.4828 [hep-th]]

3. S. Sethi, Supersymmetry breaking by fluxes. J. High Energy Phys. **1810**, 022 (2018). [arXiv:1709.03554 [hep-th]]
4. J. Moritz, A. Retolaza, A. Westphal, Toward de Sitter space from ten dimensions. Phys. Rev. D **97**(4), 046010 (2018). [arXiv:1707.08678 [hep-th]]; J. Moritz, A. Retolaza, A. Westphal, On uplifts by warped anti-D3 branes. Fortsch. Phys. **67**(1–2), 1800098 (2019). [arXiv:1809.06618 [hep-th]]; F. Carta, J. Moritz, A. Westphal, Gaugino condensation and small uplifts in KKLT. arXiv:1902.01412 [hep-th]
5. Y. Hamada, A. Hebecker, G. Shiu, P. Soler, On brane gaugino condensates in 10d. J. High Energy Phys. **1904**, 008 (2019). [arXiv:1812.06097 [hep-th]]; Y. Hamada, A. Hebecker, G. Shiu, P. Soler, Understanding KKLT from a 10d perspective. J. High Energy Phys. **1906**, 019 (2019). [arXiv:1902.01410 [hep-th]]
6. E.A. Bergshoeff, K. Dasgupta, R. Kallosh, A. Van Proeyen, T. Wrase, $\overline{\text{D3}}$ and dS. J. High Energy Phys. **1505**, 058 (2015). [arXiv:1502.07627 [hep-th]]; E. McDonough, M. Scalisi, Inflation from Nilpotent Kähler corrections. J. Cosmol. Astropart. Phys. **1611**(11), 028 (2016). [arXiv:1609.00364 [hep-th]]; R. Kallosh, A. Linde, E. McDonough, M. Scalisi, de Sitter vacua with a nilpotent superfield. Fortsch. Phys. **67**(1–2), 1800068 (2019). [arXiv:1808.09428 [hep-th]]; R. Kallosh, A. Linde, E. McDonough, M. Scalisi, 4D models of de Sitter uplift. Phys. Rev. D **99**(4), 046006 (2019). [arXiv:1809.09018 [hep-th]]; R. Kallosh, A. Linde, E. McDonough, M. Scalisi, dS vacua and the swampland. J. High Energy Phys. **1903**, 134 (2019). [arXiv:1901.02022 [hep-th]]; N. Cribiori, R. Kallosh, C. Roupec, T. Wrase, Uplifting anti-D6-brane. arXiv:1909.08629 [hep-th]; R. Kallosh, A. Linde, Mass production of type IIA dS vacua. arXiv:1910.08217 [hep-th]
7. G. Obied, H. Ooguri, L. Spodyneiko, C. Vafa, De Sitter space and the swampland. arXiv:1806.08362 [hep-th]; S.K. Garg, C. Krishnan, Bounds on slow roll and the de Sitter swampland. arXiv:1807.05193 [hep-th]; H. Ooguri, E. Palti, G. Shiu, C. Vafa, Distance and de Sitter conjectures on the swampland. Phys. Lett. B **788**, 180 (2019). [arXiv:1810.05506 [hep-th]]
8. Y. Akrami, R. Kallosh, A. Linde, V. Vardanyan, The landscape, the swampland and the era of precision cosmology. Fortsch. Phys. **67**(1–2), 1800075 (2019). [arXiv:1808.09440 [hep-th]]; L. Heisenberg, M. Bartelmann, R. Brandenberger, A. Refregier, Dark energy in the swampland. Phys. Rev. D **98**(12), 123502 (2018). [arXiv:1808.02877 [astro-ph.CO]]; L. Heisenberg, M. Bartelmann, R. Brandenberger, A. Refregier, Dark energy in the swampland II. Sci. China Phys. Mech. Astron. **62**(9), 990421 (2019). [arXiv:1809.00154 [astro-ph.CO]]
9. F. Denef, A. Hebecker, T. Wrase, de Sitter swampland conjecture and the Higgs potential. Phys. Rev. D **98**(8), 086004 (2018). [arXiv:1807.06581 [hep-th]]; M. Raveri, W. Hu, S. Sethi, Swampland conjectures and late-time cosmology. Phys. Rev. D **99**(8), 083518 (2019). [arXiv:1812.10448 [hep-th]]
10. S. Kachru, S.P. Trivedi, A comment on effective field theories of flux vacua. Fortsch. Phys. **67**(1–2), 1800086 (2019). https://doi.org/10.1002/prop.201800086. [arXiv:1808.08971 [hep-th]]
11. A. Bedroya, C. Vafa, Trans-Planckian censorship and the swampland. arXiv:1909.11063 [hep-th]; A. Bedroya, R. Brandenberger, M. Loverde, C. Vafa, Trans-Planckian censorship and inflationary cosmology. arXiv:1909.11106 [hep-th]
12. J. Martin, R.H. Brandenberger, The trans-Planckian problem of inflationary cosmology. Phys. Rev. D **63**, 123501 (2001). [hep-th/0005209]
13. K. Dasgupta, M. Emelin, M.M. Faruk, R. Tatar, de Sitter vacua in the string landscape (2019). arXiv:1908.05288 [hep-th]
14. K. Dasgupta, M. Emelin, M.M. Faruk, R. Tatar, How a four-dimensional de Sitter solution remains outside the swampland (2019). arXiv:1911.02604 [hep-th]
15. K. Dasgupta, R. Gwyn, E. McDonough, M. Mia, R. Tatar, de Sitter vacua in type IIB string theory: classical solutions and quantum corrections. J. High Energy Phys. **1407**, 054 (2014). [arXiv:1402.5112 [hep-th]]
16. K. Dasgupta, M. Emelin, E. McDonough, R. Tatar, Quantum corrections and the de Sitter swampland conjecture. J. High Energy Phys. **1901**, 145 (2019). [arXiv:1808.07498 [hep-th]]

17. G.W. Gibbons, Thoughts on tachyon cosmology. Class. Quant. Grav. **20**, S321 (2003). [hep-th/0301117]; Aspects of supergravity theories. Print-85-0061 (CAMBRIDGE); J.M. Maldacena, C. Nunez, Supergravity description of field theories on curved manifolds and a no go theorem. Int. J. Mod. Phys. A **16**, 822 (2001). [hep-th/0007018]
18. M. Motaharfar, V. Kamali, R.O. Ramos, Warm inflation as a way out of the swampland. Phys. Rev. D **99**(6), 063513 (2019). [arXiv:1810.02816 [astro-ph.CO]]; M. Motaharfar, V. Kamali, R.O. Ramos, Warm brane inflation with an exponential potential: a consistent realization away from the swampland. arXiv:1910.06796 [gr-qc]; S. Das, Warm inflation in the light of swampland criteria. Phys. Rev. D **99**(6), 063514 (2019). [arXiv:1810.05038 [hep-th]]; S. Das, Distance, de Sitter and trans-planckian censorship conjectures: the status quo of warm inflation. arXiv:1910.02147 [hep-th]
19. K. Dasgupta, C. Herdeiro, S. Hirano, R. Kallosh, D3 / D7 inflationary model and M theory. Phys. Rev. D **65**, 126002 (2002). [hep-th/0203019]; K. Dasgupta, J.P. Hsu, R. Kallosh, A.D. Linde, M. Zagermann, D3/D7 brane inflation and semilocal strings. J. High Energy Phys. **0408**, 030 (2004). [hep-th/0405247]; P. Chen, K. Dasgupta, K. Narayan, M. Shmakova, M. Zagermann, Brane inflation, solitons and cosmological solutions: 1. J. High Energy Phys. **0509**, 009 (2005). [hep-th/0501185]; R.H. Brandenberger, K. Dasgupta, A.C. Davis, A study of structure formation and reheating in the D3/D7 brane inflation model. Phys. Rev. D **78**, 083502 (2008). [arXiv:0801.3674 [hep-th]]
20. J.J. Heckman, C. Lawrie, L. Lin, G. Zoccarato, F-theory and dark energy. Fortsch. Phys. **67**(10), 1900057 (2019). [arXiv:1811.01959 [hep-th]]; J.J. Heckman, C. Lawrie, L. Lin, J. Sakstein, G. Zoccarato, Pixelated dark energy. Fortsch. Phys. **67**(11), 1900071 (2019). [arXiv:1901.10489 [hep-th]]; S. Brahma, S. Shandera, Stochastic eternal inflation is in the swampland. J. High Energy Phys. **1911**, 016 (2019). [arXiv:1904.10979 [hep-th]]; C.M. Lin, K.W. Ng, K. Cheung, Chaotic inflation on the brane and the swampland criteria. Phys. Rev. D **100**(2), 023545 (2019). [arXiv:1810.01644 [hep-ph]]
21. S. Brahma, The trans-planckian censorship conjecture from the swampland distance conjecture. arXiv:1910.12352 [hep-th]; S. Brahma, Trans-Planckian censorship, inflation and excited initial states for perturbations. arXiv:1910.04741 [hep-th]

Intensity Mapping: A New Window into the Cosmos

Hamsa Padmanabhan

Abstract The technique of intensity mapping (IM) has emerged as a powerful tool to explore the universe at $z < 6$. IM measures the integrated emission from sources over a broad range of frequencies, unlocking significantly more information than traditional galaxy surveys. Astrophysical uncertainties, however, constitute an important systematic in our attempts to constrain cosmology with IM. I describe an innovative approach which allows us to fully utilize our current knowledge of astrophysics in order to develop cosmological forecasts from IM. This framework can be used to exploit synergies with other complementary surveys, thereby opening up the fascinating possibility of constraining physics beyond ΛCDM from future IM observations.

Keywords Intensity mapping · Cosmology · Large-scale structure

1 Introduction

Intensity mapping (IM) has emerged as a novel, powerful probe of cosmology over the last decade (e.g., [4]). In this technique, the aggregate intensity of spectral line emission is mapped out to probe the underlying large-scale structure, without resolving individual systems. This offers a three-dimensional picture of the formation and growth of baryonic material (primarily neutral hydrogen—HI), with the frequency dependence tracing their redshift evolution. In contrast to traditional galaxy surveys, which reach their sensitivity limits at $z \sim 2$, probing only about a few percent [6] of the comoving volume of the observable universe, IM has the

Invited talk at the 'Quantum Theory and Symmetries' (QTS-XI) conference, Montreal, QC, Canada, July 2019.

H. Padmanabhan (✉)
Canadian Institute for Theoretical Astrophysics, Toronto, ON, Canada
e-mail: hamsa@cita.utoronto.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_42

potential to directly access the exciting 'dark ages' of the universe ($z \sim 1000 - 30$) immediately following the decoupling of radiation and matter, the formation and turning on of the first luminous sources ($z \sim 30 - 15$), and, ultimately, the *epoch of reionization*: the second major phase transition of (almost) all the cosmological baryons (believed to have completed around $z \sim 6$). IM can also provide valuable insights into astrophysical phenomena on smaller scales: probing the interstellar medium (ISM), the site of active star formation in most normal galaxies, through mapping the carbon monoxide (CO) line emission (which acts as a tracer of molecular hydrogen, H_2, e.g., Ref. [8]) and the 158 μm fine structure transition of the [CII] species (singly ionized carbon; e.g., Ref. [9]).

Besides offering exciting astrophysical insights, the unique ability of IM to efficiently probe vast volumes of the universe makes it an ideal probe of cosmology and fundamental physics—such as modified gravity and dark energy models, the nature of dark matter, inflationary scenarios and several others. However, due to the complex interplay between astrophysics and cosmology in IM surveys, this requires a precise quantification of the impact of astrophysics on the robustness of cosmological constraints. In this article, we summarize recent work exploring various facets of this inter-relationship, specifically focussing on the effect of astrophysical uncertainties on the precision and accuracy of cosmological forecasts from future IM surveys. We also describe how such analyses can open up the fascinating possibility of using IM to constrain physics beyond the standard model of ΛCDM.

2 Forecasts for Cosmology and Astrophysics

The challenge for using line-intensity mapping to constrain cosmology and astrophysics is twofold: (1) the foregrounds—both galactic and extragalactic—are orders of magnitude stronger than the signal and (2) the astrophysics of the tracer itself (such as HI) serves as an effective 'systematic' in deriving cosmological constraints from intensity maps. The former constraint may be mitigated using techniques such as *foreground avoidance and subtraction*, since the frequency structure of the foregrounds is estimated to be very different from those of the signal (e.g., [14]). The latter effect, which may be referred to as the 'astrophysical systematic', can be effectively addressed by quantifying the impact of our uncertainty in the knowledge of the tracer astrophysics, on the observable intensity fluctuations. In the case of 21 cm intensity mapping, this can be done by using a *data-driven, halo model framework* which uses a parametrized form for the HI-halo mass relation $M_{\rm HI}(M, z)$, given by [11]:

$$M_{\rm HI}(M, z) = \alpha f_{\rm H,\,c} M \left(\frac{M}{10^{11}\, M_\odot/h}\right)^{\beta} \exp\left[-\left(\frac{v_{\rm c,0}}{v_{\rm c}(M, z)}\right)^3\right] \tag{1}$$

with free parameters (1) α, the fraction of HI relative to cosmic $f_{\mathrm{H,\,c}}$, (2) β, the logarithmic slope of the HI-halo mass relation and (3) $v_{\mathrm{c,\,0}}$, the lower virial velocity cutoff below which haloes preferentially do not host HI. Similarly, the small-scale HI density profile, $\rho_{\mathrm{HI}}(r, M, z)$ can be described by:

$$\rho_{\mathrm{HI}}(r; M, z) = \rho_0 \exp\left[-\frac{r}{r_{\mathrm{s}}(M, z)}\right], \tag{2}$$

with r_{s} defined as $r_{\mathrm{s}}(M, z) \equiv R_{\mathrm{v}}(M)/c_{\mathrm{HI}}(M, z)$ and $R_{\mathrm{v}}(M)$ denoting the virial radius of the dark matter halo of mass M. The normalization ρ_0 is fixed by requiring the HI mass within the virial radius R_v to be equal to M_{HI} at each (M, z). Here, $c_{\mathrm{HI}}(M, z)$ denotes the concentration parameter defined as [7]:

$$c_{\mathrm{HI}}(M, z) = c_{\mathrm{HI,0}} \left(\frac{M}{10^{11} M_{\odot}}\right)^{-0.109} \frac{4}{(1+z)^{\eta}}, \tag{3}$$

with the free parameters $c_{\mathrm{HI,\,0}}$ and η describing the normalization and redshift evolution, respectively. Given the combination of all the data in HI available at present (DLAs, HI galaxy surveys and presently available IM observations), the best-fitting values and uncertainties for the HI astrophysical parameters are constrained to be [10, 11]: $\alpha = 0.09 \pm 0.01$, $\beta = -0.58 \pm 0.06$, $\log\left(v_{\mathrm{c,0}}/\mathrm{km\ \ s^{-1}}\right) = 1.56 \pm 0.04$, $c_{\mathrm{HI,\,0}} = 28.65 \pm 1.76$ and $\eta = 1.45 \pm 0.04$.

The full, nonlinear power spectrum of HI intensity fluctuations comprises one- and two-halo terms, which are given by:

$$P_{\mathrm{HI,\,1h}}(k, z) = \int \mathrm{d}M\, n_{\mathrm{h}}(M, z) \left[\frac{M_{\mathrm{HI}}(M)}{\bar{\rho}_{\mathrm{HI}}(z)}\right]^2 |u_{\mathrm{HI}}(k|M, z)|^2. \tag{4}$$

and

$$P_{\mathrm{HI,\,2h}}(k, z) = P_{\mathrm{lin}}(k, z) \times \left[\int \mathrm{d}M\, n_{\mathrm{h}}(M, z) b_{\mathrm{h}}(M, z, k) \frac{M_{\mathrm{HI}}(M)}{\bar{\rho}_{\mathrm{HI}}(z)} |u_{\mathrm{HI}}(k|M, z)|\right]^2, \tag{5}$$

with

$$u_{\mathrm{HI}}(k|M, z) = \frac{4\pi}{M_{\mathrm{HI}}(M, z)} \int_0^{R_{\mathrm{v}}(M)} \mathrm{d}r\, \rho_{\mathrm{HI}}(r; M, z) \frac{\sin(kr)}{kr} r^2, \tag{6}$$

and

$$P_{\mathrm{HI}}(k, z) = P_{\mathrm{HI,\,1h}}(k, z) + P_{\mathrm{HI,\,2h}}(k, z) \tag{7}$$

is the total HI power spectrum. In the above expressions, the $\bar{\rho}_{\mathrm{HI}}(z)$ denotes the average HI density at redshift z. The observable, on-sky quantity in an IM survey is the projected *angular power spectrum*, denoted by $C_{\ell}^{\mathrm{HI}}(z)$ (for the HI case), which

enables a tomographic analysis of clustering in multiple redshift bins without the assumption of an underlying cosmological model (e.g., [15]). The expression for $C_\ell^{\rm HI}(z)$ can be constructed using the Limber approximation (accurate to within 1% for scales above $\ell \sim 10$; e.g., Ref. [5]), using the angular window function, $W_{\rm HI}(z)$ of the survey, as:

$$C_\ell^{\rm HI}(z) = \frac{1}{c}\int dz \frac{W_{\rm HI}(z)^2 H(z)}{R(z)^2} P_{\rm HI}[\ell/R(z), z], \tag{8}$$

where $H(z)$ is the Hubble parameter at redshift z, and $R(z)$ is the comoving distance to redshift z. From the above angular power spectrum, and given an experimental configuration, a Fisher forecasting formalism can be used for predicting constraints on the various cosmological $[\{h, \Omega_m, n_s, \Omega_b, \sigma_8\}]$ and astrophysical $[\{c_{\rm HI}, \alpha, \beta, \eta, v_{\rm c,0}\}]$ parameters, generically denoted by p_μ. The Fisher matrix element corresponding to parameters $\{p_\mu, p_\nu\}$ at a particular redshift bin centred at z_i is given by:

$$F_{\mu\nu}(z_i) = \sum_{\ell<\ell_{\rm max}} \frac{\partial_\mu C_\ell^{\rm HI}(z_i)\partial_\nu C_\ell^{\rm HI}(z_i)}{\left[\Delta C_\ell^{\rm HI}(z_i)\right]^2}, \tag{9}$$

where ∂_μ is the partial derivative of $C_\ell^{\rm HI}$ with respect to parameter p_μ. The standard deviation, $\Delta C_\ell(z_i)$ is defined in terms of the noise of the experiment, $N_\ell^{\rm HI}$ and the sky coverage of the survey, $f_{\rm sky}$:

$$\Delta C_\ell^{\rm HI} = \sqrt{\frac{2}{(2\ell+1)f_{\rm sky}}}\left(C_\ell^{\rm HI} + N_\ell^{\rm HI}\right). \tag{10}$$

An example angular power spectrum at $z \sim 0.8$ with its associated standard deviation (illustrated by the error bars) for a SKA I MID-like experimental configuration is shown in Fig. 1. The full Fisher matrix for an experiment is constructed by summing over the individual Fisher matrices in each of the z-bins in the range covered by the survey: $F_{ij,{\rm cumul}} = \sum_{z_i} F_{ij}(z_i)$. Given the cumulative Fisher matrix, we can calculate the standard errors on the parameter p_i for the cases of fixing and marginalizing over other parameters: $\sigma^2_{i,{\rm fixed}} = F_{ii}^{-1}$, and $\sigma^2_{i,{\rm marg}} = (F^{-1})_{ii}$. This is useful to quantify the degradation in cosmological constraints when the astrophysical parameters $(c_{\rm HI}, \eta, \alpha, \beta, v_{\rm c,0})$ are marginalized over. We find that (as shown in Fig. 2) although the astrophysical information broadens the cosmological constraints, the broadening is, for the large part, mitigated by the prior information coming from the present knowledge of the astrophysics, quantified by the halo model. We also find that experiments reaching lower z-values achieve more precise cosmological constraints, as do those having a larger tomographic coverage (such as the SKA I MID; Ref. [13]).

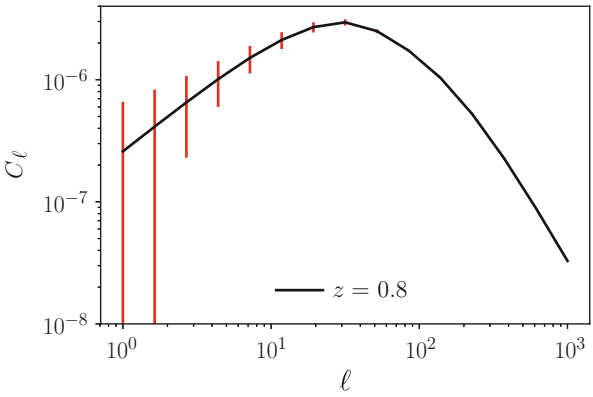


Fig. 1 Angular power spectrum $C_\ell^{\rm HI}$ at redshift 0.8, using the best-fitting HI astrophysical parameters from the halo model and a redshift binning of $\Delta z = 0.05$. The error bars shown in red represent the standard deviation $\Delta C_l^{\rm HI}$, calculated for a SKA 1 MID-like configuration

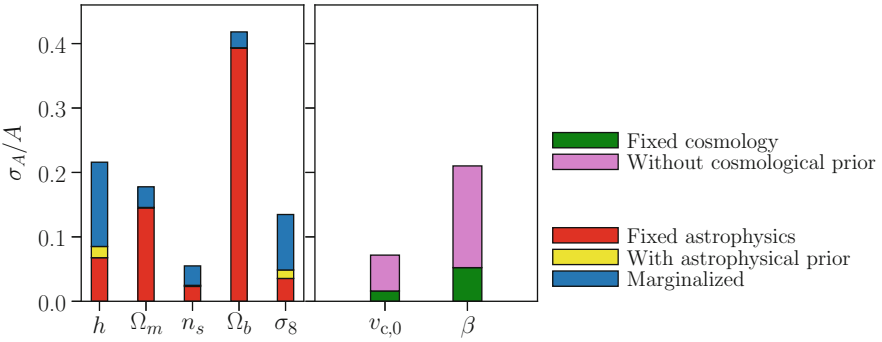


Fig. 2 *Left panel:* Constraints on the cosmological parameters from a SKA I MID-like configuration, (1) marginalized over astrophysics, (2) with fixed astrophysics and (3) with the astrophysical prior coming from the present data. *Right panel:* Astrophysical forecasts (1) without cosmological priors and (2) with fixed cosmology. Figure adapted from Ref. [13]

Cross-correlating 21 cm intensity maps with galaxy surveys (both photometric and spectroscopic) offer exciting prospects for constraining astrophysical and cosmological parameters. It can be shown [12] that cross-correlating such surveys covering similar redshift ranges and sky areas significantly improves astrophysical constraints (see Fig. 3 for an example of a CHIME-like and DESI-like survey cross-correlated in the northern hemisphere). Further, cross-correlation is a valuable tool to mitigate the effects of contaminants and foregrounds, which are expected to be significantly uncorrelated between the two surveys (e.g., Ref. [2]).

The Fisher formalism can also be used to quantify how a complementary effect, the accuracy of cosmological constraints, is affected by the astrophysical prior information. This can be done by calculating the *relative biases* on cosmological parameters, induced by a wrong assumption on the astrophysical ones. Such biases can be naturally quantified using the 'nested likelihoods' (e.g., Ref. [3]) framework in which the parameter space is split into two subsets: one containing all the parameters of interest and the other, all those deemed 'nuisance' or systematic

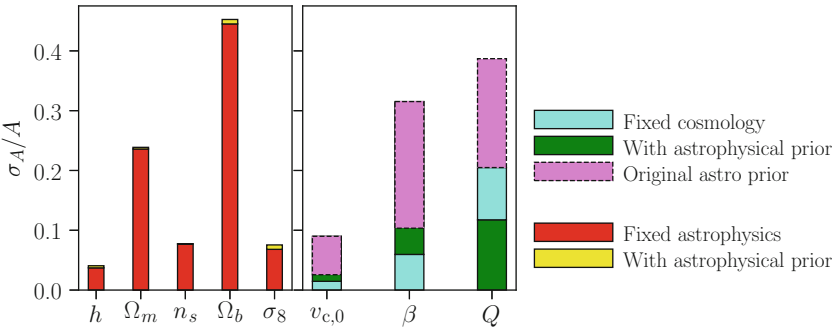


Fig. 3 Cross-correlation forecasts for a CHIME-DESI like survey combination. Fractional errors σ_A/A are plotted for $A = \{h, \Omega_m, n_s, \Omega_b, \sigma_8, v_{c,0}, \beta\}$, and an additional parameter Q which quantifies the scale dependence of the optical galaxy bias. *Left panel:* Constraints on the cosmological parameters (1) without marginalizing over astrophysics, (2) with the astrophysical prior coming from the present data. *Right panel:* Constraints on astrophysical parameters when: (1) not marginalizing over cosmology and (2) marginalizing over cosmology and adding the astrophysical prior. The violet bands in the right panel show the extent of the prior in each case. Figure adapted from Ref. [12]

for the analysis being carried out. In the present case, these two sets represent 'cosmological' and 'astrophysical' parameters, respectively. The bias on a given cosmological parameter p_a, denoted by b_{p_a}, is computed as:

$$b_{p_a} = \delta p_\alpha F_{b\alpha} \left(\mathbf{F}^{-1}\right)_{ab}. \tag{11}$$

Here, $\mathbf{F}^{-1}$ is the full Fisher matrix of astrophysical and cosmological parameters, and $F_{b\alpha}$ represents the submatrix mixing cosmological and astrophysical parameters. The term δp_α denotes the vector of shifts in the astrophysical parameters from their fiducial values:

$$\delta p_\alpha = p_\alpha^{\rm fid} - p_\alpha^{\rm true}. \tag{12}$$

An exciting science case for current and future intensity mapping surveys lies in exploring effects beyond the standard model of Λ CDM cosmology. Two widely studied examples of beyond-ΛCDM physics include (1) the existence of a nonzero $f_{\rm NL}$ parameter that quantifies the primordial non-Gaussianity and (2) incorporating the effects of modified gravity by allowing the growth parameter, γ to deviate from its fiducial value of 0.55. It can be shown [1] that these two effects lead to easily characterizable signatures on the intensity mapping power spectrum, by affecting the quantities $n_h(M)$ and $b_h(M, z)$, i.e., the abundance and bias of dark matter haloes. Thus, they can be incorporated in a straightforward manner into the angular power spectrum and as such, the Fisher forecasting formalism can be used to compute relative constraints on these observables in the presence of astrophysical uncertainties. Figure 4 shows the relative biases on all the standard cosmological

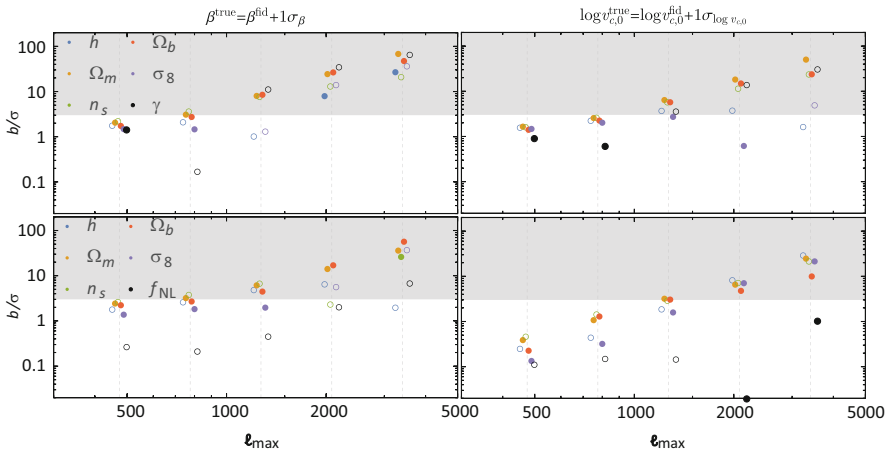


Fig. 4 Relative bias on cosmological parameters in the extended-ΛCDM framework with a SKA I MID-like experimental configuration, obtained on shifting either astrophysical parameter, β (left panels) or $\log v_{\rm c,\,0}$ (right panels), by 1σ from its fiducial value. Top panels contain the parameters in ΛCDM+γ, and lower panels contain those in ΛCDM+$f_{\rm NL}$. Empty (filled) circles indicate negative (positive) values of biases. Figure from Ref. [1]

parameters, as well as $f_{\rm NL}$ and γ, induced by a deviation of either astrophysical parameter, $v_{\rm c,\,0}$ or β from its fiducial (i.e., best-fit) value, as a function of the maximum multipole $\ell_{\rm max}$ considered in the analysis. The figure reveals that the relative biases on the standard cosmological parameters and the modified gravity parameter γ all remain within a few σ as long as the ℓ range stays below 1000. Remarkably, astrophysical uncertainties are found to have negligible effects on the $f_{\rm NL}$ parameter, despite it being strongly linked to the HI bias. Intuitively, this can be understood from the fact that $f_{\rm NL}$ measurements probe the very largest scales, and hence are not 'coupled' as strongly to the details of the HI astrophysics as other cosmological parameters.

3 Conclusions

In this article, we have explored the ability of current and future intensity mapping surveys to provide stringent constraints on cosmology and fundamental physics. A data-driven, halo model framework is well-positioned to mitigate the 'astrophysical systematic' effect on the precision and accuracy of cosmological forecasts from these surveys. Such an approach is a powerful tool to test extensions to the ΛCDM framework, such as primordial non-Gaussianity and deviations from General Relativity at cosmic scales, as well as to mitigate foregrounds through cross-correlating future intensity mapping and optical galaxy surveys. In the future, combining these datasets with more traditional probes of the high-redshift universe has the potential

to uncover fundamental physics constraints from the hitherto unexplored epochs of Cosmic Dawn and reionization.

Acknowledgments I thank the organizers of the Quantum Theory and Symmetries (QTS - XI) conference for the invitation to a productive and enriching meeting. I thank my collaborators Adam Amara, Stefano Camera and Alexandre Refregier with whom most of the work described here was done, and several others for very useful and interesting discussions.

References

1. S. Camera, H. Padmanabhan, Beyond ΛCDM with HI intensity mapping: robustness of cosmological constraints in the presence of astrophysics. MNRAS **496**, 4115 (2020). https://doi.org/10.1093/mnras/staa1663
2. S. Cunnington, L. Wolz, A. Pourtsidou, D. Bacon, Impact of foregrounds on HI intensity mapping cross-correlations with optical surveys (2019). e-prints. arXiv:1904.01479
3. A.F. Heavens, T.D. Kitching, L. Verde, On model selection forecasting, Dark Energy and modified gravity. Mon. Not. R. Astron. Soc. **380**, 1029–1035 (2007). https://doi.org/10.1111/j.1365-2966.2007.12134.x
4. E.D. Kovetz, P.C. Breysse, A. Lidz, J. Bock, C.M. Bradford, T.C. Chang, S. Foreman, H. Padmanabhan, A. Pullen, D. Riechers, M.B. Silva, E. Switzer, Astrophysics and cosmology with line-intensity mapping (2019). e-prints. arXiv:1903.04496
5. D.N. Limber, The analysis of counts of the extragalactic nebulae in terms of a fluctuating density field. Astrophys. J. **117**, 134 (1953). https://doi.org/10.1086/145672
6. A. Loeb, J.S.B. Wyithe, Possibility of precise measurement of the cosmological power spectrum with a dedicated survey of 21cm emission after reionization. Phys. Rev. Lett. **100**(16), 161301 (2008). https://doi.org/10.1103/PhysRevLett.100.161301
7. A.V. Macciò, A.A. Dutton, F.C. van den Bosch, B. Moore, D. Potter, J. Stadel, Concentration, spin and shape of dark matter haloes: scatter and the dependence on mass and environment. Mon. Not. R. Astron. Soc. **378**, 55–71 (2007). https://doi.org/10.1111/j.1365-2966.2007.11720.x
8. H. Padmanabhan, Constraining the CO intensity mapping power spectrum at intermediate redshifts. Mon. Not. R. Astron. Soc. **475**, 1477–1484 (2018). https://doi.org/10.1093/mnras/stx3250
9. H. Padmanabhan, Constraining the evolution of [C II] intensity through the end stages of reionization. Mon. Not. R. Astron. Soc. **488**(3), 3014–3023 (2019). https://doi.org/10.1093/mnras/stz1878
10. H. Padmanabhan, A. Refregier, Constraining a halo model for cosmological neutral hydrogen. Mon. Not. R. Astron. Soc. **464**(4), 4008–4017 (2017). https://doi.org/10.1093/mnras/stw2706
11. H. Padmanabhan, A. Refregier, A. Amara, A halo model for cosmological neutral hydrogen: abundances and clustering H i abundances and clustering. Mon. Not. R. Astron. Soc. **469**(2), 2323–2334 (2017). https://doi.org/10.1093/mnras/stx979
12. H. Padmanabhan, A. Refregier, A. Amara, Cross-correlating 21 cm and galaxy surveys: implications for cosmology and astrophysics (2019). e-prints. arXiv:1909.11104
13. H. Padmanabhan, A. Refregier, A. Amara, Impact of astrophysics on cosmology forecasts for 21 cm surveys. Mon. Not. R. Astron. Soc. **485**(3), 4060–4070 (2019). https://doi.org/10.1093/mnras/stz683
14. M.G. Santos, A. Cooray, L. Knox, Multifrequency analysis of 21 centimeter fluctuations from the era of reionization. Astrophys. J. **625**(2), 575–587 (2005). https://doi.org/10.1086/429857
15. S. Seehars, A. Paranjape, A. Witzemann, A. Refregier, A. Amara, J. Akeret, Simulating the large-scale structure of HI intensity maps. J. Cosmol. Astropart. Phys. **3**, 001 (2016). https://doi.org/10.1088/1475-7516/2016/03/001

Aberration in Gravito-Electromagnetism

Victor Massart and M. B. Paranjape

Abstract The dynamical effects of general relativity which go past Newtonian gravity, specifically the expectation that gravitational effects propagate with a finite velocity, have not been directly verified. In this article we use the formalism of gravito-electromagnetism to compute the first dynamical corrections to Newtonian gravity due to general relativity. We consider a system with multiple sources of dynamical gravitational effects. The main problem that arises is that there are different retarded times for the different sources. We use the Lagrange inversion theorem to express all dynamical effects in terms of the instantaneous time and then simply superpose them. We apply our results to a proposed, realizable experimental set-up and find a dynamical effect that could be observed in a LIGO-type experiment.

Keywords Gravito-electromagnetism · Aberration · Retarded time · Lagrange inversion theorem

1 Gravito-electromagnetism

Gravito-electromagnetism is a linearized, weak-field, non-relativistic approximation to general relativity. The name comes from the fact that in this approximation, gravity is almost exactly analogous to Maxwell electromagnetism. Throughout this article we will use the notation from Jackson's textbook [1] on electromagnetism,

V. Massart
Groupe de physique des particules, Département de physique, Montréal, QC, Canada
e-mail: victor.massart@umontreal.ca

M. B. Paranjape (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: paranj@lps.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_43

the review article of Mashhoon [2] and that of Ruggiero and Tartaglia [3]. This work is inspired by a previous work of M. Paranjape [4].

We begin with the full Einstein equations,

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}. \tag{1}$$

We linearize the equations by expanding the metric around the Minkowski vacuum with small perturbations : $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $|h_{\mu\nu}|\lambda 1$. We then compute the metric connection (the Christoffel symbols), the Riemann tensor, the Ricci tensor, and the curvature scalar in the approximation where we neglect terms of order h^2 and higher. We make the following gauge choice,

$$\partial^\mu (h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h^\lambda_\lambda) \equiv \partial^\mu \bar{h}_{\mu\nu} = 0, \tag{2}$$

which fixes four of the degrees of freedom of the metric. This leads to the linearized Einstein equations,

$$\Box \bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu}. \tag{3}$$

This is a wave equation with a source, it describes propagating gravitational waves in the absence of sources.

Remarkably, Eq. (3) looks essentially identical to Maxwell's equations for the electromagnetic potential in Lorenz gauge with sources,

$$\Box A^\nu = 4\pi j^\nu.$$

This similarity is very useful since electromagnetism is well understood analytically and intuitively, in particular we know the physical (retarded) solution,

$$A^\nu = 4\pi \int D_R(x - x') J^\nu(x') d^4x',$$

as well as the general form of the corresponding electric and magnetic fields, where $D_R(x - x')$ is the retarded Green function of the d'Alembertian.

We will use the non-relativistic limit for the sources in the linearized Einstein equations. The space–space part of the stress energy tensor on the right-hand side of Eq. (3) can be neglected as typically $T^{ij} \sim v^i v^j$ while $T^{0i} \sim v^i$ and $T^{00} \sim v^0$. Then in the weak-field, non-relativistic limit, keeping terms to first order in v^i, we find for the gravitational analog of the electromagnetic gauge potential $\bar{h}^{0\nu}(x)$,

$$\bar{h}_{0\nu}(x) = -16\pi G \int D_R(x - x') T^{0\nu}(x') d^4x'. \tag{4}$$

We can use the well-known electromagnetic potential of a point source, the Lienard–Wiechert potentials, to write down their gravitational counterparts. To make the comparison let us define $\bar{h}_{00} \equiv 4\Phi_G$ and $\bar{h}_{0i} = 4A^i_G$ and then we find the electromagnetic potentials are analog with the corresponding gravitational potential as:

$$\phi_{E.M.}(\boldsymbol{x}, t) = \left[\frac{e}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})R}\right]\Bigg|_{ret} \quad \Longrightarrow \quad \phi_G(\boldsymbol{x}, t) = \left[\frac{-\sqrt{G}M}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})R}\right]\Bigg|_{ret} \tag{5}$$

$$\boldsymbol{A}_{E.M.}(\boldsymbol{x}, t) = \left[\frac{e\boldsymbol{\beta}}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})R}\right]\Bigg|_{ret} \quad \Longrightarrow \quad \boldsymbol{A}_G(\boldsymbol{x}, t) = \left[\frac{-\sqrt{G}M\boldsymbol{\beta}}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})R}\right]\Bigg|_{ret}, \tag{6}$$

where $R(t_r) \equiv |\boldsymbol{x} - \boldsymbol{r}(t_r)|$ is the position of the source at the retarded time and $\beta(t_r) \equiv \frac{d}{cdt_r} r(t_r)$ is its associated (dimensionless) speed. It is clear that analog of e, the electric charge, is the gravitational charge: $\sqrt{G}M$. It is then natural to define a gravitational electric and magnetic field $\boldsymbol{E}_G = -\boldsymbol{\nabla}\Phi_G - \partial_0 \boldsymbol{A}_G$ and $\boldsymbol{B}_G = \boldsymbol{\nabla} \wedge \boldsymbol{A}_G$. Their exact form, expressed in retarded time, is

$$\boldsymbol{E}_G = -\sqrt{G}M\left[\frac{\boldsymbol{n}-\boldsymbol{\beta}}{(1-\boldsymbol{n}\cdot\boldsymbol{\beta})^3 R^2}\right]\Bigg|_{ret} - \sqrt{G}M\left[\frac{\boldsymbol{n}\wedge((\boldsymbol{n}-\boldsymbol{\beta})\wedge\dot{\boldsymbol{\beta}})}{(1-\boldsymbol{n}\cdot\boldsymbol{\beta})^3 R}\right]\Bigg|_{ret} \tag{7}$$

$$\boldsymbol{B}_G = [\boldsymbol{n}\wedge\boldsymbol{E}_G]|_{ret}\,. \tag{8}$$

These two fields are completely analogous to the electromagnetic ones and so are their Maxwell-like equations. The difference is in the minus sign which comes from the fact that gravitation is attractive while electromagnetism is not.

Nevertheless the Lorentz-like force on a particle of mass m, which comes from the geodesic equation, is

$$\frac{d\boldsymbol{p}}{dt} = \sqrt{G}m(-\boldsymbol{\nabla}\Phi_G - 4\partial_t \boldsymbol{A}_G + 2\boldsymbol{v}\wedge\boldsymbol{B}_G)). \tag{9}$$

For the purpose of computing a force, it is natural to define an effective gravito-electric field : $\boldsymbol{E}_G^{\text{Eff}} = -\boldsymbol{\nabla}\Phi_G - 4\partial_t \boldsymbol{A}_G$. Throughout this article the goal will be to compute this effective field in order to find the force on a detector due to a moving mass.

2 Lagrange Inversion Theorem and Analysis

Our aim is to propose an experiment that can be used to measure the speed of gravity. In the Newtonian limit, gravitational effects are instantaneous. By calculating dynamical correction to the Newtonian result we can identify observable effects that would allow a direct measurement of the speed of gravity. We consider a simple situation with two oscillators located on either side of a detector. The configuration is such that the ratios of their mass over the distance squared (distance to the detector), $\frac{M}{R^2}$ are equal, as shown on the LHS in Fig. 1. This makes the Newtonian gravitational contribution at the detector cancel. Then dynamical effects, which depend on the delay of the propagation of gravitational effects from the two sources to the detector, do not cancel and can be used to measure the speed of propagation of gravity, and in principle other effects due to the dynamical nature of general relativity.

We are considering two different sources of gravitational effects, and as all their dynamical gravitational effects are normally and easily expressed in terms of their respective individual retarded times, we cannot simply linearly superpose these effects at the detector. However, if we can find the dynamical effects in terms of the instantaneous time, then their linear superposition is sensible. Consider the definition of retarded time:

$$t_r = t - \frac{|R(t_r)|}{c}, \tag{10}$$

where t_r is the retarded time, t is the instantaneous time, $R(t_r)$ is the distance to the detector at the retarded time, and c is the speed of gravity (presumably equal to the speed of light). The most obvious way to rewrite a quantity (like the gravito-electric

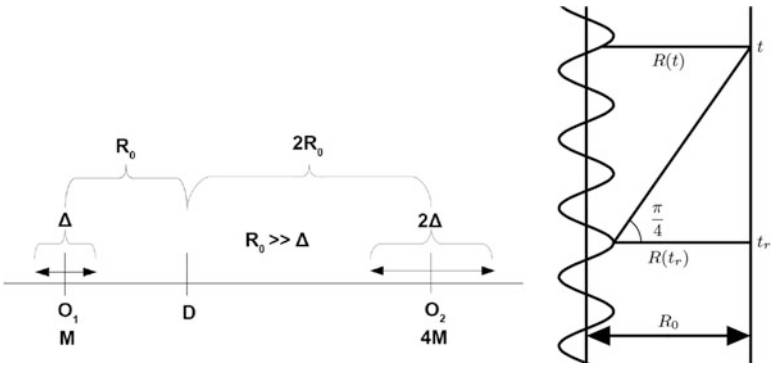


Fig. 1 LHS, a scheme of the experiment with two oscillators (O_1, O_2 with amplitudes is Δ and 2Δ), of masses M and $4M$ at distances R_0 and $2R_0$, respectively, from the central detector at D. RHS, a space-time diagram showing the position of mass M at the retarded time and at the instantaneous time from the detector

field) in terms of the instantaneous time is to find t_r as a function of t. We will then identify a small parameter in which we can perform a Taylor expansion to obtain a manageable expression.[1] A useful mathematical result is the Lagrange inversion theorem.

We define the dimensionless retarded time z as

$$z \equiv \frac{ct_r}{R_0}, \tag{11}$$

with R_0, the distance between the center of the oscillator and the detector. To make apparent the small parameter, we consider a motion of the form

$$R(z) = R_0(1 + \alpha f(z)), \tag{12}$$

where $f(z)$ describes the motion around R_0 in z and α is a small parameter. We define a dimensionless instantaneous time y in a similar way to the retarded time,

$$y \equiv \frac{ct}{R_0} - 1 \quad \text{then} \quad z = y - \alpha f(z). \tag{13}$$

The important point to note is the -1 in the definition of y it represents the time it takes for the signal to propagate the distance R_0. We then have z equals y plus a small perturbation that depends on z. However, we must in the end express our results in terms of $y + 1$ rather than just y as $y + 1$ is the actual instantaneous time (apart from multiplicative factors).

The Lagrange inversion theorem[2] says if the equation in z

$$z = y + \eta\phi(z)$$

has one root in the interior of a contour $\mathbf{C}$, then any function $g(z)$ analytic on and inside $\mathbf{C}$ can be expanded in a power series in η by the formula

$$g(z) = g(y) + \sum_{n=1}^{\infty} \frac{\eta^n}{n!} \frac{d^{n-1}}{dy^{n-1}} \left(g'(y)\phi^n(y)\right).$$

We will use the theorem with $\phi(z)$ replaced by $R(z)$ and η being replaced by $-\alpha$.

The motion of the mass on the left, O_1, is taken to be

$$\boldsymbol{R}(z) = -\boldsymbol{r}(z) = R_0(1 + \alpha \sin(vz))\hat{x}, \tag{14}$$

[1] For a non-relativistic motion, surely t_r will be equal to t plus a constant and some small perturbation that depends on t_r.

[2] For details see, for example, the book on modern analysis of Whittaker and Watson [5].

where $\alpha \equiv \frac{\Delta}{R_0}\lambda 1$ and $v \equiv \frac{\omega R_0}{c}\lambda 1$. The effective gravito-electric field reduces to

$$\boldsymbol{E}_G^{\text{Eff}} = -\hat{x}\sqrt{G}M\left(\frac{1}{\gamma^2(1-\boldsymbol{\beta}\cdot\boldsymbol{n})^2R^2} - \frac{3\beta^2}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})^3R^2} + \frac{3R\dot{\beta}}{(1-\boldsymbol{\beta}\cdot\boldsymbol{n})^2R^2}\right)\Bigg|_{ret}. \tag{15}$$

Then we apply the theorem for $g(z) = E_G(z)$, expanding in the acceleration a or velocity v and neglecting all terms of second order or higher, we obtain

$$-\hat{x}\sqrt{G}M\left(\frac{1}{R(t)^2} - \frac{4a(t)}{R(t)^2}\right)\Bigg|_{inst} \sim (-\hat{x})\sqrt{G}M\left(\frac{1}{R(t)^2} - \frac{4a(t)}{R_0^2}\right)\Bigg|_{inst},$$

where $a(t) \equiv \alpha v^2 \sin v(y+1) = \frac{R_0\Delta\omega^2}{c^2}\sin\omega t\lambda 1$. Note that this result reduces, for non-accelerated motion, to the famous Newtonian result of "instantaneous" propagation, that uniform motion is observed at the instantaneous position and not at the retarded position. (See Jackson [1], chapter 14.)

The form of the solution for the second mass is of course exactly the same except for the value of the parameters. The sum of the two gravito-electric fields will give the total field at the detector at the instantaneous time and we get

$$-4\sqrt{G}\left(\frac{Ma}{R_0^2} + \frac{M'a'}{(R_0')^2}\right) \sim -4\sqrt{G}M\frac{\omega^2}{c^2}\frac{\Delta}{R_0}\left(1 - \frac{R_0'\Delta'}{R_0\Delta}\right), \tag{16}$$

which is the first non-zero term of the force that is created at the detector by the oscillating masses in the directions $-\hat{x}$.

However, this result is not actually correct.[3] We must always consider sources which conserve energy-momentum as those which do not are not physically realizable movements. Our oscillating masses do not conserve energy-momentum, a single mass simply cannot oscillate by itself. Even taking into account the movement of the mass on the other side of the detector, energy-momentum is not conserved. We must add compensating masses attached to our masses by say springs which move in anti-synchronous motion with respect to our oscillators, allowing for the conservation of energy-momentum.

Originally we had thought that if these compensating masses were very heavy (think Earth) then the motion of these masses will be infinitesimal and will not contribute significantly to the gravitational field at the detector. This is indeed true for the higher order time dependent motion of the masses, such as the quadrupole, however, it is not true for the effective dipole. Indeed, because there are no physical, negative masses in general relativity, the net dipole moment cannot be time dependent.

[3] We thank the referee for bringing this point to our attention.

For simplicity, consider a heavy mass, M_H connected to one of our oscillating masses by a short (taken for convenience to be length zero), ideal spring. Then the equilibrium position of the two masses is at the same place, say X_0. The motion must conserve momentum, hence

$$M\frac{d\Delta(t)}{dt} + M_H\frac{d\Delta_H(t)}{dt} = 0, \tag{17}$$

where $\Delta(t)$ and $\Delta_H(t)$ are the time dependent parts of the motions of the two masses. Clearly this implies that the dipole moment $p(t)$ is actually time independent

$$\frac{d}{dt}p(t) = \frac{d}{dt}M(R_0 + \Delta(t)) + \frac{d}{dt}M_H(R_0 + \Delta_H(t)) = 0. \tag{18}$$

Even though

$$\frac{d\Delta_H(t)}{dt} = -\frac{M}{M_H}\frac{d\Delta(t)}{dt}\lambda\frac{d\Delta(t)}{dt} \tag{19}$$

as $\frac{M}{M_H} \to 0$, however, we still have

$$M\frac{d^2\Delta(t)}{dt^2} + M_H\frac{d^2\Delta_H(t)}{dt^2} = 0, \tag{20}$$

i.e.,

$$Ma + M_H a_H = 0 \tag{21}$$

for the acceleration. Thus our result Eq. (16) for each of our oscillating masses is exactly cancelled by the contribution of the necessary compensating masses. We understand that this is how it should be. However, we can now easily ascertain that the similar analysis does not apply to quadrupole time dependence. The quadrupole moment Q is constructed from

$$Q \sim MR^2. \tag{22}$$

Then

$$\dot{Q} \sim 2MR(t)\dot{R}(t) = 2MR\dot{\Delta}(t) = 2M(R_0 + \Delta(t))\dot{\Delta}(t). \tag{23}$$

Realizing $\Delta_H(t)$ for the compensating heavy mass we would find

$$\dot{Q}_H \sim 2M_H R_H(t)\dot{R}_H(t) = 2M_H R_H(t)\dot{\Delta}_H(t) = 2M_H(R_0 + \Delta_H(t))\dot{\Delta}_H(t). \tag{24}$$

Then given $M\Delta = -M_H\Delta_H$

$$\dot{Q} + \dot{Q}_H = 2M\Delta(t)\dot{\Delta}(t) + 2M_H\Delta_H(t)\dot{\Delta}_H(t) = 2M\Delta(t)(\dot{\Delta}(t) - \dot{\Delta}_H(t))$$
$$= 2M\Delta(t)\dot{\Delta}(t)\left(1 + \frac{M}{M_H}\right) \tag{25}$$

and we see the additional contribution of the compensating mass is indeed negligible.

3 Final Expression for the Force

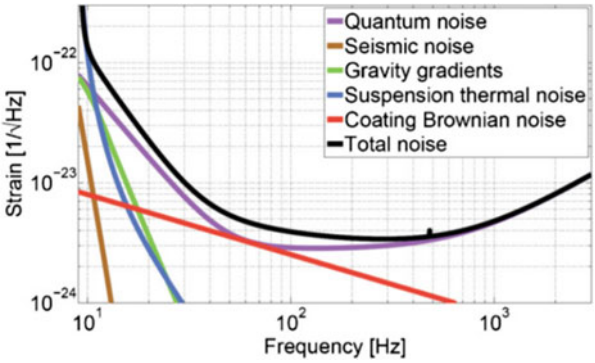
We now impose harmonic motion for the movement of all the masses as given in Eq. (14), then use the Lagrange inversion theorem to obtain a final expression for the gravitational force at the detector from all of the masses. We will give the full details in an extended article. We find for the mass m and its compensating mass M_H, dropping all terms of $\mathcal{O}\left(\frac{m}{M_H}\right)$

$$E_G^{\text{Eff}}(t) = \frac{\sqrt{G}M_H}{R_0^2} + \frac{\sqrt{G}m}{R^2(t)} - 4\sqrt{G}m\frac{\omega^2}{c^2}\frac{\Delta^2}{R_0^2}(\cos^2(\omega t) + 3\sin^2(\omega t))$$
$$= \frac{\sqrt{G}M_H}{R_0^2} + \frac{\sqrt{G}m}{R^2(t)} - 4\frac{\sqrt{G}m}{R_0^2}(a(t)\Delta(t) + 3v^2(t)). \tag{26}$$

It is clear that the compensating mass does not contribute to the time dependent oscillatory force on the detector. The system on the other side is composed of mass $4m$ at a distance $2R_0$ with oscillation amplitude 2Δ and for convenience compensatory mass $4M_H$. This last choice makes the static Newtonian force of the compensating masses also cancel at the detector. Then summing all the contributions together, we find

$$E_{G,\text{total}}^{\text{Eff}}(t) = 12\frac{\sqrt{G}m}{R_0^2}(a(t)\Delta(t) + 3v^2(t)). \tag{27}$$

We note that the result we obtain is not identical to that which is usually obtained for the field in the wave zone, which corresponds to distances much larger than the size of the source and the wavelength of the radiation produced. In the wave zone the gravitoelectromagnetic potentials drop off like $\sim\frac{1}{r}$. The gravitoelectromagnetic fields then fall off as $\sim\frac{1}{r^2}$. The approximations done to obtain those results are simply not valid for our analysis. Correspondingly, our result is not proportional to the third time derivative of the quadrupole moment squared, which is the result obtained as the leading term in the wave zone. Here we are well inside a



single wavelength, we are computing what is normally called Newtonian noise. Dimensionally of course, the result we obtain is the same as the far field result, and hence the analysis at Eq. (25) applies to our result, just as well.

4 LIGO-Type Experiment

To get an idea of experimental observability of the predicted force, suppose that we choose $4m = 1\,\mathrm{kg}$ and the factor $\alpha = \frac{\Delta}{R_0}$ to be $1/20$. Then for a frequency of 1 kHz (which is at the good precision upper limit of LIGO), the total effective gravito-electric field of the four oscillating masses is

$$E^{1\,\mathrm{kHz}}_{G,\mathrm{total}}(t) = 12\sqrt{G}m\left(\frac{\omega}{c}\right)^2\left(\frac{\Delta}{R_0}\right)^2(\cos^2(\omega t) + 3\sin^2(\omega t))$$
$$\simeq 26.4\times 10^{-18}\times(\cos^2(2000\pi t) + 3\sin^2(2000\pi t)). \tag{28}$$

Let us consider the LIGO system, here the mirror mass is 40 kg. Then the total force exerted on the LIGO mirror, obtained from the geodesic equation (Eq. (9)), would be

$$\langle F^{1\,\mathrm{kHz}}_G(t)\rangle \simeq 8.6\times 10^{-21}. \tag{29}$$

This result is very encouraging, as the amplitude of this force corresponds exactly to the sensitivity zone of LIGO [6], see accompanying figure.

Acknowledgement We thank NSERC of Canada for financial support.

References

1. J.D. Jackson, *Classical Electrodynamics* (Wiley, New York, 2007)
2. B. Mashhoon, Gravitoelectromagnetism: a brief review (2003). arXiv preprint gr-qc/0311030
3. M.L. Ruggiero, A. Tartaglia, Gravitomagnetic effects (2002). arXiv preprint gr-qc/0207065
4. M.B. Paranjape, How to measure the speed of gravity (2012). Preprint. arXiv:1208.2293. https://doi.org/10.1139/cjp-2012-0421
5. E.T. Whittaker, G.N. Watson, *A Course of Modern Analysis* (Cambridge University Press, Cambridge, 1996)
6. G.M. Harry, LIGO Scientific Collaboration, Advanced LIGO: the next generation of gravitational wave detectors. Classical Quantum Gravity **27**(8), 084006 (2010)

Stable, Thin Wall, Negative Mass Bubbles in de Sitter Space-Time

Matthew C. Johnson, M. B. Paranjape, Antoine Savard, and Natalia Tapia-Arellano

Abstract Up to now, only configurations of negative mass have been found (Belletête and Paranjape, Int J Mod Phys D22:1341017, 2013; Mbarek and Paranjape, Phys Rev D90(10):101502, 2014), no stability or dynamics was considered. Here we show the existence of stable, static, negative mass bubbles in de Sitter space-time. The bubbles have an interior that is pure de Sitter, separated by a thin wall from the exact, negative mass Schwarzschild-de Sitter space-time exterior. We use the Israel junction conditions at the wall. We find the bubbles can collapse spherically to the singular negative mass solution, violating the cosmic censorship hypothesis.

Keywords Gravity · Negative mass · de Sitter space-time

M. C. Johnson
Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada

Department of Physics and Astronomy, York University, Toronto, ON, Canada
e-mail: mjohnson@perimeterinstitute.ca

M. B. Paranjape (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada

A. Savard
Groupe de physique des particules, Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: paranj@lps.umontreal.ca

N. Tapia-Arellano
Groupe de physique des particules, Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: antoine.savard@umontreal.ca

Departamento de Física, Universidad de Santiago de Chile, Estación Central, Santiago, Chile

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_44

1 Introduction

The Schwarzschild-de Sitter metric is a singular solution of the Einstein equations with cosmological constant. The mass is a parameter that can be taken negative. Here we show that it is possible to deform the metric and smooth out the singularity with physically sensible energy-momentum, that which satisfies the dominant energy condition everywhere. This means for any future directed time-like or light-like vector u:

$$T^{0\nu}u_\nu \geq 0 \quad \text{and} \quad T^{\mu\nu}u_\nu T_{\mu\alpha}u^\alpha \geq 0. \tag{1}$$

The example of such a deformation was given in [1]. Then in [2] it was shown that with an ideal fluid, there exist bubble like configurations with the exterior space-time given exactly by the negative mass Schwarzschild-de Sitter space-time. The fluid was not dynamical, no equation of state was specified. However, it was seen that an ideal fluid, could in principle organize itself to correspond to localized regions of negative mass.

In this work we show how to obtain dynamically stable, non-singular solutions of negative mass which satisfy the dominant energy condition everywhere. They are made up of a de Sitter space-time inside a thin wall separated from negative mass Schwarzschild-de Sitter space-time on the outside. The Israel junction conditions [3, 4] express the conservation of energy-momentum across the wall.

2 Thin Wall Bubbles

The stress-energy tensor of the wall will be taken as

$$S_{\hat{a}\hat{b}} = \text{diag.}\ (\sigma, -\vartheta, -\vartheta). \tag{2}$$

The interior mass function is taken to be $m_-(r)$ which is not yet specified while the exterior mass function is taken to be

$$m_+(r) = -M + \frac{\Lambda r^3}{6} \tag{3}$$

that of a negative mass Schwarzschild-de Sitter space-time with mass $-M$ and cosmological constant Λ. The conservation of energy-momentum flux through the wall gives the equations:

$$\left(1 - \frac{2m_-(r)}{r} + \dot{r}^2\right)^{1/2} - \left(1 + \frac{2M}{r} - \frac{\Lambda r^2}{3} + \dot{r}^2\right)^{1/2} = 4\pi\sigma r \tag{4}$$

and

$$\frac{\left(1-\frac{m_-(r)}{r}-m'_-(r)+\dot{r}^2+r\ddot{r}\right)}{\left(1-\frac{2m_-(r)}{r}+\dot{r}^2\right)^{1/2}}-\frac{\left(1+\frac{M}{r}-\frac{2\Lambda r^2}{3}+\dot{r}^2+r\ddot{r}\right)}{\left(1+\frac{2M}{r}-\frac{\Lambda r^2}{3}+\dot{r}^2\right)^{1/2}}=8\pi\vartheta r. \tag{5}$$

Writing Eq. (4) as

$$\left(a+\dot{r}^2\right)^{1/2}-\left(b+\dot{r}^2\right)^{1/2}=c^{1/2} \tag{6}$$

we can solve for $\dot{r}$ as

$$\frac{\dot{r}^2}{2}+V(r)=0. \tag{7}$$

with

$$V(r)=-\tfrac{1}{2}\left(\frac{(a-b)^2}{4c}-\frac{(a+b)}{2}+\frac{c}{4}\right), \tag{8}$$

which is the potential for the radius of the wall and explicitly it is

$$V(r)=-\tfrac{1}{2}\left(\frac{\left(m_-(r)+M-\frac{\Lambda r^3}{6}\right)^2}{16\pi^2\sigma^2r^4}-\left(1+\frac{(m_-(r)-M)}{r}+\frac{\Lambda r^2}{6}\right)+4\pi^2\sigma^2r^2\right). \tag{9}$$

3 Stable Negative Mass de Sitter Bubbles

If we take $m_-(r)=\frac{\Lambda_i r^3}{6}$ with $\sigma\to(\sqrt{\Lambda/3}/4\pi)\sigma$, $r\to\sqrt{3/\Lambda}\,r$ and $M\to(1/\sqrt{3\Lambda})M$, we find

$$V(r)=-\frac{M^2}{18\sigma^2r^4}+\frac{M\left((1-(\Lambda_i/\Lambda))+\sigma^2\right)}{6\sigma^2r} -\frac{r^2\left(2(\Lambda_i/\Lambda)\sigma^2+(1-(\Lambda_i/\Lambda))^2+\sigma^4+2\sigma^2\right)}{8\sigma^2}+\frac{1}{2}. \tag{10}$$

Then if we generalize $\sigma\to\sigma(r)$ we can obtain a stable minimum. We start with $\sigma=0.04$ and then we modify it as

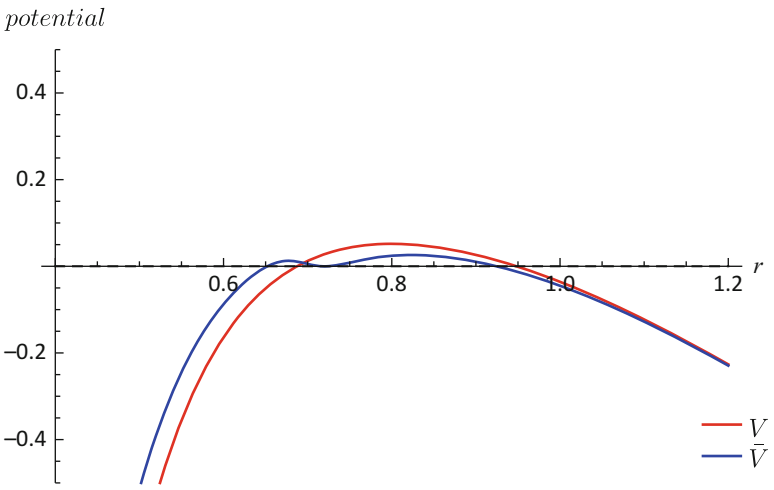


Fig. 1 The potential $V(r)$ for constant $\sigma = 0.04$ (red) and the modified potential $\bar{V}(r)$ (blue) with $\sigma(r)$ given by Eq. (11), and $M = 0.045$, $\Lambda_i/\Lambda = 0.994$

$$\sigma(r) = 0.04 - 0.0035 \tanh\left(\frac{25}{r_0}(r - r_0 + 0.105)\right) \tag{11}$$

where r_0 is the position of the maximum of the potential when $\sigma = 0.04$. We find that the potential has a stable minimum as shown in Fig. 1. The potential $V(r)$ for constant σ is shown and the modified potential with the stable minimum is called $\bar{V}(r)$.

The dominant energy condition is obviously satisfied in the bulk. On the wall, the spatial stress [4] is given by

$$\vartheta(r) = \sigma(r) + \frac{r}{2}\frac{d\sigma(r)}{dr} \tag{12}$$

and we also find numerically that the dominant energy condition on the wall, $\sigma \geqslant |\vartheta|$, is satisfied.

4 Conclusion

In conclusion, we report here that there exist dynamically stable, non-singular solutions of the Einstein equations which correspond to negative mass bubbles in an asymptotically de Sitter space-time. The import of such solutions to cosmology must be examined.

Acknowledgments We thank Emil Mottola and Edward Wilson-Ewing for useful discussions. We thank NSERC of Canada for financial support and The Perimeter Institute for Theoretical Physics for hospitality. N. T. thanks to the Conicyt scholarship 21160064 and the University of Santiago de Chile. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

References

1. J. Belletête, M.B. Paranjape, On negative mass. Int. J. Mod. Phys. **D22**, 1341017 (2013)
2. S. Mbarek, M.B. Paranjape, Negative mass bubbles in de Sitter spacetime. Phys. Rev. **D90**(10), 101502 (2014)
3. W. Israel, Singular hypersurfaces and thin shells in general relativity. Nuovo Cim. **B44**, 1 (1966)
4. M. Visser, D.L. Wiltshire, Stable gravastars: an alternative to black holes? Classical Quantum Gravity **21**, 1135–1152 (2004)

Ferromagnetic Instability in PAAI in the Sky

R. MacKenzie, M. B. Paranjape, and U. A. Yajnik

Abstract We study an idealised plasma of fermions, coupled through an abelian gauge force $U(1)_X$, and which is asymmetric in that the masses of the oppositely charged species are greatly unequal. The system is dubbed PAAI, plasma asymétrique, abélien et idéalisé. It is argued that due to the ferromagnetic instability that arises, the ground state gives rise to a complex of domain walls. This complex being held together by stresses much stronger than cosmic gravity, does not evolve with the scale factor and along with the heavier oppositely charged partners simulates the required features of Dark Energy with mass scale for the lighter fermions in the micro-eV to nano-eV range. Further, residual X-magnetic fields can mix with Maxwell electromagnetism to provide the seed for cosmic-scale magnetic fields. Thus the scenario can explain several cosmological puzzles including Dark Energy.

Keywords Itinerant ferromagnetism · Dark energy · Domain walls

1 Introduction

There are several important unresolved issues in our current understanding of cosmology. Paramount among these are the problems of Dark Matter (DM) and Dark Energy (DE). Within the Λ-CDM model DM assists in galaxy formation and

R. MacKenzie
Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: richard.mackenzie@umontreal.ca

M. B. Paranjape
Département de physique, Université de Montréal, Montréal, QC, Canada

Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: paranj@lps.umontreal.ca

U. A. Yajnik (✉)
Physics Department, Indian Institute of Technology Bombay, Mumbai, India
e-mail: yajnik@iitb.ac.in

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_45

should be a gas of non-relativistic particles, while the issue of DE is closely tied to that of the cosmological constant [1], since data [2] suggests that its energy density is constant over the epochs scanned by the cosmic microwave background (CMB). If treated as a dynamical phenomenon, DE demands an explanation for the equation of state $p = -\rho$ in terms of relativistic phenomena. From the point of view of naturalness, explaining a value of a dynamically generated quantity which is many orders of magnitude away from any of the scales of elementary particle physics or gravity is a major challenge. There are explanations that obtain such a sector as directly related to and derived from more powerful principles applicable at high scales [3–5]. On the other hand, extended and space filling objects, specifically domain walls as possible solutions to understanding Dark Energy have been proposed earlier in a variety of scenarios [6–10]. In this paper we pursue the latter approach, of invoking new species of particles and their interactions at the new low mass scale, agnostic of their connection to the known physics other than gravity. A more extensive discussion of the results reported here can be found in [11].

We consider a new sector of particles with interaction mediated by an unbroken abelian gauge symmetry denoted $U(1)_X$. The core of our mechanism involves the existence of a fermionic species that enters into a ferromagnetic state. As we will show, it is required to have an extremely small mass and hence an extremely large magnetic moment; we dub this species the *magnino*,[1] denoted M. We assume that the medium remains neutral under the X-charge due to the presence of a significantly heavier species Y of opposite charge which does not enter the collective ferromagnetic state. The wall complex resulting from the formation of magnetic domains then remains mutually bound, and due to interaction strength much larger than cosmic gravity, remains frozen. The binding of the heavier species to this complex due to the requirement of X-electrical neutrality then ensures that these particles remain unevolving, and after averaging over the large scales of the cosmic horizon act like a homogeneous space filling medium of constant density.

It is possible to explain DM within the same sector, including possible dark atoms formed by such species [14–17]. This would also solve the *concordance problem*, that is, the comparable energy densities carried in the cosmological energy budget by the otherwise-unrelated components, DM and DE. Further, the X-electromagnetism is expected to mix kinetically with the standard electromagnetism. The existence of cosmic magnetic fields at galactic and intergalactic scales [18–20] is an outstanding puzzle of cosmology. Our mechanism relying as it does on spontaneous formation of domains of X-ferromagnetism has the potential to provide the seeds needed to generate the observed fields through such mixing.

In the following, in Sect. 2 we motivate the origin of negative pressure for extended objects in cosmology. In Sect. 3 we discuss the calculation of the exchange energy for a spin-polarised PAAI. Thus we motivate the possibility of occurrence of an extended structure of domain walls, and their metastable yet long lived nature. In Sect. 4 we discuss the main results of our proposal, obtaining suggestive values for the masses and abundances for the scenario to successfully explain DE, and for

[1] The term *magnino* was earlier introduced in a different connotation in [12, 13].

the DM discussion we refer the reader to our longer paper [11]. In Sect. 5 we obtain a restriction on the length scale of the domains for successful explanation of origin of cosmic magnetic fields from mixing with standard electromagnetism. After the conclusion in Sect. 6 we also include a few salient questions from the audience and their answers in Sect. 7.

2 Cosmic Relics and the Origin of Negative Pressure

A homogeneous, isotropic universe is described by the Friedmann equation for the scale factor $a(t)$ supplemented by an equation of state relation $p = w\rho$. Extended relativistic objects in gauge theories in the cosmological setting [21] are known to lead to negative values for w [22, 23]. A heuristic argument runs as follows. In the case of a frozen-out vortex line network, the average separation between string segments scales as $1/a^3$ but there is also an increment in the energy proportional to a due to an average length of vortex network proportional to a entering the physical volume. As such, the energy density of the network has to be taken to scale as $1/a^2$, and we get the effective value $w = -1/3$. Likewise, for a domain wall complex, the effective energy density scales as $1/a$ and $w = -2/3$. By extension, for a relativistic substance filling up space homogeneously, the energy density is independent of the scale factor, and has $w = -1$. In quantum theory this arises naturally as the vacuum expectation value of a relativistic scalar field. In the following, we consider a scenario that gives rise to a complex of domain walls whose separation scale is extremely small compared to the causal horizon and which remains fixed during expansion, and hence simulates an equation of state $p = -\rho$.

3 Ferromagnetic Instability of PAAI

A system of fermions can be treated as a gas of weakly interacting quasi-particles in the presence of oppositely charged much heavier ions or protons which are mostly spectators and serve to keep the medium neutral. The total energy of such a system can be treated as a functional of electron number density, according to the Hohenberg–Kohn theorem. In a relativistic setting, it becomes a functional of the covariant 4-current, and hence also of the electron spin density [24]. In the Landau fermi liquid formalism the quasi-particle energy receives a correction from an interaction strength f with other quasi-particles which can be determined from the forward scattering amplitude $\mathcal{M}$ [25]

$$f(\mathbf{ps},\ \mathbf{p}'\mathbf{s}') = \frac{m}{\varepsilon^0(\mathbf{p})}\frac{m}{\varepsilon^0(\mathbf{p}')}\mathcal{M}(\mathbf{p}s,\ \mathbf{p}'s'), \tag{1}$$

where ε^0 is the free particle energy and $\mathcal{M}$ is the Lorentz-covariant $2 \to 2$ scattering amplitude in a specific limit not discussed here. The exchange energy

can equivalently be seen to arise as a two-loop correction to the self-energy of the fermion [26]. Using f this one can compute the exchange energy $E_{\rm xc}$, as

$$E_{\rm xc} = \sum_{\pm \mathbf{s}} \sum_{\pm \mathbf{s}'} \int \frac{d^3p}{(2\pi)^3} \frac{d^3p'}{(2\pi)^3} f(\mathbf{ps},\ \mathbf{p}'\mathbf{s}') n(\mathbf{p},\mathbf{s}) n(\mathbf{p}',\mathbf{s}') \tag{2}$$

and the effective quasi-particle energy is the kinetic energy of the quasi-particles with renormalised mass parameter $E_{\rm kin}$ plus the spin-dependent exchange energy in a spin-polarised background. For this purpose it is necessary to calculate the self-energy with a Feynman propagator in the presence of non-zero number density, and spin imbalance [27].

To set up a spin-asymmetric state, we introduce a parameter ζ such that the net density n splits up into densities of spin up and down fermions as

$$n_\uparrow = n(1+\zeta) \quad \text{and} \quad n_\downarrow = n(1-\zeta) \tag{3}$$

Correspondingly, we have Fermi momenta $p_{F\uparrow} = p_F(1+\zeta)^{1/3}$ and $p_{F\downarrow} = p_F(1-\zeta)^{1/3}$, with $p_F^3 = 3\pi^2 n$. The exchange energy was calculated in [27] and the final expression is too long to be quoted in this presentation. However the leading order expansions in $\beta = p_F/m$ for the fully polarised case $\zeta = 1$ is [11]

$$E_{\rm kin}(\zeta = 1) = m^4 \left\{ \frac{\tilde{\beta}^5}{20\pi^2} - \frac{\tilde{\beta}^7}{112\pi^2} + O\left(\beta^9\right) \right\} \tag{4}$$

$$E_{\rm xc}(\zeta = 1) = -\alpha_X m^4 \left\{ \frac{\tilde{\beta}^4}{2\pi^2} - \frac{7\tilde{\beta}^6}{27\pi^2} + O\left(\tilde{\beta}^8\right) \right\}, \tag{5}$$

where $\tilde{\beta} = 2^{1/3}\beta$. The $\zeta = 0$ case has same leading power laws with different coefficients. Thus the exchange energy tends to lower the quasi-particle energy parametrically determined by α, with either $\zeta = 0$ or $\zeta = 1$ becoming the absolute minimum depending on β. For comparison, in this notation, the rest mass energy of the degenerate gas is $E_{\rm rest} = m^4\beta^3/(3\pi^2)$.

Exploring the energy expression presents three possibilities; $\zeta = 1$ is not a minimum at all, $\zeta = 1$ is a local minimum but $E(0) < E(1)$, i.e. a metastable vacuum, and finally, $\zeta = 1$ is the absolute minimum with $\zeta = 0$ unstable vacuum. In Fig. 1 we have plotted the approximate regions of the three phases in the parameter space.

3.1 Evolution and Stability of Domain Walls

We expect domain walls to occur in this spin-polarised medium just like in ferromagnets. However due to the $SU(2)$ of spin being simply connected, the

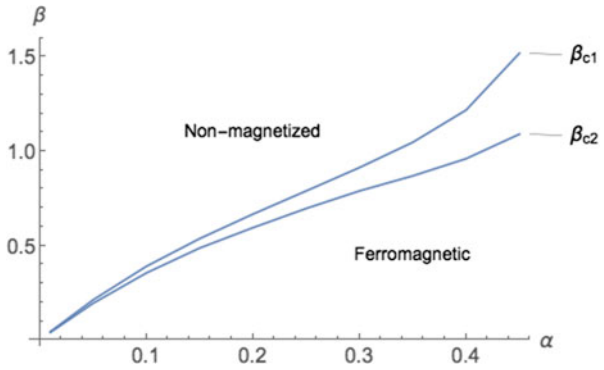


Fig. 1 Phase plot in the fine structure constant α vs $\beta = p_F/m$ plane showing the allowed region of spontaneous ferromagnetism

defects are not topologically stable and can unwind. However these processes are suppressed by a competition between the gradient energy and the extra energy stored in the domain walls, and there is a Ginzberg temperature T_G [21] below which thermal fluctuations cannot destabilise the walls trivially. The mechanism for destabilisation is then the one studied in detail in [28]. The rate for such decay is governed by an exponential factor $\exp(-B/\lambda)$ [29] where the exponent is the Euclidean action of a suitable "bounce" solution connecting the false and the true vacua [30]. On phenomenological grounds we need this complex to be stable for $\approx 10^{17}$ s. The bounce B is typically $\propto 1/\lambda$, where λ is a generic dimensionless coupling constant. Then large suppression factors $\sim 10^{-30}$ are natural for $\lambda \sim 0.01$. The other mechanism for disintegration of the DW network resides in the magnino gas becoming non-degenerate.

4 A Minimal Model for Dark Energy

We consider a hitherto unobserved sector with particle species we generically call M and Y. They are assumed to be oppositely charged under a local abelian group $U(1)_X$ with fine structure constant α_X. The mass m_M of M is assumed in the sub-eV range while the Y mass m_Y is assumed to be much larger. Charge neutrality requires that the number densities of the two species have to be equal, in turn this means that the Fermi energies are also the same. The hypothesis of larger mass is to ensure that Y with Compton wavelength $M^{-1} << p_F^{-1}$ does not enter into a collective magnetic phase.

We start our considerations at time t_1 when the temperature is just below T_G so that the wall complex has materialised. The parameters of this wall complex are ω, the thickness of individual walls and L the average separation between walls. On the scale of the horizon, the wall complex behaves just like a space filling homogeneous

substance. Further, due to the demand of neutrality, the heavier gas Y cannot expand either, although it has no condensation effects. Let us denote the number density of the magninos trapped in the walls to be n^X_{walls} and the remainder residing in the enclosed domains by n^X_{bulk}. Averaged (coarse grained) over a volume much larger than the L^3, this gives the average number density of the magninos to be

$$\langle n^X \rangle = \frac{\omega}{L} n^X_{\text{walls}} + \left(1 - \frac{\omega}{L}\right) n^X_{\text{bulk}} \tag{6}$$

And from the neutrality condition we have

$$\langle n^X \rangle = \langle n^Y \rangle \tag{7}$$

Then we can demand that PAAI in this phase acts as the DE, so that assuming Y to be non-relativistic, and ignoring other contributions,

$$\rho^Y \approx m_Y \langle n^Y \rangle = \rho_{\text{DE}} = 2.81 \times 10^{-11}\,(\text{eV})^4 \tag{8}$$

We can express the number density of Y as a ratio of the number density $n_\gamma = 3.12 \times 10^{-12}\,(\text{eV})^3$ of photons, and set $\eta^Y = \langle n^Y \rangle / n_\gamma$. Then we can obtain conditions that determine the ratio

$$\frac{m_M}{m_Y} = \frac{\beta^Y}{\beta} \approx (\eta^Y)^{4/3} \times 10^{-6} \lambda 1 \tag{9}$$

These are the essential constraints determining the key parameters of our model. Then we find that m_M ranges over 10^{-4} to 10^{-6} eV corresponding to η^Y ranging from 10^{-4} to 10^{-8} eV; and m_Y, respectively, ranges from 1 keV to 1 GeV. Further, we can also develop a corresponding multi-flavour dark sector so that Dark Matter can also be accommodated as neutral atoms of this sector. The details can be found in [11].

5 Origin of Cosmic Magnetic Fields

The origin and evolution of galactic scale magnetic fields is an open question [18, 31]. In particular the extent of seed magnetic field as against that generated by subsequent motion is probably experimentally distinguishable [19, 20]. In the present case, we can estimate the field strength of the X-magnetism in each domain, and is found to be

$$B_{\text{dom}} \approx \left(\frac{m_M}{\text{eV}}\right)^2 \left(\frac{e'}{e}\right) \left(\frac{\beta}{0.1}\right)^3 \times 2.2 \times 10^{-8} T \tag{10}$$

Since the domain structure is completely random we expect zero large-scale magnetic field on the average. Residual departure from this average can be estimated by assuming that the deviation from the mean grows as $\sqrt{N}$ as we include N domains. Thus if the X-magnetic field in individual domains has the value $B_{\rm dom}$, then on the scale of galactic clusters $L_{\rm gal}$ it possesses a root mean square value $\overline{\Delta B} \equiv B_{\rm dom}(L/L_{\rm gal})^{3/2}$.

Assuming $U(1)_X$ field mixes kinetically with standard electromagnetism through term of the form $\xi F^{\mu\nu}F^X_{\mu\nu}$, the ξ is well constrained from Supernova 1987A data to [32] $10^{-7} < \xi < 10^{-9}$. The exact value of the seed required depends on the epoch being studied and other model dependent factors [33]. We consider the possibility of a seed of $10^{-30}T$ with a coherence length of 0.1 kpc $\sim 3 \times 10^{18}$ m obtained with $\xi = 10^{-8}$.

$$\overline{\Delta B}_{\rm seed} = 10^{-30}T \sim 10^{-8} \times \left(\frac{m_M}{\rm eV}\right)^2 \left(\frac{e'}{e}\right) \beta^3 \left(\frac{L}{\rm m}\right)^{3/2} \times 10^{-40}T \tag{11}$$

From this, representative values for L for $\beta = 0.1$ are in the range 10^{11} to 10^{13} m which is solar system size. A detailed treatment to estimate the residual fluxes on large coherence length scales could trace the statistics of flux values in near neighbour domains and the rate at which the magnetic flux could undergo percolation, providing perhaps a smaller value for L, which should tally with above.

6 Conclusions

We have proposed the possibility of a negative pressure medium as arising from nothing more radical than a peculiar ground state of a pair of asymmetric fermion species interacting through an unbroken abelian gauge force. In an attempt to highlight the potential utility of the PAAI to cosmology, specifically to DE and to cosmic ferromagnetism, we have been agnostic about the earlier history of this sector. A study of temperature dependence of this phenomenon as also phenomenological inputs from the cosmic dawn data would help to sharpen this scenario.

7 Questions from the Audience

Here we address two of the more important questions raised variously by several members of the audience, which we take the liberty to recapitulate and freely paraphrase. We gratefully acknowledge these inputs as having sharpened our understanding of our proposal.

Q1 The condensed state of the magninos seems to define a special frame of reference. Does that not conflict with the standard cosmology?

A1 The magnino and accompanying particles form a homogeneous relativistic gas at a high temperature just like the visible sector in the standard model of cosmology. It thus defines its own comoving set of coordinates. Thus it will be necessary for this medium to be interacting with the standard sector during an early epoch so that the two define a common comoving frame. The new sector becomes "dark" only in the late universe. The emergent DW structure is then a comoving constant energy contribution.

Q2 What is the equation of the state of the spin-polarised ground state? Intuitively any medium consisting of ordinary quasi-particles should be subject to evolution with the corresponding equation of state and will not simulate constant energy density.

A2 The spin-polarised medium if infinite, has been estimated to satisfy $p = -0.1\rho$. However we note firstly that a negative value of effective w implies a strongly coupled medium. Further, the domain wall structure would be immune to expansion since it exists by virtue of local stresses whose strength is many orders of magnitudes greater than the local gravitational tidal force. For this reason we expect the DW complex to protect both itself and the strongly coupled quasi-particle gas from suffering tidal acceleration. Thus the energy density should remain constant, and averaged over an enormous number of domains, should be homogeneous.

Acknowledgments We thank NSERC, Canada for financial support and the Ministére des relations internationales et la francophonie of the Government of Québec for financing within the cadre of the Québec-Maharashtra exchange. RBM and MBP also thank IIT Bombay for financial support and hospitality.

References

1. S. Weinberg, Rev. Mod. Phys. **61**, 1 (1989)
2. N. Aghanim et al., (Planck) (2018), arXiv:1807.06209
3. M. Li, X.-D. Li, S. Wang, Y. Wang, Commun. Theor. Phys. **56**, 525 (2011). arXiv:1103.5870
4. U.K. Dey, T.S. Ray, U. Sarkar, Nucl. Phys. **B928**, 258 (2018). arXiv:1705.08484
5. J.I. Kapusta, Phys. Rev. Lett. **93**, 251801 (2004). arXiv:hep-th/0407164
6. R.A. Battye, M. Bucher, D. Spergel, Phys. Rev. **D60**, 043505 (1999). arXiv:astro-ph/9908047
7. R.A. Battye, A. Moss, Phys. Rev. **D76**, 023005 (2007). arXiv:astro-ph/0703744
8. L. Conversi, A. Melchiorri, L. Mersini-Houghton, J. Silk, Astropart. Phys. **21**, 443 (2004). arXiv:astro-ph/0402529
9. A. Friedland, H. Murayama, M. Perelstein, Phys. Rev. **D67**, 043519 (2003). arXiv:astro-ph/0205520
10. U.A. Yajnik, EPJ Web Conf. **70**, 00046 (2014)
11. R.B. MacKenzie, M.B. Paranjape, U.A. Yajnik (2019). arXiv:1901.00995
12. S. Raby, G. West, Phys. Lett. **B200**, 547 (1988)
13. S. Raby G. West, Phys. Lett. **B194**, 557 (1987)
14. J.L. Feng, M. Kaplinghat, H. Tu, H.-B. Yu, J. Cosmol. Astropart. Phys. **0907**, 004 (2009). arXiv:0905.3039
15. K.K. Boddy, M. Kaplinghat, A. Kwa, A.H.G. Peter, Phys. Rev. **D94**, 123017 (2016).

arXiv:1609.03592
16. J.M. Cline, Z. Liu, G. Moore, W. Xue, Phys. Rev. **D89**, 043514 (2014). arXiv:1311.6468
17. J.M. Cline, Z. Liu, W. Xue, Phys. Rev. **D85**, 101302 (2012). arXiv:1201.4858
18. R.M. Kulsrud, E.G. Zweibel, Rep. Prog. Phys. **71**, 0046091 (2008). arXiv:0707.2783
19. R. Durrer, A. Neronov, Astron. Astrophys. Rev. **21**, 62 (2013). arXiv:1303.7121
20. K. Subramanian, Rep. Prog. Phys. **79**, 076901 (2016). arXiv:1504.02311
21. T.W.B. Kibble, Phys. Rep. **67**, 183 (1980)
22. E.W. Kolb, M.S. Turner, *The Early Universe* (Addison-Wesley Pub. Co., Redwood City, 1990; revised 2003)
23. S. Dodelson, *Modern Cosmology* (Addison-Wesley Pub. Co., Boston, 2003)
24. A.K. Rajagopal, J. Callaway, Phys. Rev. B **7**, 1912 (1973)
25. G. Baym, S.A. Chin, Nucl. Phys. **A262**, 527 (1976)
26. S.A. Chin, Ann. Phys. **108**, 301 (1977)
27. B.X. Xu, A.K. Rajagopal, M.V. Ramana, J. Phys. C: Solid State Phys. **17**, 1339 (1984)
28. J. Preskill, A. Vilenkin, Phys. Rev. **D47**, 2324 (1993). arXiv:hep-ph/9209210
29. I. Yu. Kobzarev, L.B. Okun, M.B. Voloshin, Sov. J. Nucl. Phys. **20**, 644 (1975) [Yad. Fiz. **20**, 1229 (1974)]
30. S.R. Coleman, Phys. Rev. **D15**, 2929 (1977) [Erratum: Phys. Rev. **D16**, 1248(1977)]
31. R.M. Kulsrud, Ann. Rev. Astron. Astrophys. **37**, 37 (1999)
32. S. Davidson, S. Hannestad, G. Raffelt, J. High Energy Phys. **05**, 003 (2000). arXiv:hep-ph/0001179
33. L.M. Widrow, D. Ryu, D.R.G. Schleicher, K. Subramanian, C.G. Tsagas, R.A. Treumann, Space Sci. Rev. **166**, 37 (2012). arXiv:1109.4052

Three Partial Differential Equations in Curved Space and their Respective Solutions

Gopinath Kamath

Abstract The partial differential equation in curved space whose solution is the electrostatic potential A_0 is obtained following Whittaker (Whittaker, Proc R Soc A 116:720–735, 1927) in: a. $2 + 1$ dimensions with the metric $g_{\mu\nu}$ given by (a) Deser, Jackiw and 'tHooft (Deser et al., Ann Phys 152:220–235, 1984), and Clement (Clement, Int J Theor Phys 24:267, 1985) and (b) by Banados, Teitelboim and Zanelli (Banados et al., Phys Rev Lett 69:1849, 1992); and in b. $3 + 1$ dimensions with the cylindrically symmetric metric due to Trendafilova and Fulling (Trendafilova and Fulling, Eur J Phys 32:1663, 2011). While a series solution of the partial differential equation is easily obtained for two of the examples cited in $2 + 1$ and $3 + 1$ dimensions, respectively, a sensible solution eludes us with the Banados–Teitelboim–Zanelli (Banados et al., Phys Rev Lett 69:1849, 1992) black hole metric, inviting an alternative to Whittaker's approach (Whittaker, Proc R Soc A 116:720–735, 1927) that is exact and sketched here briefly.

Keywords Partial differential equations in curved spaces · Whittaker's method · $2 + 1$ and $3 + 1$ dimensions

1 Introduction

The method of images has been a go-to for solutions to problems in electrostatics; while the theoretical aspects of image charges have been discussed at length by Muller-Kirsten [1], its application to several problems has been dealt with in detail both by Jackson [2] and Muller-Kirsten [1]. Parenthetically, from the standpoint of a stress-tensor the method has also been adopted for the Casimir effect [3] for the parallel plate problem at zero and finite temperatures by Brown and Maclay [4]. Going further, a matrix solution to the Maxwell's equations in flat space that yields both the static and time-dependent solutions has been presented by Moses [5]

G. Kamath (✉)
Department of Mathematics, Indian Institute of Technology Tirupati, Tirupati, India

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_46

recently. The aforesaid methods motivate this report in an effort to determine the electrostatic and exact solutions to Maxwell's equation in curved space; while the former are easier to obtain as solutions of second-order partial differential equations using Whittaker's method [6] in two examples in 2 + 1 and 3 + 1 dimensions, respectively, and this will be dealt with first in the sequel, a sensible answer with Ref. [6] eludes us for the third example with the metric of Banados, Teitelboim and Zanelli [7], thus motivating—in the latter part of this report—an answer via the first-order partial differential equations of Moses [5] but with the orientation suggested by Plebanski [8].

We begin with Eq. (5) in Whittaker's paper [6], namely,

$$\sum_{q,r} g^{qr}\phi_{prq} = j_p \tag{1}$$

Editing it here with a change of notation to

$$g^{\mu\nu}A_{0:\mu;\nu} = j_0 \tag{2}$$

helps, as for electrostatic phenomena in a gravitational field $j_i, A_i\ i \neq 0$ are zero and the time derivatives of A_0 are zero. In 2 + 1 dimensions following Deser, Jackiw and 'tHooft [9] and Clement [10] one has

$$ds^2 = (cdt + \lambda d\theta)^2 - dr^2 - r^2 d\theta^2 \tag{3}$$

and Eq. (2) becomes

$$r^2\frac{\partial^2 A_0}{\partial r^2} + r\frac{\partial A_0}{\partial r} + \frac{\partial^2 A_0}{\partial \theta^2} = j_0 \tag{4}$$

where the $g_{\mu\nu}$ are given with $\lambda = \frac{kJ}{2\pi}, k = 8\pi G, J = \left|\overrightarrow{J}\right|$ by

$$g_{00} = 1, g_{10} = 0, g_{20} = \lambda, g_{11} = -1, g_{22} = -r^2 + \lambda^2, \tag{5}$$

G being the gravitational constant and J the spin of the massless particle. The counterpart of (3) in 3 + 1 dimensions for the cylindrically symmetric metric defined by

$$ds^2 = \left(\frac{a}{r}\right)c^2dt^2 - \frac{r}{a}dr^2 - r^2d\theta^2 - 0\left(\frac{r}{a}\right)^2 dz^2, a \neq 0, r^2 = x^2 + y^2 \tag{6}$$

is

$$r^2\frac{\partial^2 A_0}{\partial r^2} + 2r\frac{\partial A_0}{\partial r} + \frac{r}{a}\frac{\partial^2 A_0}{\partial \theta^2} + ar\frac{\partial^2 A_0}{\partial z^2} = j_0 \tag{7}$$

Eq. (7) bears comparison with Eq. (25) in Ref. [6], namely

$$\left(1-\frac{\alpha}{r}\right)\frac{\partial}{\partial r}\left(r^2\frac{\partial\phi_0}{\partial r}\right)+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\phi_0}{\partial\theta}\right)+\frac{1}{\sin^2\theta}\frac{\partial^2\phi_0}{\partial\phi^2}=0 \tag{8}$$

it being obtained on using the Schwarzschild metric with $g^{\mu\nu}$ given by

$$g^{00}=\frac{r}{r-\alpha},\ g^{11}=-c^2\frac{r-\alpha}{r},\ g^{22}=-\frac{c^2}{r^2},\ g^{33}=-\frac{c^2}{r^2\sin^2\theta} \tag{9}$$

A general solution to (8) has been given in Eq. (31) of that paper in terms of the product

$$P_n^1(z)\,P_m^n(\cos\theta)\,(a_m\cos m\ \phi+b_m\sin m\ \phi)\,,\alpha\,(z+1)\equiv 2r \tag{10}$$

and can similarly be determined here for the homogeneous versions of Eqs. (4) and (7).

While presenting the solutions to Eqs. (4) and (7) below we note that Eq. (6) is a special case of

$$\begin{aligned}&ds^2=\left(\tfrac{r}{a}\right)^j c^2dt^2-\left(\tfrac{r}{a}\right)^{j+k}dr^2-r^2d\theta^2-\left(\tfrac{r}{a}\right)^k dz^2,\\&a\neq 0,2\,(j+k)=-jk\end{aligned} \tag{11}$$

that was discussed by Trendafilova and Fulling [11] recently.

2 The Solutions to Eqs. (4) and (7)

A general solution to the homogeneous version of (4) is easily obtained as

$$A_0\,(r,\theta)=\sum_{m=0}^{\infty}\left(a_m r^m+b_m r^{-m}\right)(c_m\sin m\theta+d_m\cos m\theta) \tag{12}$$

Equally for Eq. (7) the radial part will be a solution of

$$r\frac{d^2R}{dr^2}+2\frac{dR}{dr}+\nu^2R=0,a\nu^2=a^2\ p^2-m^2 \tag{13}$$

with the θ and z dependence of the solution written in terms of the product

$$(a_m\cos m\theta+b_m\sin m\theta)\left(c_n e^{-pz}+d_n e^{pz}\right) \tag{14}$$

in the general solution to the homogeneous version of (7). Eq. (13) has the general solution

$$R(r) = \frac{1}{\nu\sqrt{r}}\left(C_1 J_1\left(2\nu\sqrt{r}\right) + C_2 Y_1\left(2\nu\sqrt{r}\right)\right) \tag{15}$$

with J_1 and Y_1 being the Bessel functions of the first and second kind, respectively, of order 1; and the general solution to the homogeneous version of Eq. (7) will, therefore, be

$$\begin{array}{l}\phi\,(r,\theta,z) = \\ \int\limits_{-\infty}^{\infty} dp \sum\limits_{m=0}^{\infty} \left\{ \begin{array}{l} \frac{1}{\nu\sqrt{r}}\left(C_1 J_1\left(2\nu\sqrt{r}\right) + C_2 Y_1\left(2\nu\sqrt{r}\right)\right) \\ \left(a_m \cos m\,\theta + b_m \sin m\,\theta\right)\left(c_p\theta(p)e^{-pz} + d_p\theta\,(-p)\,e^{pz}\right) \end{array} \right\}\end{array} \tag{16}$$

While Eqs. (12) and (16) are the general solutions, one also has the radial functions of $\phi(r) = \ln r$ and $\phi(r) = \frac{1}{r}$ associated with the Coulomb potential as obvious solutions to Eqs. (4) and (7), respectively; in this context, there have been several calculations [12–14] on the correction to the hydrogen spectrum from the inclusion of the Schwarzschild metric, for example, as a perturbation. However, a complementary investigation involving the quantum mechanics of a hydrogen ion, for instance, subject to the Coulomb-like potential expressed by (12) and (16) is clearly preempted as the solutions are infinite series rather than a sum; to this end it clearly pays to obtain the required answer from the method of Moses [5] tempered with the orientation due to Plebanski [8] that will be introduced in the latter part of this paper, if only because it has none of the infirmities that are evident in this section.

2.1 A Third Example

In 2 + 1 dimensions the metric of Banados, Teitelboim and Zanelli [7] is got from

$$\begin{array}{l} ds^2 = -\left(N^{\perp}\right)^2 dt^2 + \frac{1}{f^2}dr^2 + r^2\left(d\theta + N^{\theta}dt\right)^2, \\ 2r^2 N^{\theta} = -J,\ \left(N^{\perp}\right)^2 = f^2 = -M + \frac{r^2}{l^2} + \frac{J^2}{4r^2} \end{array} \tag{17}$$

and in polar coordinates it works to

$$g_{00} = M - \frac{r^2}{l^2},\ g_{11} = \frac{1}{f^2},\ g_{02} = -\frac{J}{2},\ g_{22} = r^2 \tag{18}$$

The counterpart of Eqs. (4) and (7) is now given by

$$f^2\frac{\partial^2 A_0}{\partial r^2}+\frac{1}{r}\left(f^2-\frac{J^2}{2r^2}\right)\frac{\partial A_0}{\partial r}-\frac{M-\frac{r^2}{l^2}}{f^2r^2}\frac{\partial^2 A_0}{\partial\theta^2}-\left(\frac{2}{l^2}+\frac{J^2}{4l^2f^2r^2}\right)A_0=j_0 \tag{19}$$

the homogeneous version of which becomes

$$\frac{d^2R}{dr^2}+\frac{1}{r}\left(1-\frac{J^2}{2f^2r^2}\right)\frac{dR}{dr}+\frac{1}{f^2}\left(\frac{m^2}{f^2r^2}\left(M-\frac{r^2}{l^2}\right)-\left(\frac{2}{l^2}+\frac{J^2}{4l^2f^2r^2}\right)\right)R=0 \tag{20}$$

with $A_0(r,\theta)=R(r)H(\theta)$ and m an integer following from

$$\frac{\partial^2 H}{\partial\theta^2}+m^2H=0 \tag{21}$$

Eq. (20) has a regular singular point at $r=0$ and with $h(r)=r^2f^2$ one gets

$$h^2\frac{d^2R}{dr^2}+\frac{h^2}{r}\left(1-\frac{J^2}{2h}\right)\frac{dR}{dr}+r^2\left[m^2\left(M-\frac{r^2}{l^2}\right)-\frac{2}{l^2}\left(h+\frac{J^2}{8}\right)\right]R=0 \tag{22}$$

Eq. (22) is solved through transformations following Polyanin and Zaitsev [15] so as to reconcile it with

$$4g^2\frac{d^2y}{dx^2}-\left(2g\frac{d^2g}{dx^2}-\left(\frac{dg}{dx}\right)^2+b\right)y=0 \tag{23}$$

the solution to which is an elementary function whose form depends on the constant b. Reworking Eq. (22) to

$$\begin{aligned}&4h^2\frac{d^2K}{dt^2}-\left(m^2\frac{t}{l^2}+\frac{t^2}{l^4}-a\right)K=0,\\&4a\equiv 7M^2-\frac{5J^2}{l^2}+4Mm^2,\ t\equiv r^2\end{aligned} \tag{24}$$

and writing the solution to (24) as

$$K(t)=h^c\exp\frac{b}{2}\int\frac{dt}{h} \tag{25}$$

the constants c, b work to

$$\begin{aligned} 4c &= 1 \pm \sqrt{2}, \\ 2b &= \left(m^2 + M\right)\left(1 \pm \sqrt{2}\right) \end{aligned} \tag{26}$$

with the parameter m^2 hitherto free, as a solution to

$$4c^2 m^4 + (4c+1)\, m^2 M + \frac{1}{4}\left(7M^2 + \left(-3 \pm 2\sqrt{2}\right)\frac{J^2}{l^2}\right) = 0 \tag{27}$$

Clearly, $m = 0$ is not a root of this equation as it would imply that

$$J^2 = 7\left(-3 \mp 2\sqrt{2}\right) M^2 l^2 \tag{28}$$

Thus it would seem that there are no θ-dependent solutions to (19) with the metric given by (18). That this conclusion is wrong is easily checked from a Frobenius approach to the differential equation; as an alternative therefore to Whittaker's method for Eq. (17), the sequel suggests a way out may consist in solving Maxwell's equations as first-order partial differential equations with constant coefficients in curved space following Moses [5] and Plebanski [8]. To this end we define following Ref. [8],

$$D_a = \sqrt{-g}\, f^{0a},\ H = \sqrt{-g}\, f^{21},\ E_a = f_{a0},\ B = f_{21} \tag{29}$$

to obtain

$$\begin{aligned} D_{i,i} &= -\frac{4\pi}{c} j^0 \\ D_{1,0} + H_{,2} &= \frac{4\pi}{c} j^1 \\ D_{2,0} - H_{,1} &= \frac{4\pi}{c} j^2 \\ -B_{,0} + E_{2,1} - E_{1,2} &= 0 \end{aligned} \tag{30}$$

as Maxwell's equations, noting that in Cartesian coordinates (see Eq. (43) below) $\sqrt{-g} = 1$. The definitions

$$\Psi \equiv (B + iH \;\; E_1 + iD_1 \;\; E_2 + iD_2 \;\; 0)^T,$$
$$\Gamma \equiv \left(0 \;\; \frac{4\pi i}{c} j^1 \;\; \frac{4\pi i}{c} j^2 \;\; \frac{4\pi i}{c} j^0\right)^T \tag{31}$$

now enable one to rewrite Eqs. (30) as a matrix equation with 4×4 matrices as

$$(I\partial_0 + \alpha_1 \partial_1 + \alpha_2 \partial_2)\, \Psi = \Gamma \tag{32}$$

with I being the unit matrix and the α's defined by

$$\alpha_1 = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \alpha_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \tag{33}$$

The operator $\vec{\alpha} \cdot \vec{\partial}$ has eigenvalues $\pm p, \pm ip$ with $p = \left|\vec{p}\,\right|$, with the respective linearly independent set of orthonormal eigenvectors given by

$$\begin{aligned} \sqrt{2}pX_+ &= e^{i\vec{p}\cdot\vec{x}} \begin{pmatrix} 0 & ip_1 & ip_2 & p \end{pmatrix}^T, \; \sqrt{2}pX_- = e^{i\vec{p}\cdot\vec{x}} \begin{pmatrix} 0 & p_1 & p_2 & ip \end{pmatrix}^T \\ \sqrt{2}pY_+ &= e^{i\vec{p}\cdot\vec{x}} \begin{pmatrix} p & p_2 & -p_1 & 0 \end{pmatrix}^T, \; \sqrt{2}pY_- = e^{i\vec{p}\cdot\vec{x}} \begin{pmatrix} -p & p_2 & -p_1 & 0 \end{pmatrix}^T \end{aligned} \tag{34}$$

A diligent application of the work by Moses [5] now helps to determine the matrix elements in

$$\Psi = (B + iH \;\; E_1 + iD_1 \;\; E_2 + iD_2 \;\; 0)^T$$

in terms of those in

$$\Gamma = \left(0 \;\; \frac{4\pi i}{c} j^1 \;\; \frac{4\pi i}{c} j^2 \;\; \frac{4\pi i}{c} j^0\right)^T$$

exactly. For time-independent solutions one easily obtains the answer

$$\Psi(x) = \frac{4\pi i}{c} \int_p \int_z \frac{e^{i\vec{p}\cdot(\vec{x}-\vec{z})}}{2p^3}$$
$$\left(-2ip\left(p_2 j^1 - p_1 j^2\right) \;\; 2ipp_1 j^0 \;\; 2ipp_2 j^0 - 2ip\left(p_1 j^1 + p_2 j^2\right)\right)^T \tag{35}$$

The last term in (35) will be zero by current conservation and the momentum integration yields

$$\Psi(x) = \frac{2i}{c}\int_z \left(\frac{\vec{j}\times(\vec{x}-\vec{z})}{|\vec{x}-\vec{z}|^2} \; -\frac{j^0}{2}\partial_x \log|\vec{x}-\vec{z}|^2 \; -\frac{j^0}{2}\partial_y \log|\vec{x}-\vec{z}|^2 \; 0 \right)^T$$
$$\equiv \left(B + iH \;\; E_1 + iD_1 \;\; E_2 + iD_2 \;\; 0 \right)^T \tag{36}$$

The second equality in (36) yields

$$\vec{E} + i\vec{D} = -i \,\text{grad}\left(\log\left|\vec{x} - \vec{R}\right|^2 \right) \tag{37}$$

when $j^0 = c\delta\left(\vec{z} - \vec{R}\right)$ reflecting an unit charge at $\vec{z} = \vec{R}$. By extension one also has

$$B + iH = \frac{2i}{c}\int_z \frac{\vec{j} \times \left(\vec{x} - \vec{z}\right)}{\left|\vec{x} - \vec{z}\right|^2} \tag{38}$$

The relation between $\left(B, \vec{E}\right)$ and $\left(H, \vec{D}\right)$ follows from

$$f_{a0} = g_{ab}g_{0c}f^{ab},\; f_{21} = g_{2\mu}g_{1\nu}f^{\mu\nu} \tag{39}$$

as $D_a = f^{0a}$, $H = f^{21}$, $E_a = f_{a0}$, $B = f_{21}$ vide Eq. (27) above. To elaborate one has for Eq. (39)

$$f_{a0} = (g_{a0}g_{01} - g_{a1}g_{00})\, f^{01} + (g_{a0}g_{02} - g_{a2}g_{00})\, f^{02} + (g_{a2}g_{01} - g_{a1}g_{20})\, f^{21}$$
$$f_{21} = (g_{20}g_{11} - g_{21}g_{10})\, f^{01} + (g_{20}g_{12} - g_{22}g_{10})\, f^{02} + (g_{22}g_{11} - g_{21}g_{12})\, f^{21} \tag{40}$$

and with a little effort one gets from Eqs. (40)

$$\begin{pmatrix} B + iH \\ E_1 + iD_1 \\ E_2 + iD_2 \end{pmatrix} = K \begin{pmatrix} f^{01} \\ f^{02} \\ f^{21} \end{pmatrix} \tag{41}$$

where

$$K = \begin{pmatrix} (g_{20}g_{11} - g_{21}g_{10}) & (g_{20}g_{12} - g_{22}g_{10}) & (g_{22}g_{11} - g_{21}g_{12} + i) \\ (g_{10}g_{01} - g_{11}g_{00} + i) & (g_{10}g_{02} - g_{12}g_{00}) & (g_{12}g_{01} - g_{11}g_{20}) \\ (g_{20}g_{01} - g_{21}g_{00}) & (g_{20}g_{02} - g_{22}g_{00} + i) & (g_{22}g_{01} - g_{21}g_{20}) \end{pmatrix} \tag{42}$$

One can now determine through matrix inversion the column vector on the right-hand side of (41) as Eqs. (38) and (39) define the left-hand side of Eq. (41) for an unit charge at $\overrightarrow{z} = \overrightarrow{R}$.

This is a lengthy calculation that will be pursued in detail elsewhere, but to wrap up this paper we give below in Cartesian coordinates the $g_{\mu\nu}$ that is the counterpart of the polar versions given by Eqs. (18) and (5), respectively, with $f^2 = -M + \frac{r^2}{l^2} + \frac{J^2}{4r^2}$, $r^2 = x^2 + y^2$:

$$\begin{gathered} g_{00} = M - \frac{r^2}{l^2},\ g_{01} = \frac{jy}{2r^2},\ g_{02} = -\frac{jx}{2r^2}, \\ g_{11} = \frac{x^2+y^2f^2}{r^2f^2},\ g_{12} = \frac{xy}{r^2}\left(\frac{1}{f^2-1}\right),\ g_{22} = \frac{y^2+x^2f^2}{r^2f^2} \end{gathered} \tag{43}$$

and

$$\begin{gathered} g_{00} = 1,\ g_{01} = -\frac{\lambda y}{x^2+y^2},\ g_{02} = \frac{\lambda x}{x^2+y^2}, \\ g_{11} = -1 + \left(\frac{\lambda y}{z^2+y^2}\right)^2,\ g_{12} = -\frac{\lambda^2 xy}{\left(x^2+y^2\right)^2},\ g_{22} = -1 + \left(\frac{\lambda x}{x^2+y^2}\right)^2 \end{gathered} \tag{44}$$

Each of Eqs. (43) and (44) leads through Eqs. (41) and (42) to the time-independent $\overrightarrow{D}(x)$, $H(x)$ for the metric given in (18) and (5), respectively, and they will be the counterparts in curved space of the Coulomb potential of $\phi(r) = \ln r$ in $2 + 1$ dimensional flat space. Needless to say so and in conclusion, one can simply repeat the above steps to obtain for the cylindrically symmetric metric of Eq. (6) the counterpart of the Coulomb potential of $\phi(r) = \frac{1}{r}$.

References

1. J.W.H. Muller-Kirsten, *Electrodynamics: An Introduction Including Quantum Effects* (World Scientific Publishing, Singapore, 2004)
2. J.D. Jackson, *Classical Electrodynamics*, 8th edn. (Wiley Eastern, New Delhi, 1989)
3. H.B.G. Casimir, On the attraction between two perfectly conducting plates. Proc. Kon. Ned. Akad. Wetensch. **51**, 793 (1948)
4. L.S. Brown, G.J. Maclay, Vacuum stress between conducting plates: an image solution. Phys. Rev. **184**, 1272 (1969). https://doi.org/10.1103/PhysRev.184.1272
5. H.E. Moses, Solution of Maxwell's equations in terms of a Spinor notation: the direct and inverse problem. Phys. Rev. **113**, 1670 (1959). https://doi.org/10.1103/PhysRev.113.1670
6. E.T. Whittaker, On electric phenomena in gravitational fields. Proc. R. Soc. A **116**, 720–735 (1927). https://doi.org/10.1098/rspa.1927.0160
7. M. Banados, C. Teitelboim, J. Zanelli, Black hole in three-dimensional spacetime. Phys. Rev. Lett. **69**, 1849 (1992). https://doi.org/10.1103/PhysRevLett.69.1849
8. J. Plebanski, Electromagnetic waves in gravitational fields. Phys. Rev. **118**, 1396 (1960). https://doi.org/10.1103/PhysRev.118.1396
9. Deser S, Jackiw R and 'tHooft G (1984), Three-dimensional Einstein gravity: dynamics of flat space, Ann. Phys. 152:220–235. https://doi.org/10.1016/0003-4916(84)90085-X

10. G. Clement, Stationary solutions in three-dimensional general relativity. Int. J. Theor. Phys. **24**, 267 (1985). https://doi.org/10.1007/BF00669791
11. C.S. Trendafilova, S.A. Fulling, Static solutions of Einstein's equations with cylindrical symmetry. Eur. J. Phys. **32**, 1663 (2011). https://doi.org/10.1088/0143-0807/32/6/020
12. Z.-H. Zhao, Y.-X. Liu, X.-G. Li, The energy-level shifts of a stationary hydrogen atom in the static external gravitational field with Schwarzschild geometry. Phys. Rev. **D76**, 064016 (2007). https://doi.org/10.1103/PhysRevD.76.064016
13. L. Parker, L.O. Pimentel, Gravitational perturbation of the hydrogen spectrum. Phys. Rev. **D25**, 3180 (1982). https://doi.org/10.1103/Phys.Rev.D.25.3180
14. E. Fischbach, B.S. Freeman, C. Wen-Kwei, General relativistic effects in hydrogenic systems. Phys. Rev. **D23**, 2157 (1981). https://doi.org/10.1103/Phys.Rev.D23.2157
15. A.D. Polyanin, V.F. Zaitsev, *Handbook of Exact Solutions for Ordinary Differential Equations*, 2nd edn. (Chapman and Hall/CRC, Boca Raton, FL, 2003)

Part V
Integrability

What Does the Central Limit Theorem Have to Say About General Relativity?

Réjean Plamondon

Abstract In this paper, we speculate on a possible connection between the Bayes's law and the Einstein's general relativity equation to support the use of a metric based on an *erfc* gravitational potential that has been recently proposed to provide some cues to open problems in the solar systems. Starting from a basic interdependence premise, an analogy between Einstein's equation and Bayes's law is used to analyze the linear case of a weak field static symmetric massive object, providing a probabilistic context that takes into account the probability of presence of a given energy density in its corresponding 4D curved space-time manifold. Using the Central Limit Theorem to model globally the very slow process of star formation and mathematically express the corresponding probability density, the new framework provides a rationale for the emergence of a weighted Newton's law of gravitation. One key feature of this modified gravity model is that it relies on the existence of an intrinsic emergent physical constant σ, a star-specific proper length that scales all its surroundings.

Keywords Modified gravity · Weighted Newton's law · Central limit theorem · Bayes's law · Star proper length · *erfc* potential · Emergence · Self-organization

In its most general configuration, a complex system is a network of heterogeneous and usually simpler subsystems that interact among each other to give rise to emergent features that guides its self-organization into a more complex system. The description of the whole process simplifies at a given level of representation, leading to some emergent properties [1]. These emergent systems are omnipresent in physics, chemistry, and biology [2].

Among the tools that can be used to study such systems and their convergence is the Central Limit Theorem [3]. This theorem has been developed over four centuries in the context of searching for the asymptotic probabilistic behavior of a

R. Plamondon (✉)
Département de génie électrique, Polytechnique Montréal, Montréal, QC, Canada
e-mail: rejean.plamondon@polymtl.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_47

sum of independent or quasi-independent random variables. The key feature of this theorem, which makes it practical for the study of complex systems, is that although the details of the individual sub-processes are unknown, the behavior of the whole system can be predicted, under some non-restrictive conditions, to converge toward multivariate Gaussian functions.

In this paper, we use this modeling approach to conjecture about emergent gravity, speculating from a possible connection between the Bayes's law and the Einstein's general relativity equation. The idea of modifying gravity to come up with new relativistic field descriptions has been proposed time and again in the last decades to provide among other things alternative explanations to some open problems in astronomy and astrophysics [4, 5, 6]. These extensions aimed at correcting and enlarging Einstein's theory to encompass several shortcomings when cosmological, astrophysical, mathematical, and quantum mechanical observations and objections are taken into account [7, 8, 9]. In this mindset, in a recent paper [10], the static non-empty symmetric geometry described by a metric based on an *erfc* gravitational potential has been proposed and studied in detail. This new metric provides a consistent set of predictions and interpretations regarding some open problems in the solar system, like the fly-by anomalies, the secular increase of the astronomical unit, the residual Pioneers' delays [11].

In the present manuscript, a fundamental question is addressed: can we lay the foundations for an emergent model that predicts the existence of an *erfc* potential using the central limit theorem? In the next section, starting from a complementarity that has been pointed out by Wheeler, we propose a comparison between Einstein's equation and Bayes's law of conditional probabilities and use it to support our analogical and speculative argumentation. The whole framework relies on a global probabilistic description of a star formation from which a fundamental law of gravitation comes out as a consequence of an asymptotic convergence predicted by the Central Limit Theorem. In Sect. 2, we put the general relativity in a probabilistic context and in Sect. 3, we present the conditions under which a weighted Newton's law automatically emerges from this new scheme and then conclude.

1 Introducing a Probabilistic Context in General Relativity

Einstein's gravitation equation, which links the space-time curvature tensor G to the energy-momentum tensor T,

$$G = KT \tag{1}$$

has been encapsulated by Wheeler as, "Space-time tells matter how to move; matter tells space-time how to curve." [12]. This points out an interesting interdependence that can be used to put general relativity into a probabilistic context if this assertion is converted into a general and fundamental premise:

"Space-time curvature (S) and energy-momentum (E) are two inextricable descriptive approaches to define the physically observable probabilistic universe (U); they must be mutually exploited to describe any subset U_i of this universe. The probability of observing and describing a given subset of the universe $\mathrm{P}(U_i)$, i.e. the joint probability $\mathrm{P}(S_i,E_i)$, can be studied from two equivalent methods: either by analyzing the curvature of space-time S_i after hypothesizing a given energy-momentum E_i or by analyzing the energy-momentum E_i under the hypothesis of a given space-time curvature S_i. In terms of conditional probabilities, this leads to two equivalent descriptions:

$$\mathrm{P}\left(U_i\right) = \mathrm{P}\left(S_i,\, E_i\right) = \mathrm{P}\left(S_i/E_i\right)\mathrm{P}\left(E_i\right) = \mathrm{P}\left(E_i/S_i\right)\mathrm{P}\left(S_i\right) \tag{2}$$

Using the corresponding probability density function $f()$ of these conditional probabilities $P()$ and rewriting (2) in a 4D Bayesian format, we get:

$$\begin{gathered} f\left(S_{\mu\nu}/E_{\mu\nu}\right) f\left(E_{\mu\nu}\right) = f\left(E_{\mu\nu}/S_{\mu\nu}\right) f\left(S_{\mu\nu}\right) \\ f\left(S_{\mu\nu}/E_{\mu\nu}\right) = \frac{f\left(S_{\mu\nu}\right)}{f\left(E_{\mu\nu}\right)} \times f\left(E_{\mu\nu}/S_{\mu\nu}\right) \end{gathered} \tag{3}$$

In other words, we consider the space-time curvature and the energy-momentum tensors as continuous 4D random variables and the values of their probability density functions define the probability that these random variables have a particular range of values within an infinitesimal space-time interval, providing an estimate of the relative likelihood that these random variables have these values in this interval.

This latter equation can be linked to Einstein's Eq. (1) through the following analogy:

$$G_{\mu\nu} \iff f\left(S_{\mu\nu}/E_{\mu\nu}\right) \tag{4}$$

and

$$KT_{\mu\nu} \iff \frac{f\left(S_{\mu\nu}\right)}{f\left(E_{\mu\nu}\right)} \times f\left(E_{\mu\nu}/S_{\mu\nu}\right) \tag{5}$$

In other words, $f(S_{\mu\nu}/E_{\mu\nu})$ can be interpreted as describing the probability of space-time to be curved under the conditional probability of observing a given energy-momentum density $\left[\frac{f\left(S_{\mu\nu}\right)}{f\left(E_{\mu\nu}\right)} \times f\left(E_{\mu\nu}/S_{\mu\nu}\right)\right]$, which can be linked to $G_{\mu\nu}$ and $T_{\mu\nu}$, respectively.

2 Emergence of a Weighted Newton's Law of Gravitation

Under weak field, low speed, classical conditions, only the 00-component of (1) is significant:

$$G_{00} = R_{00} - \frac{1}{2} g_{00} R = K T_{00} \tag{6}$$

Applying the previous analogy to such a system and using dimensional analysis, one can associate $f(S_{00}/E_{00})$, in $(\mathrm{E/L}^{-4})$, with the probability density of the space-time subset S_{00} to be curved given a matter-energy E_{00}, to G_{00}, a curvature component in (L^{-2}):

$$f\left(S_{00}/E_{00}\right) = k_1 G_{00} \tag{7}$$

Similarly,

$$\frac{f\left(S_{00}\right)}{f\left(E_{00}\right)} f\left(E_{00}/S_{00}\right) = k_2 K T_{00} \tag{8}$$

where the coefficients k_1 and k_2 are unit-balancing constants.

A way to perform estimate $f(E_{00}/S_{00})$ is to analyze the very slow process of star formation using a simple stochastic model. Assuming that in a remote and isolated part of the Universe, a star is slowly building up from the gradual agglomeration of chunks of matter-energy. Considering these chunks as random variables described by their own density functions, this process, which involves hydrodynamics, thermodynamics, radiation transport, etc., is equivalent, from a global probabilistic point of view, to adding up random variables, i.e. making the convolution of their corresponding probability density functions. Since these densities respect the Lindeberg conditions [13], in the sense that they are real, normalized, non-negative functions with a finite third moment and a scaled dispersion, then the Central Limit Theorem applies and predicts that in a flat Euclidean space-time, when the number of random chunks is very large $(\mathrm{N} \to \infty)$,

$$f\left(E_{00}/S_{00}\right) \propto f\left(\mathbf{x}\right) = \lim_{N \to \infty} \left[f_1\left(\mathbf{x}\right) * f_2\left(\mathbf{x}\right) * \cdots * f_N\left(\mathbf{x}\right)\right] \tag{9}$$

and the ideal form of the global probability density $f(\mathbf{x})$ will be a Gaussian multivariate and will tend to the following general form:

$$f\left(\mathbf{x}\right) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right] \tag{10}$$

where $\mathbf{x}$ is an n dimension random vector measuring the distance from the mean vector $\boldsymbol{\mu}$ of the distribution, and $\Sigma = \mathrm{E}[(\mathbf{x} - \boldsymbol{\mu}), (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}}]$ is the statistical covariance matrix measuring the expected (E) dispersion of this distribution.

For a 4D pseudo-Euclidean static system $n = 4$, with the quadri-vector $\mathbf{x} = \left(c\bar{t}, x, y, z\right)$, centered at $\boldsymbol{\mu} = \left(c\bar{t}, 0, 0, 0\right)$, Eq. (10) can be rewritten as:

$$f\left(\overline{r}\right) = \frac{1}{4\pi^2\sigma^4}\exp\left(-\frac{\overline{r}^2}{2\sigma^2}\right) \tag{11}$$

where $\overline{r}$ is the Euclidian distance $\left(\overline{r}^2 = x^2 + y^2 + z^2\right)$ from the zero-centered mean of the $f\left(\overline{r}\right)$ density. The diagonal covariance matrix (Σ) reduces to σ^2, a weighting parameter that scales the norm of the quadri-vector **x**, a Lorentz invariant. This scalar σ^2 is de facto a Lorentz invariant, an intrinsic and emergent feature of the central limit process. It reflects the system intrinsic proper length. This specific scale is the basic feature that can be used to get a curved space description $f\left(\widehat{r}\right)$ of the star's probability density and to point out some of its specific inherent properties.

Indeed, Eq. (11) is not practically useful in its present form, since it is only valid in a flat space-time that is, when an observer is at infinity from the star or locally, on a geodesic.

In other words, $\overline{r}$ defines the distance from the apparent star center as seen from infinity in a hypothetical flat space while $\widehat{r}$ defines the physical curvilinear distance from the star center in the curved space-time.

In their simplest algebraic form, the relationship between $\overline{r}$ and $\widehat{r}$ can be summarized as follows:

$$\left[\begin{array}{l}\overline{r} = 0 \text{ at } \widehat{r} = \infty \\ \overline{r} = \infty \text{at } \widehat{r} = 0\end{array}\right] \Rightarrow \overline{r} = \frac{s}{\widehat{r}} \tag{12}$$

where s is a scale parameter that can be determined from the invariance of σ:

$$\overline{r} = \sigma = \widehat{r} \Rightarrow s = \sigma^2 \tag{13}$$

This leads to making the following change of coordinates:

$$\frac{\overline{r}}{\sigma} = \frac{\sigma}{\widehat{r}} \tag{14}$$

to get $f\left(\widehat{r}\right)$, a projection of $f\left(\overline{r}\right)$ on a manifold of variable curvature described locally by the coordinate $\widehat{r}$. Making this change of coordinates making sure that the normalization of the probability densities in both representation spaces is maintained, this leads to:

$$f\left(\widehat{E}_{00}/\widehat{S}_{00}\right) = k_3 f\left(\widehat{r}\right) = \frac{k_3}{4\pi^2\sigma^2\widehat{r}^2}\exp\left(\frac{-\sigma^2}{2\widehat{r}^2}\right) \tag{15}$$

Equation (15) expresses the probability density of finding the star within an equivalent 3-ball of radius $\widehat{r}$, in a curved manifold under static, symmetric, weak field, and low speed conditions.

Pursing on this analogy, one can define the energy priori probability density by:

$$f\left(E_{00}\right) = \frac{1}{M_{\text{tot}}c^2} \tag{16}$$

Taking into account the mapping defined by Eq. (14) $f(S_{00})$ can be estimated in two steps. First, an invariant reference surface, valid both in the flat and curved descriptions, must be established. This is done assuming that the total energy of the star is distributed on the reference 2-spheres of constant curvature $1/\sigma^2$ in both representations:

$$f\left(S_{00(\bar{r}=\sigma)}\right) = \frac{1}{4\pi\sigma^2} = f\left(S_{00\left(\widehat{r}=\sigma\right)}\right) \tag{17}$$

Second, a value of $f\left(S_{00(\bar{r})}\right)$ valid all over the flat space is obtained by weighting by the corresponding 3-ball volumes defined at $\bar{r}$ and σ, respectively:

$$f\left(S_{00(\bar{r})}\right) = \frac{1}{4\pi\sigma^2}\frac{V_{3\text{b}.\bar{r}}}{V_{3\text{b}.\sigma}} = \frac{1}{4\pi\sigma^2}\frac{\frac{4\pi}{3}\bar{r}^3}{\frac{4\pi}{3}\sigma^3} = 4\pi\sigma^2\frac{\bar{r}^3}{\sigma^3} \tag{18}$$

which, using Eq. (14), leads to an expression for $f\left(S_{00\left(\widehat{r}\right)}\right)$ valid at any corresponding $\bar{r}$ position:

$$f\left(S_{00\left(\widehat{r}\right)}\right) = \frac{1}{4\pi\sigma^2}\frac{\sigma^3}{\widehat{r}^3} = \frac{\sigma}{4\pi\,\widehat{r}^3} \tag{19}$$

In other words, the mapping resulting from Eq. (14) guarantees that the energy-momentum tensor $T_{00\left(\widehat{r}\right)}$ affecting the curvature at a radial distance $\widehat{r}$ in the curved space is consistent with the energy-momentum tensor component $T_{00(\bar{r})}$ that would be measured in an ideal flat space at the corresponding distance $\bar{r}$ from the Gaussian density center.

Substituting Eqs. (15) and (19) in Eq. (6) leads to the description of a central force field as a function of the curvilinear distance from the star center:

$$\nabla^2\Phi\left(\text{r}\right) = \frac{2Kc^2\sigma^2Mc^2}{(4\pi\sigma)^3r^5}\exp\left(-\frac{\sigma^2}{2r^2}\right), \tag{20}$$

where $K = k_3/k_1$ (with k_1 and k_2 in L^{-2} and k_3 in $\text{L}^4\,\text{E}^{-1}$) and where, from now on, the curved hat over the coordinate r is omitted $\left(\widehat{r} \rightarrow r\right)$.

This Laplacian can be solved to get an expression for the magnitude of the linear radial gravitational field:

$$g(r) = -\left|\vec{\nabla}\,\Phi\,(\mathrm{r})\right| = -\frac{2Kc^2Mc^2}{(4\pi\sigma)^3 r^2}\exp\left(-\frac{\sigma^2}{2r^2}\right) \tag{21}$$

which reduces to the Newton description for large r values.

The gravitational potential can be obtained, if Eq. (21) is integrated:

$$\Phi(r) = \int_0^r g(r)dr = \frac{2Kc^2Mc^2}{(4\pi\sigma)^3}\left(\frac{\sqrt{\pi}}{\sqrt{2}\sigma}\right) erfc\left(\frac{\sigma}{\sqrt{2}r}\right) = \Phi_{erfc}(r) \tag{22}$$

where the integration limits, from 0 to r, are consistent, according to Eq. (14), with integration from $\bar{r} = \infty$ to 0 in the corresponding flat space representation, which leads to an *erfc* potential. In other words, in a flat space, the gravitational potential is fixed to zero at $\bar{r} = \infty$ and the Minkowskian metric is recovered under this condition. But when the flat model is projected into a curved space, according to the inverse relationship (14) between $\bar{r}$ and $\widehat{r}$, Eq. (22) predicts a constant potential at r = infinity. This is the particular feature of an *erfc* potential which leads to an original description of the space-time surrounding a massive object as previously published [7, 8]. Eq. (22) also converges toward the Newton limit, if the constant term included in the *erfc* function is arbitrarily subtracted, which leads to an *erf* potential that tends toward a $1/r$ behavior at large r values.

3 Conclusion

Under the paradigm of a self-organizing universe, the laws of physics should emerge from the space-time and matter-energy distribution. From a global perspective, the analogical and speculative approach presented in this paper can be seen as a heuristic strategy to mathematically take into account Mach's principle [9]. On top of Einstein's arguments [14], Eq. (3) provides a rationale based on a fundamental law of probabilities, the Bayes's law [15]. Applying this interdependence principle and the Central Limit Theorem, we have pinpointed a possible explanation for the emergence of a weighted Newton's law of gravitation in such a system. One key feature of the present theory [16] is that it is based on the existence of an intrinsic star-specific physical constant, the parameter σ^2, which automatically emerges from a convergence process described by the Central Limit Theorem. As reported in [7, 8], the new *erfc* potential once incorporated into a spherically symmetric metric, describes various features of the resulting modified Schwarzschild geometry. For example, computing the systematic errors that emerge when the effect of σ is

neglected, the Hubble constant H_0 can be linked to σ_{Sun} and the secular increase of the astronomical unit V_{AU} to σ_{Earth}, which leads to accurate numerical predictions: $H_0 = 74.42(.02)$(km/s)/Mpc and $V_{\mathrm{AU}} \cong 7.8$cm - yr^{-1}. Moreover, investigating the expected impacts of the *erfc* potential on the fly-bys anomalies and the residual Pioneers' delay lead to corrections for the osculating asymptotic velocity of a fly-by at the perigee of the order of 10 mm/s and an inward radial acceleration of 8.34×10^{-10}m/s^2 affecting the Pioneer spacecrafts.

To the best of our knowledge, the Bayesian paradigm proposed in this paper has never been investigated. Bayesian approaches have extensively been used to explore complex problems from a probabilistic point of view in numerous fields of science [17, 18], including quantum physics [19], astronomy [20], artificial intelligence [21], and neuroscience [22], to name a few examples. The present model adds up to this exhaustive list.

Acknowledgements This work has been partly supported over the years by RGPIN-2015-06409 discovery grants from NSERC Canada to Réjean Plamondon, as well as by the Department of Electrical Engineering, Polytechnique Montréal. The author thanks Claudéric Ouellet-Plamondon and André Laveau for their kind help with computer simulations and Richard Gagnon for his constructive comments.

References

1. M. Mitchelle, *Complexity: A Guided Tour* (Oxford University Press, Oxford, 2009), 349p
2. I. Licata, A. Sakaji (eds.), *Physics of Emergence and Organization* (World Scientific, Singapore, 2008)
3. W.J. Adams, *The Life and Times of the Central Limit Theorem*, vol 35, 2nd edn. (American Mathematical Society, London Mathematical Society, History of Mathematics, Providence, 2009), 195p
4. C. Barcelo, S. Liberati, M. Visser, Analogue gravity. Living Rev. Relativ. **14**(3), 1–15 (2011) http://www.livingreviews.org/lrr-2011-3
5. V. Faraoni, *Cosmology in Scalar-Tensor Gravity (Fundamental Theories of Physics)* (Springer Sciences and Business Media, Dordrecht, 2004)
6. T.P. Sotiriou, V. Faraoni, f(R) theories of gravity. Rev. Mod. Phys. **82**, 451 (2010)
7. S. Capoziello, V. Faraoni, *Beyond Einstein Gravity, a Survey of Gravitational Theories for Cosmology and Astrophysics* (Springer, Dorrdrecht, 2010)
8. A. De Felice, S. Tsujikawa, Living Rev. Relativ. **13**, 3 (2010). https://doi.org/10.12942/lrr-2010-3
9. T. Clifton, P.G. Ferreira, A. Padilla, C. Skordis, Modified gravity and cosmology. Phys. Rep. **513**, 1–189 (2012)
10. R. Plamondon, General relativity: An erfc metric. Results Phys. **9**, 456–462 (2018)
11. R. Plamondon, Solar system anomalies: revisiting Hubble's law. Phys. Essays **30**(4), 403–411 (2017)
12. J.A. Wheeler, K.W. Ford, *Geons, black holes, and quantum foam: a life in physics* (W. W. Norton & Company, New York, 2000)
13. W.J. Lindeberg, Eine neue Herleitung des Exponential-gesetze in der Wahscheinlichkeitsrechnung. Math Zeit **15**, 211 (1922)
14. A. Einstein, Die Grundlage der Allegemeinen Relativitatstheorie. Ann. Phys. **49**, 769 (1916)

15. W. Feller, *An Introduction to Probability Theory and its Applications volume II (Wiley Series in Probability and Mathematical Statistics)* (Wiley, New York, 1966)
16. R. Plamondon, C. O'Reilly, Ouellet-Plamondon, Strokes against Stroke—Strokes for Strides. Pattern Recogn. **47**, 929–944 (2014)
17. Statistical Science, Big Bayes Stories: A Collection of Vignettes, **29**(1), (2014)
18. B. Efron, Bayes' theorem in the 21st century. Science **340**, 1177–1178 (2013)
19. C.A. Fuchs, Quantum mechanics as quantum information (and only a little more), in *Quantum Theory: Reconsideration of Foundations*, ed. by A. Khrennikov, (Växjö University Press, Växjö, 2002), pp. 463–543
20. E.H. Feigelson, G.J. Babu, *Statistical Methods in Astronomy, in Planets, Stars, and Stellar Systems,* ed. by T. Ostwalt, Chief Ed., vol. 2, *Astronomical Techniques, Software, and Data*, ed. by H. Bond, (Springer, Netherlands, 2013)
21. D. Barber, *Bayesian Reasoning and Machine Learning* (Cambridge University Press, Cambridge, 2012)
22. K. Doya, S. Ishii, A. Pouget, R. P. N. Rao (eds.), *Bayesian Brain: Probabilistic Approaches to Neural Coding* (MIT Press, Cambridge, 2007)

Dressing for a Vector Modified KdV Hierarchy

Panagiota Adamopoulou and Georgios Papamikos

Abstract We present the hierarchy and soliton solutions associated with a multi-component generalisation of the modified Korteweg–de Vries equation. A recursive formula for obtaining the Lax operators associated with the higher flows of the hierarchy is provided. Using the method of rational dressing and the symmetries of the Lax operators we obtain the one-soliton solution. We also derive the general rank one-breather solution and express it in terms of determinants. We present the simplest case of the one-breather solution, which is parameterised by two constant unit vectors normal to each other. Finally, we obtain the associated Bäcklund transformation for the hierarchy.

Keywords Multi-component integrable hierarchies · Soliton solutions · Darboux-dressing transformations · Lax representation

1 Introduction

Since the establishment of the modern theory of integrable systems there have appeared many generalisations of soliton equations in both independent and dependent variables, such as extensions in $2+1$ variables, systems of equations, and equations with non-commutative variables. Multi-component integrable equations and their hierarchies have also attracted much attention due to their rich mathematical structure and appearance in applications. In the current paper we are concerned with the hierarchy associated with the following multi-component generalisation of the modified KdV (mKdV) equation:

P. Adamopoulou (✉)
Heriot-Watt University, Edinburgh, UK
e-mail: p.adamopoulou@hw.ac.uk

G. Papamikos
University of Essex, Colchester, UK
e-mail: g.papamikos@essex.ac.uk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_48

$$\mathbf{u}_t + \mathbf{u}_{xxx} + \frac{3}{2}\|\mathbf{u}\|^2\mathbf{u}_x = 0\,, \quad \text{with} \quad \mathbf{u} = \mathbf{u}(x,t) \in \mathbb{R}^N\,, \tag{1}$$

and its soliton solutions obtained via the Darboux-dressing scheme. In (1), $\|\mathbf{u}\|$ denotes the standard Euclidean norm in $\mathbb{R}^N$ and the subscripts denote partial differentiation with respect to the corresponding variables. For convenience, in what follows we will denote by $\mathbf{u}_j$ the jth partial derivative with respect to x, with $\mathbf{u}_0 = \mathbf{u}$. We remark that Eq. (1), which has also appeared in [11], is not the vector modified KdV (vmKdV) equation associated with the vector NLS hierarchy. Several other multi-component generalisations of the mKdV equation have also been studied in e.g. [3, 5]. In [1] the authors derived the hierarchy of commuting flows associated with Eq. (1) based on the Drinfel'd–Sokolov scheme [4], and further presented the soliton solution as well as the breather solutions of general rank for the hierarchy. Here, we additionally present the Bäcklund transformation for the hierarchy, as well as an explicit formula for the simplest one-breather solution.

2 Higher Vector mKdV Flows

Equation (1) admits an infinite number of conservation laws, as shown in [1] and also discussed in [6]. Additionally, the vmKdV is invariant under the Lie symmetries

$$\begin{aligned} \widetilde{x} &= x + \alpha\,, \\ \widetilde{t} &= t + \beta\,, \\ (\widetilde{x}, \widetilde{t}, \widetilde{\mathbf{u}}) &= (e^{\epsilon}x, e^{3\epsilon}t, e^{-\epsilon}\mathbf{u})\,, \end{aligned}$$

and is also O_N-invariant, i.e. invariant under $\widetilde{\mathbf{u}} = A\mathbf{u}$ with $A \in O_N$. A hierarchy of commuting generalised symmetries can be recursively constructed $\mathbf{u}_{t_{2n+1}} = \mathcal{R}\mathbf{u}_{t_{2n-1}}$, $n = 1, 2, \ldots$ using the recursion operator $\mathcal{R}$, given in [1, 2, 10],

$$\mathcal{R}\mathbf{f} = -D_x^2\mathbf{f} - \|\mathbf{u}\|^2\mathbf{f} - \mathbf{u}_1 D_x^{-1}\left(\mathbf{u}^T\mathbf{f}\right) - D_x^{-1}(\mathbf{u}_1\mathbf{f}^T - \mathbf{f}\mathbf{u}_1^T)\mathbf{u}\,, \tag{2}$$

starting from $\mathbf{u}_{t_1} = \mathbf{u}_1$. For example, we obtain the vmKdV equation (1) by acting with $\mathcal{R}$ on $\mathbf{u}_{t_1} = \mathbf{u}_1$

$$\mathbf{u}_{t_3} = \mathcal{R}\mathbf{u}_{t_1} = -\mathbf{u}_3 - \frac{3}{2}\|\mathbf{u}\|^2\mathbf{u}_1\,,$$

and similarly from $\mathbf{u}_{t_5} = \mathcal{R}\mathbf{u}_{t_3}$ we obtain

$$\mathbf{u}_{t_5} = \mathbf{u}_5 + \frac{5}{2}\|\mathbf{u}\|^2\mathbf{u}_3 + \frac{5}{2}\|\mathbf{u}_1\|^2\mathbf{u}_1 + 5\left(\mathbf{u}^T\mathbf{u}_1\right)\mathbf{u}_2 + 5\left(\mathbf{u}^T\mathbf{u}_2\right)\mathbf{u}_1 + \frac{15}{8}\|\mathbf{u}\|^4\mathbf{u}_1\,.$$

All members of the hierarchy $\mathbf{u}_{t_{2n+1}} = \mathcal{R}^n \mathbf{u}_1$ admit a zero curvature representation $[\mathcal{L}(\lambda), \mathcal{A}_{2n+1}(\lambda)] = 0$, where the linear operators $\mathcal{L}(\lambda)$ and $\mathcal{A}_{2n+1}(\lambda)$ are of the form

$$\mathcal{L}(\lambda) = D_x - \mathcal{U}(\lambda)\,, \quad \mathcal{A}_{2n+1}(\lambda) = D_{t_{2n+1}} - \mathcal{V}_{2n+1}(\lambda)\,, \quad n = 1, 2, \ldots\,. \tag{3}$$

Here λ is a spectral parameter, D_x and $D_{t_{2n+1}}$ denote differentiation with respect to x and t_{2n+1}, respectively, and $\mathcal{U}(\lambda)$ and $\mathcal{V}_{2n+1}(\lambda)$ take values in $so_{N+2}(\mathcal{D})[\lambda]$, where $\mathcal{D}$ is the differential ring $\mathbb{R}[\mathbf{u}, \mathbf{u}_1, \ldots]$, with $D_x \mathbf{u}_j = \mathbf{u}_{j+1}$. The Lax operators (3) are invariant under a $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ reduction group (see e.g. [7]), and in particular they satisfy

$$\mathcal{X}(\lambda)^\dagger = -\mathcal{X}(\lambda)\,, \quad \mathcal{X}(\lambda^*)^* = \mathcal{X}(\lambda)\,, \quad Q\mathcal{X}(-\lambda)Q^{-1} = \mathcal{X}(\lambda)\,, \tag{4}$$

where † denotes the formal adjoint operator, * denotes complex conjugation, and $Q = \mathrm{diag}(-1, 1, \ldots, 1)$. The matrix $\mathcal{U}(\lambda)$ in (3) is of the form $\mathcal{U}(\lambda) = \lambda J + U$ with [10]

$$J = \begin{pmatrix} 0 & 1 & \mathbf{0}^T \\ -1 & 0 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & 0_N \end{pmatrix}, \quad U = \begin{pmatrix} 0 & 0 & \mathbf{0}^T \\ 0 & 0 & \mathbf{u}^T \\ \mathbf{0} & -\mathbf{u} & 0_N \end{pmatrix}, \tag{5}$$

where $\mathbf{0}$ is the N-dimensional zero vector and 0_N the $N \times N$ zero matrix. The $\mathcal{V}_{2n+1}(\lambda)$ are found recursively starting from $\mathcal{V}_1(\lambda) = \mathcal{U}(\lambda)$ according to

$$\mathcal{V}_{2n+1}(\lambda) = \lambda^2 \mathcal{V}_{2n-1} + \lambda A_{2n-1} + B_{2n-1}\,, \quad n = 1, 2, \ldots\,, \tag{6}$$

where

$$\begin{aligned} A_{2n-1} &= -[J, U_{t_{2n-1}}] - D_x^{-1}(\mathbf{u}^T \mathbf{u}_{t_{2n-1}})J\,, \\ B_{2n-1} &= -D_x U_{t_{2n-1}} - D_x^{-1}(\mathbf{u}^T \mathbf{u}_{t_{2n-1}})U + D_x^{-1}\left[D_x U_{t_{2n-1}}, U\right]. \end{aligned}$$

For example, we have that

$$\mathcal{V}_3(\lambda) = \lambda^2 \mathcal{U}(\lambda) - \lambda\left([J, D_x U] + \frac{\|\mathbf{u}\|^2}{2} J\right) - D_x^2 U - \frac{\|\mathbf{u}\|^2}{2} U + [D_x U, U]\,, \tag{7}$$

such that the compatibility condition $[\mathcal{L}(\lambda), \mathcal{A}_3(\lambda)] = 0$ is equivalent to vmKdV equation (1).

3 Soliton Solutions for the Vector mKdV Hierarchy

Employing the method of rational dressing (see for example [8, 12]), we construct soliton and breather solutions to the vmKdV hierarchy. Given that $\Psi(x, t_{2n+1}, \lambda)$ is the fundamental solution to the overdetermined compatible system of linear equations

$$\mathcal{L}(\lambda)\Psi = 0\,, \quad \mathcal{A}_{2n+1}(\lambda)\Psi = 0\,, \quad n = 1, 2, \dots\,, \tag{8}$$

then a Darboux transformation

$$\Psi \to \Phi = M(\lambda)\Psi\,, \quad \text{with} \quad \det M(\lambda) \neq 0\,,$$

is a gauge transformation that leaves the linear equations form invariant

$$\widetilde{\mathcal{L}}(\lambda)\Phi = 0\,, \quad \widetilde{\mathcal{A}}_{2n+1}(\lambda)\Phi = 0\,, \tag{9}$$

where $\widetilde{\mathcal{L}}(\lambda) = D_x - \widetilde{\mathcal{U}}(\lambda)$ with $\widetilde{\mathcal{U}}(\lambda) = \mathcal{U}(\widetilde{\mathbf{u}}, \lambda)$, and similarly for $\widetilde{\mathcal{A}}_{2n+1}(\lambda)$. Here $\widetilde{\mathbf{u}}$ is a new exact solution of the vmKdV hierarchy. The Darboux matrix $M(\lambda)$ can be a constant matrix with respect to the independent variables and it will result to the O_N-invariance of the vmKdV hierarchy, or it can be a non-constant matrix which leads to Darboux-dressing relations or to a dressing chain (Bäcklund transformation). From the compatibility condition of (8) and (9) it follows that the Darboux matrix $M(\lambda)$ satisfies the Lax–Darboux equations

$$D_x M = \widetilde{\mathcal{U}} M - M\mathcal{U}\,, \quad D_{t_{2n+1}} M = \widetilde{\mathcal{V}}_{2n+1} M - M\mathcal{V}_{2n+1}\,, \tag{10}$$

or equivalently,

$$\widetilde{\mathcal{L}} = M\mathcal{L}M^{-1}\,, \quad \widetilde{\mathcal{A}}_{2n+1} = M\mathcal{A}_{2n+1}M^{-1}\,. \tag{11}$$

We further require that the Darboux matrix also satisfies the relations

$$M(\lambda)M(\lambda)^T = 1\,, \quad M(\lambda) = M(\lambda^*)^*\,, \quad M(\lambda) = QM(-\lambda)Q^{-1}\,, \tag{12}$$

which follow from the symmetries (4) of the Lax operators. The Darboux matrices that lead to the one-soliton and one-breather solutions of the vmKdV hierarchy are parameterised by a complex number and an element of the complex Grassmannian $Gr(s, \mathbb{C}^{N+2})$, with $s = 1, 2, \dots, N+1$ in the case of a breather solution, and $s = 1$ for a soliton solution. Further details are presented in [1], see also [8] in relation to the dressing for the vector sine-Gordon equation.

3.1 Soliton Solution: Dressing and Bäcklund Transformations

We consider a Darboux matrix with rational dependence on the spectral parameter λ, satisfying relations (12). The soliton solution corresponds to a Darboux matrix with two simple poles at $\lambda = \pm i\mu$ of the form [1, 8, 9]

$$M(\lambda) = 1 + \frac{2i\mu}{\lambda - i\mu} P - \frac{2i\mu}{\lambda + i\mu} QPQ\,, \quad \text{with} \quad P^* = QPQ\,, \tag{13}$$

where the matrix P is a projector of $\text{rank}(P) = 1$ of the form

$$P = \frac{Q\mathbf{q}\mathbf{q}^T}{\mathbf{q}^T Q\mathbf{q}} \quad \text{with} \quad \mathbf{q}^T\mathbf{q} = 0\,, \tag{14}$$

and $\mathbf{q}(x, t_{2n+1}) \in \mathbf{CP}^{N+1}$. From Eqs. (11) we obtain explicit expressions for $\mathbf{q}$ in terms of the fundamental solution of the linear systems (8), as well as the dressing transformation $\mathbf{u} \to \widetilde{\mathbf{u}}$ that leads to soliton solutions for the vmKdV hierarchy. In particular, from the constant in λ terms of the first equation in (11) and the double pole at $\lambda = i\mu$ of both equations in (11) we obtain respectively

$$\widetilde{U} = U - 2i\mu\,[J, P - P^*]\,, \quad P\mathcal{L}(i\mu)P^T = 0\,, \quad P\mathcal{A}_{2n+1}(i\mu)P^T = 0\,. \tag{15}$$

Using the first equation in (15) to express $\widetilde{U}$ in terms of U and P leads to the dressing transformation for the hierarchy

$$\widetilde{u}_j = u_j - 4i\mu\,\frac{q_1\, q_{j+2}}{-q_1^2 + \sum_{k=2}^{N+2} q_k^2}\,, \quad j = 1, 2, \ldots, N\,, \tag{16}$$

with q_j the components of $\mathbf{q}$, while the other two equations provide conditions for $\mathbf{q}$. More precisely, we obtain that the complex vector $\mathbf{q}$ satisfies

$$\mathcal{L}(i\mu)\mathbf{q} = 0, \quad \mathcal{A}_{2n+1}(i\mu)\mathbf{q} = 0\,, \tag{17}$$

which imply that

$$\mathbf{q} = \Psi(i\mu)\mathbf{C}\,, \quad \text{with} \quad \mathbf{C}^T\mathbf{C} = 0\,, \tag{18}$$

where $\Psi(i\mu)$ is the fundamental solution of the linear systems (8) at $\lambda = i\mu$ and $\mathbf{C} \in \mathbb{C}^{N+2}$ constant.

Example: One-Soliton Solution Starting from the trivial solution $\mathbf{u} = \mathbf{0}$ the fundamental solution of (8) is given by

$$\Psi = \exp(\xi J)\,, \quad \text{with} \quad \xi(t_{2n+1}, \mu) = \sum_{n=0}^{\infty} (-1)^n \mu^{2n+1} t_{2n+1}\,,$$

with $t_1 = x$. The solution Ψ takes the form

$$\Psi = \begin{pmatrix} \cosh\xi & i\sinh\xi & \mathbf{0}^T \\ -i\sinh\xi & \cosh\xi & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & 1_N \end{pmatrix}.$$

Further, the conditions $P^* = QPQ$ and $\mathbf{C}^T\mathbf{C} = 0$ in (13) and (18) imply that $\mathbf{C} = (i, c_0, \mathbf{c}^T)^T$, with $c_0 \in \mathbb{R}$, and $\mathbf{c} \in \mathbb{R}^N$ a constant vector such that $c_0^2 + \|\mathbf{c}\|^2 = 1$, see [1, 9]. Thus, the vector $\mathbf{q}$ takes the form

$$\mathbf{q} = (i\cosh\xi + ic_0\sinh\xi, c_0\cosh\xi + \sinh\xi, \mathbf{c}^T)^T\,. \tag{19}$$

Then, the dressing formula (16) leads to the one-soliton solution for the vmKdV hierarchy

$$\widetilde{\mathbf{u}} = \frac{2\mu\,\mathbf{c}}{\cosh\xi + c_0\sinh\xi}\,.$$

We can also use the Darboux matrix (13) in order to derive the Bäcklund transformation for the vmKdV hierarchy. To this end, we first express $\mathbf{q}$ in the form

$$\mathbf{q} = (i, a_0, \mathbf{a}^T)^T\,, \quad \text{with} \quad a_0^2 + \|\mathbf{a}\|^2 = 1\,, \tag{20}$$

and $a_0 \in \mathbb{R}$, $\mathbf{a} \in \mathbb{R}^N$. Using the first equation in (15) to express the components of matrix P in terms of the components of U and $\widetilde{U}$ and, given the form of P in (14), we obtain the following relation:

$$\mathbf{a} = \frac{\widetilde{\mathbf{u}} - \mathbf{u}}{2\mu}\,. \tag{21}$$

Then, from the residue of the simple pole at $\lambda = i\mu$ of the Lax–Darboux equations (10) we find that $\mathbf{a}_x = -\mu a_0\mathbf{a} - a_0\mathbf{u}$, and thus combining with (21) we obtain

$$(\widetilde{\mathbf{u}} - \mathbf{u})_x = -\mu a_0(\widetilde{\mathbf{u}} + \mathbf{u})\,, \quad a_0^2 + \frac{1}{4\mu^2}\|\widetilde{\mathbf{u}} - \mathbf{u}\|^2 = 1\,. \tag{22}$$

3.2 Breather Solution

The breather solution corresponds to a Darboux matrix $M(\lambda)$ with four poles at $\lambda = \pm\mu, \pm\mu^*$, with μ a generic complex number, of the form [1, 8]

$$M(\lambda) = 1 + \frac{M_0}{\lambda - \mu} - \frac{QM_0Q}{\lambda + \mu} + \frac{M_0^*}{\lambda - \mu^*} - \frac{QM_0^*Q}{\lambda + \mu^*}, \tag{23}$$

where

$$M_0 = q^* B q^T + Qq C q^T + Qq^* D q^T\,, \qquad q^T q = 0\,, \tag{24}$$

and $q \in Gr(s, \mathbb{C}^{N+2}) \simeq M_{N+2,s}(\mathbb{C})/GL_s(\mathbb{C})$, for $s = 1, 2, \ldots, N+1$. Here, $B, C, D \in M_{s,s}(\mathbb{C})$ are of the form

$$B = DG^* H^{*-1},\ C = -D^* F^* H^{*-1},\ D = -\left(F H^{-1} F^* + G^* H^{*-1} G^* - H^*\right)^{-1}, \tag{25}$$

where the matrices F, G, H are given by

$$F = \frac{1}{2\mu} q^T Q q\,, \quad G = \frac{1}{\mu - \mu^*} q^{*T} q\,, \quad H = \frac{1}{\mu + \mu^*} q^{*T} Q q\,. \tag{26}$$

The double pole at $\lambda = \mu$ of the Lax–Darboux equations (11) leads to relations

$$\mathcal{L}(\mu) q = 0, \quad \mathcal{A}_{2n+1}(\mu) q = 0\,,$$

thus, we can express the matrix q in terms of the fundamental solution of the linear problems $\mathcal{L}(\mu)\Psi = 0\,,\ \mathcal{A}_{2n+1}(\mu)\Psi = 0$ as

$$q = \Psi(\mu) C\,, \quad \text{with} \quad C^T C = 0\,, \tag{27}$$

and $C \in M_{N+2,s}(\mathbb{C})$ a constant matrix. The constant in λ terms of the first equation in (11) provide the following dressing transformation leading to breather-type solutions of the vmKdV hierarchy:

$$\widetilde{u}_j = u_j - 4\mathrm{Re} \sum_{k,l=1}^{s} \begin{vmatrix} q_1^k & 0 & 0 \\ 0 & q_{j+2}^l & B_{kl}^* - D_{kl}^* \\ 0 & q_{j+2}^{*l} & C_{kl} \end{vmatrix}, \quad j = 1, 2, \ldots, N\,, \tag{28}$$

with B, C, D given in (25), see [1] for details. Hence, the breather-type solutions for the hierarchy are characterised by the rank s of matrix q as well as the position μ of the pole of the Darboux matrix $M(\lambda)$ in (23). In the following example we derive the simplest ($s = 1$) one-breather solution, and show that it can be expressed

as a ratio of determinants and that it is parameterised by two constant unit vectors normal to each other.

Example: Rank One One-Breather Solution In the case $s = 1$, the dressing transformation (28) takes the form

$$\widetilde{u}_j = u_j - 4\mathrm{Re}\frac{\Delta_j}{\Delta}, \quad j = 1, 2, \ldots, N, \tag{29}$$

where Δ_j and Δ are the determinants

$$\Delta_j = \begin{vmatrix} q_1 & 0 & 0 \\ 0 & q_{j+2} & H - G \\ 0 & q_{j+2}^* & F^* \end{vmatrix}, \quad \Delta = \begin{vmatrix} F & H - G \\ G + H & F^* \end{vmatrix},$$

and F, G, H are now scalar quantities. Starting with the trivial solution $\mathbf{u} = \mathbf{0}$, the fundamental solution of the linear system (8) at $\lambda = \mu$ takes the form

$$\Psi(\mu) = \begin{pmatrix} \cos\xi & \sin\xi & \mathbf{0}^T \\ -\sin\xi & \cos\xi & \mathbf{0}^T \\ \mathbf{0} & \mathbf{0} & 1_N \end{pmatrix}, \quad \text{with} \quad \xi = \sum_{n=0}^{\infty} \mu^{2n+1} t_{2n+1}.$$

The one-breather solution can be written as

$$\widetilde{\mathbf{u}} = -\frac{4}{\Delta}\mathrm{Re}\big((C_1 \cos\xi + C_2 \sin\xi)(F^*\mathbf{c} + (G - H)\mathbf{c}^*)\big), \tag{30}$$

where $\mathbf{C} = (C_1, C_2, \mathbf{c}^T)^T$ such that $\mathbf{C}^T\mathbf{C} = 0$. The latter condition implies that the real and imaginary parts of vector $\mathbf{C}$ have the same length, and furthermore they are normal to each other. Using the fact that $\mathbf{C}$ is in $\mathbf{CP}^{N+1}$ we can normalise its real and imaginary parts and assume their length is equal to one. It follows that the one-breather solution for the vmKdV hierarchy is parameterised by a complex number (the pole of the Darboux matrix (23)) and an element of the unit tangent bundle $T_1\left(\mathbf{S}^{N+1}\right)$ of the sphere

$$T_1\left(\mathbf{S}^{N+1}\right) = \{(\mathbf{v}_1, \mathbf{v}_2) \in \mathbb{R}^{2(N+2)} |\ \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0,\ \|\mathbf{v}_1\| = \|\mathbf{v}_2\| = 1\}.$$

For example, in the case where the real and imaginary parts of vector $\mathbf{C}$ are given by

$$\mathbf{C}_R = (1, 0, \mathbf{0})^T, \quad \mathbf{C}_I = (0, \ldots, 0, 1, 0, \ldots, 0)^T,$$

respectively, where 1 appears in the $(j + 2)$ position in $\mathbf{C}_I$, then, starting from the trivial solution $\mathbf{u} = \mathbf{0}$ the breather solution (30) takes the form

$$\widetilde{u}_k = \begin{cases} -\frac{4}{\Delta}\mathrm{Re}\big(i\cos\xi(F^* + H - G)\big)\,, & k = j\,, \\ 0\,, & k \neq j\,. \end{cases} \tag{31}$$

The denominator in the above expression can be written in the form

$$\Delta = -\frac{1}{r^2}\left(\tan\theta\sin^2 A + \frac{\cosh^2 B}{\tan\theta}\right)^2, \tag{32}$$

where $r = |\mu|$, $\theta = \arg(\mu)$, $A = \mathrm{Re}(\xi)$ and $B = \mathrm{Im}(\xi)$. Finally, for the particular choice of $\mathbf{C}$ given above, we obtain the following expression for the rank one one-breather solution:

$$\widetilde{u}_j = 4r\frac{\sin\theta\sin A\sinh B - \cos\theta\cos A\cosh B}{\tan\theta\sin^2 A + \frac{\cosh^2 B}{\tan\theta}}\,, \tag{33}$$

which, as expected, is a breather solution of the scalar mKdV.

4 Conclusions

In this paper we presented a vectorial generalisation of the well-known mKdV equation, namely Eq. (1). It is important to note that this equation is not the third flow of the vector NLS hierarchy, since its Lax pair is characterised by relations (4), and additionally J in (5) is not a regular element of the underlying Lie algebra, contrary to the Drinfel'd–Sokolov construction [4]. For Eq. (1) we presented its recursion operator which can be used to construct all higher flows of the integrable hierarchy, and also a recursive formula that produces the Lax operator associated with each member of the hierarchy. We used the structure of the Lax operators to construct two Darboux transformations and we obtained the one-soliton solution and the general rank one-breather solution. The breather solution, in the simplest case, is a breather solution of the scalar mKdV equation. We leave the study of the more general breather solutions, as well as the n-soliton and n-breather solution for future work. As part of this goal, we aim to study Hirota's direct method as well as the general scattering problem and its inversion through a Riemann–Hilbert problem.

References

1. P. Adamopoulou, G. Papamikos, Drinfel'd-Sokolov construction and exact solutions of vector modified KdV hierarchy. Nucl. Phys. B **952**, 114933 (2020)
2. S.C. Anco, Hamiltonian flows of curves in $G/SO(N)$ and vector soliton equations of mKdV and Sine-Gordon type. Symmetry Integr. Geom. Methods Appl. **2**, 044 (2006)

3. C. Athorne, A.P. Fordy, Generalised KdV and mKdV equations associated with symmetric spaces. J. Phys. A: Math. Gen. **20**, 1377–1386 (1987)
4. V.G. Drinfel'd, V.V. Sokolov, Lie algebras and equations of Korteweg-de Vries type. J. Sov. Math. **30**, 1975–2036 (1985)
5. M. Iwao, R. Hirota, Soliton solutions of a coupled modified KdV equations. J. Phys. Soc. Jpn. **66**, 577–588 (1997)
6. J. Jackaman, G. Papamikos, T. Pryer, The design of conservative finite element discretisations for the vectorial modified KdV equation. Appl. Num. Math. **137**, 230–251 (2019)
7. A.V. Mikhailov, Reduction in integrable systems. The reduction group. JETP Lett. **32**, 187–192 (1980)
8. A.V. Mikhailov, G. Papamikos, J.P. Wang, Dressing method for the vector sine-Gordon equation and its soliton interactions. Physica D **325**, 53–62 (2016)
9. A.V. Mikhailov, G. Papamikos, J.P. Wang, Darboux transformation for the vector sine-Gordon equation and integrable equations on a sphere. Lett. Math. Phys. **106**, 973–996 (2016)
10. J.A. Sanders, J.P. Wang, Integrable systems in n-dimensional Riemannian geometry. Mosc. Math. J. **3**, 1369–1393 (2003)
11. V.V. Sokolov, T. Wolf, Classification of integrable polynomial vector evolution equations. J. Phys. A: Math. Gen. **34**, 11139–11148 (2001)
12. V.E. Zakharov, A.B. Shabat, Integration of nonlinear equations of Mathematical Physics by the Method of Inverse Scattering. II. Funct. Anal. Appl. **13**, 166–174 (1979)

Time Evolution in Quantum Systems and Stochastics

Anastasia Doikou, Simon J. A. Malham, and Anke Wiese

Abstract The time evolution problem for non-self-adjoint second order differential operators is studied by means of the path integral formulation. Explicit computation of the path integral via the use of certain underlying stochastic differential equations, which naturally emerge when computing the path integral, leads to a universal expression for the associated measure regardless of the form of the differential operators. The discrete non-linear hierarchy (DNLS) is then considered and the corresponding hierarchy of solvable, in principle, SDEs is extracted. The first couple members of the hierarchy correspond to the discrete stochastic transport and heat equations. The discrete stochastic Burgers equation is also obtained through the analogue of the Cole–Hopf transformation. The continuum limit is also discussed.

Keywords Time evolution · Non-self-adjoint operators · Quantum integrability · Hierarchy of SDEs

One of our main aims here is the solution of the time evolution problem associated with non-self-adjoint operators using the path integral formulation. We consider the general second order differential operator $\hat{L}_0$, and the associated time evolution problem:

$$-\partial_t f(\mathrm{x},t) = \hat{L} f(\mathrm{x},t) = \Big(\hat{L}_0 + u(\mathrm{x})\Big) f(\mathrm{x},t), \tag{1}$$

$$\hat{L}_0 = \frac{1}{2}\sum_{i,j=1}^{M} g_{ij}(\mathrm{x}) \frac{\partial^2}{\partial \mathrm{x}_i \partial \mathrm{x}_j} + \sum_{j=1}^{M} b_j(\mathrm{x}) \frac{\partial}{\partial \mathrm{x}_j}, \qquad g(\mathrm{x}) = \sigma(\mathrm{x})\sigma^T(\mathrm{x}), \tag{2}$$

where the diffusion matrix $g(\mathrm{x})$ and the matrix $\sigma(\mathrm{x})$ are in general dynamical (depending on the fields x_j) $M \times M$ matrices, while x and the drift $b(\mathrm{x})$ are M vector

A. Doikou (✉) · S. J. A. Malham · A. Wiese
School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK
e-mail: a.doikou@hw.ac.uk; s.j.a.malham@hw.ac.uk; a.wiese@hw.ac.uk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_49

fields with components x_j, b_j, respectively, and ^{T} denotes usual transposition. The operator $\hat{L}_0$ is not in general self-adjoint (Hermitian), therefore we also introduce the adjoint operator defined for any suitable function $f(\mathrm{x}, t)$ as

$$\hat{L}_0^\dagger f(\mathrm{x}, t) = \frac{1}{2} \sum_{i,j=1}^{M} \frac{\partial^2}{\partial \mathrm{x}_i \partial \mathrm{x}_j} \Big(g_{ij}(\mathrm{x}) f(\mathrm{x}, t) \Big) - \sum_{j=1}^{M} \frac{\partial}{\partial \mathrm{x}_j} \Big(b_j(\mathrm{x}) f(\mathrm{x}, t) \Big). \tag{3}$$

Then, two distinct time evolution equations emerge.

1. The Fokker–Planck equation:

$$\partial_{t_1} f(\mathrm{x}, t_1) = \hat{L}_0^\dagger f(\mathrm{x}, t_1) \tag{4}$$

   $t_1 \geq t_2$, with known initial condition $f(\mathrm{x}, t_2) = f_0(\mathrm{x})$.
2. The Kolmogorov backward equation:

$$-\partial_{t_2} f(\mathrm{x}, t_2) = \hat{L}_0 f(\mathrm{x}, t_2) \tag{5}$$

   $t_2 \leq t_1$, with known final condition $f(x, t_1) = f_f(x)$.

To simplify the complicated situation of non-constant diffusion coefficients we employ a local change of frame, which reduces the $\hat{L}$ operator to the simpler form with constant diffusion coefficients, and an effective drift. We then compute the generic path integral by requiring that the fields involved satisfy discrete time analogues of stochastic differential equations (SDEs). These equations naturally emerge when computing the path integral via the time discretization scheme. This leads to the computation of the measure, which turns out to be an infinite product of Gaussians.

Our second objective is to explore links between SDEs, and quantum integrable systems. To illustrate these associations we discuss a typical exactly solvable discrete quantum system, the discrete non-linear Schrödinger hierarchy. We express the quantum integrals of motion as second order differential operators after a suitable rescaling of the fields and we then extract a hierarchy of associated SDEs, which can be in principle solved by means of suitable integrator factors. The first two non-trivial members of the hierarchy correspond to the discrete stochastic transport and heat equations. The discrete stochastic Burgers equation is also obtained from the discrete stochastic heat equation through the analogue of the Cole–Hopf transformation (see also relevant [1]). More details on the derivation of the reported results can be found in [2].

1 Time Evolution and the Feynman–Kac Formula

Before we compute the solution of the time evolution problem via the path integral formulation we shall implement the quantum canonical transform that turns the dynamical diffusion matrix in (2) into identity at the level of the PDEs. This result will be then used for the explicit computation of the general path integral, and the derivation of the Feynman–Kac formula [2].

1.1 The Quantum Canonical Transformation

We will show in what follows that the general $\hat{L}$ operator can be brought into the less involved form:

$$\hat{L} = \frac{1}{2}\sum_{j=1}^{M} \frac{\partial^2}{\partial \mathrm{y}_j{}^2} + \sum_{j=1}^{M} \tilde{b}_j(\mathrm{y}) \frac{\partial}{\partial \mathrm{y}_j} + u(\mathrm{y}) \tag{6}$$

with an induced drift $\tilde{b}(\mathrm{y})$. This can be achieved via a simple change of the parameters x_j, which geometrically is nothing but a change of frame. Indeed, let us introduce a new set of parameters y_j such that [2]:

$$d\mathrm{y}_i = \sum_j \sigma_{ij}^{-1}(\mathrm{x})\, d\mathrm{x}_j, \quad \det \sigma \neq 0, \tag{7}$$

then $\hat{L}$ can be expressed in the form (6), and the induced drift components are given as

$$\tilde{b}_k(\mathrm{y}) = \sum_j \sigma_{kj}^{-1}(\mathrm{y}) b_j(\mathrm{y}) + \frac{1}{2}\sum_{j,l} \sigma_{jl}(\mathrm{y}) \partial_{\mathrm{y}_l} \sigma_{kj}^{-1}(\mathrm{y}). \tag{8}$$

Bearing also in mind that $\sum_j \sigma_{jl}\sigma_{kj}^{-1} = \delta_{kl}$, we can write in the compact vector/matrix notation:

$$\tilde{b}(\mathrm{y}) = \sigma^{-1}(\mathrm{y})\Big(b(\mathrm{y}) - \frac{1}{2}(\nabla_{\mathrm{y}}\sigma^T(\mathrm{y}))^T\Big), \qquad \nabla_{\mathrm{y}} = \Big(\partial_{\mathrm{y}_1}, \ldots, \partial_{\mathrm{y}_M}\Big), \tag{9}$$

where one first solves for $\mathrm{x} = \mathrm{x}(\mathrm{y})$ via (7). The transformation discussed above corresponds to a generalization of the so-called Lamperti transform at the level of SDEs (we refer the interested reader to [2] and the references therein).

1.2 The Path Integral: Feynman–Kac Formula

We are now in the position to solve the time evolution problem for the considerably simpler operator (6). Our starting point is the time evolution equation (3), (4), (6):

$$\partial_t f(\mathrm{y}, t) = \hat{L}^{\dagger} f(\mathrm{y}, t),$$

we then explicitly compute the propagator $K(\mathrm{y}_f, \mathrm{y}_i|t, t')$:

$$f(\mathrm{y}, t) = \int \prod_{j=1}^{M} dy'_j \; K(\mathrm{y}, \mathrm{y}'|t, t') f(\mathrm{y}', t') \tag{10}$$

$$= \int \prod_{n=1}^{N} \prod_{j=1}^{M} dy_{jn} \prod_{n=1}^{N} K(\mathrm{y}_{n+1}, \mathrm{y}_n|t_{n+1}, t_n) f(\mathrm{y}_1, t_1). \tag{11}$$

We employ the standard time discretization scheme as shown above, (see also, for instance, [3]), we insert the unit N times, $(\frac{1}{2\pi} \int d\mathrm{y}_{jn} \; dp_{jn} \; e^{\mathrm{i}p_{jn}(\mathrm{y}_{jn}-a)} = 1)$, for each component y_j, and we perform the Gaussian integrals with respect to each p_{jn} parameter. We then conclude that the path integral can be expressed as

$$K(\mathrm{y}_f, \mathrm{y}_i|t, t') = \int d\mathbf{q} \; \exp\left[-\sum_j \sum_n \frac{\left(\Delta \mathrm{y}_{jn} - \delta \tilde{b}_{jn}(\mathrm{y})\right)^2}{2\delta} + \delta \sum_n u_n(\mathrm{y}) \right] \tag{12}$$

$$d\mathbf{q} = \frac{1}{(2\pi\delta)^{\frac{NM}{2}}} \prod_{n=2}^{N} \prod_{j=1}^{M} d\mathrm{y}_{jn}, \tag{13}$$

where $f_n = f_n(\mathrm{y}_n)$ and $\Delta \mathrm{y}_{jn} = \mathrm{y}_{jn+1} - \mathrm{y}_{jn}$, where $\delta = t_{n+1} - t_n$ and with boundary conditions: $\mathrm{y}_f = \mathrm{y}_{N+1}$, $\mathrm{y}_i = \mathrm{y}_1$, $t_i = t' = 0$ (t' will be dropped henceforth for brevity), $t_f = t$.

We recall expression (12) and we make the fundamental assumption [2]:

$$\Delta \mathrm{y}_n - \delta \tilde{b}_n(\mathrm{y}) = \Delta \mathrm{w}_n \tag{14}$$

assuming also that w_{nj} are Brownian paths (see, for instance, [4] on Wiener processes), i.e., (14) is the discrete time analogue of an SDE. After a change of the volume element in (12), subject to (14), we conclude (see [2] for the detailed computation):

$$K(\mathrm{y}_f, \mathrm{y}_i|t) = \int d\mathbf{M} \; e^{\int_0^t u(\mathrm{y}_s) ds}, \tag{15}$$

$$d\mathbf{M} = \lim_{\delta\to 0}\lim_{N\to\infty} \frac{1}{(2\pi\delta)^{\frac{NM}{2}}} \exp\Big[-\frac{1}{2\delta}\sum_{n=1}^{N} \Delta \mathrm{w}_n^T \Delta \mathrm{w}_n\Big] \prod_{n=2}^{N}\prod_{j=1}^{M} d\mathrm{w}_{jn}. \tag{16}$$

We may now evaluate the measure: in the continuum time limit (16), we consider the Fourier representation on $[0,\ t]$ for w_s, i.e., Wiener's representation of the Brownian path [4]:

$$\mathrm{w}_s = \frac{\mathrm{f}_0}{\sqrt{t}} s + \sqrt{\frac{2}{t}} \sum_{k>0} \frac{\mathrm{f}_k}{\omega_k} \sin\omega_k s, \quad \omega_k = \frac{2\pi k}{t}. \tag{17}$$

$\mathrm{f}_0 = \frac{\mathrm{w}_t}{\sqrt{t}}$ and $\mathrm{f}_k,\ k \in \{0,\ 1, \ldots\}$ are M vectors with components $\mathrm{f}_{kj},\ j \in \{1,\ 2, \ldots, M\}$ being standard normal variables. We are interested in the computation of the measure in the continuum limit $N \to \infty,\ \delta \to 0$, and we also recall the following boundary conditions: $\mathrm{w}(s=0) = 0,\ \mathrm{w}(s=t) = \mathrm{w}_t$, then

$$\begin{aligned} d\mathbf{M} &= \frac{e^{-\frac{1}{2t}\mathrm{w}_t^T \mathrm{w}_t}}{(2\pi t)^{\frac{M}{2}}}\, d\mathbf{M}_0 \\ d\mathbf{M}_0 &= \prod_{k\geq 1}\prod_{j=1}^{M} \frac{d\mathrm{f}_{kj}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\sum_{k\geq 1}\sum_{j} \mathrm{f}_{kj}^2\right]. \end{aligned} \tag{18}$$

The measure naturally is expressed as an infinite product of Gaussians regardless of the specific forms of the diffusion coefficients and the drift.

Having computed the propagator explicitly (15) we conclude that Eq. (11) can be then expressed as

$$f(\mathrm{x}_f, t_f) = \int d\mathbf{M}\, e^{\int_0^t u(\mathrm{x}_s) ds} f_0(\mathrm{x}_0), \qquad f_0(\mathrm{x}_0) = f(\mathrm{x}_0, t_0)$$

which is precisely the Feynman–Kac formula, and describes the time evolution of a given initial profile $f_0(\mathrm{x}_0)$ to $f(\mathrm{x}_f, t_f)$ a solution of the Fokker–Planck equation. One could have started from the Kolmogorov backward equation and computed the path integral backwards in time:

$$f(\mathrm{x}_0, t_0) = \int d\mathbf{M}\, e^{\int_0^t u(\mathrm{x}_s) ds} f_f(\mathrm{x}_f), \qquad f_f(\mathrm{x}_f) = f(\mathrm{x}_f, t_f).$$

In this case the Feynman–Kac formula describes the reversed time evolution of a given final state $f_f(\mathrm{x}_f)$, to a previous state $f(\mathrm{x}_0, t_0)$ a solution of the Kolmogorov backward equation.

One of the main aims is the computation of expectation values:

$$\langle \mathcal{O}(\mathrm{x}_s) \rangle = \frac{\mathbf{E}_t\left(\mathcal{O}(\mathrm{x}_s)\, e^{\int_0^t u(\mathrm{x}_s) ds}\right)}{\mathbf{E}_t\left(e^{\int_0^t u(\mathrm{x}_s) ds}\right)}, \quad 0 \leq s \leq t, \tag{19}$$

where we define via (15), (18)

$$\mathbf{E}_t\Big(\mathcal{O}(\mathrm{x}_s)\Big) = \int d\mathbf{w_t}\, d\mathbf{M}\, \mathcal{O}(\mathrm{x}_s) \quad 0 \leq s \leq t. \tag{20}$$

Equation (19) can be used provided that solutions of the associated SDEs are available, so that the fields x_{tj} are expressed in terms of the variables w_{tj}.

2 The Quantum (D)NLS and a Hierarchy of S(P)DEs

We start our analysis with the DNLS model (see, for instance [5–7], [9]), with the corresponding quantum Lax operator given by [5, 6],

$$L_j(\lambda) = \begin{pmatrix} \lambda + \Theta_j + z_j Z_j & z_j \\ Z_j & 1 \end{pmatrix}$$

$z_j,\ Z_j$ are canonical $[z_i,\ Z_j] = -\delta_{ij}$, and we consider the map:

$$z_j \mapsto \mathrm{x}_j, \qquad Z_j \mapsto \partial_{\mathrm{x}_j}. \tag{21}$$

Let us now define the generating function of the integrals of motion of the system:

$$\mathbf{t}(\lambda) = tr\Big(L_M(\lambda) \ldots L_2(\lambda) L_1(\lambda)\Big). \tag{22}$$

Indeed, the expansion of $\ln \mathbf{t}(\lambda) = \sum_{k=0}^{M} \frac{I_k}{\lambda^k}$ provides the local integrals of motion (see, e.g., [8]). We keep here terms up to third order in the expansion of $\ln \mathbf{t}$ and by suitably scaling the involved fields, we obtain the first three local integrals of motion of the quantum DNLS hierarchy (keeping the suitably scaled terms) [5]:

$$H_1 = \sum_j^M \mathrm{x}_j \partial_{\mathrm{x}_j}$$

$$H_2 = \frac{1}{2} \sum_{j=1}^{M} \mathrm{x}_j^2 \partial_{\mathrm{x}_j}^2 - \sum_{j=1}^{M} \Delta^{(1)}(\mathrm{x}_j) \partial_{\mathrm{x}_j}$$

$$H_3 = \frac{1}{2}\sum_{j=1}^{M} \mathrm{x}_j^2 \partial_{\mathrm{x}_j}^2 - \nu \sum_{j=1}^{M} \Delta^{(2)}(\mathrm{x}_j)\partial_{\mathrm{x}_j} + (\text{higher order terms})\ldots, \tag{23}$$

where we have chosen $\Theta_j = 1$, and $H_1 = I_1,\ H_2 = -I_2 + \frac{1}{2}I_1,\ H_3 = -\frac{1}{3}\big(I_3 + I_2 - \frac{1}{2}I_1\big)$, $\nu = \frac{1}{3}$. We also define: $\Delta^{(1)} z_j = z_{j+1} - z_j,\ \Delta^{(2)} z_j = z_{j+2} - 2z_{j+1} + z_j$. The next order in the expansion provides H_4, which is the Hamiltonian of the quantum version of complex mKdV system and so on. The equations of motion (classical and quantum) associated, e.g., with $H^{(2)}$ can be derived via the zero curvature condition or Heisenberg's equation (recall also (21)):

$$\frac{dz_j}{dt} = -\Delta^{(1)} z_j + z_j^2 Z_j. \tag{24}$$

Similar equations can be obtained for H_3, but are omitted here for brevity. The Hamiltonians $H_{2,3}$ are of the form (2), and the corresponding set of SDEs are [2]

$$d\mathrm{x}_{tj} = -\nu_k \Delta^{(k-1)} \mathrm{x}_t dt + \mathrm{x}_{tj} d\mathrm{w}_{tj}, \tag{25}$$

where $k \in \{2,\ 3\}$ and $\nu_2 = 1,\ \nu_3 = \frac{1}{3}$. ν_k can be set equal to one henceforth, after suitably rescaling time. By comparing (24) and (25) ($k = 2$) we observe that the non-linearity appearing in (24) is replaced by the multiplicative noise in (25).

Let us now derive the solution of the set of SDEs (25) introducing suitable integrator factors (see, e.g., [10]). Let us consider the general set of SDES

$$d\mathrm{x}_{tj} = b_j(\mathrm{x}_t)dt + \mathrm{x}_{tj} d\mathrm{w}_{tj}.$$

We introduce the following set of integrator factors:

$$\mathcal{F}_j(t) = \exp\Big(-\int_0^t d\mathrm{w}_{sj} + \frac{1}{2}\int_0^t ds \Big) \tag{26}$$

and define the new fields: $\mathrm{y}_{tj} = \mathcal{F}_j(t)\mathrm{x}_{tj}$, then one obtains a differential equation for the vector field y:

$$\frac{d\mathrm{y}_t}{dt} = \mathcal{A}(t)\mathrm{y}_t \ \Rightarrow \mathrm{y}_t = \mathcal{P}\exp\Big(\int_0^t \mathcal{A}(s)ds \Big)\mathrm{y}_0. \tag{27}$$

For instance, in the case of (25), for $k = 2$, the $M \times M$ matrix $\mathcal{A}$ is given as

$$\mathcal{A}(t) = \sum_{j=1}^{M} \Big(e_{jj} - \mathcal{B}_j(t) e_{jj+1} \Big), \qquad \mathcal{B}_j(t) = \exp\Big(\Delta^{(1)}(\mathrm{w}_{tj}) \Big),$$

where e_{ij} are $M \times M$ matrices with entries $(e_{ij})_{kl} = \delta_{ik}\delta_{jl}$. For $k = 3$, the $\mathcal{A}$ matrix involves also terms e_{jj+2}, and so on. The solution (27) can be expressed as a formal series expansion

$$\mathcal{P}\exp\Big(\int_0^t \mathcal{A}(s)ds\Big) = \sum_{n=0}^{\infty}\int_0^t \int_0^{t_n} \dots \int_0^{t_2} dt_n dt_{n-1}\dots dt_1 \mathcal{A}(t_n)\mathcal{A}(t_{n-1})\dots\mathcal{A}(t_1),$$
$$t \geq t_n \geq t_{n-1} \dots \geq t_2.$$

Remark 1 The discrete version of the stochastic Burgers equation can be obtained from the discrete stochastic heat equation through the analogue of the Cole–Hopf transformation. Indeed, by setting $\mathrm{x}_j = e^{\mathrm{y}_j}$, in (25) ($k = 3$):

$$d\mathrm{y}_j = -\Big(e^{\Delta \mathrm{y}_j}\big(e^{\Delta \mathrm{y}_{j+1}} - 1\big) - \big(e^{\Delta \mathrm{y}_j} + 1\big)\Big)dt + d\mathrm{w}_j, \tag{28}$$

where for simplicity we have set $\Delta^{(1)} = \Delta$. By also setting $u_j = \Delta \mathrm{y}_j$, we obtain a discrete version of the stochastic Burgers equation

$$du_j = -\Big(e^{u_{n+1}}\big(e^{u_{j+2}} - e^{u_j}\big) - 2\big(e^{u_{j+1}} - e^{u_j}\big)\Big)dt + \Delta d\mathrm{w}_j. \tag{29}$$

Assuming the scaling $\Delta y_j \sim \delta$, we expand the exponentials and keep up to second order terms in (28), (29):

$$d\mathrm{y}_j = -\Big(\Delta^{(2)}\mathrm{y}_j + \big(\Delta \mathrm{y}_j\big)^2 + \mathcal{O}(\delta^3)\Big) + d\mathrm{w}_j \tag{30}$$

$$du_j = -\Big(\Delta^{(2)}u_j + \Delta u_j^2 + \mathcal{O}(\delta^3)\Big) + \Delta d\mathrm{w}_j. \tag{31}$$

The second of the equations above provides a good approximation for the discrete viscous Burgers equation, as will be also clear in the next subsection.

2.1 The Continuum Models and SPDEs

It will be instructive to consider the continuum limits of the Hamiltonians H_2, H_3 (23) and the respective SDEs. After considering the thermodynamic limit $M \to \infty$, $\delta \to 0$ ($\delta \sim \frac{1}{M}$) we obtain

$$\mathrm{x}_{tj} \to \varphi(x,t), \quad \frac{\mathrm{x}_{tj+1} - \mathrm{x}_{tj}}{\delta} \to \partial_x \varphi(x,t),$$

$$\delta \sum_j f_j \ \to \ \int dx \ f(x), \qquad \mathrm{w}_{tj} \ \to \ W(x,t), \tag{32}$$

where the Wiener field or Brownian sheet $W(x,t)$ is periodic and square integrable in $[-L,\ L]$, and is represented as [4]

$$W(x,t) = \frac{\sqrt{L}}{\pi} \sum_{n\geq 1} \frac{1}{n} \left(X_t^{(n)} \cos \frac{n\pi x}{L} + Y_t^{(n)} \sin \frac{n\pi x}{L} \right), \tag{33}$$

$X_t^{(n)},\ Y_t^{(n)}$ are independent Brownian motions. In the continuum limit the Hamiltonians (23) become the Hamiltonians of quantum NLS hierarchy:

$$H_c^{(k)} = \int dx \left(\frac{1}{2} \varphi^2(x) \hat{\varphi}^2(x) - \partial_x^{(k-1)} \varphi(x) \hat{\varphi}(x) \right), \quad k = 2,\ 3, \tag{34}$$

where $\left[\varphi(x),\ \hat{\varphi}(y)\right] = \delta(x-y)$, $(\hat{\varphi}(x) \sim \frac{\partial}{\partial \varphi(x)})$ and the SDEs (25) become the stochastic transport ($k=2$) and heat equation ($k=3$) with multiplicative noise:

$$\partial_t \varphi(x,t) = -\partial_x^{k-1} \varphi(x,t) + \varphi(x,t) \dot{W}(x,t).$$

The stochastic heat equation can be mapped to the stochastic Hamilton–Jacobi and viscous Burgers equations [1]. Indeed, we set: $\varphi = e^h,\ u = \partial_x h$ then (35):

$$\begin{aligned} \partial_t h(x,t) &= -\partial_x^2 h(x,t) - (\partial_x h(x,t))^2 + \dot{W}(x,t) \\ \partial_t u(x,t) &= -\partial_x^2 u(x,t) - 2u(x,t) \partial_x u(x,t) + \partial_x \dot{W}(x,t). \end{aligned} \tag{35}$$

Connections between the SDEs and the quantum Darboux transforms [11, 12] can be also studied. The classical Darboux–Bäcklund transformation [9, 13, 14] provides an efficient way to find solutions of integrable PDEs. The key question is how this transformation can facilitate the solution of SDEs [1, 15, 16].

Acknowledgments AD acknowledges support from the EPSRC research grant: EP/R009465/1.

References

1. I. Corwin, Exactly solving the KPZ equation (2018). arXiv.1804.05721 [math.PR]
2. A. Doikou, S.J.A. Malham, A. Wiese, Nucl. Phys. **B945**, 114658 (2019)
3. B. Simon, *Functional Integration and Quantum Physics* (AMS Publishing, Providence, 2005)
4. C. Prevôt, M. Röckner, *A Concise Course on Stochastic Partial Differential Equations* (Springer, Berlin, 2007)
5. A. Kundu, O. Ragnisco, J. Phys. **A27**, 6335 (1994)

6. E. Sklyanin, *In Integrable Systems: From Classical to Quantum*. Proc. Lecture Notes, vol. 26 (CRM, Montreal, 1999), pp. 227–250. arXiv:nlin/0009009
7. A.C. Scott, J.C. Eilbeck, Phys. Lett. **A119**, 60 (1986)
8. V.E. Korepin, N.M. Bogoliubov, A.G. Izergin, *Quantum Inverse Scattering Method and Correlation Functions* (Cambridge University Press, Cambridge, 1983)
9. A. Doikou, I. Findlay, S. Sklaveniti, Nucl. Phys. **B941**, 376 (2019)
10. B. Oksendal, *Stochastic Differential Equations: Introduction and Applications* (Springer, Berlin, 2003)
11. C. Korff, J. Phys. **A49**, 104001 (2016)
12. A. Doikou, I. Findlay, The quantum auxiliary linear problem & quantum Darboux-Bäcklund transformations (2017). arXiv:1706.06052 [math-ph]
13. V.E. Zakharov, A.B. Shabat, Funct. Anal. Appl. **13**, 166 (1979)
14. V.B. Matveev, M.A. Salle, *Darboux Transformations and Solitons* (Springer, Berlin, 1991)
15. N. O'Connell, *In Memoriam Marc Yor - Séminaire de Probabilités XLVII*. Lecture Notes in Mathematics, vol. 2137 (Springer, Cham, 2015)
16. A. Doikou, S.J.A. Malham, I. Stylianidis, A. Wiese, Applications of Grassmannian and graph flows to nonlinear systems (2019). arxiv:1905.0535 [math.AP]

Part VI
Particle Physics

Solvable Models of Magnetic Skyrmions

Bernd Schroers

Abstract We give a succinct summary of the recently discovered solvable models of magnetic skyrmions in two dimensions, and of their general solutions. The models contain the standard Heisenberg term, the most general translation invariant Dzyaloshinskii–Moriya (DM) interaction term and, for each DM term, a particular combination of anisotropy and Zeeman potentials. We argue that simple mathematical features of the explicit solutions help understand general qualitative properties of magnetic skyrmion configurations in more generic models.

Keywords Magnetic skyrmions · Gauged sigma models

1 Introduction

Magnetic skyrmions are topological solitons in two-dimensional field theories for the magnetisation field $\boldsymbol{n}$ of a magnetic material [1, 2]. In the continuum version, the energy functional consists of the Dirichlet energy (quadratic in derivatives), a potential which includes anisotropy terms and a Zeeman contribution (no derivatives), and the crucial Dzyaloshinskii–Moriya (DM) interaction term (linear in derivatives) [3, 4]. Such an energy functional has stationary configurations which are stable under Derrick scaling provided the DM term is negative for those configurations.

Magnetic skyrmions have been the subject of intensive experimental and numerical studies in recent years because they combine interesting physics with potential technological applications in magnetic information storage and manipulation [5]. More recently, rigorous analytical studies have established conditions for the existence of solutions as well as energy bounds in different topological sectors [6, 7], and have clarified the interesting way in which the relative energy of skyrmions and anti-skyrmions depends on the DM term [8].

B. Schroers (✉)
Department of Mathematics and Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh, UK
e-mail: b.j.schroers@hw.ac.uk

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_50

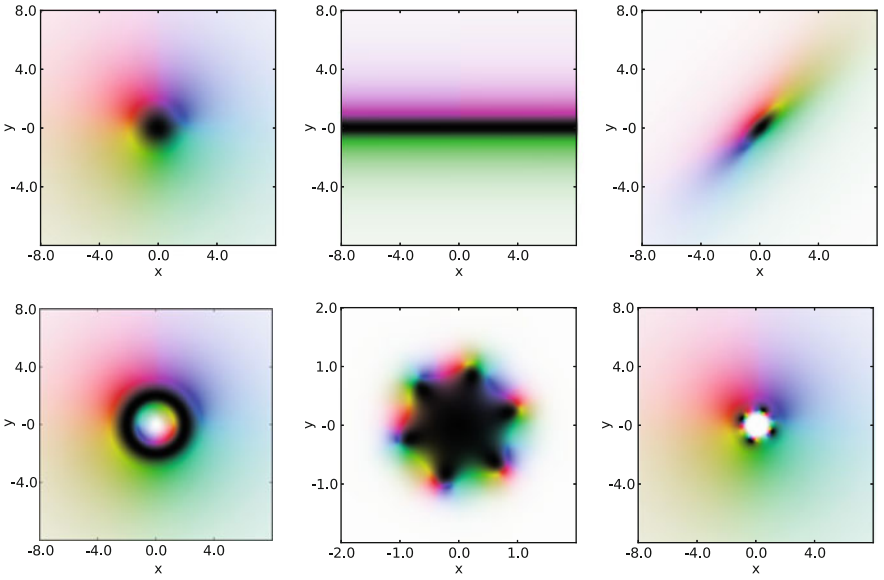


Fig. 1 Magnetic skyrmion solutions in the model with DM term $(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ and the potential $\frac{1}{2}(1 - n_3)^2$. The value of the magnetisation vector $\boldsymbol{n}$, assumed to be of unit length, is shown in terms of the Runge colour sphere: the region near the south pole $n_3 = -1$ is shown in black, and the 'vacuum region' near the north pole $n_3 = 1$ is shown in white. Elsewhere, the longitudinal angle $\arctan(n_2/n_1)$ is mapped onto the colours red, green, blue, with intermediate colours interpolating. Solutions are determined by a choice of a holomorphic function h and the formula (23). Top from left to right: the Bloch skyrmion with $h(z) = 0$, the line defect with $h(z) = \frac{i}{2}z$ and the anti-skyrmion with $h(z) = z$. Bottom from left to right: the 'empty bag' with $h(z) = 2i/z$, the anti-skyrmion of charge $Q = 5$ with $h(z) = z^5$ and the anti-skyrmion of charge $Q = 4$ with $h(z) = 1/z^5$

This talk is about the critically coupled models recently proposed in [9] and [10]. These models require a particular choice of potential for any given DM term, but with this choice they can be viewed as a gauged version of the Belavin–Polyakov $O(3)$ sigma model [11]. In particular, solutions can be obtained explicitly in terms of holomorphic functions to the Riemann sphere $\mathbb{C} \cup \{\infty\}$. In Fig. 1, we show examples of such solutions in a model with standard DM term $(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ and the potential $\frac{1}{2}(1 - n_3)^2$. They include the axisymmetric skyrmion configuration (which has topological charge $Q = -1$ in our conventions), a line defect ($Q = 0$), an anti-skyrmion configuration ($Q = 1$) as well as bags and multi-(anti)-skyrmion configurations which show qualitative features of the configurations studied numerically in [12] and [13].

This talk is designed to explain the models and the construction of their solutions from holomorphic data as simply and directly as possibly. We sum up the method of solution as a four-step recipe in Sect. 3. For details we refer the reader to the papers [9, 10].

2 Magnetic Skyrmions and Gauged Sigma Models

2.1 Formulating Magnetic Skyrme Models as Gauged Sigma Models

The most general energy functional for the magnetisation field $\boldsymbol{n} : \mathbb{R}^2 \to S^2$ which we consider has the form

$$E[\boldsymbol{n}] = \int_{\mathbb{R}^2} \left(\frac{1}{2}(\nabla \boldsymbol{n})^2 + \sum_{a=1}^{3} \sum_{i=1}^{2} \mathcal{D}_{ai} (\partial_i \boldsymbol{n} \times \boldsymbol{n})_a + V(\boldsymbol{n}) \right) dx_1 dx_2, \tag{1}$$

where V is a potential which may include a Zeeman term and anisotropy terms, and $\mathcal{D}$ is the spiralization tensor parametrising the DM interaction.

In the following we will use the complex coordinate $z = x_1 + i x_2$ in the plane, and define associated derivatives in the standard way, so $\partial_z = \frac{1}{2}(\partial_1 - i\partial_2)$ and $\partial_{\bar{z}} = \frac{1}{2}(\partial_1 + i\partial_2)$. As shown in [10], one can view the expression (1) as the energy functional of a gauged sigma model with a fixed $SU(2)$ background gauge field. To see this, it is convenient to think of an $SU(2)$ gauge field in the plane simply as a pair of vectors $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ in $\mathbb{R}^3$, one for each Cartesian direction in the plane, which act on the magnetisation field $\boldsymbol{n} \in \mathbb{R}^3$ via the vector product (the commutator of the $su(2)$ Lie algebra) so that the covariant derivative and curvature are

$$D_i \boldsymbol{n} = \partial_i \boldsymbol{n} + \boldsymbol{A}_i \times \boldsymbol{n}, \quad i = 1, 2, \quad \boldsymbol{F}_{12} = \partial_1 \boldsymbol{A}_2 - \partial_2 \boldsymbol{A}_1 + \frac{1}{2} \boldsymbol{A}_1 \times \boldsymbol{A}_2. \tag{2}$$

The energy functional of the gauged non-linear sigma models studied in [10] is

$$E_A[\boldsymbol{n}] = \int_{\mathbb{R}^2} \left(\frac{1}{2}|D_1 \boldsymbol{n}|^2 + \frac{1}{2}|D_2 \boldsymbol{n}|^2 - (\boldsymbol{F}_{12}, \boldsymbol{n}) \right) dx_1 dx_2. \tag{3}$$

In the application to magnetic skyrmions, the gauge field is fully determined by the spiralization tensor: the Cartesian components $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are simply the negatives of the column vectors which make up the 3×2 spiralization matrix $\mathcal{D}$. In symbols

$$(\boldsymbol{A}_i)_a = -\mathcal{D}_{ai}, \tag{4}$$

so that the DM term can be written as

$$\sum_{a=1}^{3} \sum_{i=1}^{2} \mathcal{D}_{ai} (\partial_i \boldsymbol{n} \times \boldsymbol{n})_a = \sum_{i=1}^{2} (\boldsymbol{A}_i \times \boldsymbol{n}, \partial_i \boldsymbol{n}). \tag{5}$$

We now need to pick a particular form of the potential V in (1) to obtain a solvable model, namely

$$V_A(\boldsymbol{n}) = \frac{1}{2}|\boldsymbol{A}_1 \times \boldsymbol{n}|^2 + \frac{1}{2}|\boldsymbol{A}_2 \times \boldsymbol{n}|^2 - (\boldsymbol{n}, \boldsymbol{A}_1 \times \boldsymbol{A}_2). \tag{6}$$

With the choice $V = V_A$, the magnetic skyrmion energy (1) equals the energy (3) of the gauged sigma model with the gauge field (4). This is easily checked, and makes use of the fact that, for constant gauge fields, one has $\boldsymbol{F}_{12} = \boldsymbol{A}_1 \times \boldsymbol{A}_2$. Note that the energy functional (3) is invariant under $SU(2)$ gauge transformations, but that this gauge invariance is broken by the gauge choice (4) to obtain the magnetic skyrme energy functional (1) at critical coupling. The residual symmetries of the critically model are discussed in [9] and [10].

For a simple illustration, consider

$$\boldsymbol{A}_1 = -\kappa \boldsymbol{e}_1, \boldsymbol{A}_2 = -\kappa \boldsymbol{e}_2, \tag{7}$$

where $\boldsymbol{e}_1 = (1, 0, 0)^t$ and $\boldsymbol{e}_2 = (0, 1, 0)^t$ are the first two elements of the canonical frame for $\mathbb{R}^3$, and $\kappa > 0$ is a real parameter. This produces the standard DM term $\kappa(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ and the potential $V_A = \frac{\kappa^2}{2}(1-n_3)^2$. Expanding the square, the potential is seen to be a particular linear combination of an easy-plane anisotropy potential with a Zeeman potential, see [9]. This model with $\kappa = 1$ is the one whose solutions are shown in Fig. 1.

2.2 A Bogomol'nyi Equation for Gauged Sigma Models

Returning now to the case of a general gauge field, we use various gauge-theoretical identities [9, 10] to write the energy (3) as

$$E_A[\boldsymbol{n}] = \frac{1}{2}\int_{\mathbb{R}^2} (D_1\boldsymbol{n} + \boldsymbol{n} \times D_2\boldsymbol{n})^2 dx_1 dx_2 + 4\pi(Q + \Omega_A), \tag{8}$$

where Q is the integral expression for the degree of $\boldsymbol{n}$

$$Q[\boldsymbol{n}] = \frac{1}{4\pi}\int_{\mathbb{R}^2} (n, \partial_1\boldsymbol{n} \times \partial_2\boldsymbol{n}) dx_1 dx_2, \tag{9}$$

and Ω_A is a generalised version of what was called total vortex strength in [9]:

$$\Omega_A[\boldsymbol{n}] = -\frac{1}{4\pi}\int_{\mathbb{R}^2} (\partial_1(\boldsymbol{A}_2, \boldsymbol{n}) - \partial_2(\boldsymbol{A}_1, \boldsymbol{n})) dx_1 dx_2. \tag{10}$$

If these integrals are well-defined, they only depend on global properties of $\boldsymbol{n}$ and on its boundary behaviour. If the latter is kept fixed, the energy is therefore minimised when the square in (8) vanishes, i.e. when the Bogomol'nyi equation holds. This is a gauged version of the Bogomol'nyi equations in the standard Belavin–Polyakov model [11], but with a definite sign:

$$\boldsymbol{n} \times D_1 \boldsymbol{n} = D_2 \boldsymbol{n}. \tag{11}$$

Ths Bogomol'nyi equation implies the variational equation of the energy functional (3), see [10].

In the context of magnetic skyrmions, Eq. (11) first appeared in [7] where it was noticed that, for a certain family of potentials V, it characterises the energy minimisers in the $Q = -1$ sector of the theory with the standard DM term $(\boldsymbol{n}, \nabla \times \boldsymbol{n})$. The role of this equation in critically coupled magnetic skyrme models for arbitrary degree $Q \geq -1$ was observed and explored in [9]. Its role and solvability in the more general gauged sigma model and the associated magnetic skyrmion models are the subject of [10]. Generalised versions of this equation have been studied in differential geometry as vortex equations for maps from Riemann surfaces into Kähler manifolds which permit the action of a Lie group, but in that case the gauge field typically obeys a second, coupled equation [14]. The relation between these vortex equations and Eq. (11) with a fixed background as proposed in [10] was clarified in [15].

2.3 Boundary Contributions to the Energy

As far we are aware it is not known for which class of magnetisation fields $\boldsymbol{n}$ the general energy functional (1) is well-defined and finite. For the standard DM term $(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ and a certain class of potentials V, this question is answered in [6] and [7], where it was also pointed that, for analytical reasons, it is preferable to modify the energy functional by adding the boundary term

$$E_{\kappa,\infty}[\boldsymbol{n}] = -\kappa \int_{\mathbb{R}^2} (\partial_1 n_2 - \partial_2 n_1) dx_1 dx_2. \tag{12}$$

Adding this term to the energy effectively modifies the DM term: $\kappa(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ is replaced by $\kappa((\boldsymbol{n} - \boldsymbol{e}_3), \nabla \times \boldsymbol{n})$ where $\boldsymbol{e}_3 = (0, 0, 1)^t$, and this is the term considered in [6, 7].

In the context of gauged sigma models, it was proposed in [10] that one should more generally add the boundary term

$$E_{A,\infty}[\boldsymbol{n}] = -4\pi \Omega_A[\boldsymbol{n}] = \int_{\mathbb{R}^2} (\partial_1 (\boldsymbol{A}_2, \boldsymbol{n}) - \partial_2 (\boldsymbol{A}_1, \boldsymbol{n})) dx_1 dx_2 \tag{13}$$

to obtain a well-defined variational problem. Clearly, (13) reduces to (12) for the simple gauge field (7).

Adding the term (13) to the energy (3) has a number of advantages, at least from an analytical point of view. It does not change the Euler–Lagrange equation one obtains for variations which vanish rapidly at infinity, but its inclusion means that one can allow for variation with a slower fall-off. We refer the reader to [10] for details. Furthermore, the study of solutions of arbitrary degree in [9] shows that the modified energy is well-defined for some solutions for which the unmodified energy integral (3) is not.

Geometrically, the unmodified energy (3) has a natural interpretation when evaluated on a solution of the Bogomol'nyi equation as the equivariant degree of that solution [15]. By contrast, the modified energy evaluated on a solution $\boldsymbol{n}$ of the Bogomol'nyi equation is equal to the integral expression for the degree:

$$E_A[\boldsymbol{n}] + E_{A,\infty}[\boldsymbol{n}] = 4\pi\, Q[\boldsymbol{n}] \quad \text{if} \quad \boldsymbol{n} \times D_1\boldsymbol{n} = D_2\boldsymbol{n}. \tag{14}$$

3 Exact Magnetic Skyrmions

3.1 The General Solution in Four Easy Steps

Since the magnetic skyrmion energy functional (1) with the potential (6) is a particular example of the energy for a gauged sigma model of the form (3), we can obtain an infinite family of solutions of the variational equations by solving (11). Here we focus on the formula needed for magnetic skyrmions, so for constant gauge fields. In that case, the solution of (11) can be obtained via the following recipe. For details we again refer to [10].

(I) *Complex coordinate for the magnetisation:* In order to write down the solution, one needs to work in terms of a complex stereographic coordinate for the magnetisation field $\boldsymbol{n}$. It is given by stereographic projection from the south pole, or algebraically by

$$w = \frac{n_1 + in_2}{1 + n_3}. \tag{15}$$

(II) *Complexified gauge field:* Next, one needs to write the gauge field explicitly as an $su(2)$ matrix-valued gauge field on $\mathbb{R}^2$ according to

$$A_i = \sum_{a=1}^{3} A_i^a t_a, \quad i = 1, 2, \tag{16}$$

where $t_a = -\frac{i}{2}\tau_a$, and τ_a are the Pauli matrices. In fact we require the complex linear combination

$$A_{\bar{z}} = \frac{1}{2}(A_1 + iA_2). \tag{17}$$

In the case at hand, this is a constant, complex and traceless 2×2 matrix, so generically an element of $sl(2, \mathbb{C})$.

(III) *Solution in complex coordinates:* The solution of the Bogomol'nyi equation is given in terms of the exponential

$$g(\bar{z}) = \exp\left(-\frac{1}{2}(A_1 + iA_2)\bar{z}\right) = \begin{pmatrix} a(\bar{z}) & b(\bar{z}) \\ c(\bar{z}) & d(\bar{z}) \end{pmatrix}, \tag{18}$$

which is a 2×2 matrix function of $\bar{z}$ with determinant one. The general solution of (11) in stereographic coordinates is

$$w(z, \bar{z}) = \frac{c(\bar{z}) + d(\bar{z})f(z)}{a(\bar{z}) + b(\bar{z})f(z)}, \tag{19}$$

where f is an arbitrary holomorphic function from $\mathbb{C}$ into $\mathbb{C}P^1 \simeq \mathbb{C} \cup \{\infty\}$ (in particular it is allowed to take the value ∞).

(IV) *Translating back into Cartesian coordinates:* Substitution of the general solution (19) into the inverse of (15)

$$n_1 + in_2 = \frac{2w}{1 + |w|^2}, \quad n_3 = \frac{1 - |w|^2}{1 + |w|^2}, \tag{20}$$

yields an explicit (but possibly complicated) formula for the magnetisation field.

The energy density of Bogomol'nyi solutions is either the degree density or the sum of the degree density and the vorticity, depending on the choice of energy functional, see our discussion in Sect. 2.3. Expressions for both directly in terms the stereographic coordinates are given in [9, 10].

3.2 Examples

Axisymmetric DM Terms As discussed in [10], the DM term is invariant under rotations in the plane and simultaneous rotations of the magnetisation field about a suitable axis if and only if $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are orthogonal and have the same length. In that case $(\boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{A}_1 \times \boldsymbol{A}_2)$ is an oriented and (up to scaling) orthonormal basis of $\mathbb{R}^3$. With $|\boldsymbol{A}_1| = |\boldsymbol{A}_2| = \kappa$, the potential for the solvable model is conveniently expressed in terms of $\hat{\boldsymbol{A}} := \boldsymbol{A}_1 \times \boldsymbol{A}_2/\kappa^2$ as

$$V_A(n) = \frac{\kappa^2}{2}\left(1 - (\boldsymbol{n}, \hat{\boldsymbol{A}})\right)^2 = \frac{\kappa^2}{8}(\boldsymbol{n} - \hat{\boldsymbol{A}})^4. \tag{21}$$

The DM term (5) and the integrand of the boundary term (13) combine neatly into $\sum_{a=1}^{3}\sum_{i=1}^{2}\mathcal{D}_{ai}(\partial_i \boldsymbol{n} \times (\boldsymbol{n} - \hat{\boldsymbol{A}}))_a$ in this case. For the simple case (7) with DM term $\kappa(\boldsymbol{n}, \nabla \times \boldsymbol{n})$ and potential $V(\boldsymbol{n}) = \frac{\kappa^2}{2}(1 - n_3)^2$, one checks that the matrix (18) is

$$g = \begin{pmatrix} 1 & -\frac{i}{2}\kappa\bar{z} \\ 0 & 1 \end{pmatrix}. \tag{22}$$

The solution (19) is best written in terms of the inverse coordinate $v = 1/w$ as

$$v = -\frac{i}{2}\kappa\bar{z} + h, \tag{23}$$

where $h = 1/f$ is, like f, an arbitrary holomorphic map $\mathbb{C} \to \mathbb{C}P^1$. The simplest choice $h = 0$ leads to the Bloch hedgehog skyrmion with $Q = -1$ and the Belavin–Polyakov profile function $\theta(r) = 2\arctan\left(\frac{r}{2\kappa}\right)$, as already noticed in [7]. Many other solutions are discussed in [9], and our Fig. 1 shows the solutions one obtains for different choices of rational functions h. It follows from the calculations in [9] that the degree of a skyrmion configuration depends on the parameter $L = \lim_{|z|\to\infty} |(2h)/(\kappa\bar{z})|$. For configurations determined by (23) with rational $h(z) = p(z)/q(z)$, where p and q are polynomials of degree M and N, it is

$$Q[\boldsymbol{n}] = \begin{cases} M & \text{if} \quad L > 1 \\ N & \text{if} \quad L = 1 \\ N - 1 & \text{if} \quad L < 1. \end{cases} \tag{24}$$

This shows in particular that in this model there are infinitely many solutions of the Bogomol'nyi equation (11) for each integer degree $Q \geq -1$. The modified energy (14) takes the values $4\pi Q$ on these solutions.

Rank One DM Interactions The spiralization tensor has rank one when $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are linearly dependent, so $\boldsymbol{A}_1 \times \boldsymbol{A}_2 = 0$. In this case, the curvature $\boldsymbol{F}_{12}$ vanishes and the gauge field (and therefore the DM interaction) can be removed by an $SU(2)$ gauge transformation. The solvable model with the potential (6) can therefore be mapped into the standard Belavin–Polyakov $O(3)$ sigma model [11]. It follows immediately that solutions of (11) exist for any integer degree Q in these models, and that their energy (14) is $4\pi|Q|$. In particular it follows that skyrmions and anti-skyrmions of equal and opposite degree have the same energy. This result was derived in [8] for $|Q| = 1$ in more general rank one models. Example solutions of solvable rank one models and their properties are discussed in [10] and also [16]. We note that a similar reformulation in terms of a flat gauge field was recently applied to a rather different ferromagnetic model in [17].

4 Conclusion

We have shown that solvable models of magnetic skyrmions exist for any DM interaction term. Even though they require fine-tuning of the potential, their exact solutions shed light on qualitative properties of solutions in more general models. These include general features of multi-(anti)-skyrmion configurations such as the appearance of $Q + 1$ maxima in the energy density of certain charge $Q > 0$ configurations (as shown in the $Q = 5$ solution in Fig. 1), or the deformation of a skyrmion to an anti-skyrmion via a line defect, as shown in the top row of Fig. 1 and discussed in some detail in [9].

The solvable models also shed light on the crucial influence of the DM interaction on the relative energy of skyrmions compared to anti-skyrmions. Our short discussion illustrates the more general findings of [8]. In the axisymmetric models of Sect. 3.2, $Q = -1$ skyrmions have energy -4π, whereas $Q = 1$ anti-skyrmions have the opposite energy 4π (this can be reversed by a different choice of solvable model, see [9]). In rank one models, by contrast, skyrmions and anti-skyrmions have the same energy. It was shown in [8] that models with generic spiralization tensors should interpolate between these two extremes, and it would be interesting to explore this in the solvable models with generic DM terms.

To end, we note that the language of gauged non-linear sigma models provides a rare and rather beautiful link between pure mathematics and real physics by connecting the geometry of holomorphic maps and vortex equations as discussed in [14] with magnetic skyrmions. In fact, simply allowing the gauge field to depend non-trivially on space may provide further applications, for example, to the study of impurities as discussed in [18] and [10].

Acknowledgments I thank Bruno Barton-Singer for sharing his Python code for generating the plots in Fig. 1, Calum Ross for pointing out an error in an earlier version of this paper, and the referee for constructive comments.

References

1. A.N. Bogdanov, D.A. Yablonskii, Thermodynamically stable 'vortices' in magnetically ordered crystals. The mixed state of magnets. Zh. Eksp. Teor. Fiz. **95**, 178 (1989)
2. N. Nagaosa, Y. Tokura, Topological properties and dynamics of magnetic skyrmions. Nat. Nanotechnol. **8**, 899 (2013). https://doi.org/10.1038/NNANO.2013.243
3. I. Dzyaloshinskii, A thermodynamic theory of 'weak' ferromagnetism of antiferromagnetics. J. Phys. Chem. Solids **4**, 241 (1958). https://doi.org/10.1016/0022-3697(58)90076-3
4. T. Moriya, Anisotropic superexchange interaction and weak ferromagnetism. Phys. Rev. **120**, 91 (1960). https://doi.org/10.1103/PhysRev.120.91
5. A. Fert, V. Cros, J. Sampaio, Skyrmions on the track. Nat. Nanotechnol. **8**, 152 (2013). https://doi.org/10.1038/nnano.2013.29
6. C. Melcher, Chiral skyrmions in the plane. Proc. R. Soc. A **470**, 20140394 (2014). https://doi.org/10.1098/rspa.2014.0394

7. L. Döring, C. Melcher, Compactness results for static and dynamic chiral skyrmions near the conformal limit. Calc. Var. **56**, 60 (2017). https://doi.org/10.1007/s00526-017-1172
8. M. Hoffmann, B. Zimmermann, G.P. Müller, D. Schürhoff, N.S. Kiselev, C. Melcher, S. Blügel, Antiskyrmions stabilized at interfaces by anisotropic Dzyaloshinskii-Moriya interactions. Nat. Commun. **8**, 308 (2017). https://doi.org/10.1038/s41467-017-00313-0
9. B. Barton-Singer, C. Ross, B.J. Schroers, Magnetic skyrmions at critical coupling. Commun. Math. Phys. (2020). https://doi.org/10.1007/s00220-019-03676-1, https://arxiv.org/abs/1812.07268
10. B.J. Schroers, Gauged sigma models and magnetic skyrmions. SciPost Phys. **7**, 030 (2019). https://doi.org/10.21468/SciPostPhys.7.3.030
11. A.A. Belavin, A.M. Polyakov, Metastable states of two-dimensional isotropic ferromagnets. JETP Lett. **22**, 245 (1975)
12. F.N. Rybakov, N.S. Kiselev, Chiral magnetic skyrmions with arbitrary topological charge ('skyrmionic sacks'). Phys. Rev. B **99**, 064437 (2019). https://doi.org/10.1103/PhysRevB.99.064437
13. D. Foster, C. Kind, P.J. Ackerman, J.S.B. Tai, M.R. Dennis, I.I. Smalyukh, Two-dimensional skyrmion bags in liquid crystals and ferromagnets. Nat. Phys. **15**, 655 (2019). https://doi.org/10.1038/s41567-019-0476-x
14. K. Cieliebak, A.R. Gaio, D.A. Salamon, J-holomorphic curves, moment maps, and invariants of Hamiltonian group actions. Int. Math. Res. Notices **10**, 831 (2000). https://doi.org/10.1155/S1073792800000453
15. E. Walton, Some exact skyrmion solutions on curved thin films. https://arxiv.org/abs/1908.08428
16. M. Hongo, T. Fujimori, T. Misumi, M. Nitta, N. Sakai, Instantons in chiral magnets. https://arxiv.org/abs/1907.02062
17. M. Speight, T. Winyard, Skyrmions and spin waves in frustrated ferromagnets at low applied magnetic field. https://arxiv.org/abs/1909.07970
18. C. Adam, J. M. Queiruga, A. Wereszczynski, BPS soliton-impurity models and supersymmetry. J. High Energy Phys. **2019**, 164 (2019). https://doi.org/10.1007/JHEP07(2019)164

Applications of Symmetry to the Large Scale Structure of the Universe (Scale Invariance) and to the Hadronic Spectrum (Heavy Diquark Symmetry)

Mark B. Wise

Abstract Two applications of symmetries of quantum theories that are relevant in very different circumstances, early universe cosmology and hadronic physics, are discussed. The first concerns constraints on the form of correlations of the density perturbations in the very early universe. The second is about the hadronic spectrum, in particular states with flavor quantum numbers corresponding to two heavy bottom quarks and two light anti-quarks (tetraquarks with two heavy quarks).

Keywords Symmetry · Scale invariance · Hadronic spectrum

1 Scale Invariance and Inflationary Cosmology

In this section I will not provide references to the early pioneering work on inflation. They can be found in the excellent review [1].

The inflationary cosmology is a hypothetical era in the very early universe when the energy density was dominated by potential energy (i.e., an effective cosmological constant). We view this as arising from the evolution of a scalar field (the inflation) which temporarily is moving slowly along a flat part of its potential with almost constant energy density V. If there is a large enough smooth patch of space time the metric in that patch becomes de Sitter space time, where

$$ds^2 = -dt^2 + a(t)^2 dx^i dx^i, \quad a(t) = e^{Ht} \text{ and } H^2 = \frac{8\pi}{3} GV. \tag{1}$$

With enough of this de Sitter exponential expansion this smooth patch can expand to cover the whole visible universe. Eventually the scalar field enters a part of its potential where it moves more rapidly towards the minimum and the universe

M. B. Wise (✉)
Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA, USA
e-mail: wise@theory.caltech.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_51

reheats because of the inflations coupling to other fields, for example, the standard model, Higgs field. I will not dwell on how hard it actually is to dynamically realize this situation. The upshot is that we end up with an extremely smooth (the exponential expansion dilutes all classical inhomogeneities) region that includes our whole visible universe (and perhaps much more).

This scenario provides a mechanism to dynamically generate the density perturbations through quantum fluctuations in a scalar degree of freedom. In different gauges precisely how the scalar degree of freedom is defined takes different forms. We prefer to work in a gauge where the inflation field is spatially homogeneous. In this gauge it is the curvature perturbation $\delta_\zeta(\mathbf{x}, t)$ during the inflationary era that is the scalar degree of freedom that characterizes the fluctuations about pure de Sitter space time. Just after inflation the energy density perturbations δ_ρ that cause the microwave background anisotropies and the correlations of galaxy positions (on large scales) have wavelengths well outside the horizon. They reenter the horizon during the matter dominated era and on long wavelengths are related to the curvature perturbations generated during the inflationary era by,

$$\delta_\rho \propto \nabla^2 \delta_\zeta . \tag{2}$$

1.1 Symmetries of Space Time Metrics

Let us start with the metric for flat Minkowski space time, $ds^2 = -dt^2 + dx^i dx^i$. Some symmetries of this metric are translations of time $t \to t + d$, translations of space $x^i \to x'^i = x^i + c^i$ and rotations $x^i \to x'^i = R^{ij} x^j, \quad RR^T = I$. These symmetries have implications for the equal time correlations of a scalar quantum field $\phi(t, \mathbf{x})$ in a Minkowski space background. Fourier transforming,

$$\phi(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{\phi}(t, \mathbf{k}) \tag{3}$$

one has that

$$\langle \tilde{\phi}(t, \mathbf{k}) \tilde{\phi}(t, \mathbf{k}') \rangle = P_\phi(k) (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}'). \tag{4}$$

The Dirac delta function arises from the spatial translation invariance, and the fact that the power spectrum $P_\phi(k)$ only depends on k, the magnitude of the wave vector $\mathbf{k}$ follows from rotational invariance. Time translation invariance did not play a role since we evaluated the equal time two-point correlation.

For a de Sitter space time background the metric is no longer invariant under time translations. Instead a combined time translation and scale transformation

$$t \to t' = t + d, \; x^i \to x'^i = \lambda x^i, \quad \lambda = e^{-Hd} \tag{5}$$

is an invariance. At infinite time this just becomes scale invariance. Now let the scalar field $\phi(t, \mathbf{x})$ be propagating in a de Sitter background space time. Fourier transforming with respect to the comoving coordinate $\mathbf{x}$,

$$\phi(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{x}} \tilde{\phi}(t, \mathbf{k}). \tag{6}$$

In free quantum field theory it is modes of fixed comoving wave vector $\mathbf{k}$ that evolve independently in a simple fashion. The physical wave vector is $\mathbf{p} = \mathbf{k}/a(t)$ is time dependent. In this case the symmetries of the metric imply that,

$$\langle \tilde{\phi}(\infty, \mathbf{k}) \tilde{\phi}(\infty, \mathbf{k}') \rangle = P_\phi(k)(2\pi)^3 \delta(\mathbf{k} + \mathbf{k}'), \quad P_\phi(\mathbf{k}) \propto 1/k^3. \tag{7}$$

Note that the scale invariance fixes the k dependence of the power spectrum. It is only the overall normalization of the power spectrum that is not determined by the symmetries.

1.2 Mass Density Perturbations in the Inflationary Cosmology

During the inflationary era there are quantum fluctuations of the scalar curvature field ζ and since the metric is approximately de Sitter and as $t \to \infty$ in the inflationary era,

$$P_\zeta(k) \propto \frac{1}{k^3}. \tag{8}$$

Today ($t = t_0$[1]) we are interested in fluctuations in the mass density ρ. At very small wave vectors

$$\tilde{\delta}_\rho(\mathbf{k}; t_0) \propto \tilde{\delta}_\zeta(\mathbf{k}) k^2, \tag{9}$$

where the factor of k^2 arises from the Laplacian in Eq. (2). This implies that the power spectrum for the mass density fluctuations has the form

$$P_\rho(k; t_0) \propto \frac{k^2 k^2}{k^3} = k \tag{10}$$

at small k. This is called a Harrison–Zeldovich spectrum.

Observationally the measured power spectrum determined by measurements of the cosmic microwave background (CMB) anisotropy is quite close to this. At low k the Planck collaboration found that [2],

$$P_\rho(k; t_0) \propto k^{n_s}, \quad n_s = 0.9655 \pm 0.0062. \tag{11}$$

[1] We take $a(t_0) = 1$ so physical and comoving wave vectors coincide.

In the inflationary cosmology scale invariance is not exact. It is broken by a small amount because the inflaton potential is not exactly flat and the Hubble constant decreases somewhat as the inflaton evolves along its potential. The measured tilt $(n_s - 1)$ is consistent with what inflationary models give. Moving away from very small k the density perturbations take the form

$$\tilde{\delta}_\rho(\mathbf{k}; t_0) = \tilde{\delta}_\zeta(\mathbf{k}) k^2 T(k), \tag{12}$$

where the transfer function $T(k)$ goes to a constant as $k \to 0$ but cuts off the growth at large k.

Non-Gaussianities in the curvature probability distribution imply the existence of higher point connect correlations of the curvature scalar. For the n-point correlation

$$\langle \zeta(\infty, \mathbf{k}_1), \ldots, \tilde{\zeta}(\infty, \mathbf{k}_n)\rangle_c = P_\zeta^{(n)}(\mathbf{k}_1, \ldots, \mathbf{k}_n)(2\pi)^3 \delta(\mathbf{k}_1 + \cdots + \mathbf{k}_n). \tag{13}$$

The higher point spectra[2] are also constrained by the symmetries of de Sitter space. The $P^{(n)}$'s are rotationally invariant functions of the wave vectors so they must depend only on the magnitudes and dot products of the wave vectors. Furthermore scale invariance implies that

$$P_\zeta^{(n)}(\lambda \mathbf{k}_1, \ldots, \lambda \mathbf{k}_n) = \lambda^{3-3n} P_\zeta^{(n)}(\mathbf{k}_1, \ldots, \mathbf{k}_n), \tag{14}$$

where the $+3$ in the power of λ arises because of the Dirac delta function on the sum of the wave vectors.

An example of a curvature bispectrum $P^{(3)}$ consistent with this scaling is

$$P^{(3)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \propto \left[\left(\frac{1}{k_1^3}\right)\left(\frac{1}{k_2^3}\right) + \left(\frac{1}{k_1^3}\right)\left(\frac{1}{k_3^3}\right) + \left(\frac{1}{k_2^3}\right)\left(\frac{1}{k_3^3}\right) \right], \tag{15}$$

where $k_j = |\mathbf{k}_j|$. This form occurs in a phenomenological model called local non-Gaussianity.

Measurements of the CMB anisotropy place constraints on primordial non-Gaussianity [3]. Those constraints are usually written as limits on the quantities f_{NL} and τ_{NL} where schematically,

$$P_\zeta^{(3)} \sim f_{NL} P_\zeta^{(2)} P_\zeta^{(2)}, \qquad P_\zeta^{(4)} \sim \tau_{NL} P_\zeta^{(2)} P_\zeta^{(2)} P_\zeta^{(2)}. \tag{16}$$

Since the microwave background temperature anisotropy is very small $\delta T/T \sim 10^{-5}$ it implies very schematically that ζ is of that order. $P_\zeta^{(3)}$ has three ζs in its definition, while $P_\zeta^{(2)} P_\zeta^{(2)}$ has four. Hence, if non-Gaussianity was order one, we

[2]Note $P^{(2)} = P$.

would expect that $f_{NL} \sim 10^5$. But the constraint from the CMB is (for local non-Gaussianity) roughly $f_{NL} < 10$. Clearly this is a very strong constraint and the fact that single field inflation is consistent with this is a success of that model. Similar remarks hold for τ_{NL}.

Even though in standard single field inflation the perturbations in the curvature and hence the mass density are almost Gaussian [5], more complicated models with additional fields that are not much heavier than the Hubble constant during inflation and couple to the inflation can give rise to non-Gaussianities that are consistent with CMB limits and may be measured in the future by galaxy surveys.

The simplest model of this type is called quasi single field inflation (QSFI) [6] and it has one additional scalar field. If the mass of the additional scalar m is small compared with the Hubble constant during inflation, then the bispectrum in the so-called squeezed limit (i.e., $q = |\mathbf{k}_1| = |\mathbf{k}_2 + \mathbf{k}_3| \ll |\mathbf{k}_{2,3}|$) can have dramatic physical consequences that we discuss later. In the squeezed limit $\mathbf{k}_2 \simeq -\mathbf{k}_3$ (in configuration space it corresponds to the points x_2 and x_3 close together and well separated from x_3) QSFI gives for the bispectrum,

$$P^{(3)} \propto \frac{k^{-\alpha}}{q^{3-\alpha}k^3}, \tag{17}$$

where $\alpha = m^2/(3H)^2$ with m the mass of the additional scalar. This form holds when α is small compared with unity. There is a similarly simple expression for the trispectrum $P^{(4)}$ in what is called the compressed limit where $\mathbf{k_1} \simeq -\mathbf{k}_2$ $(k \simeq k_{1,2})$ $\mathbf{k_3} \simeq -\mathbf{k}_4$ $(k' \simeq k_{3,4})$ and $q = |\mathbf{k}_1 + \mathbf{k}_2| = |\mathbf{k}_3 + \mathbf{k}_3| << k, k'$,

$$P^{(4)} \propto \frac{(kk')^{-\alpha}}{q^{3-2\alpha}(kk')^3}. \tag{18}$$

Note in coordinate space the compressed limit corresponds to the spatial points $x_{1,2}$ close to each other and $x_{3,4}$ close to each other with these pairs of points far apart. The form of the bispectrum and trispectrum in the squeezed and compressed limits in QSFI has been discussed several times in the literature, see, for example, [7] and the references therein.

1.3 Non-Gaussianity and the Galaxy Power Spectrum at Small Wave Vectors

Galaxies do not trace the mass. Rather their smoothed number density n_g is a continuous field in some function of the mass density fluctuations averaged over a region of size R that corresponds to the size of a region that collapses to form a galaxy when the fluctuations in the mass density go nonlinear. We denote this averaged mass density fluctuation by $\delta_{\rho,R}(\mathbf{x})$. In Fourier space $\tilde{\delta}_{\rho,R}(\mathbf{k}) =$

$W(kR)\tilde{\delta}_\rho(\mathbf{k})$ where $W(kR)$ is a window function that goes to zero for $k >> 1/R$ and a constant for $k << 1/R$. For simplicity we imagine the galaxy number density fluctuations $(\delta_g(\mathbf{x}) = (n_g(\mathbf{x}) - \overline{n})/\overline{n})$ are related to the smeared mass density perturbations by the expansion

$$\delta_g(\mathbf{x}) = b_1\delta_{\rho,R}(\mathbf{x}) + b_2\left(\delta_{\rho,R}(\mathbf{x})^2 - \langle(\delta_{\rho,R}(\mathbf{x}))^2\rangle\right) + \cdots \tag{19}$$

This relation is sometimes called a bias expansion. The values of the bias coefficients b_j are independent of $\delta_{\rho,R}$ and fix what a galaxy is in terms of the smeared primordial fluctuations. Of course more complicated expansions involving derivatives would be more realistic but this will do for our purposes. Fourier transforming gives a power spectrum for galaxies

$$\langle\tilde{\delta}_g(\mathbf{q}_1)\tilde{\delta}_g(\mathbf{q}_2)\rangle = P_g(q)(2\pi)^3\delta(\mathbf{q}_1 + \mathbf{q}_2), \tag{20}$$

where in the argument of the power spectrum $q = |\mathbf{q}_1| = |\mathbf{q}_2|$. The two-point correlation for the fluctuations in the galaxy number density then has an expansion

$$\begin{aligned}\langle\delta_g(\mathbf{x})\delta_g(\mathbf{y})\rangle = {} & b_1^2\langle\delta_{\rho,R}(\mathbf{x})\delta_{\rho,R}(\mathbf{y})\rangle \\ & + b_2 b_1\left(\langle\delta_{\rho,R}(\mathbf{x})^2\delta_{\rho,R}(\mathbf{y})\rangle_c + \langle\delta_{\rho,R}(\mathbf{x})\delta_{\rho,R}(\mathbf{y})^2\rangle_c\right) \\ & + b_2^2\langle\delta_{\rho,R}(\mathbf{x})^2\delta_{\rho,R}(\mathbf{y})^2\rangle_c + \cdots\end{aligned} \tag{21}$$

The term proportional to b_1^2 gives a contribution to the power spectrum for the galaxy number density perturbations that, at small q, is linear in the wave vector q since it is proportional to the power spectrum for the smeared mass density fluctuations. The other terms can give very different behavior at small q if there is primordial non-Gaussianity. Using Eq. (17) at low q the primordial three-point function gives a contribution to the galaxy power spectrum that is proportional to $b_1 b_2$ and has an enhanced behavior at small q (recall α is small),

$$P_g(q) \sim b_1 b_2\left[\int \frac{d^3k}{(2\pi)^3}T(k)^2(k^2)^2W(kR)^2(1/k^{3+\alpha})\right]q^2/q^{3-\alpha} \propto 1/q^{1-\alpha}. \tag{22}$$

Similarly the primordial trispectrum in Eq. (18) gives a contribution to the galaxy power spectrum

$$P_g(q) \sim b_2^2\left[\int \frac{d^3k}{(2\pi)^3}T(k)^2(k^2)^2W(kR)^2(1/k^{3+\alpha})\right]^2 1/q^{3-2\alpha} \propto 1/q^{3-2\alpha}. \tag{23}$$

Enhancements by $1/q^2$ and $1/q^4$ of the power spectrum for fluctuations in the galaxy number density over the power spectrum for the mass density fluctuations (e.g., Eqs. (22) and (23) when $\alpha = 0$) were originally discovered, respectively,

in [8] and [9]. Similar enhancements occur in the primordial bispectrum for galaxies in QSFI [10].

The enhancements in Eqs. (22) and (23) occur in other inflationary models. In QSFI they arise from tree Feynman diagrams in de Sitter space. There are simple models where it is one loop diagrams [11] that give rise to these effects.

The SPHEREx all sky survey should be able to observe primordial non-Gaussianity if f_{NL} is greater than unity (http://spherex.caltech.edu). It may also see these enhancements in the galaxy power spectrum at low q.

2 Heavy Diquark and Quark Spin Flavor Symmetry

In this section I will mostly give references for more recent work. See [12] for references to the earlier work on the applications of heavy quark symmetry and the development of the $1/m_Q$ expansion.

Heavy quarks Q have $m_Q \gg \Lambda_{\rm QCD}$. In a meson or baryon that contains a single heavy quark and other light degrees of freedom (i.e., light quarks and gluons) the interactions of the light degrees of freedom with the heavy quark and between themselves involve momentum transfers of order the non-perturbative scale associated with the strong interactions, $\Lambda_{\rm QCD}$. Changes in the heavy quark velocity caused by these interactions $\Delta v \sim \Lambda_{\rm QCD}/m_Q \to 0$ as $m_Q \to \infty$. In this limit the strong interactions of light degrees of freedom with the heavy quark are independent of the heavy quark's mass and spin. Hence for charm and bottom quarks with the same four velocity, in the limit $m_{c,b} \to \infty$, there is an $SU(4)$ spin flavor symmetry.[3] This symmetry has physical implications that are important for understanding the properties of hadrons containing a single heavy quark and for extracting the elements of the CKM matrix V_{cb} and V_{ub} from experimental data. Note that the heavy quark flavor symmetry arises not because the charm and bottom quarks have almost equal mass ($m_c/m_b \sim 1/3$) but rather because both their masses are large when compared with $\Lambda_{\rm QCD}$.

In the limit $m_Q \to \infty$, for a given spin of the light degrees of freedom $s_l \neq 0$, hadrons with a single heavy quark come in degenerate multiplets with total spin $s = s_l - 1/2$ and $s = s_l + 1/2$. For $s_l = 1/2$ and $Q = b$ an example is the $\bar{B}$ ($s = 0$) and $\bar{B}^*$ ($s = 1$) mesons. Hadrons with $s_l = 0$ (e.g., for $Q = c$ the Λ_c) have total spin 1/2 and are not members of a doublet with different total spins.

The effective theory for a heavy quark Q moving with four velocity v replaces the full QCD heavy quark field Q with Q_v where

$$Q(x) = \exp[-im_Q v \cdot x]\,(Q_v(x) + \cdots)\,, \tag{24}$$

[3] The top quark is the heaviest of the quarks but since its mass is greater than the W boson mass its lifetime is comparable to time scales associated with non-perturbative strong interactions and so it never forms a hadron.

and $v_\mu \gamma^\mu Q_v = Q_v$. The exponential factor scales out the large part of the heavy quark's momentum proportional to its mass. Using this definition one can derive an effective theory for a heavy quark where the Lagrange density takes the form

$$\mathcal{L} = i\bar{Q}_v v \cdot D Q_v - \frac{1}{2m_Q}\bar{Q}_v D^2 Q_v - c(m_Q/\mu)\frac{g}{4m_Q}\bar{Q}_v \sigma \cdot G Q_v + \cdots. \tag{25}$$

The leading interactions, which are not suppressed by powers of $1/m_Q$, are independent of m_Q and display the spin flavor symmetry. In the above c is a constant and μ is the subtraction point. Dependence on the subtraction point cancels between the matching coefficient c and matrix elements of the chromo magnetic moment operator it multiplies. At leading log level, $c(1) = 1$.

2.1 Evidence for the Applicability of Heavy Quark Symmetry

Heavy quark spin flavor symmetry is an approximate symmetry and so it is useful before we apply a variant of it to tetraquarks to see how well some of the predictions work for ordinary mesons that contain a heavy quark. Since the term in Eq. (25) that breaks the spin symmetry (i.e., the chromomagnetic operator term) and induces a splitting between members of the meson spin doublets has a coefficient with heavy quark mass dependence $1/m_Q$ (neglecting the m_Q dependence in c) one has the approximate relation $m^2_{B^*} - m^2_B = m^2_{D^*} - m^2_D$. This works reasonably well (especially when you take into account that the charm quark is not so heavy compared with the scale of the strong interactions). Experimentally $m^2_{B^*} - m^2_B = 480 \times 10^3$ MeV^2, while $m^2_{D^*} - m^2_D = 550 \times 10^3$ MeV^2. But there is more going here than heavy quark symmetry since it also works for the hadrons with only light quarks, i.e., $m^2_\rho - m^2_\pi = 570 \times 10^3$ MeV^2 and $m^2_{K^*} - m^2_K = 550 \times 10^3$ MeV^2.

Next consider the pseudoscalar meson decay constants f_M defined by,

$$\langle 0|\bar{q}\gamma_\mu Q|M(p)\rangle = f_M p_\mu. \tag{26}$$

Here M is a pseudoscalar meson with $Q\bar{q}$ flavor quantum numbers and p is its four momentum. Treating the bottom and charm quark masses as very large, when compared with the scale of the non-perturbative strong interactions, heavy quark flavor symmetry predicts that

$$\frac{f_B}{f_D} = \left(\frac{m_D}{m_B}\right)^{1/2}\left(\frac{\alpha_s(m_b)}{\alpha_s(m_c)}\right)^{-6/25} \simeq 0.67. \tag{27}$$

For this ratio lattice QCD calculations are more accurate than the data and give [13], $f_B/f_D \simeq 184/213 \simeq 0.86$. This is not a huge success of heavy quark methods, but perhaps reasonable agreement with the prediction of heavy quark symmetry in Eq. (27) given that there are $\Lambda_{\rm QCD}/m_c$ corrections.

The next topic we discuss is the semileptonic form factors for $\bar{B} \to D\ell\nu_\ell$ and $\bar{B} \to D^*\ell\nu_\ell$ decay. The D and D^* mesons are members of the charm quark ground state heavy meson doublet with $s_l = 1/2$. The Lorentz invariant form factors that determine the invariant matrix elements for these decays (and hence the differential decay distributions) are defined by,

$$\langle D(v')|\bar{c}\gamma^\mu b|\bar{B}(v)\rangle = (m_B m_D)^{1/2}\left[h_+(v+v')^\mu + h_-(v-v')^\mu\right], \tag{28}$$

$$\langle D^*(v'\epsilon)|\bar{c}\gamma^\mu b|\bar{B}(v)\rangle = (m_B m_{D^*})^{1/2} h_V \epsilon^{\mu\nu\alpha\beta}\epsilon^*_\nu v'_\alpha v_\beta, \tag{29}$$

and

$$\langle D^*(v'\epsilon)|\bar{c}\gamma^\mu\gamma_5 b|\bar{B}(v)\rangle = (m_B m_{D^*})^{1/2}\left[-ih_{A_1}(w+1)\epsilon^{*\mu} + ih_{A_2}((w)(\epsilon^*\cdot v)v^\mu + ih_{A_3}(\epsilon^*\cdot v)v'^\mu\right]. \tag{30}$$

In the equations above v is the four velocity of the decaying $\bar{B}$ meson, v' is the four velocity of the final state D or D^* meson, and $w = v\cdot v'$. The point $w = 1$ is special since at that point the final state charmed meson is at rest in the rest frame ($v = (1, 0, 0, 0)$), of the decaying $\bar{B}$ meson. Because heavy quark flavor symmetry relates particles at the same four velocity and not momentum it is convenient to use initial and final four velocities to parameterize the matrix elements. The six Lorentz scalar form factors h_+, h_-, h_V, h_{A_1}, h_{A_2}, and h_{A_3} are functions of w. The form factors have a weak dependence on the charm and bottom quark masses that can be calculated by perturbatively matching QCD onto the heavy quark effective theory. Neglecting matching corrections, heavy quark spin symmetry allows one to relate all the form factors to a single universal function $\xi(w)$,

$$h_+(w) = h_V(w) = h_{A_1}(w) = h_{A_3}(w) = \xi(w), \quad h_-(w) = h_{A_2}(w) = 0. \tag{31}$$

Heavy quark flavor symmetry determines that $\xi(1) = 1$ and furthermore there are no $\Lambda_{QCD}/m_{c,b}$ corrections to this normalization condition for the zero recoil matrix elements of the currents. The absence of $\Lambda_{QCD}/m_{c,b}$ corrections to the zero recoil matrix elements is known as Luke's theorem.

There is evidence from lattice QCD simulations that heavy quark symmetry predictions for the semileptonic form factors in the D meson case work well, in particular h_- is much smaller than h_+ and $h_+(1)$ is close to unity. That is illustrated by the plots of recent lattice results [14] for $h_\pm$ versus w in Figs. 1 and 2.

Armed with some confidence that heavy quark spin flavor symmetry is a quantitatively useful tool we now discuss its application to tetra quarks that contain two heavy quarks.

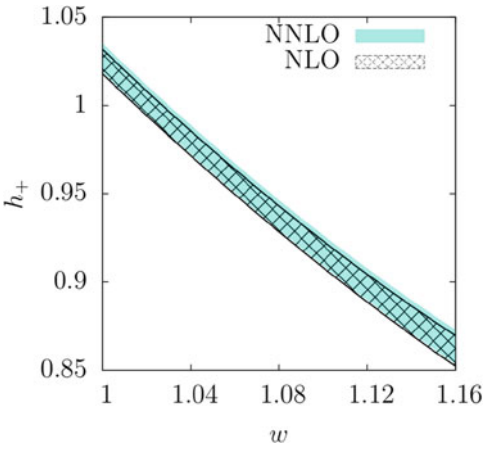


Fig. 1 Comparison of the NLO (hatched) and NNLO (solid) chiral-continuum fits for h_+ versus w

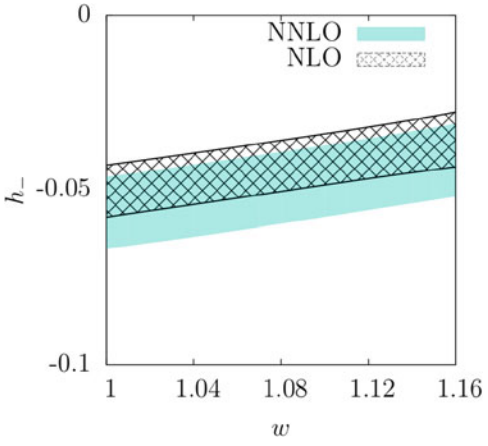


Fig. 2 Comparison of the NLO (hatched) and NNLO (solid) chiral-continuum fits for h_- versus w

2.2 $T_{QQ\bar{q}\bar{q}}$ Tetraquarks

For very heavy quarks Q the lowest lying $T_{QQ\bar{q}\bar{q}}$ tetraquark states are stable with respect to the strong and electromagnetic interactions [16, 17]. The reason for this is quite simple. If $m_Q \gg \Lambda_{\rm QCD}$, then the heavy QQ diquark in the tetraquark will be in a color $\bar{\mathbf{3}}$ configuration since in the color channel $\mathbf{3} \times \mathbf{3} \to \bar{\mathbf{3}}$ the one gluon color Coulombic potential,

$$V_c(r) = -\frac{2}{3}\frac{\alpha_s}{r} \tag{32}$$

is attractive. Hence the diquark has a large Coulombic binding energy[4] (compared with $\Lambda_{\rm QCD}$) and a small size[5] (compared with $1/\Lambda_{\rm QCD}$). Strong decay of the lowest lying $T_{QQ\bar{q}\bar{q}}$ tetraquark states to a baryon with two heavy quarks and an anti-nucleon (when $q = u, d$), $\Xi_{QQq} + \bar{N}_{qqq}$, is kinematically forbidden since the final state has an additional $q\bar{q}$ pair, which costs an additional ~600 MeV of mass. Strong decay to two heavy mesons $M_{Q\bar{q}} + M_{Q\bar{q}}$ does not require an additional $q\bar{q}$ pair but now the final state does not have the large color Coulombic binding energy proportional to m_Q that the tetraquark state does and so this channel is also kinematically forbidden. For very heavy quarks the diquark inside a tetraquark is a small color anti-triplet that can be treated as a point like object.

There has been some recent theoretical work [18–20] that indicates that the bottom quark is heavy enough for the ground state $T_{bb\bar{q}\bar{q}}$ tetraquarks to be stable with respect to the strong and electromagnetic interactions with a mass that is around 100 MeV below the $\bar{B}\bar{B}$ threshold. Such tetraquarks produced at a collider will be long lived with a lifetime approximately half the B meson lifetime. The three papers supporting this conclusion on the stability of tetraquarks with two heavy bottom quarks are a lattice calculation, a fit using heavy quark methods, and a fit/derivation using more phenomenological methods.

Treating the color triplet $\bar{Q}\bar{Q}$ diquark as point like there is heavy quark-diquark spin flavor symmetry and it implies (neglecting terms suppressed by the heavy quark mass) the mass relation,

$$m_{T_{QQ\bar{q}\bar{q}}} - m_{\Lambda_{Qqq}} = m_{\Xi_{QQq}} - m_{M_{Q\bar{q}}}, \tag{33}$$

which was used in [19]. Here I have written the mass relation when the two light anti-quarks in the tetraquark and baryon are in a light degree of freedom spin-zero configuration (this is appropriate when the two light quarks are an up and a down quark). There is an analogous relation when $s_l = 1$. It is easy to understand why Eq. (33) holds in the heavy quark and diquark limit. On both sides of the equation the mass of the light degrees of freedom cancels out and the difference between the diquark and quark masses is the same on the left- and right-hand sides of the equation. There are other interesting applications of heavy quark-diquark symmetry, see, for example, [21].

We are interested in the case where the diquark consists of two heavy b quarks. Then the heavy diquark has spin one. The heavy flavor symmetry is not between and bottom and a charm quark but between a b quark and a bb diquark. Hence there are no Λ_{QCD}/m_c corrections to Eq. (33). The notation for the light quarks is a little schematic in Eq. (33). For example, as we already mentioned the subscript qq could correspond to ud. Furthermore we have neglected the light quark masses.

This symmetry relation neglects the finite size of the diquark QQ configuration and so relies on the diquark being small, with a size much less than $1/\Lambda_{QCD}$. A

[4]The binding energy is of order $\alpha_s^2 m_Q$.

[5]The size is of order $1/(\alpha_s m_Q)$.

contribution to the shift in the mass of the tetraquark due to the radius of the diquark was estimated in [22] to be of order

$$\Delta m_{T_{bb\bar{q}\bar{q}}} \sim \left(\frac{\pi \alpha_s(m_b v_{\rm rel}) \langle r^2 \rangle}{54} \right) f_B^2 m_B , \tag{34}$$

where $\langle r^2 \rangle$ is the radius squared of the non-relativistic diquark state. This mass shift is only about 10 MeV for[6] $\langle r^2 \rangle = 3.2\ {\rm GeV}^{-2}$ and so treating the diquark as a point object is likely to be a good approximation. Finite diquark size corrections are not always proportional to $\alpha_s \langle r^2 \rangle \propto 1/(\alpha_s m_Q^2)$. A recent calculation [23] found a finite diquark size correction to hyperfine splittings that goes as $(1/m_Q^2)\alpha_s \langle 1/r \rangle \sim \alpha_s^2/m_Q$.

No stable (with respect to the strong and electromagnetic interactions) hadrons have been discovered that contain two heavy bottom quarks. In 2017, the doubly charmed baryon Ξ_{cc}^{++} (or in the notation used in this paper Ξ_{ccu}) was discovered at LHCb [24]. It has been observed in the exclusive weak decay modes, $\Xi_{cc}^{++} \to \Lambda_c^+ K^- \pi^+ \pi^-$ (the discovery mode) and $\Xi_{cc}^{++} \to \Xi_c^+ \pi^+$ [25]. There is considerable interest in the detection of the analogous baryons containing two heavy bottom quarks Ξ_{bbq}, $q = u, d$, partly because it would be the first step towards observing the tetraquark states, $T_{bb\bar{q}\bar{q}}$.

Recently, Gershen and Poluektov [26] proposed the inclusive decay mode Ξ_{bbq} or $T_{bb\bar{q}\bar{q}} \to \bar{B}_c + X_{c,s,q}$ as a potential discovery channel for the doubly bottom weakly decaying hadrons at the LHC. They made the clever observation that $\bar{B}_c$'s that do not point back to the collision interaction point, but instead are displaced from it, can only arise from the weak decay of a hadron with two bottom quarks. They also note that the decay chain, $\bar{B}_c \to J/\psi \pi^- \to \mu^+ \mu^- \pi^-$, is a reasonable mode to detect a displaced $\bar{B}_c$ meson. Ordinary $\bar{B}$ mesons that are displaced will not do for this purpose since they can result from the weak decay of a long lived $\bar{B}_c$ meson. The branching ratio for this $\bar{B}_c$ decay is not small and there will be many more $\bar{B}_c$'s produced at the interaction point by hadronization then there are baryons with two heavy bottom quarks. The method of Gershen and Poluektov would not tell you if it was a tetraquark or baryon with two heavy bottom quarks that was produced or the mass of the hadron with two heavy bottom quarks that was decaying to the displaced $\bar{B}_c$ but it would represent the first step along the journey to discover tetraquarks with two heavy bottom quarks that are stable with respect to the strong and electromagnetic interactions.

Acknowledgments This work was supported by the by DOE Grant DE-SC0011632. I am also grateful for the support provided by the Walter Burke Institute for Theoretical Physics.

[6] This is the radius squared for the diquark bound by the Cornell potential $V = -(2/3)0.3/r + (1/2)(0.2\ {\rm GeV}^2) r$.

References

1. D. Baumann, PoS TASI **2017**, 009 (2018). arXiv:1807.03098 [hep-th]
2. Y. Akrami et al. [Planck Collaboration], arXiv:1807.06211 [astro-ph.CO]
3. Y. Akrami et al. [Planck Collaboration], arXiv:1905.05697 [astro-ph.CO]
4. P.A.R. Ade et al. [Planck Collaboration], Astron. Astrophys. **594**, A17 (2016). arXiv:1502.01592 [astro-ph.CO]
5. J.M. Maldacena, J. High Energy Phys. **0305**, 013 (2003). astro-ph/0210603
6. X. Chen, Y. Wang, J. Cosmol. Astropart. Phys. **1004**, 027 (2010). arXiv:0911.3380 [hep-th]
7. D. Baumann, S. Ferraro, D. Green, K.M. Smith, J. Cosmol. Astropart. Phys. **1305**, 001 (2013). arXiv:1209.2173 [astro-ph.CO]
8. N. Dalal, O. Doré, D. Huterer, A. Shirokov, Phys. Rev. D **77**, 123514 (2008). arXiv:0710.4560 [astro-ph]
9. T.J. Allen, B. Grinstein, M.B. Wise, Phys. Lett. B. **197**, 66 (1987)
10. H. An, M. McAneny, A.K. Ridgway, M.B. Wise, Phys. Rev. D **97**(12), 123528 (2018). arXiv:1711.02667 [hep-ph]
11. H. An, M.B. Wise, Z. Zhang, Phys. Rev. D **99**(5), 056007 (2019). arXiv:1806.05194 [hep-ph]
12. A.V. Manohar, M.B. Wise, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **10**, 1 (2000)
13. S. Aoki et al., Review of lattice results concerning low-energy particle physics (2016). arXiv:1607.00299 [hep-latt]
14. J.A. Bailey et al. [MILC Collaboration], Phys. Rev. D **92**(3), 034506 (2015)
15. A. Vaquero Avilés-Casco, C. DeTar, D. Du, A. El-Khadra, A.S. Kronfeld, J. Laiho, R.S. Van de Water, EPJ Web Conf. **175**, 13003 (2018). arXiv:1710.09817 [hep-lat]
16. J. Carlson, L. Heller, J.A. Tjon, Phys. Rev. D **37**, 744 (1988)
17. A.V. Manohar, M.B. Wise, Nucl. Phys. B **399**, 17 (1993)
18. A. Francis, R.J. Hudspith, R. Lewis, K. Maltman, Phys. Rev. Lett. **118**, 142001 (2017). arXiv:1607.05214 [hep-lat]; EPJ Web Conf. **175**, 05023 (2018). arXiv:1711.03380 [hep-lat]
19. E.J. Eichten, C. Quigg, Phys. Rev. Lett. **119**(20), 202002 (2017). arXiv:1707.09575 [hep-ph]
20. M. Karliner, J.L. Rosner, Phys. Rev. Lett. **119**(20), 202001 (2017). arXiv:1707.07666 [hep-ph]
21. T. Mehen, Phys. Rev. D **96**(9), 094028 (2017). arXiv:1708.05020 [hep-ph]
22. H. An, M.B. Wise, The direct coupling of light quarks to heavy di-quarks. Phys. Lett. B **788**, 131 (2019). arXiv:1809.02139 [hep-ph]
23. T.C. Mehen, A. Mohapatra, Phys. Rev. D **100**(7), 076014 (2019). arXiv:1905.06965 [hep-ph]
24. LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett **119**, 112001 (2017). arXiv:1707.01621.06657 [hep-ex]
25. LHCb Collaboration, R. Aaij et al., Phys. Rev. Lett **121**, 162002 (2018). arXiv:1806.01621.01919 [hep-ex]
26. T. Gershon, A. Poluektov, J. High Energy Phys. **1901**, 019 (2019). arXiv:1810.06657 [hep-ph]

Leptophobic Z' in Supersymmetry and Where to Find Them

Jack Y. Araz

Abstract This report presents possible loopholes in heavy Z' searches which are concentrated around high-mass resonances. We present a scenario where massive neutral vector boson Z', predicted by E_6-inspired $U(1)'$ models, can become leptophobic due to kinetic mixing between $Z - Z'$. Such a scenario can avoid high-mass dilepton resonance searches. Furthermore, we propose a channel through supersymmetric cascade decays to observe leptophobic Z'. We show that it is possible to reach up to 8σ at $\sqrt{s} = 14$ TeV and 3 ab^{-1} integrated luminosity. This report follows the steps of the work that has been conducted in Araz et al. (J High Energy Phys 02:092, 2018).

Keywords Supersymmetry · Z'

1 Introduction

Heavy neutral gauge boson, Z', can be obtained by extending the Standard Model (SM) gauge structure with a $U(1)$ gauge group. Such models have been widely studied under Grand Unified Scheme [1]. Similarly, under supersymmetric realisations, Z' can emerge from $U(1)$ extended minimal supersymmetric extension of the SM (UMSSM). It can also be achieved in the so-called Sequential Standard Model (SSM) where Z', and W', bosons have same couplings to fermions as their SM counterparts.

Search for heavy neutral vector bosons in LHC has been extensively studied by experimental collaborations due to its relatively familiar resonance signature [2–7]. These searches set the mass bound for the Z' assuming that it can only decay through the SM particles. These studies focus on high-mass dilepton or dijet resonances.

The original version of this chapter was revised: Figures 1–4 have been updated. A correction to this chapter is available at https://doi.org/10.1007/978-3-030-55777-5_61

J. Y. Araz (✉)
Concordia University, Montréal, QC, Canada
e-mail: jack.araz@concordia.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_52

In case of the dijet resonance studies at $\sqrt{s} = 13$ TeV, ATLAS collaboration [5] presented the Z' mass bounds around 2.1–2.9 TeV for 37 fb^{-1} integrated luminosity and CMS collaboration [3] around 2.7 TeV for 12.9 fb^{-1}. In case of dilepton resonance studies, ATLAS collaboration presented the mass bounds at 3.8–4.1 TeV for 36.1 fb^{-1} [2] and CMS collaboration presented at 3.5 TeV for 13 fb^{-1} integrated luminosity.[1] We investigate possible loopholes within the specific scenarios to lower these mass limits and also propose a leptophobic framework for Z' to search for a possible signal through supersymmetric decay channels in HL- & HE-LHC.

The rest of this report is organised as follows; in Sect. 2, we introduced the theoretical properties of UMSSM and discussed current mass bounds on Z' at the time when this study conducted. In the following Sect. 3, we discussed possible supersymmetric signatures to observe a leptophobic Z' in HL- & HE-LHC and finally we conclude in Sect. 4.

2 Theoretical Framework and Z' Mass Bounds

Within a possible minimal representation in supersymmetry (SUSY), Z' bosons can arise from the breaking of rank-6, E_6 group, which is greatly motivated by Grand Unified Theories. Among the various possible breaking schemes of E_6 exceptional group, the so-called UMSSM framework being the $U(1)'$ extended MSSM, arises from the breaking through $SO(10)$. The two-step breaking structure creates separate $U(1)$ gauge groups [8–10]

$$E_6 \to SO(10) \otimes U(1)_\psi \to (SU(5) \otimes U(1)_\chi) \otimes U(1)_\psi ,$$

where these $U(1)$'s can mix via θ_{E_6} angle and generate $U(1)'$,

$$U(1)' = U(1)_\psi \cos\theta_{E_6} - U(1)_\chi \sin\theta_{E_6} .$$

Here $SU(5)$ group further breaks into MSSM gauge structure. Alongside with such vector boson, extending MSSM framework with a $U(1)$ gauge provides an anomaly free gauge structure. This can prevent the R-parity violating terms from the Lagrangian and eliminates short-lived proton possibility which occurs due to $SU(5)$ breaking.

In E_6, the matter sector of UMSSM arises from the breaking of **27**-plet vector representation in to $\mathbf{16} \oplus \overline{\mathbf{10}} \oplus 1$, where **16** further decomposes to the matter fields, $\overline{\mathbf{10}}$ to electroweak Higgs fields and exotic quarks and singlet, 1, becomes $U(1)'$ singlet superfield, S. In addition to the MSSM's large particle content, UMSSM accommodates extended slepton, sneutrino, and squark sector up to 6 supersymmetric particles in each family. Also, breaking $SO(10)$'s **16**-plet vector

[1] It is important to note that these limits are recently updated. See [6] for dijet resonances at 137 fb^{-1} and [7] for dilepton resonances at 139 fb^{-1}.

representation provides a right-handed neutrino superfield which generates right-handed neutrino and its supersymmetric partner in this framework. The Higgs sector consists of two charged $H^{\pm}$ and four neutral scalar bosons, which are a pseudoscalar A and three neutral scalars h, H, and H', where h and H are being MSSM-like Higgs. This spectrum also creates an extended gaugino sector with a total of six neutralinos where chargino sector remains unchanged.

At tree level, the squared mass of Z' boson is given by

$$M_{Z'}^2 = g'({Q'_S}^2 v_S^2 + {Q'_{H_u}}^2 v_{H_u}^2 + {Q'_{H_d}}^2 v_{H_d}^2) \,,$$

where g' being the coupling constant of the $U(1)'$ gauge group, Q'_i is corresponding $U(1)'$ charge of singlet s, doublet H_u, and doublet H_d superfields, and v_i being their respective vacuum expectation values. The existence of the $U(1)'$ group leads to extra D- & F-terms which contributes to the masses of the supersymmetric partners and Higgses. Large bounds of Z' boson requires large v_S which further leads to a heavy particle spectrum especially in the Higgs sector. Thus, requiring massive Z' also causes massive particle spectrum, which can be tuned via corresponding softmasses.

The interaction Lagrangian of the fields $\hat{B}_\mu$ and $\hat{Z}'_\mu$ with the fermions is suppressed by their corresponding coupling and charge, $g_1 Y_i$ and $g' Q'_i$, respectively, where i represents the corresponding fermionic field in the interaction, g_1 and g' are $U(1)_Y$ and $U(1)'$ couplings, respectively. The typical E_6-inspired relation between $U(1)_Y$ and $U(1)'$ couplings is taken to be $g'/g_1 = \sqrt{5/3}$. In [11], it has been shown that shifting this relation from GUT scale to Z'-mass scale, and abandoning grand unification assumption, and adding the possibility to decay through supersymmetric particles, can reduce Z' mass limit up to 300 GeV depending on θ_{E_6} angle [12, 13]. Breaking SUSY sector in a higher scale generates enough room for $U(1)'$ coupling to grow via renormalisation group equations which leads to larger g' and higher production amplitudes for Z'. Thus, breaking SUSY in a lower energy scale limits the growth in g' and reduces the production cross section, which leads to relatively less stringent mass bounds for Z'. In Fig. 1 we compared high-mass dilepton limit from ATLAS collaboration [2] to our results, sampling θ_{E_6} at -0.79π $(U(1)'_\eta)$ and 0 $(U(1)'_\psi)$, where Z' production has been calculated in next-to-leading order (NLO) precision. In both panels, the y-axis shows the production cross section with its branching fraction to dilepton final states where Z' mass varied between 2 and 5 TeV within the bounds of narrow width approximation. Panels show the effect of different breaking scales where for the left panel g'/g_1 relation is taken at GUT scale, and in the right panel, it downgraded to Z' mass scale. The dashed line shows the case where Z' is only allowed to decay into the SM particles (USM), and the dotted line shows the case with SUSY particles added (UMSSM). The error bars represent the theoretical uncertainties, scale and parton distribution function (PDF) uncertainties added in quadrature. The calculation has been performed via MadGraph [14]. The left panel of Fig. 1 reveals that, when the relation between couplings are assumed to be taken at GUT scale, the addition of supersymmetric partners to the possible decay channels improves the mass limits around 20% and

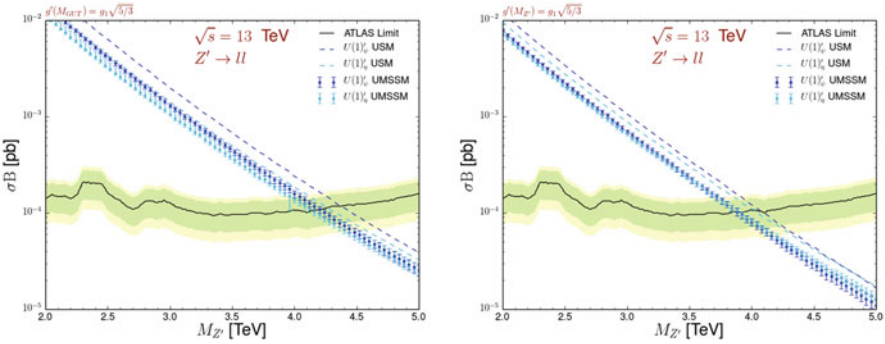


Fig. 1 The effect of symmetry breaking scale in both inclusive (UMSSM) and noninclusive (USM) supersymmetric particle cases is presented. The left panel shows Z' mass versus its corresponding cross section times branching ratio to dilepton final states where SUSY breaking occurring at GUT scale. The dotted line represents UMSSM and dashed line USM realisations, and the error bars show the variation caused by theoretical uncertainties. The right panel shows the same for SUSY breaking occurring at Z' mass scale. These results are presented in [11]

reduces it up to 4 TeV. Moving this relation to a lower scale in the right panel reduces the mass limits up to 3.8 TeV, where still preserving the difference between USM and UMSSM.

Although reducing production amplitudes and adding supersymmetric particles as possible decay channels create a considerable deviation from the current bounds, it is also possible to propose an altogether leptophobic Z' to avoid the bounds coming from Drell–Yan modes. Leptophobic Z' can be achieved by mixing $U(1)_Y$ and $U(1)'$ gauge fields. The mixing can be delivered in twofolds. First, there can be mixing between Z and Z' mass eigenstates. This, however, is constrained by electroweak precision tests to be at the $\mathcal{O}(10^{-3})$ [15]. Secondly, a kinematical mixing between Z and Z' can be introduced through an angle χ,

$$\mathcal{L}_{\text{kin}} \subset -\frac{\sin\chi}{2}\hat{B}^{\mu\nu}\hat{Z}'_{\mu\nu}\,,$$

where $\hat{B}^{\mu\nu}$ and $\hat{Z}'_{\mu\nu}$ are the $U(1)_Y$ and $U(1)'$ boson field strength tensors, respectively. This mixing further modifies the interaction vertices between Z' and other fermions. The interaction term can be written in terms of canonical (diagonal) kinetic terms, Z'_μ, as,

$$\mathcal{L}_{\text{int}} \subset -g'\bar{\psi}_i\gamma^\mu\left(\frac{Q'_i}{\cos\chi} - \frac{g_1}{g'}Y_i\tan\chi\right)Z'_\mu\psi_i\,,$$

where ψ_i is corresponding fermionic field and Y_i and Q'_i being their $U(1)_Y$ and $U(1)'$ charges. For simplicity, we limited the Lagrangian with the Z' portion. The term in the parenthesis can be represented as modified $U(1)'$ charge, $\bar{Q}_i$. To end up with a leptophobic Z', $\bar{Q}_i$ needs to be zero for left and right lepton superfields

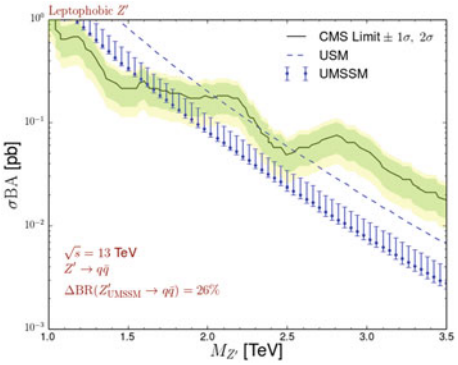


Fig. 2 Plot shows leptophobic Z' mass versus production cross section and dijet branching ratio with corresponding acceptance rate. The dashed line shows the USM sample, which assumes that Z' can only decay through the SM particles, and the dotted line shows UMSSM, including all supersymmetric particles. Error bars include scale and PDF variations as well as the variation in the dijet decay rate per sample. The CMS limit is taken from [3]

[16]. As shown in [11], left and right leptonic superfield charges can be set to zero, $\bar{Q}_L = \bar{Q}_E = 0$, only around

$$\theta_{E_6} = (n - 0.79)\pi \quad \text{for} \quad n = 0, 1, 2 \ldots \quad , \tag{1}$$

which requires $\sin\chi \cong (-1)^{n+1}0.3$, assuming $g'/g_1 = \sqrt{5/3}$ ratio is satisfied at Z' mass scale.

Such leptophobic Z' can only be bounded via high-mass dijet searches. It has been shown, in [3], that these bounds are much weaker compared to Drell–Yan signature. In Fig. 2 we performed a similar comparison as above where this time the leptophobic Z' mass varied to compare with CMS collaboration's dijet limits [3]. In order to zoom into the respectively low mass region, we varied Z' mass between 1 and 3.5 TeV and plotted against the multiplication of its production cross section, branching ratio to dijet final states and the acceptance rate stated in [3]. Again, the dashed line assumes that Z' can only decay to the SM particles (USM) and dotted line includes its possible decay channels to SUSY particles as well (UMSSM). In addition to the theoretical uncertainties, the error bars also show the variation in the total decay rate to dijet final states. Such construction reduces the Z' mass limits to 1.6 TeV and creates a considerably large deviation between USM and UMSSM depending on the dijet branching fraction.

3 Leptophobic Z' in HL- & HE-LHC

A leptophobic Z' can still be observed through leptonic final states, but instead of direct decays, leptons can be produced through a cascade decay of supersymmetric particles. In Fig. 3, we proposed a possible scenario where two charginos are pair

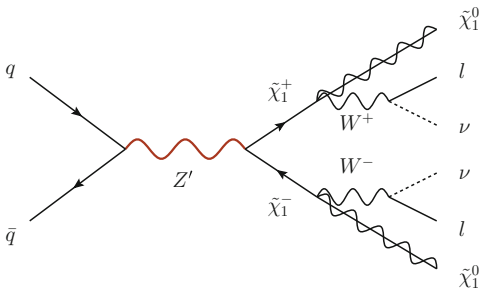


Fig. 3 Chosen analysis signal where Z' production and its decay through two lightest charginos. The figure has been produced with the help of the JaxoDraw package [17]

Table 1 UMSSM parameters for the reference points **BM I** and **BM II**

Parameter	$\tan\beta$	μ_{eff} (GeV)	M_0 (TeV)	M_1 (GeV)	$M_{\tilde{\chi}_1^\pm}$ (GeV)
BM I	9.11	218.9	2.6	106.5	344
BM II	16.08	345.3	1.9	186.7	178

Parameter	M_2 (GeV)	M_3 (TeV)	M_1' (GeV)	A_0 (TeV)	A_λ (TeV)
BM I	230.0	3.6	198.9	2	5.9
BM II	545.5	5.5	551.7	1.5	5.1

produced via on-shell Z' and they further decay into missing energy and dileptons through W. Such scenario might generate its background through heavy Higgs production and their decay to charginos. However, we observed that the branching fraction of such decay is negligible compared to heavy Z boson's decay rate. Also, neutralino pair production or chargino-neutralino associative production through Z' has been observed to have small branching fraction compared to chargino pair production. To test this hypothesis, we choose two UMSSM benchmarks to study possible observability of such leptophobic scenarios. Following [11], we chose two benchmarks using Eq. (1), where for **BM I** n chosen to be 0 and for **BM II** n chosen to be 1. A conservative, 2.5 TeV Z' mass has been chosen for both benchmarks where other parameters are shown in Table 1. The particle mass spectrum has been calculated in SARAH [18] interfaced with SPheno [19, 20]. The particle masses and decay tables can be found in [11] which are not quoted here. The Z' branching ratio to chargino pairs is about 2% (6%) for **BM I** (**BM II**). To ensure 100% branching ratio of chargino to LSP, the mass splitting between two particles is chosen to be slightly larger than W mass.

The amplitudes of the process are computed at NLO precision for a centre-of-mass energy $\sqrt{s} = 14$ TeV, using MadGraph5_aMC@NLO [14] with the NLO set of NNPDF 2.3 parton densities. The production cross section of Z' calculated as 120 fb. Prepared matrix-level events are hadronized and showered in Pythia 8 [21] and the detector response has been simulated in Delphes [22] using the Snowmass parametrization [23, 24]. The jets are clustered using FastJet program [25] with `antikT` algorithm using $R = 0.6$.

For the preselection, jets (leptons) are required to have minimum 40 (20) GeV transverse momentum and $|\eta| < 2.4$ (1.5). Leptons within $\Delta R < 0.4$ to a hard jet are omitted. Only two muons are allowed in the final state with 15% hadronic activity within a cone radius $R = 4$ and the angular separation in the transverse plane between two leptons is required to be greater than 2.5. To isolate the dilepton final state, all jets are vetoed. The leading and second-leading leptons are required to have at least 300 and 200 GeV transverse momentum, respectively. Finally, the missing transverse energy has been required to be greater than 100 GeV. We calculated the significance of the signal over the background using two different approaches. One being standard ($s = S/\sqrt{B+\sigma_B^2}$) and other begin Asimov significance [26] given as;

$$Z_A = \sqrt{2\left((S+B)\ln\left[\frac{(S+B)(S+\sigma_B^2)}{B^2+(S+B)\sigma_B^2}\right] - \frac{B^2}{\sigma_B^2}\ln\left[1+\frac{\sigma_B^2 S}{B(B+\sigma_B^2)}\right]\right)},$$

where S being the number of signal events, B number of the background events, and σ_B is the uncertainty over the background. For 10% systematic uncertainty on the background, we calculated $s = 4.2\sigma$, $Z_A = (3.5 \pm 0.9)\sigma$ for **BM I**, and $s = 8.14\sigma$, $Z_A = (6.06 \pm 1.18)\sigma$ for **BM II** at 3 ab^{-1}. The details of the analysis alongside with the differential distributions can be found in [11].

It is possible to project these results into higher energies with different luminosity values to estimate the effect of HL- & HE-LHC. In Fig. 4, we calculated Z' production cross section at $\sqrt{s} = 27$ TeV with its mass varying between 1.5 and 4 TeV. Using a naive assumption that all the kinematic structure and cut efficiencies will propagate in the same way, we projected our results to 3 ab^{-1} and 15 ab^{-1} and presented them with various systematic uncertainties on the background from 10% to 30% where the error bars showing the theoretical uncertainties. Our results shows

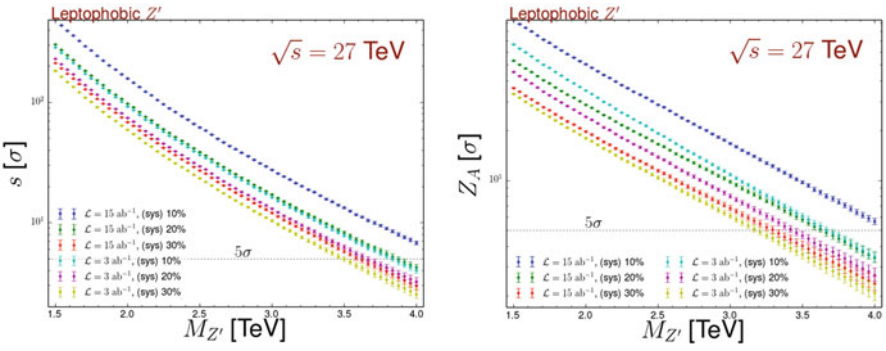


Fig. 4 Left panel shows the standard significance against Z' mass plotted for various luminosity and systematic uncertainty values. Right panel shows the same figure for Asimov significance, Z_A. These results are presented in [27]

that 5σ significance can be reached using standard (Z_A) significance for a Z' at 3.5 (3.2) TeV assuming 30% systematic uncertainty and 3 ab^{-1} integrated luminosity. We showed that, due to the linear dependence of the systematic uncertainties to the luminosity, its effect increases with increasing luminosity.

4 Conclusion

We investigate possible loopholes in E_6-inspired UMSSM searches at LHC, specifically through Z' production. We showed that the inclusion of supersymmetric particles could cause a 20% reduction in Z' mass limits, and even those limits can be reduced further by changing the SUSY breaking scale. Throughout our analysis, we continually compared UMSSM with USM to emphasise the difference that inclusion of supersymmetric particles can create.

Furthermore, we examined the possibility of having a leptophobic Z' by adding kinetic mixing between $Z - Z'$ to avoid high-mass leptonic resonance searches. We showed that the bounds from dijet resonances are considerably small and allow a light Z' up to 1.6 TeV. To back up such claim, we proposed a supersymmetry driven channel to observe such leptophobic Z'. We showed that it is possible to reach 8σ sensitivity at 14 TeV centre-of-mass energy with 3 ab^{-1} integrated luminosity using chargino production through Z' where charginos further decayed into two leptons and missing energy.

Assuming all cut efficiencies and kinematic structures will be preserved, we extrapolated our results to higher energies and luminosities. We showed the observable mass range for Z' up to 5σ sensitivity for 27 TeV centre-of-mass energy. All results are compared with different systematic uncertainties on the background and different luminosity values to address the importance of understanding the background as well as the signal itself. These results show that investigating such scenarios is both feasible and worthwhile at HL- & HE-LHC.

References

1. P. Langacker, The physics of heavy $Z^{'}$ gauge bosons. Rev. Mod. Phys. **81**, 1199–1228 (2009). http://dx.doi.org/10.1103/RevModPhys.81.1199
2. ATLAS Collaboration, M. Aaboud et al., Search for new high-mass phenomena in the dilepton final state using 36.1 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV with the ATLAS detector (2017). http://arxiv.org/abs/1707.02424
3. CMS Collaboration, A.M. Sirunyan et al., Search for dijet resonances in proton–proton collisions at $\sqrt{s}$ = 13 TeV and constraints on dark matter and other models. Phys. Lett. **B769**, 520–542 (2017). http://dx.doi.org/10.1016/j.physletb.2017.02.012, http://arxiv.org/abs/1611.03568
4. CMS Collaboration, Search for a high-mass resonance decaying into a dilepton final state in 13 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV. Tech. Rep. CMS-PAS-EXO-16-031, CERN, Geneva (2016)

5. ATLAS Collaboration, M.E.A. Aaboud, Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s} = 13$ TeV with the ATLAS detector. Phys. Rev. D **96**, 052004 (2017). http://dx.doi.org/10.1103/PhysRevD.96.052004
6. CMS Collaboration, A search for dijet resonances in proton-proton collisions at $\sqrt{s} = 13$ TeV with a new background prediction method. Tech. Rep. CMS-PAS-EXO-19-012, CERN, Geneva (2019)
7. ATLAS Collaboration, G. Aad et al., Search for high-mass dilepton resonances using 139 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector. Phys. Lett. **B796**, 68–87 (2019). http://dx.doi.org/10.1016/j.physletb.2019.07.016, http://arxiv.org/abs/1903.06248
8. R. Slansky, Group theory for unified model building. Phys. Rep. **79**, 1–128 (1981). http://dx.doi.org/https://doi.org/10.1016/0370-1573(81)90092-2
9. R.H. Benavides, L. Muoñz, W.A. Ponce, O. Rodríguez, E. Rojas, Electroweak couplings and LHC constraints on alternative Z' models in E_6. Int. J. Mod. Phys. **A33**, 1850206 (2018). http://dx.doi.org/10.1142/S0217751X18502068. http://arxiv.org/abs/1801.10595
10. J.Y. Araz, M. Frank, B. Fuks, Differentiating $U(1)'$ supersymmetric models with right sneutrino and neutralino dark matter. Phys. Rev. **D96**, 015017 (2017). http://dx.doi.org/10.1103/PhysRevD.96.015017, http://arxiv.org/abs/1705.01063
11. J.Y. Araz, G. Corcella, M. Frank, B. Fuks, Loopholes in Z' searches at the LHC: exploring supersymmetric and leptophobic scenarios. J. High Energy Phys. **02**, 092 (2018). http://dx.doi.org/10.1007/JHEP02(2018)092, http://arxiv.org/abs/1711.06302
12. G. Corcella, S. Gentile, Heavy neutral gauge bosons at the LHC in an extended MSSM. Nucl. Phys. **B866**, 293–336 (2013). http://dx.doi.org/10.1016/j.nuclphysb.2012.09.009, http://dx.doi.org/10.1016/j.nuclphysb.2012.11.024, http://arxiv.org/abs/1205.5780
13. G. Corcella, Phenomenology of supersymmetric Z' decays at the Large Hadron Collider. Eur. Phys. J. **C75**, 264 (2015). http://dx.doi.org/10.1140/epjc/s10052-015-3459-9, http://arxiv.org/abs/1412.6831
14. J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. J. High Energy Phys. **07**, 079 (2014). http://dx.doi.org/10.1007/JHEP07(2014)079, http://arxiv.org/abs/1405.0301
15. J. Erler, P. Langacker, T.-J. Li, The Z-Z' mass hierarchy in a supersymmetric model with a secluded $U(1)'$ breaking sector. Phys. Rev. **D66**, 015002 (2002). http://dx.doi.org/10.1103/PhysRevD.66.015002, http://arxiv.org/abs/hep-ph/0205001
16. K.S. Babu, C.F. Kolda, J. March-Russell, Leptophobic $U(1)'$s and the R(b) - R(c) crisis. Phys. Rev. **D54**, 4635–4647 (1996). http://dx.doi.org/10.1103/PhysRevD.54.4635, http://arxiv.org/abs/hep-ph/9603212
17. D. Binosi, J. Collins, C. Kaufhold, L. Theussl, JaxoDraw: a graphical user interface for drawing Feynman diagrams. Version 2.0 release notes. Comput. Phys. Commun. **180**, 1709–1715 (2009). http://dx.doi.org/10.1016/j.cpc.2009.02.020, http://arxiv.org/abs/0811.4113
18. F. Staub, SARAH 4: a tool for (not only SUSY) model builders. Comput. Phys. Commun. **185**, 1773–1790 (2014). http://dx.doi.org/10.1016/j.cpc.2014.02.018, http://arxiv.org/abs/1309.7223
19. W. Porod, F. Staub, SPheno 3.1: extensions including flavour, CP-phases and models beyond the MSSM. Comput. Phys. Commun. **183**, 2458–2469 (2012). http://dx.doi.org/10.1016/j.cpc.2012.05.021, http://arxiv.org/abs/1104.1573
20. W. Porod, SPheno, a program for calculating supersymmetric spectra, SUSY particle decays and SUSY particle production at e^+e^- colliders. Comput. Phys. Commun. **153**, 275–315 (2003). http://dx.doi.org/10.1016/S0010-4655(03)00222-4, http://arxiv.org/abs/hep-ph/0301101
21. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten et al., An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015). http://dx.doi.org/10.1016/j.cpc.2015.01.024, http://arxiv.org/abs/1410.3012

22. DELPHES 3 Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens et al., DELPHES 3, A modular framework for fast simulation of a generic collider experiment. J. High Energy Phys. **02**, 057 (2014). http://dx.doi.org/10.1007/JHEP02(2014)057, http://arxiv.org/abs/1307.6346
23. M. Berggren, A. Cakir, D. Krücker, J. List, A. Lobanov, I.A. Melzer-Pellmann, Non-simplified SUSY: stau-Coannihilation at LHC and ILC, in *Proceedings, Community Summer Study 2013: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA*, 29 July–6 Aug 2013. http://arxiv.org/abs/1307.8076
24. J. Anderson et al., Snowmass energy Frontier simulations, in *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA*, 29 July–Aug 2013. http://arxiv.org/abs/1309.1057
25. M. Cacciari, G.P. Salam, G. Soyez, FastJet user manual. Eur. Phys. J. **C72**, 1896 (2012). http://dx.doi.org/10.1140/epjc/s10052-012-1896-2, http://arxiv.org/abs/1111.6097
26. G. Cowan, K. Cranmer, E. Gross, O. Vitells, Asymptotic formulae for likelihood-based tests of new physics. Eur. Phys. J. **C71**, 1554 (2011). http://dx.doi.org/10.1140/epjc/s10052-011-1554-0, http://dx.doi.org/10.1140/epjc/s10052-013-2501-z, http://arxiv.org/abs/1007.1727
27. Working Group 3 Collaboration, X. Cid Vidal et al., Beyond the standard model physics at the HL-LHC and HE-LHC. http://arxiv.org/abs/1812.07831

Axion-Like Particles, Magnetars, and X-ray Astronomy

Jean-François Fortin and Kuver Sinha

Abstract If they exist, axion-like particles (ALPs) can be produced in the core of magnetars. As they propagate, they then convert to photons in the magnetosphere, leading to possible signatures in the X-ray photon spectrum. Moreover, since ALPs only mix with the parallel mode of the photon, interesting modifications to the photon polarization are also possible. We determine these effects by computing the Stokes parameters in the presence of ALPs and show that the relevant quantity factorizes into a product of the ALP-to-photon conversion probability and a factor that depends on the initial mixture of ALPs and photons at the surface of the magnetar.

Keywords Astroparticle · Axions · Magnetars · X-rays

1 Introduction

Although the Standard Model (SM) of particle physics explains a wealth of phenomena, several questions remain unanswered. One of the outstanding problems of the SM concerns dark matter (DM). DM—non-baryonic matter not included in the SM—represents approximatively 85% of the total matter content of the Universe. It was introduced to explain discrepancies between theoretical predictions and direct observations of galaxy clusters and galactic rotation curves. The strong CP problem—another issue of the SM—relates to the smallness of the neutron electric dipole moment and can be explained by the introduction of axions [1, 2]. Axions, and more generically ALPs, could play the role of DM, thus simultaneously solving two important questions left unanswered by the SM.

J.-F. Fortin (✉)
Université Laval, Québec, QC, Canada
e-mail: Jean-Francois.Fortin@phy.ulaval.ca

K. Sinha
University of Oklahoma, Norman, OK, USA
e-mail: Kuver.Sinha@ou.edu

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_53

ALPs are ubiquitous in string theory [3]. Contrary to vanilla axions that couple to the SM through coupling constants dependent on their mass, ALPs in the string theory axiverse have independent parameters: their coupling constants to the SM are unrelated to their mass. Moreover, ALPs are generically very light and weakly coupled. Hence, they are notoriously hard to produce and detect. Their existence might nonetheless be inferred from indirect detection. Indeed, in dense stellar objects, ALP production can be significant and, once produced, they can easily escape, leading to an extra energy sink. Thus, ALP production cannot be too large otherwise the extra energy sink would ruin the agreement between standard astrophysical evolution scenarios and observations. The upper bound is usually taken as follows: the ALP emissivity should not be larger than the neutrino emissivity. This condition leads to one constraint on the ALP parameter space.

Moreover, in background magnetic fields, oscillations between ALPs and photons occur [4]. Such oscillations can lead to significant changes to the photon spectrum. Since the ALP emissivity can be as large as the neutrino emissivity, and the latter is usually much larger than the photon emissivity, a second constraint on the ALP parameter space appears. For the oscillations to be substantial, it is necessary to have extreme magnetic fields. Therefore magnetars—neutron stars with extreme magnetic fields—are the natural stellar candidates to focus on in such scenarios. Moreover, since the dominant ALP production mechanism in the core of magnetars is nucleon–nucleon bremsstrahlung, the ALP emissivity peaks in the X-ray range, thus ALP oscillations will predominantly occur with X-ray photons. One interesting effect of the oscillations is therefore related to X-ray polarization. Indeed, ALPs couple to the parallel mode of the photon, hence the photon spectrum and the photon polarization, described by the Stokes parameters, are modified by the presence of ALPs. This brief communication presents the modifications to the Stokes parameters originating from the production of ALPs in magnetars and their subsequent conversion in the magnetosphere [5, 6].

2 Oscillation Equations

The interacting Lagrangian between ALPs and the SM particles of interest for magnetars is given by

$$\mathcal{L} \supset -\frac{g}{4} a F_{\mu\nu} \tilde{F}^{\mu\nu} + g_{aN} (\partial_\mu a) \bar{N} \gamma^\mu \gamma_5 N. \tag{1}$$

Here the first term is responsible for ALP–photon oscillations in the magnetosphere while the second term leads to ALP production through bremsstrahlung in the magnetar core. Hence the usual energy sink argument constrains g_{aN} while the photon spectrum argument constrains g (more precisely a combination of g and g_{aN}).

At energies below the electron mass, only photons are relevant and their Lagrangian is obtained by integrating out the electron loop, leading to non-linear quantum electrodynamics (QED). With the non-linear QED Lagrangian and the ALP–photon interaction (1) the evolution equations for the ALP and photons are (with negligible plasma contributions) [4, 7]

$$i\frac{d}{dx}\begin{pmatrix} a \\ E_\parallel \\ E_\perp \end{pmatrix} = \begin{pmatrix} \omega r_0 + \Delta_a r_0 & \Delta_M r_0 & 0 \\ \Delta_M r_0 & \omega r_0 + \Delta_\parallel r_0 & 0 \\ 0 & 0 & \omega r_0 + \Delta_\perp r_0 \end{pmatrix} \begin{pmatrix} a \\ E_\parallel \\ E_\perp \end{pmatrix}, \tag{2}$$

where

$$\Delta_\parallel = \frac{1}{2} q_\parallel \omega \sin^2\theta, \quad \Delta_\perp = \frac{1}{2} q_\perp \omega \sin^2\theta, \quad \Delta_a = -\frac{m_a^2}{2\omega}, \quad \Delta_M = \frac{1}{2} g B \sin\theta. \tag{3}$$

Here the ALP, parallel, and perpendicular photon electric fields are $a(x)$, $E_\parallel(x)$, and $E_\perp(x)$, respectively. The dimensionless evolution parameter is given by $x = r/r_0$ where r is the distance from the center of the magnetar to the propagating fields and r_0 is the magnetar's radius. The propagating fields have energies ω while the angle between the magnetic field and the direction of propagation is θ. Finally, the ALP mass is given by m_a while $q_\parallel$ and $q_\perp$ are dimensionless functions of the background magnetic field B given by [4, 7]

$$q_\parallel = \frac{7\alpha}{45\pi} b^2 \hat{q}_\parallel, \qquad \hat{q}_\parallel = \frac{1 + 1.2b}{1 + 1.33b + 0.56b^2}, \tag{4}$$

$$q_\perp = \frac{4\alpha}{45\pi} b^2 \hat{q}_\perp, \qquad \hat{q}_\perp = \frac{1}{1 + 0.72b^{5/4} + (4/15)b^2}, \tag{5}$$

with $b = B/B_c$ and $B_c = m_e^2/e$ the critical QED field strength.

Hence the three-state oscillation system (2) decouples into a two-state oscillation system for the ALP and the parallel photon mode, and a one-state system for the perpendicular photon mode. Using the probability conservation property $\frac{d}{dx}[|a(x)|^2 + |E_\parallel(x)|^2] = 0$ it is easy to recast the evolution equations (2) in the form [5, 6]

$$\frac{d\chi(x)}{dx} = -\Delta_M r_0 \cos[\Delta\phi(x)], \tag{6}$$

$$\frac{d\Delta\phi(x)}{dx} = (\Delta_a - \Delta_\parallel) r_0 + 2\Delta_M r_0 \cot[2\chi(x)] \sin[\Delta\phi(x)], \tag{7}$$

$$\frac{d\Sigma\phi(x)}{dx} = (2\omega + \Delta_a + \Delta_\parallel) r_0 - 2\Delta_M r_0 \csc[2\chi(x)] \sin[\Delta\phi(x)], \tag{8}$$

$$\frac{d\phi_\perp(x)}{dx} = (\omega + \Delta_\perp) r_0, \tag{9}$$

where $a(x) = A\cos[\chi(x)]e^{-i\phi_a(x)}$ and

$$E_\parallel(x) = iA\sin[\chi(x)]e^{-i\phi_\parallel(x)}, \qquad E_\perp(x) = A_\perp e^{-i\phi_\perp(x)}. \tag{10}$$

Here the angle $\chi(x)$ determines the relative ALP–photon mixture, $\Delta\phi(x) = \phi_a(x) - \phi_\parallel(x)$ is the phase difference between the ALP field and the parallel photon field, $\Sigma\phi(x) = \phi_a(x) + \phi_\parallel(x)$ is the sum of the ALP field and the parallel photon field phases, while $\phi_\perp(x)$ is the phase of the perpendicular photon field.

Since the quantities of interest in the following are the intensities

$$I_a(x) = A^2\cos^2[\chi(x)], \qquad I_\parallel(x) = A^2\sin^2[\chi(x)], \qquad I_\perp(x) = A_\perp^2, \tag{11}$$

for the ALP field, the parallel photon field, and the perpendicular photon field, respectively, only the evolution equations for $\chi(x)$ and $\Delta\phi(x)$ are relevant. The evolution equations for $\Sigma\phi(x)$ and $\phi_\perp(x)$ can be discarded, greatly simplifying the analysis.

3 Stokes Parameters

Before discussing polarization, it is important to note that the dominant ALP production mechanism and the dominant hard X-ray photon production mechanism are completely independent. Hence, the initial phase difference $\Delta\phi_0 = \Delta\phi(1)$ is effectively random and one should average over it, leading to the averaged intensities

$$\bar{I}_a(\chi_0, x) = \int_0^{2\pi} \frac{d\Delta\phi_0}{2\pi} I_a(\chi_0, \Delta\phi_0, x), \quad \bar{I}_\parallel(\chi_0, x) = \int_0^{2\pi} \frac{d\Delta\phi_0}{2\pi} I_\parallel(\chi_0, \Delta\phi_0, x). \tag{12}$$

Therefore, at a distance xr_0 from the magnetar, the relevant Stokes parameters are

$$I(\chi_0, x) = \bar{I}_\perp(x) + \bar{I}_\parallel(\chi_0, x), \qquad Q(\chi_0, x) = \bar{I}_\perp(x) - \bar{I}_\parallel(\chi_0, x), \tag{13}$$

and depend on the initial relative ALP–photon mixture $\chi_0 = \chi(1)$. Here $I(\chi_0, x)$ and $Q(\chi_0, x)$ are the sum and difference between the (averaged) perpendicular photon intensity and the averaged parallel photon intensity, respectively. By subtracting the Stokes parameters at a distance xr_0 from the magnetar (which are ultimately observed on Earth) by their values at the surface of the magnetar (which are determined by the astrophysics of magnetars), the surface-subtracted Stokes parameters

$$\Delta I(\chi_0, x) = I(\chi_0, x) - I(\chi_0, 1) = \bar{I}_\parallel(\chi_0, x) - A^2\sin^2(\chi_0), \tag{14}$$

$$\Delta Q(\chi_0, x) = Q(\chi_0, x) - Q(\chi_0, 1) = -\Delta I(\chi_0, x) \tag{15}$$

characterize the effects of ALP–photon oscillations in the magnetosphere. Indeed, in the absence of ALPs, the surface-subtracted Stokes parameters are $\Delta I(\chi_0, x) = \Delta Q(\chi_0, x) = 0$. However, non-vanishing surface-subtracted Stokes parameters satisfying the relation $\Delta Q(\chi_0, x) = -\Delta I(\chi_0, x)$ can be seen as a smoking gun signature for the presence of ALPs.

4 ALP-to-Photon Conversion Probability

By studying the normalized surface-subtracted Stokes parameter

$$R(\chi_0, x) = \frac{\Delta I(\chi_0, x)}{A^2} \tag{16}$$

$$= \cos(2\chi_0) \int_0^{2\pi} \frac{d\Delta\phi_0}{2\pi} \frac{1}{2} \left\{ 1 - \frac{\cos\left[2\chi(x)|_{\chi(1)=\chi_0, \Delta\phi(1)=\Delta\phi_0} \right]}{\cos(2\chi_0)} \right\} \tag{17}$$

it is possible to show that $R(\chi_0, x)$ factorizes as

$$R(\chi_0, x) = P_{a\to\gamma}(x) \cos(2\chi_0), \tag{18}$$

where

$$P_{a\to\gamma}(x) = \sin^2\left[\chi(x)|_{\chi(1)=0, \Delta\phi(1)=0} \right], \tag{19}$$

is the ALP-to-photon conversion probability at a distance xr_0 for pure ALP initial state [6]. This remarkable factorization property can be traced back to the averaging over the initial phase difference $\Delta\phi_0$ and would not necessarily occur if the dominant ALP and photon production mechanisms were related.

Hence, the Stokes parameters (13) and the averaged intensities (12) at a distance xr_0 from the magnetar are given explicitly in terms of the initial photon and ALP intensities through $A_\perp$, A, and χ_0 (which depend on the production mechanisms) and the ALP-to-photon conversion probability through $P_{a\to\gamma}(x)$ (which depends on the ALP and magnetar parameters through the evolution equations). They are given respectively by

$$I(\chi_0, x) = A_\perp^2 + \frac{A^2}{2} \left\{ 1 + [2P_{a\to\gamma}(x) - 1] \cos(2\chi_0) \right\}, \tag{20}$$

$$Q(\chi_0, x) = A_\perp^2 - \frac{A^2}{2} \left\{ 1 + [2P_{a\to\gamma}(x) - 1] \cos(2\chi_0) \right\}, \tag{21}$$

for the Stokes parameters and

$$\bar{I}_a(\chi_0, x) = \frac{A^2}{2}\left\{1 - [2P_{a\to\gamma}(x) - 1]\cos(2\chi_0)\right\}, \tag{22}$$

$$\bar{I}_\parallel(\chi_0, x) = \frac{A^2}{2}\left\{1 + [2P_{a\to\gamma}(x) - 1]\cos(2\chi_0)\right\}, \tag{23}$$

$$\bar{I}_\perp(\chi_0, x) = A_\perp^2, \tag{24}$$

for the averaged intensities.

Under some assumptions (mainly usual perturbation theory and magnetic field dipolar approximation), the ALP-to-photon conversion probability can also be computed analytically, bypassing the evolution equations, as [6]

$$P_{a\to\gamma}(x) = \left(\frac{\Delta_{M0} r_0^3}{r_{a\to\gamma}^2}\right)^2 \times \begin{cases} \frac{\pi}{3|\Delta_a r_{a\to\gamma}|} e^{\frac{6\Delta_a r_{a\to\gamma}}{5}} & |\Delta_a r_{a\to\gamma}| > 0.45 \\ \frac{\Gamma\left(\frac{2}{5}\right)^2}{5^{\frac{6}{5}}|\Delta_a r_{a\to\gamma}|^{\frac{4}{5}}} & |\Delta_a r_{a\to\gamma}| < 0.45 \end{cases}, \tag{25}$$

where the conversion radius $r_{a\to\gamma}$ is given by

$$x_{a\to\gamma} = \frac{r_{a\to\gamma}}{r_0} = \left(\frac{7\alpha}{45\pi}\right)^{1/6}\left(\frac{\omega}{m_a}\frac{B_0}{B_c}|\sin\theta|\right)^{1/3}. \tag{26}$$

Hence the Stokes parameters and all averaged intensities have been determined in terms of the ALP and magnetar parameters.

5 Discussion and Conclusion

The introduction of ALPs can lead to extra cooling of stellar objects. This extra cooling can be used to restrict the ALP parameter space. Due to ALP–photon oscillations in background magnetic fields, ALPs propagating away from stellar objects like magnetars (which have magnetospheres with extreme magnetic fields) leads to modifications of the photon spectrum and polarization. By demanding that the photon luminosity derived from ALP–photon oscillations does not exceed the observed photon luminosity, extra constraints on the ALP parameter space appear. Moreover, since ALP–photon oscillations occur only with the parallel mode of the photon, the presence of ALPs could be inferred from modifications to the Stokes parameters.

Starting from the fact that the dominant production mechanisms for ALPs and photons are independent, we computed the Stokes parameters in the presence of ALPs in terms of the intensities at the magnetar's surface (determined by astrophysics) and the ALP-to-photon conversion probability (determined by the ALP and magnetar parameters through the evolution equations). In the process, we

uncovered an interesting factorization property. Using reasonable assumptions, we then computed the ALP-to-photon conversion probability analytically. Therefore, the Stokes parameters in the presence of ALPs were completely determined by the parameters of the theory, i.e. in terms of the astrophysics of magnetars and in terms of the ALP parameters. It is then straightforward to use these results to constrain the ALP parameter space for specific magnetars [5, 6]. Such an analysis shows that the Stokes parameter Q is modified by the presence of ALP oscillations and can even change sign, leading to a clear effect on X-ray polarization.

References

1. R.D. Peccei, H.R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977). https://doi.org/10.1103/PhysRevLett.38.1440
2. R.D. Peccei, H.R. Quinn, Phys. Rev. D **16**, 1791 (1977). https://doi.org/10.1103/PhysRevD.16.1791
3. A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, J. March-Russell, Phys. Rev. D **81**, 123530 (2010). https://doi.org/10.1103/PhysRevD.81.123530
4. G. Raffelt, L. Stodolsky, Phys. Rev. D **37**, 1237 (1988). https://doi.org/10.1103/PhysRevD.37.1237
5. J.F. Fortin, K. Sinha, J. High Energy Phys. **1806**, 048 (2018). arXiv:1804.01992 [hep-ph]. https://doi.org/10.1007/JHEP06(2018)048
6. J.F. Fortin, K. Sinha, J. High Energy Phys. **1901**, 163 (2019) arXiv:1807.10773 [hep-ph]. https://doi.org/10.1007/JHEP01(2019)163
7. D. Lai, J. Heyl, Phys. Rev. D **74**, 123003 (2006) [astro-ph/0609775]. https://doi.org/10.1103/PhysRevD.74.123003

Anomalies in B Decays: A Sign of New Physics?

David London

Abstract At the present time, there are a number of measurements of B-decay observables that disagree with the predictions of the standard model. These discrepancies have been seen in processes governed by two types of decay: (1) $b \to s\mu^+\mu^-$ and (2) $b \to c\tau^-\bar{\nu}$. In this talk, I review the experimental results, as well as the proposed new-physics explanations. We may be seeing the first signs of physics beyond the standard model.

Keywords Anomalies in B decays · New physics

1 Introduction

The development of the standard model (SM) in particle physics is one of the great triumphs in all of physics. The SM has made a great many predictions, almost all of which have been verified, including the existence of the Higgs boson. There is no question that the SM is correct.

However, there are many reasons to believe it is not complete, such as the large number of arbitrary parameters, the hierarchy problem, the matter-antimatter asymmetry in the universe, dark matter, etc. In order to address these issues, there must be physics beyond the SM. We do not know what the new physics (NP) is, nor where it is, so we have to search for it in all possible ways:

- Direct searches: in high-energy experiments, one task is to look for the production of new particles. Unfortunately, to date, such searches have revealed nothing. No SUSY, no direct dark matter detection, no new particles.
- Indirect searches: here the idea is to look for virtual effects of new particles. This method has been more promising.

D. London (✉)
Physique des Particules, Université de Montréal, Montréal, QC, Canada
e-mail: london@lps.umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_54

2 *B*-Decay Anomalies

2.1 $b \to s\mu^+\mu^-$

$b \to s$ transitions, which have $\Delta Q_{em} = 0$, are flavor-changing neutral-current (FCNC) processes. In the SM, these can arise only at loop level. One such FCNC decay is $b \to s\mu^+\mu^-$. This process receives several SM contributions, one of which is shown in Fig. 1. The SM amplitude is suppressed by loop factors and small elements in the Cabibbo–Kobayashi–Maskawa (CKM) quark mixing matrix:

$$A \sim \frac{1}{16\pi^2}\frac{g^4}{M_W^2}\frac{m_t^2}{M_W^2}V_{tb}V_{ts}^* \,. \tag{1}$$

Processes whose rates are small in the SM are excellent places to search for NP. Indeed, there are a number of measurements of observables involving $b \to s\mu^+\mu^-$ that disagree with the predictions of the SM:

- The measured branching ratios of $B \to K^*\mu^+\mu^-$ [1] and $B_s \to \phi\mu^+\mu^-$ [2] have been found to be smaller than the predictions of the SM. Here there are significant theoretical uncertainties, related to the poorly known values of the hadronic form factors [3–5].
- Deviations from the SM expectations have been found in measurements of the angular distribution of $B \to K^*\mu^+\mu^-$ [6–9], particularly in the angular observable P_5' [10]. Here, the form-factor uncertainties are smaller than for the branching ratios [11, 12], but they are still important.
- LHCb has measured

$$R_K \equiv \frac{\mathcal{B}(\bar{B} \to K\mu^+\mu^-)}{\mathcal{B}(\bar{B} \to Ke^+e^-)} \,. \tag{2}$$

Using the Run 1 data (2014) [13], for $1 \leqslant q^2 \leqslant 6\ \text{GeV}^2$, where q^2 is the dilepton invariant mass-squared, it was found that

$$R_{K,\text{Run 1}}^{\text{old}} = 0.745^{+0.090}_{-0.074}\ (\text{stat}) \pm 0.036\ (\text{syst}) \,. \tag{3}$$

The SM prediction is $R_K^{\text{SM}} = 1 \pm 0.01$ [14]. This measurement disagrees with the SM at the level of 2.6σ, suggesting a violation of lepton universality.

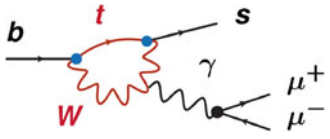


Fig. 1 One SM contribution to $b \to s\mu^+\mu^-$

At the Rencontres de Moriond, 2019, LHCb presented new R_K results [15]: (1) the Run 1 data were reanalyzed using a new reconstruction selection method, and (2) the Run 2 data were analyzed. The results are

$$\begin{aligned} R^{\text{new}}_{K,\text{Run 1}} &= 0.717^{+0.083}_{-0.071}\,(\text{stat})^{+0.017}_{-0.016}\,(\text{syst})\,, \\ R_{K,\text{Run 2}} &= 0.928^{+0.089}_{-0.076}\,(\text{stat})\,\pm^{+0.020}_{-0.017}\,(\text{syst})\,. \end{aligned} \tag{4}$$

Combining the Run 1 and Run 2 results gives

$$R_K = 0.846^{+0.060}_{-0.054}\,(\text{stat})^{+0.016}_{-0.014}\,(\text{syst})\,. \tag{5}$$

The central value is closer to the SM prediction, but, due to the smaller errors, the discrepancy with the SM is still $\sim 2.5\sigma$.

LHCb has also measured

$$R_{K^*} \equiv \frac{\mathcal{B}(\bar{B} \to K^* \mu^+ \mu^-)}{\mathcal{B}(\bar{B} \to K^* e^+ e^-)}\,, \tag{6}$$

finding [16]

$$R_{K^*} = \begin{cases} 0.660^{+0.110}_{-0.070}\,(\text{stat}) \pm 0.024\,(\text{syst})\,, & 0.045 \leqslant q^2 \leqslant 1.1\ \text{GeV}^2\,, \\ 0.685^{+0.113}_{-0.069}\,(\text{stat}) \pm 0.047\,(\text{syst})\,, & 1.1 \leqslant q^2 \leqslant 6.0\ \text{GeV}^2\,. \end{cases}$$

Compared to SM predictions, these correspond to discrepancies of 2.4σ and 2.5σ.

At the Rencontres de Moriond, 2019, Belle announced its measurement of R_{K^*} [17]:

$$R_{K^*} = \begin{cases} 0.52^{+0.36}_{-0.26} \pm 0.05\,, & 0.045 \leqslant q^2 \leqslant 1.1\ \text{GeV}^2\,, \\ 0.96^{+0.45}_{-0.29} \pm 0.11\,, & 1.1 \leqslant q^2 \leqslant 6.0\ \text{GeV}^2\,, \\ 0.90^{+0.27}_{-0.21} \pm 0.10\,, & 0.1 \leqslant q^2 \leqslant 8.0\ \text{GeV}^2\,, \\ 1.18^{+0.52}_{-0.32} \pm 0.10\,, & 15.0 \leqslant q^2 \leqslant 19.0\ \text{GeV}^2\,, \\ 0.94^{+0.17}_{-0.14} \pm 0.08\,, & 0.045 \leqslant q^2\,. \end{cases}$$

Although the central values are closer to the SM predictions, the errors are considerably larger than in the LHCb measurement.

- On average, older measurements of the branching ratio of $B_s \to \mu^+\mu^-$ were in agreement with the prediction of the SM [18–20]. However, a new measurement by ATLAS disagrees with SM by 2.4σ [21]. Combining all results leads to tension of $\sim 2\sigma$ with the SM.

There are therefore quite a few measurements of observables that are in disagreement with the predictions of the SM. All of these involve the decay $b \to s\mu^+\mu^-$,

which suggests trying to explain the data by allowing NP to contribute to this decay. The model-independent starting point is the effective Hamiltonian

$$H_{\rm eff} = -\frac{\alpha G_F}{\sqrt{2}\pi} V_{tb} V_{ts}^* \sum_{a=9,10} (C_a O_a + C'_a O'_a) \,, \tag{7}$$

where $O_{9(10)} = [\bar{s}\gamma_\mu P_L b][\bar{\mu}\gamma^\mu(\gamma_5)\mu]$, and the primed operators have $L \to R$. The Wilson coefficients include both SM and NP contributions: $C_X = C_{X,\rm SM} + C_{X,\rm NP}$.

Performing a combined fit to all the data, in the simplest scenarios it is found that the data can be explained if[1]

$$\begin{aligned} &\text{(i)} \quad C^{\mu\mu}_{9,\rm NP} = -1.10 \pm 0.16 \,, \\ &\text{(ii)} \quad C^{\mu\mu}_{9,\rm NP} = -C^{\mu\mu}_{10,\rm NP} = -0.53 \pm 0.08 \,, \end{aligned} \tag{8}$$

with a pull of close to 6σ(!). (I note in passing that scenario (ii) involves purely left-handed NP.)

2.2 $b \to c\tau^-\bar{\nu}$

There is another set of observables whose measurements also exhibit discrepancies with the SM. They involve the decay $b \to c\tau^-\bar{\nu}$. This is a $\Delta Q_{em} = 1$ process, and proceeds in the SM via tree-level W exchange, see Fig. 2. The amplitude is given by

$$A \sim \frac{g_2^2}{M_W^2} V_{cb} \,, \tag{9}$$

where $|V_{cb}| \simeq 0.04$.

Before the Rencontres de Moriond, 2019, BaBar, Belle, and LHCb measured the quantities

$$R_{D^{(*)}} \equiv \frac{\mathcal{B}(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau)}{\mathcal{B}(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell)} \quad , \quad (\ell = e, \mu) \,. \tag{10}$$

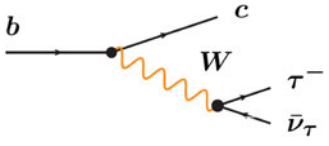


Fig. 2 SM diagram for $b \to c\tau^-\bar{\nu}$

[1] These numbers are taken from Ref. [22]. Other analyses [23–27] obtain similar results.

Their measurements exhibited discrepancies with the predictions of the SM. Combining R_D and R_{D^*}, the deviation was $\sim 3.8\sigma$ [28]. At Moriond, 2019, Belle announced new results [29]:

$$R_{D^*}^{\tau/\ell}/(R_{D^*}^{\tau/\ell})_{\rm SM} = 1.10 \pm 0.09\ ,$$
$$R_{D}^{\tau/\ell}/(R_{D}^{\tau/\ell})_{\rm SM} = 1.03 \pm 0.13\ . \tag{11}$$

These results are in better agreement with the SM, so that the deviation from the SM in R_D and R_{D^*} (combined) has been reduced from $\sim 3.8\sigma$ to 3.1σ [28].

LHCb has also measured

$$R_{J/\psi} \equiv \frac{\mathcal{B}(B_c^+ \to J/\psi \tau^+ \nu_\tau)}{\mathcal{B}(B_c^+ \to J/\psi \mu^+ \nu_\mu)}\ , \tag{12}$$

finding [30]

$$\frac{R_{J/\psi}}{(R_{J/\psi})_{\rm SM}} = 2.51 \pm 0.97\ . \tag{13}$$

Here the discrepancy with the SM is 1.7σ [31].

The discrepancies in R_D, R_{D^*}, and $R_{J/\psi}$ are hints of τ-μ and τ-e universality violation in $b \to c\ell^-\bar{\nu}$, and suggest the presence of NP in $b \to c\tau^-\bar{\nu}$ decays.

3 Models of New Physics

For the $b \to s\mu^+\mu^-$ anomalies, there are two classes of NP models that contribute to the decay at tree level, and can explain the data.

The first class involves a new Z' boson (see Fig. 3). The Z' must have a FCNC coupling to $\bar{s}b$ and must couple to $\mu^+\mu^-$. The model can follow scenarios (i) or (ii) [Eq. (8)]. A great many Z' models have been proposed (far too many to list here). Some combine explanations of the B anomalies with other weaknesses of the SM, such as dark matter, $(g-2)_\mu$, and neutrino masses.

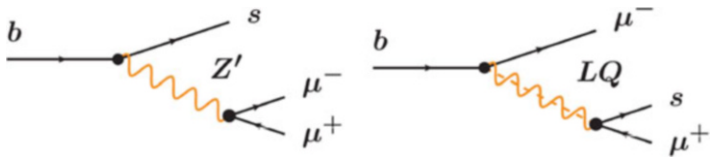


Fig. 3 Z' (left) and LQ (right) contributions to $b \to s\mu^+\mu^-$

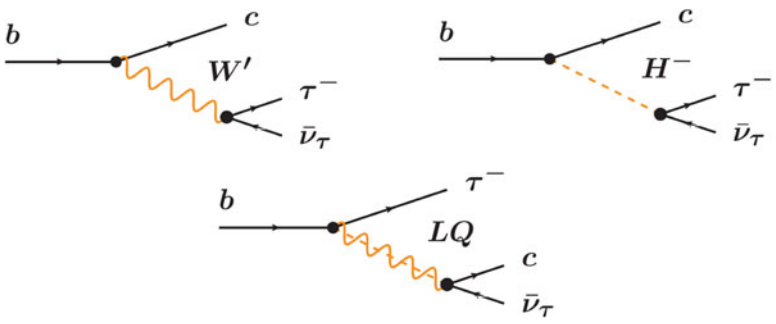


Fig. 4 W' (top left), charged-Higgs (top right), and LQ (bottom) contributions to $b \to c\tau^-\bar{\nu}$

The second class of NP models involves leptoquark (LQ) exchange (see Fig. 3). There are several different types of LQ that can explain the $b \to s\mu^+\mu^-$ data. All fit within scenario (ii) of Eq. (8) (purely LH NP) [32].

Turning to $b \to c\tau^-\bar{\nu}$, there are three types of NP whose contributions to this decay could explain the data: (1) a new W' boson, (2) a charged-Higgs boson, and (3) a leptoquark (see Fig. 4). However, the H^- is disfavored by (theoretical) constraints from $B_c^- \to \tau^-\bar{\nu}_\tau$ [33], leaving the W' or (several different types of) LQ as NP explanations. Here, the NP couplings can be left-handed (LH) and/or right-handed (RH).

3.1 Distinguishing NP Explanations

As we have seen, there are several NP explanations for the anomalies in $b \to s\mu^+\mu^-$. But this raises the question: how can we distinguish among them? One way is to look at CP violation in $B \to K^*\mu^+\mu^-$ [32]. Now, CP violation is generated by the interference of (at least) two amplitudes with different weak phases. In the presence of NP, this can arise due to SM-NP interference. Here the signal is not direct CP violation, but rather CP asymmetries in the angular distribution. The key point here is that SM-Z' and SM-LQ interferences are different, leading to different CP-violating effects. Thus, by measuring CP violation in $B \to K^*\mu^+\mu^-$, one can differentiate the NP models.

The situation is similar for the NP explanations of the $b \to c\tau^-\bar{\nu}$ anomalies. By looking at CP violation in $B \to D^*\tau^-\bar{\nu}$ (and also in $B \to D^*\mu^-\bar{\nu}$), one can distinguish NP models [34]. Once again, the signal involves CP asymmetries in the angular distribution. The measurement of CP violation in these decays allows us to differentiate the W' and LQ models. It also provides information about the LH/RH NP couplings.

3.2 Simultaneous Explanations of $b \to s\mu^+\mu^-$ and $b \to c\tau^-\bar{\nu}$

Now, $(c, s)_L$ is a doublet of $SU(2)_L$. This suggests that, if the NP coupling is purely LH, $b \to s$ and $b \to c$ transitions are related. It should therefore be possible to find NP that can simultaneously explain both the $b \to s\mu^+\mu^-$ and $b \to c\tau^-\bar{\nu}$ anomalies [35].

There are two classes of models that, in principle, can do this:

- A new triplet of vector bosons $(W'^{\pm}, Z'^0)$. The W' and Z' contribute respectively to $b \to c\tau^-\bar{\nu}$ and $b \to s\mu^+\mu^-$.
- A LQ of charge $Q_{em} = \frac{2}{3}$. It couples to $\bar{b}\mu^+$ and $\bar{s}\mu^+$ (for $b \to s\mu^+\mu^-$) and to $\bar{b}\tau^+$ and $c\bar{\nu}_\tau$ (for $b \to c\tau^-\bar{\nu}$).

It is found [36–39] that, when all constraints are taken into account, including those from direct searches at the LHC, the $(W'^{\pm}, Z'^0)$ model is excluded. But the LQ model is viable!

4 Summary

The SM is certainly correct, but it is not complete: there must be physics beyond the SM. Recently, there have been several measurements of observables that are in disagreement with the predictions of the SM:

- $b \to s\mu^+\mu^-$: These include many observables involving this decay. Some are clean, while others have important theoretical uncertainties. Global fits allowing for NP in $b \to s\mu^+\mu^-$ find improvements over the SM at the level of close to 6σ. NP Models with an extra Z' or with different types of LQs have been proposed as explanations.
- $b \to c\tau^-\bar{\nu}$: here there are several clean observables, with a net deviation from the SM of $\sim 3\sigma$. These can be explained in models with an extra W' or with different types of LQs.

It is of course possible that these discrepancies with the SM are all statistical fluctuations, and will go away with more data. This said, their combined statistical significance is sizeable ($\gtrsim 4\sigma$), so they will not disappear soon. Hopefully, we are indeed seeing the first experimental signals of NP.

Acknowledgement This work was financially supported in part by NSERC of Canada.

References

1. R. Aaij et al. [LHCb Collaboration], Differential branching fractions and isospin asymmetries of $B \to K^{(*)}\mu^+\mu^-$ decays. J. High Energy Phys. **1406**, 133 (2014). arXiv:1403.8044 [hep-ex]. https://doi.org/10.1007/JHEP06(2014)133
2. R. Aaij et al. [LHCb Collaboration], Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$. J. High Energy Phys. **1509**, 179 (2015). arXiv:1506.08777 [hep-ex]]. https://doi.org/10.1007/JHEP09(2015)179
3. A. Bharucha, D.M. Straub, R. Zwicky, $B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules. J. High Energy Phys. **1608**, 098 (2016). arXiv:1503.05534 [hep-ph]. https://doi.org/10.1007/JHEP08(2016)098
4. R.R. Horgan, Z. Liu, S. Meinel, M. Wingate, Rare B decays using lattice QCD form factors. PoS LATTICE **2014**, 372 (2015). arXiv:1501.00367 [hep-lat]. https://doi.org/10.22323/1.214.0372
5. N. Gubernari, A. Kokulu, D. van Dyk, $B \to P$ and $B \to V$ form factors from B-meson light-cone sum rules beyond leading twist. J. High Energy Phys. **1901**, 150 (2019). arXiv:1811.00983 [hep-ph]. https://doi.org/10.1007/JHEP01(2019)150
6. R. Aaij et al., LHCb Collaboration, Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity. J. High Energy Phys. **1602**, 104 (2016). arXiv:1512.04442 [hep-ex]. https://doi.org/10.1007/JHEP02(2016)104
7. ATLAS Collaboration, Angular analysis of $B_d^0 \to K^*\mu^+\mu^-$ decays in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector, Tech. Rep. ATLAS-CONF-2017-023, CERN, Geneva, 2017
8. CMS Collaboration, Measurement of the P_1 and P_5' angular parameters of the decay $B^0 \to K^{*0}\mu^+\mu^-$ in proton-proton collisions at $\sqrt{s} = 8$ TeV, Tech. Rep. CMS-PAS-BPH-15-008, CERN, Geneva, 2017
9. V. Khachatryan et al. [CMS Collaboration], Angular analysis of the decay $B^0 \to K^{*0}\mu^+\mu^-$ from pp collisions at $\sqrt{s} = 8$ TeV. Phys. Lett. B **753**, 424 (2016). arXiv:1507.08126 [hep-ex]. https://doi.org/10.1016/j.physletb.2015.12.020
10. S. Descotes-Genon, T. Hurth, J. Matias and J. Virto, Optimizing the basis of $B \to K^*ll$ observables in the full kinematic range, J. High Energy Phys. **1305**, 137 (2013). arXiv:1303.5794 [hep-ph]. https://doi.org/10.1007/JHEP05(2013)137
11. A. Khodjamirian, T. Mannel, A.A. Pivovarov, Y.-M. Wang, Charm-loop effect in $B \to K^{(*)}\ell^+\ell^-$ and $B \to K^*\gamma$. J. High Energy Phys. **1009**, 089 (2010). arXiv:1006.4945 [hep-ph]. https://doi.org/10.1007/JHEP09(2010)089
12. C. Bobeth, M. Chrzaszcz, D. van Dyk, J. Virto, Long-distance effects in $B \to K^*\ell\ell$ from analyticity. Eur. Phys. J. C **78**(6), 451 (2018). arXiv:1707.07305 [hep-ph]. https://doi.org/10.1140/epjc/s10052-018-5918-6
13. R. Aaij et al. [LHCb Collaboration], Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays. Phys. Rev. Lett. **113**, 151601 (2014). arXiv:1406.6482 [hep-ex]. https://doi.org/10.1103/PhysRevLett.113.151601
14. M. Bordone, G. Isidori, A. Pattori, On the Standard Model predictions for R_K and R_{K^*}. Eur. Phys. J. C **76**(8), 440 (2016). arXiv:1605.07633 [hep-ph]. https://doi.org/10.1140/epjc/s10052-016-4274-7
15. T. Humair (for the LHCb Collaboration), "Lepton Flavor Universality tests with heavy flavour decays at LHCb," talk given at Moriond, March 22 2019. See also R. Aaij et al. [LHCb Collaboration], Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays. Phys. Rev. Lett. **122**(19), 191801 (2019). arXiv:1903.09252 [hep-ex]. https://doi.org/10.1103/PhysRevLett.122.191801
16. R. Aaij et al. [LHCb Collaboration], Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays. J. High Energy Phys. **1708**, 055 (2017). arXiv:1705.05802 [hep-ex]. https://doi.org/0.1007/JHEP08(2017)055

17. M. Prim (for the Belle Collaboration), "Search for $B \to \ell\nu\gamma$ and $B \to \mu\nu_\mu$ and Test of Lepton Universality with $R(K^*)$ at Belle," talk given at Moriond, March 22 2019. See also A. Abdesselam et al. [Belle Collaboration], Test of lepton flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle. arXiv:1904.02440 [hep-ex]
18. S. Chatrchyan et al. [CMS Collaboration], Measurement of the $B_s^0 \to \mu^+\mu^-$ Branching Fraction and Search for $B^0 \to \mu^+\mu^-$ with the CMS Experiment. Phys. Rev. Lett. **111**, 101804 (2013). arXiv:1307.5025 [hep-ex]. https://doi.org/10.1103/PhysRevLett.111.101804
19. V. Khachatryan et al. [CMS and LHCb Collaborations], Observation of the rare $B_s^0 \to \mu^+\mu^-$ decay from the combined analysis of CMS and LHCb data. Nature **522**, 68 (2015). arXiv:1411.4413 [hep-ex]. https://doi.org/10.1038/nature14474
20. R. Aaij et al. [LHCb Collaboration], Measurement of the $B_s^0 \to \mu^+\mu^-$ branching fraction and effective lifetime and search for $B^0 \to \mu^+\mu^-$ decays. Phys. Rev. Lett. **118**(19), 191801 (2017). arXiv:1703.05747 [hep-ex]. https://doi.org/10.1103/PhysRevLett.118.191801
21. M. Aaboud et al. [ATLAS Collaboration], Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector. J. High Energy Phys. **1904**, 098 (2019). arXiv:1812.03017 [hep-ex]. https://doi.org/10.1007/JHEP04(2019)098
22. A. Datta, J. Kumar, D. London, The B anomalies and new physics in $b \to se^+e^-$. Phys. Lett. B **797**, 134858 (2019). arXiv:1903.10086 [hep-ph]. https://doi.org/10.1016/j.physletb.2019.134858
23. M. Algueró, B. Capdevila, A. Crivellin, S. Descotes-Genon, P. Masjuan, J. Matias, J. Virto, Emerging patterns of New Physics with and without Lepton Flavour Universal contributions. Eur. Phys. J. C **79**(8), 714 (2019). arXiv:1903.09578 [hep-ph]. https://doi.org/10.1140/epjc/s10052-019-7216-3
24. M. Ciuchini, A.M. Coutinho, M. Fedele, E. Franco, A. Paul, L. Silvestrini, M. Valli, New physics in $b \to s\ell^+\ell^-$ confronts new data on Lepton Universality. Eur. Phys. J. C **79**(8), 719 (2019). arXiv:1903.09632 [hep-ph]. https://doi.org/10.1140/epjc/s10052-019-7210-9
25. J. Aebischer, W. Altmannshofer, D. Guadagnoli, M. Reboud, P. Stangl, D.M. Straub, B-decay discrepancies after Moriond 2019. arXiv:1903.10434 [hep-ph]
26. K. Kowalska, D. Kumar, E.M. Sessolo, Implications for new physics in $b \to s\mu\mu$ transitions after recent measurements by Belle and LHCb. Eur. Phys. J. C **79**(10), 840 (2019). https://doi.org/10.1140/epjc/s10052-019-7330-2. arXiv:1903.10932 [hep-ph]
27. A. Arbey, T. Hurth, F. Mahmoudi, D.M. Santos, S. Neshatpour, Update on the $b \to s$ anomalies. Phys. Rev. D **100**(1), 015045 (2019). https://doi.org/10.1103/PhysRevD.100.015045. arXiv:1904.08399 [hep-ph]
28. The Heavy Flavor Averaging Group, https://hflav.web.cern.ch/
29. A. Abdesselam et al. [Belle Collaboration], Measurement of $\mathcal{R}(D)$ and $\mathcal{R}(D^*)$ with a semileptonic tagging method. arXiv:1904.08794 [hep-ex]
30. R. Aaij et al. [LHCb Collaboration], Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi\tau^+\nu_\tau)/\mathcal{B}(B_c^+ \to J/\psi\mu^+\nu_\mu)$. Phys. Rev. Lett. **120**, no. 12, 121801 (2018). arXiv:1711.05623 [hep-ex]. https://doi.org/10.1103/PhysRevLett.120.121801
31. R. Watanabe, New Physics effect on $B_c \to J/\psi\tau\bar{\nu}$ in relation to the $R_{D^{(*)}}$ anomaly. Phys. Lett. B **776**, 5 (2018). arXiv:1709.08644 [hep-ph]. https://doi.org/10.1016/j.physletb.2017.11.016
32. A.K. Alok, B. Bhattacharya, D. Kumar, J. Kumar, D. London, S.U. Sankar, New physics in $b \to s\mu^+\mu^-$: Distinguishing models through CP-violating effects. Phys. Rev. D **96**(1), 015034 (2017). arXiv:1703.09247 [hep-ph]. https://doi.org/10.1103/PhysRevD.96.015034
33. R. Alonso, B. Grinstein, J. Martin Camalich, Lifetime of B_c^- Constrains Explanations for Anomalies in $B \to D^{(*)}\tau\nu$. Phys. Rev. Lett. **118**(8), 081802 (2017). arXiv:1611.06676 [hep-ph]. https://doi.org/10.1103/PhysRevLett.118.081802
34. B. Bhattacharya, A. Datta, S. Kamali, D. London, CP violation in $\bar{B}^0 \to D^{*+}\mu^-\bar{\nu}_\mu$. J. High Energy Phys. **1905**, 191 (2019). arXiv:1903.02567 [hep-ph]. https://doi.org/10.1007/JHEP05(2019)191

35. B. Bhattacharya, A. Datta, D. London, S. Shivashankara, Simultaneous explanation of the R_K and $R(D^{(*)})$ puzzles. Phys. Lett. B **742**, 370 (2015). arXiv:1412.7164 [hep-ph]. https://doi.org/10.1016/j.physletb.2015.02.011
36. B. Bhattacharya, A. Datta, J.P. Guévin, D. London, R. Watanabe, Simultaneous explanation of the R_K and $R_{D^{(*)}}$ puzzles: a model analysis. J. High Energy Phys. **1701**, 015 (2017). arXiv:1609.09078 [hep-ph]. https://doi.org/10.1007/JHEP01(2017)015
37. D. Buttazzo, A. Greljo, G. Isidori, D. Marzocca, B-physics anomalies: a guide to combined explanations. J. High Energy Phys. **1711**, 044 (2017). arXiv:1706.07808 [hep-ph]. https://doi.org/10.1007/JHEP11(2017)044
38. J. Kumar, D. London, R. Watanabe, Combined explanations of the $b \to s\mu^+\mu^-$ and $b \to c\tau^-\bar{\nu}$ anomalies: a general model analysis. Phys. Rev. D **99**(1), 015007 (2019). arXiv:1806.07403 [hep-ph]. https://doi.org/10.1103/PhysRevD.99.015007
39. C. Cornella, J. Fuentes-Martin, G. Isidori, Revisiting the vector leptoquark explanation of the B-physics anomalies. J. High Energy Phys. **1907**, 168 (2019). arXiv:1903.11517 [hep-ph]. https://doi.org/10.1007/JHEP07(2019)168

Loopholes in W_R Searches at the LHC

Özer Özdal

Abstract We present mass bounds of the W_R gauge boson in generic left-right symmetric models. Assuming that the gauge bosons couple universally to quarks and leptons, we allow different gauge couplings $g_R \neq g_L$ and mass mixing, $V_{CKM}^L \neq V_{CKM}^R$ in the left and right sectors. We investigate scenarios where W_R is lighter, or heavier than the right-handed neutrino ν_R. In these scenarios, W_R mass bounds can be considerably relaxed, while Z_R mass bounds are much more stringent. In the case where $M_{W_R} \leq M_{\nu_R}$, the experimental constraints come from $W_R \to tb$ and $W_R \to jj$ channels, while if $M_{W_R} \geq M_{\nu_R}$, the dominant constraints come from $W_R \to \ell\ell jj$. Our results show that W_R mass bounds can be relaxed significantly.

Keywords Left-right symmetric models · Exclusion limits for heavy right-handed W boson

Left-right symmetric models (LRSMs) [1–3] provide a natural explanation for neutrino masses without resorting to higher scales. Based on the $SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$, these models restore parity symmetry, which is conserved at high energy, and spontaneously broken at some energy scale connected to the $SU(2)_R$ breaking scale [4]. Small neutrino masses are induced by heavy (and most often, Majorana) right-handed neutrinos through the phenomena known as the seesaw mechanism [5, 6]. Within the framework described here both Type I and Type II seesaw mechanisms can be naturally imbedded in the model.

In making any specific predictions, there are several sources of uncertainty in the model. The most important source of uncertainty comes from the right-handed quark mixing matrix, similar to that of the left-handed quark Cabibbo–Kobayashi–Maskawa (CKM) mixing. It is also commonly assumed that left-right symmetry imposes $g_L = g_R$. If breaking of $SU(2)_R \otimes U(1)_{B-L}$ occurs at a high scale, at that

Ö. Özdal (✉)
Department of Physics, Concordia University, Montreal, QC, Canada
e-mail: ozer.ozdal@concordia.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_55

scale $g_L = g_R$, but below that, the couplings g_L and g_R could evolve differently, and would be different at low energy scales.

Bounds on extra particle masses depend strongly on the above assumptions, in particular on the size of the right-handed gauge coupling and/or the right-handed CKM matrix elements, and none more so than the charged gauge boson W_R. This boson is interesting for several reasons. First, regardless of the other details of the spectrum of left-right models, discovery of a charged gauge boson will be an indication of the presence of an additional $SU(2)$ symmetry group, and if testing its decay products would indicate that it is right-handed, this will be an unambiguous signal for left-right symmetry.[1] Second, the LHC has significantly improved searches for W_R bosons, and the limits on their masses are becoming more stringent.

Both left- and right-handed fermions are doublets under the extended $SU(3)_c \otimes SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$ gauge group. The electroweak symmetry is broken by the bi-doublet Higgs field. In addition, to break the $SU(2)_R \otimes U(1)_{B-L}$ gauge symmetry and to provide Majorana mass terms for neutrinos (the right-handed neutrino is automatically included in the right-handed lepton doublet), we introduce the Higgs triplets under $SU(2)_L \otimes SU(2)_R \otimes U(1)_{B-L}$:

$$\Delta_L \equiv \begin{pmatrix} \delta_L^+/\sqrt{2} & \delta_L^{++} \\ \delta_L^0 & -\delta_L^+/\sqrt{2} \end{pmatrix} \sim (\mathbf{3}, \mathbf{1}, \mathbf{2}), \quad \Delta_R \equiv \begin{pmatrix} \delta_R^+/\sqrt{2} & \delta_R^{++} \\ \delta_R^0 & -\delta_R^+/\sqrt{2} \end{pmatrix} \sim (\mathbf{1}, \mathbf{3}, \mathbf{2}). \tag{1}$$

While only Δ_R is needed for symmetry breaking, Δ_L is included to preserve left-right symmetry. After symmetry breaking, the most general vacuum is

$$\langle \Phi \rangle = \begin{pmatrix} \kappa_1/\sqrt{2} & 0 \\ 0 & \kappa_2 e^{i\alpha}/\sqrt{2} \end{pmatrix}, \quad \langle \Delta_L \rangle = \begin{pmatrix} 0 & 0 \\ v_L e^{i\theta_L}/\sqrt{2}\ 0 \end{pmatrix}, \quad \langle \Delta_R \rangle = \begin{pmatrix} 0 & 0 \\ v_R/\sqrt{2}\ 0 \end{pmatrix}. \tag{2}$$

For the LRSM model to break down to the SM, the hierarchy of the vacuum expectation values (VEVs) must be, $v_R \gg (\kappa_1,\ \kappa_2) \gg v_L$, and $\sqrt{\kappa_1^2 + \kappa_2^2} = v = 246$ GeV. Here the presence of non-zero VEV of Δ_R, v_R breaks the $SU(2)_R \otimes U(1)_{B-L}$ to $U(1)_Y$, while the presence of bi-doublet VEVs κ_1 and κ_2 break the electroweak symmetry, at the same time inducing a VEV for Δ_L denoted by v_L.

In the charged sector, the left and right gauge bosons mix to give the mass eigenstates, W_1 and W_2;

$$\begin{pmatrix} W_1 \\ W_2 \end{pmatrix} = \begin{pmatrix} \cos\xi & -\sin\xi \\ \sin\xi & \cos\xi \end{pmatrix} \begin{pmatrix} W_L \\ W_R \end{pmatrix}, \tag{3}$$

[1] Unlike the discovery of a Z' which can indicate the presence of any variety of additional $U(1)'$ groups.

where the mixing angle ξ and the mass eigenvalues are given by

$$\tan 2\xi = \frac{4g_R g_L \kappa_1 \kappa_2}{2g_R^2 v_R^2 + (g_R^2 - g_L^2)v^2} \approx \frac{2g_L \kappa_1 \kappa_2}{g_R v_R^2} \tag{4}$$

$$M_{W_1}^2 = \frac{1}{4}\left[g_L^2 v^2 \cos^2\xi + g_R^2(2v_R^2 + v^2)\sin^2\xi - 2g_R g_L \kappa_1 \kappa_2 \cos\xi \sin\xi\right]$$

$$M_{W_2}^2 = \frac{1}{4}\left[g_L^2 v^2 \sin^2\xi + g_R^2(2v_R^2 + v^2)\cos^2\xi + 2g_R g_L \kappa_1 \kappa_2 \cos\xi \sin\xi\right]. \tag{5}$$

The $SU(2)_R$ breaking scale v_R and mixing angle ξ are restricted from low energy observables, such as $K_L - K_S$, ϵ_K, $B^0 - \bar{B}^0$ mixings and $b \to s\gamma$ processes, which constrain the right scale through the charged right-handed W_R boson mass as well as the triplet Higgs masses. Taking into account the smallness of the mixing angle ξ, in what follows we shall refer to W_1 as W_L and W_2 as W_R, as to be able to compare with experimental results and nomenclature.[2]

In most analyses, W_R bosons are expected to be heavy. However all of the analyses assume that the model is manifestly left-right symmetric, that is, the coupling constants are the same for the left and right gauge sectors, $g_L = g_R$ *and* that the quark mass mixing in the right-handed sector V_{CKM}^R is either diagonal, or equal to the one in the left-handed sector (the Cabibbo–Kobayashi–Maskawa matrix, V_{CKM}^L). This does not have to be the case, and analyses of a more general model, the so-called asymmetric left-right model exist [7, 8].

We take this general approach here. We calculate the production cross section and decays of the W_R bosons in the LRSM with $g_L \neq g_R$ and allowing for general entries in the mixing matrix for the right-handed quarks, V_{CKM}^R, parametrized as the left-handed matrix, but allowing the elements to vary independently:

$$V_{\text{CKM}}^R = \begin{bmatrix} c_{12}^R c_{13}^R & s_{12}^R c_{13}^R & s_{13}^R e^{i\delta_R} \\ -s_{12}^R c_{23}^R - c_{12}^R s_{23}^R s_{13} e^{i\delta_R} & c_{12}^R c_{23}^R - s_{12}^R s_{23}^R s_{13}^R e^{i\delta_R} & s_{23}^R c_{13}^R \\ s_{12}^R s_{23}^R - c_{12}^R c_{23}^R s_{13}^R e^{i\delta_R} & -c_{12}^R c_{23}^R - s_{12}^R c_{23}^R s_{13}^R e^{i\delta_R} & c_{23}^R c_{13}^R \end{bmatrix} \tag{6}$$

We then proceed as follows:

- We first choose one value for M_{W_R}. Then we vary the parameters c_{12}^R, c_{13}^R, and c_{23}^R in the range $[-1, 1]$. The phase δ_R is set to zero (as we are not concerned with CP violation), and we impose matrix unitary condition. For each set of the randomly chosen V_{CKM}^R parameters as above, we impose the theoretical and experimental constraints including the mass bounds and flavor constraints from K and B mesons. This ensures, for instance, that the non-SM neutral bi-doublet Higgs boson is very heavy (>10 TeV), as required to suppress flavor-violating effects.

[2] Note however that they both contain a non-zero left and right $SU(2)$ gauge component.

- For each value of M_{W_R} we obtain many solutions for V_{CKM}^R consistent with all bounds. Of these solutions, we choose the one yielding the smallest branching ratio for $W_R \to t\bar{b}$. [3]
- When $M_{\nu_R} > M_{W_R}$, solutions emerge allowing for low values of $BR(W_R \to t\bar{b})$, which vary from about 23.3% for high M_{W_R}, to about 29% for low M_{W_R}, while when $M_{\nu_R} < M_{W_R}$, this ratio changes from 15.7% for high M_{W_R}, to about 24.7% for low M_{W_R}.

We performed random scans over the parameter space and imposed the mass bounds on all the particles, as well as other constraints as given in Table I [9]. In a general left-right model, the masses of the Z_R and W_R gauge bosons are related, but the mass ratios depend sensitively on the values of the coupling constants g_R and g_{B-L}. While the mass of W_R is proportional to g_R, the mass of the Z_R boson is proportional to $\sqrt{g_R^2 + g_{B-L}^2}$. Lowering g_R will decrease the production cross section for W_R, while having no effect on its branching ratios. Breaking the symmetry to $U(1)_{\rm EM}$ imposes that couplings constants are related through

$$\frac{1}{e^2} = \frac{1}{g_L^2} + \frac{1}{g_R^2} + \frac{1}{g_{B-L}^2}\,, \tag{7}$$

and therefore decreasing g_R results in increasing g_{B-L}. Note also that these couplings are related, through $SU(2)_R \otimes U(1)_{B-L} \to U(1)_Y$ symmetry breaking to the coupling of the hypercharge group, g_Y, through

$$\frac{1}{g_Y^2} = \frac{1}{g_R^2} + \frac{1}{g_{B-L}^2}\,. \tag{8}$$

This means, using the usual definition of $\sin\phi = \dfrac{g_{B-L}}{\sqrt{g_R^2 + g_{B-L}^2}}$ and $\sin\theta_W = \dfrac{g_Y}{\sqrt{g_L^2 + g_Y^2}}$, we obtain the following theoretical limit:

$$\tan\theta_W = \frac{g_R \sin\phi}{g_L} \leqslant \frac{g_R}{g_L}\,, \tag{9}$$

showing clearly that we cannot lower g_R below its minimum value, $g_L \tan\theta_W$. We analyze the case where $g_R = g_L$, as well as when $g_L \neq g_R = 0.37$, its allowed minimal value.

[3] We have also tried to minimize $BR(W_R \to jj)$ but found that this choice yields a more restrictive lower mass bound for W_R.

We investigate the case where the right-handed neutrino is heavier than W_R, (so the on-shell decay $W_R \to \nu_R \ell$ is disallowed) and where the possible decay channels for W_R are

$$W_R \to jj(q\bar{q}'),\ W_R \to W_L h,\ W_R \to W_L Z \text{ and } W_R \to W_L hh\,,$$

all the other Higgs states being very heavy. Of these, the three-body decay $W_L hh$ is very weak with $\frac{\Gamma(W_R \to W_L h)}{\Gamma(W_R \to W_L hh)} \sim 1/v$, while the decay $W_R \to W_L h$ depends on $\tan\beta$. Most analyses assume $\tan\beta$ to be very small (~ 0.01), yielding $BR(W_R \to W_L h)$ to be negligible. Perturbativity bounds alone require $\tan\beta < 0.8$; however, the mass of the SM-like Higgs boson h also depends on $\tan\beta$, and values of $\tan\beta > 0.6$ result in the instability of the h mass. To keep our analysis general, we investigated the production and decays of the W_R mass for two cases: small $\tan\beta = 0.01$, and large $\tan\beta = 0.5$. In addition, we allow for two values of g_R, viz. $g_R = g_L$, and $g_L \neq g_R = 0.37$, as well as vary matrix elements of $V^R_{\rm CKM}$ to show how the results are affected. The resulting plots for $\sigma(pp \to W_R) \times BR(W_R \to t\bar{b})$ are given in the left plane of Fig. 1, where we compare our four different cases (different color-coded, as indicated in the panel insertion) with the CMS result [10] using collision data collected at $\sqrt{s}$ =13 TeV with $\mathcal{L}$=35.9 fb^{-1}. We show the observed and expected limit curves for the combined electron and muon final states. For $g_L = g_R$, $V^L_{\rm CKM} = V^R_{\rm CKM}$, and $\tan\beta = 0.01$, we note that the branching ratio W_R into $t\bar{b}$ ranges from 32% to 33%, as W_R decays into $q\bar{q}'$ pairs democratically. We confirm the CMS result [10] and exclude W_R boson mass up to 3.6 TeV. For the case where $g_R \neq g_L$ and $V^L_{\rm CKM} = V^R_{\rm CKM}$, represented by the blue line, we set $g_R \simeq 0.37$ and $\tan\beta = 0.01$. The W_R production cross section decreases due to relatively small g_R, and the exclusion limit for W_R masses can be reduced to 2.7 TeV in that scenario. Increasing $\tan\beta$ to 0.5 and $g_R \simeq 0.37$ enhances the branching ratios of $W_R \to W_L h$ to about $\sim 1.95\%$ and the branching ratio of $W_R \to W_L Z_L$ to about $\sim$2.0%. In this case, the branching ratio of $W_R \to t\bar{b}$ is reduced slightly, to 31.0–31.8%, as shown by the pink line. As can be read from the plot, this reduces the W_R mass limits only slightly, to 2675 GeV. However, when we allow $V^L_{\rm CKM} \neq V^R_{\rm CKM}$, $\tan\beta$= 0.5, and $g_R \simeq 0.37$, this maximizes decays of W_R into other final states, and the branching ratio of $W_R \to t\bar{b}$ is reduced substantially: from about 20% for high M_{W_R} (4 TeV) to about 29% for low M_{W_R} (1.5 TeV). The orange line in the left plane of Fig. 1 represents our result for this scenario, and the exclusion limit is reduced to 2360 GeV with respect to observed limit, whereas it can be estimated at 1940 GeV based on the expected limit.

In the right panel of Fig. 1 we plot the cross section of $\sigma \times BR(W_R \to jj)$ vs W_R mass, and compare it to the ATLAS result [11] at $\sqrt{s}$ =13 TeV for $\mathcal{L}$=37 fb^{-1}. (For comparison, we included their acceptance factor A.) The red curve represents the exclusion limit for W_R mass when the gauge couplings $g_L = g_R$, and $\tan\beta$ = 0.01. The branching fraction of $W_R \to jj$ varies slightly with mass. We keep the same color coding for curves as in the previous panel. The mass restrictions are comparable, but slightly weaker than those for the $W_R \to t\bar{b}$ decay, ranging

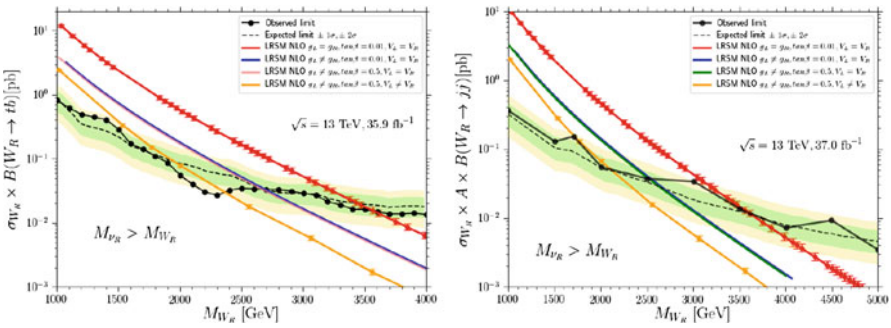


Fig. 1 (left): The cross section of $W_R \to tb$ vs W_R mass for different values of $\tan\beta$ and g_R; (right): The cross section of $W_R \to jj$ vs W_R mass, for different values of $\tan\beta$ and g_R, compared to ATLAS data [11]

from $M_{W_R} \geqslant 3625$ GeV when $g_L = g_R$, $\tan\beta = 0.01$, and $V^R_{\rm CKM} = V^L_{\rm CKM}$, to $M_{W_R} \geqslant 2.0$ TeV when $g_R = 0.37$, $\tan\beta = 0.5$, and $V^R_{\rm CKM} \neq V^L_{\rm CKM}$. Neither results are particularly sensitive to values of $\tan\beta$, but depend on choices for g_R and $V^R_{\rm CKM}$.

The case where the W_R is heavier than the right-handed neutrino opens the possibility of $W_R \to \nu_R \ell$, followed by decay of $\nu_R \to \ell W^*_R \to \ell jj$ giving rise to an $\ell\ell jj$ signature. The Majorana nature of the right-handed neutrino allows for probing both the same-sign and opposite-sign dilepton channels [12]. Both the ATLAS [13] and CMS [14] collaborations have looked for such a W_R signal, excluding W_R masses up to about 4.7 TeV for right-handed (muon or electron) neutrino masses up to 3.1 TeV [14]. For lower right-handed neutrino masses (below 200 GeV), the bound is less restrictive than the one originating from dijet searches. For the tau channel, it is even much weaker, with M_{W_R} constrained to be only smaller than 2.9 TeV [15].

In addition to $W_R \to t\bar{b}$ and $W_R \to jj$ channels, the most stringent bounds come from the channel in which the W_R boson decays to a first or second generation charged lepton and a heavy neutrino of the same lepton flavor. Both ATLAS [13] and CMS [14] assume that the heavy neutrino further decays to another charged lepton of the same flavor and a virtual $W^\star_R$ with a 100% rate. Then, the virtual $W^\star_R$ then decays into two light quarks, producing the decay chain. However, this is true only for small $\tan\beta$ values, where the corresponding mixing angle ξ between the two charged gauge bosons is extremely small. In the large $\tan\beta$ case, the mixing between W_L and W_R, although small enough to satisfy flavor and CP bounds, becomes important.

The planes of Fig. 2 show the result of the analysis of the $\ell\ell jj$ final states. The left panel plots the result for decays into $\ell\ell = ee$ channel whereas the right one is for the $\ell\ell = \mu\mu$ channel. For the scenarios where $\tan\beta = 0.01$, the contribution through W_L bosons is suppressed. Therefore, the main contribution to $\ell\ell jj$ final states comes via the virtual $W^\star_R$ boson. However, for consistency we sum up the contributions through the virtual $W^\star_R$ and real W_L boson, and our graphs for large $\tan\beta$ values in

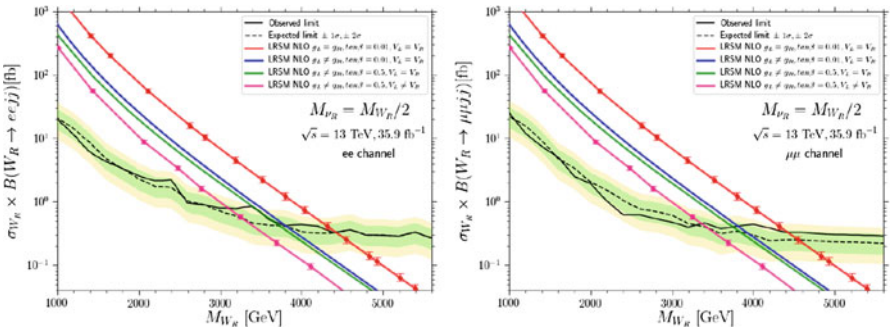


Fig. 2 (left) The cross section of $W_R \to eejj$ vs W_R mass, for the case $M_{\nu_R} = M_{W_R}/2$, compared to the CMS data [14]; (right) The cross section of $W_R \to \mu\mu jj$ vs W_R mass, for the case where $M_{\nu_R} = M_{W_R}/2$, compared to [14]

the bottom planes represent the combined contribution. The most stringent bounds occur for $g_L = g_R$, $\tan\beta = 0.01$, and $V^L_{\rm CKM} = V^R_{\rm CKM}$. We confirmed in that case the CMS result [14] and find out that the observed (expected) 95 % CL lower limit on W_R masses is 4420 (4420) GeV in the ee channel and 4420 (4500) GeV in the $\mu\mu$ channel.

The observed (expected) limit is reduced to 3800 (3800) GeV in the ee channel and 3800 (3950) GeV in the $\mu\mu$ channel when $g_L \neq g_R$, $\tan\beta = 0.01$, and $V^L_{\rm CKM} = V^R_{\rm CKM}$. In the scenarios where $\tan\beta = 0.5$, contributions to $\ell\ell jj$ final states proceed through both virtual $W^\star_R$ and W_L bosons. When the W_R mass is about 1 TeV, approximately 90.5% of the combined contribution is obtained through W_L bosons, limiting the virtual $W^\star_R$ boson contribution to 9.5%. However, this relation is flipped when W_R mass is about 4 TeV, where 77.5% of the combined contribution to $\ell\ell jj$ final states is obtained through the virtual $W^\star_R$ boson, leaving the W_L boson to contribute at 20.6%. In the scenario where we $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, and $V^L_{\rm CKM} = V^R_{\rm CKM}$, we improve the bounds to where the observed (expected) 95% CL lower limit is 3725 (3720) GeV in the ee channel, and 3750 (3900) GeV in the $\mu\mu$ channel. In addition to lowering g_R and increasing $\tan\beta$, we verified the effect of different CKM matrices, allowing $V^L_{\rm CKM} \neq V^R_{\rm CKM}$ in our final scenario. The partial contributions through virtual $W^\star_R$ and W_L in this scenario are very close to the one where $V^L_{\rm CKM} = V^R_{\rm CKM}$ and $g_R \neq g_L$, and $\tan\beta = 0.5$. In this case the results are least constraining and we find that the observed (expected) 95% CL lower limit is 3100 (3300) GeV in the ee channel and 3350 (3400) GeV in the $\mu\mu$ channel.

In Fig. 3, we analyze the correlations in the two dimensional $M_{W_R} - M_{\nu_R}$ mass plane, covering a range of neutrino masses both below and above the W_R boson mass. Contrary to the CMS analysis [14], which assumes that only one heavy neutrino flavor ν_R contributes significantly to the W_R decay width, in our analysis all three heavy right-handed neutrino flavors contribute democratically. The W_R production cross section is calculated for each solution in this 2D plane using MG5_AMC@NLO, and the observed (expected) 95% CL limits obtained from our

analysis are applied to explore excluded regions. The expected and observed upper limits on the cross section for different W_R and ν_R mass hypotheses are compared with the latest CMS results [14] @ $\mathcal{L} = 35.9$ fb^{-1} and ATLAS results [13] @ $\mathcal{L} = 36.1$ fb^{-1}, as seen in Fig. 3. Note that we generate our results using the CMS [14] data, as this is available. The ATLAS analysis, although more recent and at a slightly higher luminosity, is able to rule out a small subset of parameter points in the $M_{W_R} < M_{\nu_R}$ region. However they do not share their observed (expected) cross section results publicly. Because of that, when we extrapolate our results for slightly higher luminosity in that region, our cross sections are very small, and we do not obtain any restrictions. Thus, we decided it safer to compare our analysis with the existing data points provided by CMS, while indicating restrictions from both experimental analyses on the plots.

In the $M_{\nu_R} < M_{W_R}$ case, we assume that the contribution comes through the following decay chain:

$$W_R \to \ell \nu_R \to \ell\ell W_R^\star \to \ell\ell q q', \quad \ell = e \text{ or } \mu\,, \tag{10}$$

while in the $M_{\nu_R} > M_{W_R}$ case, we assume that the contribution comes through the following decay chain:

$$W_R^\star \to \ell \nu_R \to \ell\ell W_R \to \ell\ell q q', \quad \ell = e \text{ or } \mu\,. \tag{11}$$

In our analysis, there is no excluded solution in the $M_{\nu_R} > M_{W_R}$ region since the corresponding cross section in that region is below the experimental limits, as can be read from the color bars in Fig. 3. This is understood from our previous analysis, as the production cross section of $\nu_R \ell$ through W_R bosons dominates the one obtained through the virtual $W_R^\star$ bosons. The top planes of Fig. 3 represent the results of the exclusion in the two dimensional $M_{W_R} - M_{\nu_R}$ mass plane based on the scenario where $g_L = g_R$, $\tan\beta = 0.01$, and $V_{\text{CKM}}^L = V_{\text{CKM}}^R$, whereas middle and bottom planes show the same exclusion for the scenario where $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, $V_{\text{CKM}}^L = V_{\text{CKM}}^R$, and $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, $V_{\text{CKM}}^L \neq V_{\text{CKM}}^R$, respectively. For the scenario where $g_L \neq g_R$, $\tan\beta = 0.5$, and $V_{\text{CKM}}^L = V_{\text{CKM}}^R$, W_R bosons with masses up to 3.7 (3.7) TeV in the ee channel and up to 3.7 (3.9) TeV in the $\mu\mu$ channel are excluded at 95% CL, for M_{ν_R} up to 2.8 (2.9) TeV in the ee channel, and 3.1 (3.0) TeV in the $\mu\mu$ channel. The 2D exclusion limits are less stringent in the ee channel, where W_R boson masses are excluded up to 3.0 TeV for ν_R masses close to the $M_{W_R} = M_{\nu_R}$ degeneracy line. On the other hand, we exclude less parameter space in the two dimensional $M_{W_R} - M_{\nu_R}$ mass plane when $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, and $V_{\text{CKM}}^L \neq V_{\text{CKM}}^R$. As seen from the bottom planes of Fig. 3, W_R bosons with masses up to 3.1 (3.3) TeV are excluded at 95% CL, for M_{ν_R} up to 2.1 TeV, in the ee channel whereas W_R bosons with masses up to 3.3 (3.4) TeV are excluded at 95% CL, for M_{ν_R} up to 2.6 (2.5) TeV in the $\mu\mu$ channel. Here again, the 2D exclusion limits are less stringent in the ee channel, where W_R boson masses are excluded up to 2.0 TeV for M_{ν_R} masses close to the $M_{W_R} = M_{\nu_R}$ degeneracy line.

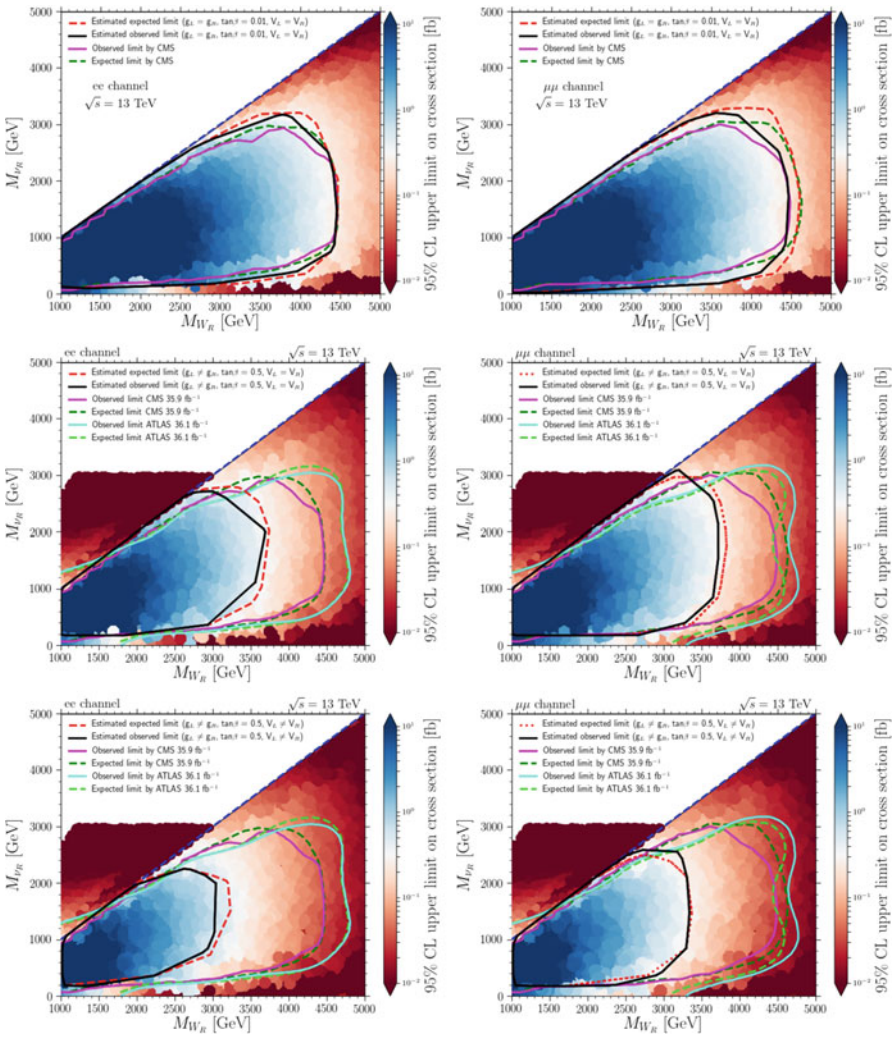


Fig. 3 Observed (continuous lines) and expected (dashed lines) 95% CL exclusion contours in the $M_{W_R} - M_{\nu_R}$ plane in the *ee* (left columns) and $\mu\mu$ (right columns) channels for Majorana ν_R neutrinos. The dashed blue line in the each plane indicates the region where $M_{W_R} = M_{\nu_R}$. We also show observed and expected limits by ATLAS [13] @ $\mathcal{L} = 36.1$ fb^{-1}, and CMS [14] @ $\mathcal{L} = 35.9$ fb^{-1}, obtained assuming $g_L = g_R$, $\tan\beta = 0.01$, and $V^L_{\text{CKM}} = V^R_{\text{CKM}}$. The cross section values are indicated by the colors in the legend of the planes. In the top planes, the continuous (dashed) black (red) line shows the estimated observed (expected) limit for the scenario where $g_L = g_R$, $\tan\beta = 0.01$, and $V^L_{\text{CKM}} = V^R_{\text{CKM}}$. In the middle planes, $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, and $V^L_{\text{CKM}} = V^R_{\text{CKM}}$, whereas the same limits in the bottom planes are analyzed for the scenario where $g_L \neq g_R = 0.37$, $\tan\beta = 0.5$, and also $V^L_{\text{CKM}} \neq V^R_{\text{CKM}}$

Table 1 Related branching ratios and cross sections for **BM I** and **BM II**

	BM I : $M_{\nu_R} > M_{W_R}$	**BM II** : $M_{\nu_R} < M_{W_R}$
m_{W_R} [GeV]	2557	3689
m_{ν_R} [GeV]	16797	1838
$\sigma(\text{pp} \to W_R)$ [fb] @13 TeV	48.7	3.98
$\sigma(\text{pp} \to W_R)$ [fb] @27 TeV	478.0	77.3
BR($W_R \to t\bar{b}$) [%]	26.3	19.9
BR($W_R \to jj$) [%]	58.6	45.8
BR($W_R \to \nu_R \ell$) [%]	–	6.5 (each family)
BR($W_R \to h_1 W_L$) [%]	1.8	1.5
BR($W_R \to W_L Z$) [%]	2.0	1.6
BR($\nu_R \to \ell q q'$) [%]	–	65.3
BR($\nu_R \to W_L \ell$) [%]	1.1×10^{-4}	33.1
BR($\nu_R \to W_R \ell$) [%]	99.9	–

Finally, allowing $g_L \neq g_R = 0.37$ and $V^L_{\text{CKM}} \neq V^R_{\text{CKM}}$ will not have significant consequences on other sectors of the model. Both the singly charged and doubly charged Higgs bosons δ^+_R and δ^{++}_R are expected to be heavy. Even so, their production mechanism is dominated by photon-mediated Drell Yan, or $\gamma\gamma$ fusion, and their branching ratios are independent of g_R. This leads further support to W_R production and decay as being most promising signal to test this scenario. To assess properties and differences between the two scenarios, we give two representative benchmarks, one for the first scenario, $M_{\nu_R} > M_{W_R}$ and one for the second scenario, $M_{\nu_R} < M_{W_R}$, with the W_R masses, cross sections and branching ratios given in Table 1, for LHC at $\sqrt{s} = 13$ and 27 TeV. Overall our results yield weaker limits on W_R mass, yielding hope that W_R could be discovered at HL-HE LHC.

References

1. J.C. Pati, A. Salam, Lepton number as the fourth color. Phys. Rev. **D10**, 275–289 (1974). [Erratum: Phys. Rev. **D11**, 703 (1975)]
2. R.N. Mohapatra, J.C. Pati, A natural left-right symmetry. Phys. Rev. **D11**, 2558 (1975)
3. R.N. Mohapatra, J.C. Pati, Left-right gauge symmetry and an isoconjugate model of CP violation. Phys. Rev. **D11**, 566–571 (1975)
4. G. Senjanovic, R.N. Mohapatra, Exact left-right symmetry and spontaneous violation of parity. Phys. Rev. **D12**, 1502 (1975)
5. R.N. Mohapatra, G. Senjanovic, Neutrino mass and spontaneous parity violation. Phys. Rev. Lett. **44**, 912 (1980). [231 (1979)]
6. R.N. Mohapatra, G. Senjanovic, Neutrino masses and mixings in gauge models with spontaneous parity violation. Phys. Rev. **D23**, 165 (1981)
7. M. Frank, A. Hayreter, I. Turan, Production and decays of W_R bosons at the LHC. Phys. Rev. **D83**, 035001 (2011)
8. P. Langacker, S. Uma Sankar, Bounds on the mass of W(R) and the W(L)-W(R) mixing angle xi in general SU(2)-L x SU(2)-R x U(1) models. Phys. Rev. **D40**, 1569–1585 (1989)

9. M. Frank, Ö. Özdal, P. Poulose, Relaxing LHC constraints on the W_R mass. Phys. Rev. **D99**(3), 035001 (2019)
10. A.M. Sirunyan et al., Search for heavy resonances decaying to a top quark and a bottom quark in the lepton+jets final state in proton-proton collisions at 13TeV. Phys. Lett. **B777**, 39–63 (2018)
11. M. Aaboud et al., Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector. Phys. Rev. **D96**(5), 052004 (2017)
12. W.-Y. Keung, G. Senjanovic, Majorana neutrinos and the production of the right-handed charged gauge boson. Phys. Rev. Lett. **50**, 1427 (1983)
13. M. Aaboud et al., Search for heavy Majorana or dirac neutrinos and right-handed W gauge bosons in final states with two charged leptons and two jets at $\sqrt{s}$=13 TeV with the ATLAS detector (2018)
14. A.M. Sirunyan et al., Search for a heavy right-handed W boson and a heavy neutrino in events with two same-flavor leptons and two jets at $\sqrt{s}$ = 13 TeV. J. High Energy Phys. **05**(05), 148 (2018)
15. A.M. Sirunyan et al., Search for third-generation scalar leptoquarks and heavy right-handed neutrinos in final states with two tau leptons and two jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV. J. Heat Energy Phys. **07**, 121 (2017)

t-t-h, Top & Bottom Partners, and the Brane Higgs Limit

Manuel Toharia

Abstract Higgs boson production in association with top quarks (tth) is one of the most interesting Higgs production mechanisms, as it probes the top quark Yukawa coupling. The ATLAS experiment announced a combined *observed* significance of 6.3 sigma for this process along with a combined *expected* significance of 5.1 sigma (both relative to the background-only hypothesis) (Aaboud et al. [ATLAS], Phys Lett B 784:173–191, 2018). The CMS experiment announced a combined *observed* significance of 5.2 sigma with a combined *expected* significance of 4.2 sigma (Sirunyan et al. (CMS Collaboration), Phys Rev Lett 120:23180, 2018). We explore the possibility of deviations in Higgs Yukawa couplings that might lead to enhancements in tth production, within models with vector-like partners of the top and bottom quarks, with masses of order 1–2 TeV. In a certain limit of parameter space, Higgs production through gluon fusion can remain under control even with large deviations of the top quark Yukawa coupling. We call this scenario the "Brane Higgs Limit", as it can emerge naturally in models of warped extra-dimensions with all matter fields in the bulk (except the Higgs). The scenario is predictive and easily falsifiable for all Higgs production and decay modes.

Keywords Collider physics · Higgs boson · Warped extra-dimensions

1 Introduction

RUN 1 of the LHC culminated in the discovery of a new particle consistent with the Standard Model (SM) Higgs boson with a mass of 125 GeV. Extensions to the SM could manifest themselves indirectly within the Higgs sector, and in particular they could affect the production cross section and decay rates of the Higgs boson. After

M. Toharia (✉)
Dawson College, Montreal, QC, Canada
e-mail: mtoharia@dawsoncollege.qc.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_56

the end of Run II of the LHC, both ATLAS and CMS report that the main production channel of the Higgs (gluon fusion) is consistent with the SM expectation.

As mentioned in the abstract, both ATLAS and CMS experiments have announced the observation of tth channel with a combined $observed$ significance of 6.3 sigma and 5.2 sigma respectively [1, 2]. In both results the $expected$ significance is slightly lower (5.1 sigma for ATLAS and 4.2 for CMS) in both cases caused by mild excesses in the $tt \to H \to \gamma\gamma$ channel (very mild excesses with measured cross sections of $\sigma(t\bar{t}H) \times B(H \to \gamma\gamma) = 1.59^{+0.43}_{-0.39}$ fb for ATLAS [3] and $\sigma(t\bar{t}H) \times B(H \to \gamma\gamma) = 1.56^{+0.34}_{-0.32}$ fb for CMS [4], to be compared with the SM expected cross section of $1.13^{+0.08}_{-0.11}$ fb). Of course, measurements in the tth channel are hindered by poor statistics, and moreover, not all decay channels give slightly enhanced signal strengths. Nevertheless, should these tantalizing signals survive more precise measurements at higher luminosity, they would provide the much awaited signals for new physics.

One possibility to explain a possible tth enhancement is the modification of the Higgs coupling to the top quark. A simple explanation would be to invoke the presence of vector-like quarks which will mix with the SM top quark and thus can modify their couplings to the Higgs, and in particular it could enhance them (see for example [5] for an early proposal or the more recent [6] and references therein). In this framework, the rates for loop-induced processes (specially gluon fusion Higgs production) could remain SM-like by compensating effects between SM fermions and the new vector-like fermions in the loops [6]. The masses of these vector-like quarks must be of order 1–2 TeV in order to modify significantly the top Yukawa coupling; such masses may still be safe from the current LHC lower limits on their masses, $m_{\mathrm{VLQ}} \gtrsim 1000$ GeV [7].

2 Top and Bottom Partners: A Doublet and Two Singlets

The simple scenario that we consider contains the usual SM gauge groups and matter fields, with the addition of a vector-like quark $SU(2)$ doublet and two vector-like quarks $SU(2)$ singlets, one with up-type gauge charge and another with down-type gauge charge. We denote $q^0_L \equiv \begin{pmatrix} t^0_L \\ b^0_L \end{pmatrix}$ as the SM third generation doublet, and t^0_R as the SM right handed top. Using similar notation we define $Q_{L,R} \equiv \begin{pmatrix} Q^t_{L,R} \\ Q^b_{L,R} \end{pmatrix}$ as the new vector-like quark doublet, $T_{R,L}$ as the new vector-like up-type quark singlet, and $B_{L,R}$ as the new vector-like down-type singlet. We can write down the Yukawa couplings between the SM top quark and the new vector-like quarks, which will lead to 5×5 fermion mass matrices. The mass and interaction Lagrangian in the top sector can be then written as

$$\mathcal{L}_{mass} = Y_t^0\, \overline{q}_L \tilde{H} t_R + Y_{qT}\, \overline{q}_L \tilde{H} T_R + Y_{Qt}\, \overline{Q}_L \tilde{H} t_R + Y_1 \overline{Q}_L \tilde{H} T_R + Y_2 \overline{Q}_R \tilde{H} T_L + M_Q \overline{Q}_L Q_R + M_T \overline{T}_L T_R, \quad (1)$$

with a similar expression for the bottom sector. After electroweak symmetry breaking, the Yukawa couplings induce off-diagonal terms into the fermion mass matrix. In the basis defined by the vectors $(\overline{q}_L, \overline{Q}_L, \overline{T}_L)$ and (t_R, Q_R, T_R) we can write the heavy quark mass matrix as

$$\mathbf{M_t} = \begin{pmatrix} vY_t^0 & 0 & vY_{qT} \\ vY_{Qt} & M_Q & vY_1^t \\ 0 & vY_2^t & M_T \end{pmatrix}, \quad (2)$$

where, in general, all entries are complex and where v is the Higgs vacuum expectation value (VEV). A similar expression will hold for the bottom sector heavy quark mass matrix. The associated top and bottom sector Yukawa coupling matrices are

$$\tilde{\mathbf{Y}}_\mathbf{t} = \begin{pmatrix} Y_t^0 & 0 & Y_{qT} \\ Y_{Qt} & 0 & Y_1^t \\ 0 & Y_2^t & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{Y}}_\mathbf{b} = \begin{pmatrix} Y_b^0 & 0 & Y_{qB} \\ Y_{Qb} & 0 & Y_1^b \\ 0 & Y_2^b & 0 \end{pmatrix}. \quad (3)$$

The mass matrices $\mathbf{M_t}$ and $\mathbf{M_b}$ are diagonalized by bi-unitary transformations, $\mathbf{V_{t}}_L^\dagger \mathbf{M_t} \mathbf{V_t}_R = \mathbf{M_t}^{diag}$, and $\mathbf{V_{b}}_L^\dagger \mathbf{M_b} \mathbf{V_b}_R = \mathbf{M_b}^{diag}$. At the same time, the Higgs Yukawa couplings are obtained after transforming the Yukawa matrices into the physical basis, $\mathbf{V_{t}}_L^\dagger \tilde{\mathbf{Y}}_\mathbf{t} \mathbf{V_t}_R = \mathbf{Y_t}^{phys}$ and $\mathbf{V_{b}}_L^\dagger \tilde{\mathbf{Y}}_\mathbf{b} \mathbf{V_b}_R = \mathbf{Y_b}^{phys}$.

In the physical basis, the top quark mass and the top Yukawa coupling (the first entries in the physical mass matrix and the physical Yukawa matrix) are not related anymore by the SM relationship $m_t^{phys} = v y_t^{SM}$ (see for example [8] and references therein) (with v normalized to $v = 174$ GeV for simplicity). The same goes for the bottom quark, and we thus define the shifts, δy_t and δy_t, between the SM and the physical Yukawa couplings, due to the diagonalization, as

$$y_t^{phys} = y_t^{SM} - \delta y_t \qquad \text{and} \qquad y_b^{phys} = y_b^{SM} - \delta y_b. \quad (4)$$

2.1 Higgs Production in the Brane Higgs Limit

Let us now consider the radiative coupling of the Higgs boson to gluons. This coupling depends on the physical Yukawa couplings y_{nn} of all the fermions running in the loop and on their physical masses m_n. The real and imaginary parts of the couplings (the scalar and pseudo-scalar parts) contribute to the cross section through different loop functions, $A_{1/2}^S$ and $A_{1/2}^P$, as they generate the two operators $hG_{\mu\nu}G^{\mu\nu}$ and $hG_{\mu\nu}\tilde{G}^{\mu\nu}$.

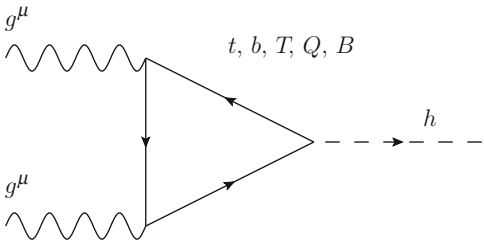


Fig. 1 Feynman diagram for the production cross section $gg \to h$ in a setup with new vector-like fermions Q, T and B

The cross section for the process depicted in Fig. 1, is

$$\sigma_{gg\to h} = \frac{\alpha_s^2 m_h^2}{576\pi}\left[|c^S_{ggh}|^2 + |c^P_{ggh}|^2\right]\delta(s - m_h^2) \tag{5}$$

where

$$c^S_{ggh} = \sum_{n=1}^{3} Re\left(\frac{y_{nn}}{m_n}\right) A^S_{1/2}(\tau_f) \qquad \text{and} \qquad c^P_{ggh} = \sum_{n=1}^{3} Im\left(\frac{y_{nn}}{m_n}\right) A^P_{1/2}(\tau_f) \tag{6}$$

with $\tau = m_h^2/4m_n^2$ and with the loop functions $A^S_{1/2}(\tau)$ and $A^P_{1/2}(\tau)$ as defined in [9].

The amplitudes c^S_{ggh} and c^P_{ggh} can then be written in terms of traces involving the fermion mass and Yukawa matrices involving top and vector-like up-quarks, and bottom and any vector-like down quarks, $\mathbf{M}_i$ and $\mathbf{Y}_i$ with $i = t, b$, so that we obtain

$$c^S_{ggh} = \sum_n Re\left(\frac{y^u_{nn}}{m^u_n}\right) + \sum_n Re\left(\frac{y^d_{nn}}{m^d_n}\right) - Re\left(\frac{y_b}{m_b}\right) + Re\left(\frac{y_b}{m_b}\right) A^S_{1/2}(\tau_b) \tag{7}$$

where we have added and subtracted the bottom quark loop contribution in order to keep the dependence in $A_{1/2}(\tau_b)$, and with a similar expression holding for c^P_{ggh}. We evaluate exactly the sums in the top sector and find

$$\sum_n \left(\frac{y^u_{nn}}{m^u_n}\right) = \frac{1}{v}\frac{1 + 3\varepsilon_{Q_t}\varepsilon_T \frac{|Y^t_2|}{|Y^0_t|} e^{i\theta^t_2}\left(1 - e^{i\theta^t_1}\frac{|Y^t_1||Y^0_t|}{|Y_{Qt}||Y_{qT}|}\right)}{1 + \varepsilon_{Q_t}\varepsilon_T \frac{|Y^t_2|}{|Y^0_t|} e^{i\theta^t_2}\left(1 - e^{i\theta^t_1}\frac{|Y^t_1||Y^0_t|}{|Y_{Qt}||Y_{qT}|}\right)}, \tag{8}$$

where we have defined the small parameters $\varepsilon_T = \dfrac{v|Y_{qT}|}{|M_T|}$, $\varepsilon_{Q_t} = \dfrac{v|Y_{Qt}|}{|M_Q|}$, and θ^t_1 and θ^t_2 are relative phases between the entries of the Yukawa matrix. Similar expressions will hold in the bottom sector.

In the SM limit, the expression in Eq. (7) must tend to $\sim \frac{1}{v}(1 + A^S_{1/2}(\tau_b))$, and we would like to find a way to ensure that the overall contribution from the top partners to Higgs production (in gluon fusion) is kept under control. We will require that the Yukawa coupling matrices involving the new vector-like quarks are such that $\det \tilde{\mathbf{Y}}_\mathbf{t} = \det \tilde{\mathbf{Y}}_\mathbf{b} = 0$ with the matrices $\tilde{\mathbf{Y}}_\mathbf{t}$ and $\tilde{\mathbf{Y}}_\mathbf{b}$ defined in Eq. (3). This constraint implies that $Y^t_2 \left(1 - e^{i\theta_1} \frac{|Y^t_1||Y^0_t|}{|Y_{Qt}||Y_{qT}|}\right) = 0$ and thus ensures that the top sector contribution to Higgs production, given in Eq. (8), gives the same result as the SM top quark contribution to the same process.

The above condition of vanishing determinant could simply come from a specific flavor structure in the Yukawa matrix but it turns out that the flavor structure required can also be obtained naturally in models of warped extra-dimensions. In particular we think of the usual Randall–Sundrum warped extra-dimensional scenario [10] with matter fields in bulk. It is necessary, that the Higgs be sufficiently localized towards the brane and that the KK modes of the top quark (and bottom quark) be much lighter than the KK partners of the rest of quarks. Because of this connection to extra-dimensions, we call this parameter space region of our simple model with vector-like fermions the *Brane Higgs Limit* [6].

Finally we can write the ggh couplings as

$$c^S_{ggh} = \frac{1}{v}\left(1 + A^S_{1/2}(\tau_b)\right) + Re\left(\frac{\delta y_b}{m_b}\right)\left(1 - A^S_{1/2}(\tau_b)\right), \quad (9)$$

and

$$c^P_{ggh} = Im\left(\frac{\delta y_b}{m_b}\right)\left(\frac{3}{2} - A^P_{1/2}(\tau_b)\right), \quad (10)$$

where we have used the definitions of the Yukawa coupling shifts in Eq. (4).

2.2 Higgs Phenomenology

We will analyze the new physics in the Higgs sector by computing the signal strengths of different processes, i.e. the ratio of the event cross section with the expected SM cross section.

For gluon fusion production, we obtain in our scenario

$$\frac{\sigma_{gg\to h}}{\sigma^{SM}_{gg\to h}} = \frac{\Gamma_{h\to gg}}{\Gamma^{SM}_{h\to gg}} = (1 + \Delta_{gg}), \quad (11)$$

where $\Delta_{gg} = 2.13v\left(Re\frac{\delta y_b}{m_b}\right) + 1.13v^2\left(Re\frac{\delta y_b}{m_b}\right)^2 + 2.51v^2\left(Im\frac{\delta y_b}{m_b}\right)^2$. This result links in a simple and nontrivial way Higgs production through gluon fusion to the bottom quark Yukawa coupling (or more precisely to its relative shift $v\delta y_b/m_b$).

For tth production cross section, we obtain using Eq. (4),

$$\frac{\sigma_{pp\to tth}}{\sigma^{SM}_{pp\to tth}} = \left|\frac{y_t}{y_{SM}}\right|^2 = (1+\Delta_{tt}) \quad \text{where} \quad \Delta_{tt} = -2vRe\left(\frac{\delta y_t}{m_t}\right) + v^2\left|\frac{\delta y_t}{m_t}\right|^2. \tag{12}$$

Once the Higgs is produced it will decay and so we need to obtain the different partial decay widths. For b-decays we find

$$\frac{\Gamma_{h\to bb}}{\Gamma^{SM}_{h\to bb}} = \left|\frac{y_b}{y_{SM}}\right|^2 = (1+\Delta_{bb}) \quad \text{where} \quad \Delta_{bb} = -2vRe\left(\frac{\delta y_b}{m_b}\right) + v^2\left|\frac{\delta y_b}{m_b}\right|^2. \tag{13}$$

We can also obtain the correction of Higgs decay into $\gamma\gamma$ within the *Brane Higgs Limit*, since the fermion loop is the same as the gluon fusion loop (although there is an additional W loop contribution in this case). We find

$$\frac{\Gamma_{h\to\gamma\gamma}}{\Gamma^{SM}_{h\to\gamma\gamma}} = (1+\Delta_{\gamma\gamma}), \tag{14}$$

where $\Delta_{\gamma\gamma} = -0.14v\left(Re\frac{\delta y_b}{m_b}\right) + 0.005v^2\left(Re\frac{\delta y_b}{m_b}\right)^2 + 0.01v^2\left(Im\frac{\delta y_b}{m_b}\right)^2$.

The signal strength of a particular production and decay mode is defined as $\mu^{ii}_{ggh} = \sigma(gg\to h)/\sigma_{SM}(gg\to h)\times Br(h\to ii)/Br^{SM}(h\to ii)$ with a similar expression for $\mu^{ii}_{t\bar{t}h}$.

With all these ingredients, we find the $t\bar{t}h$ production and decay strengths

$$\mu^{VV}_{t\bar{t}h} = \frac{(1+\Delta_{tt})}{(1+0.58\Delta_{bb}+0.086\Delta_{gg})}, \tag{15}$$

$$\mu^{bb}_{t\bar{t}h} = \mu^{VV}_{t\bar{t}h}(1+\Delta_{bb}), \tag{16}$$

$$\mu^{\gamma\gamma}_{t\bar{t}h} = \mu^{VV}_{t\bar{t}h}(1+\Delta_{\gamma\gamma}), \tag{17}$$

as well as the ggh strengths

$$\mu^{VV}_{ggh} = \frac{(1+\Delta_{gg})}{(1+0.58\Delta_{bb}+0.086\Delta_{gg})}, \tag{18}$$

$$\mu^{\gamma\gamma}_{ggh} = \mu^{VV}_{ggh}(1+\Delta_{\gamma\gamma}), \tag{19}$$

with the corrections Δ_{ii} depending only on top or bottom quark Yukawa coupling shifts and where we have taken into account the numerical values for the main SM Higgs branching ratios.

2.3 Higgs Phenomenology

Within the *Brane Higgs Limit* region of parameter space, all signal strengths associated with the Higgs will deviate from the SM values *only* due to shifts in the top and bottom quark Yukawa couplings. This means that ratios of Higgs signal strengths involving electroweak production processes, and decays through the same channels "ii", should be equal to one, i.e. $\mu^{ii}_{VBF}/\mu^{ii}_{Wh} = \mu^{ii}_{VBF}/\mu^{ii}_{Zh} = 1$. Also signal strengths involving decays into WW should be equal to signals with decays into ZZ, i.e. $\mu^{WW}_{ggh}/\mu^{ZZ}_{ggh} = \mu^{WW}_{tth}/\mu^{ZZ}_{tth} = \mu^{WW}_{Vh}/\mu^{ZZ}_{Vh} = \mu^{WW}_{VBF}/\mu^{ZZ}_{VBF} = 1$.

The corrections to the rest of Higgs signal strengths depend only on two complex parameters, i.e. the relative top and bottom Yukawa coupling deviations δy_t and δy_b. We start exploring the dominant Higgs production mechanism, the gluon fusion process, paying particular attention to the signal strengths $\mu^{\gamma\gamma}_{ggh}$ and $\mu^{WW,ZZ}_{ggh}$. These depend only on the deviation of the bottom quark coupling (magnitude and phase). It is therefore possible to study the relationship between these two signal strengths, for different values of δy_b. This is plotted in Fig. 2, where we show that only a specific region in the $(\mu^{\gamma\gamma}_{ggh}, \mu^{WW,ZZ}_{ggh})$ plane is allowed, due to the *Brane Higgs Limit* constraint.

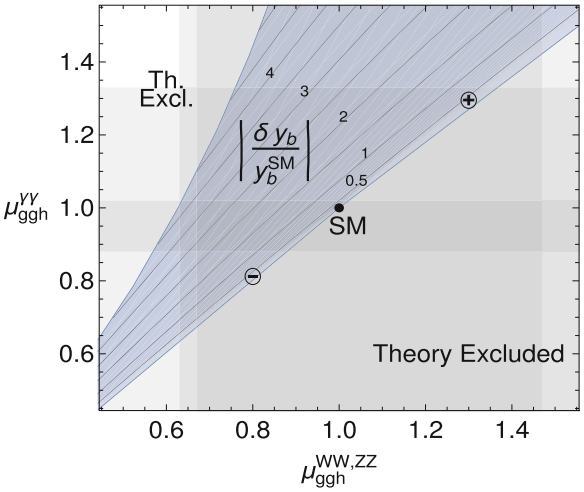


Fig. 2 Contours of the bottom quark Yukawa correction $\left|\delta y_b/y_b^{SM}\right|$ with respect to the gluon fusion signal strengths μ^{VV}_{ggh} and $\mu^{\gamma\gamma}_{ggh}$. The Gray bands represent the experimental bounds set by the LHC RUNS 1 and 2. The "Theory Excluded" regions are points excluded by the *Brane Higgs Limit* constraint. Each contour is traced by varying the phase of δy_b and we include two parameter space points as example limits, marked by a $\oplus$ and a $\ominus$, representing, respectively, an overall enhancement or suppression with respect to the SM predictions. Note that the SM point lies exactly at the boundary between the "Theory Excluded" region and the "allowed" region (blue shaded) and note that the (blue) boundary line separating both regions does not represent a contour of constant $\left|\delta y_b/y_b^{SM}\right|$

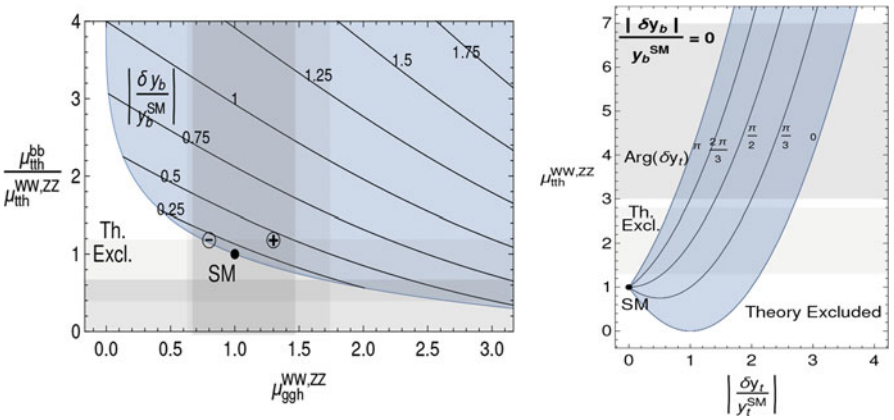


Fig. 3 Contours of the bottom quark Yukawa correction $\left|\delta y_b / y_b^{SM}\right|$ with respect to the gluon fusion signal strength μ_{ggh}^{VV} and the ratio of the $t\bar{t}h$ signal strengths $\mu_{tth}^{bb}/\mu_{tth}^{VV}$ (left panel); and dependence of the Higgs signal strength μ_{tth}^{VV} with respect to the top quark Yukawa coupling correction $\left|\delta y_t / y_t^{SM}\right|$ (right panel). The Gray bands represent the experimental bounds set by the LHC RUNS 1 and 2. The "Th. Excl." region comprises all points excluded by the *Brane Higgs Limit* constraint and we included (left) the two same parameter space points, marked by $\oplus$ and $\ominus$, as in the previous figure

After exploring the gluon fusion signals, we study the effect of small shifts in the bottom Yukawa coupling in tth signals. In particular we focus on the behavior of the ratios $\mu_{tth}^{bb}/\mu_{tth}^{VV} = \mu_{Vh}^{bb}/\mu_{Vh}^{VV}$ as a function of μ_{ggh}^{VV} (all top quark Yukawa dependence cancels out in the ratio). This is shown on the left panel of Fig. 3, where we consider variations of the ratio $\mu_{tth}^{bb}/\mu_{tth}^{VV}$ with respect to the gluon fusion strength μ_{tth}^{bb}.

We now investigate the signals that do depend on the top Yukawa coupling deviations. In the right panel of Fig. 3 we study the variation of μ_{tth}^{VV} with respect to the top Yukawa deviation $|\delta y_t|$. The rest of $t\bar{t}h$ signals strengths can be obtained from ratios of other Higgs production signals strengths, since for example $\mu_{tth}^{\gamma\gamma} = \frac{\mu_{ggh}^{\gamma\gamma}}{\mu_{ggh}^{VV}}\,\mu_{tth}^{VV}$. We conservatively assume no bottom quark Yukawa coupling shifts for this plot ($\delta y_b = 0$), in which we see that the parameter space region is a diagonal band. We also show contours of the phase of the top Yukawa shift δy_t, tracing the band diagonally. The dependence is very sensitive to variations in the phase of the shift of the top Yukawa coupling. We can clearly see that if the magnitude of the top Yukawa deviation is less than 1 (the natural expectation for heavy KK top partners), in order to obtain a signal enhancement (as hinted by LHC data), the phase must be close to π.

3 Outlook

We have presented a simple explanation of possible enhancements in the $t\bar{t}h$ associated production, with the effect decoupled from deviations in Higgs production through gluon fusion at the LHC. This was possible by the addition of vector-like top and bottom partners whose main effect is to modify the top and bottom quark Yukawa couplings and thus potentially modify the cross section for $t\bar{t}h$ production.

In general the effect should also affect the cross section for Higgs production through gluon fusion but we showed that working in a particular limit of parameter space, the *Brane Higgs limit*, the corrections to gluon fusion caused by the top Yukawa coupling variations are exactly offset by the contributions of the new top partners, thus keeping the cross section under control (and only depending on variations of the bottom quark Yukawa coupling!).

Essentially the scenario manages to decouple possible deviations on gluon fusion production from deviations in tth production, which are usually expected to be directly connected. We showed how the scenario links and constrains theoretically all the possible deviations making the scenario easily falsifiable in the coming high luminosity runs at the LHC.

Acknowledgments I would like to thank FRQNT for partial financial support under grant number PRCC-191578 as well as Dawson College for support from the professional development fund. I would also like to thank Gilles Couture, Mariana Frank, and Cherif Hamzaoui for fruitful discussions and collaborations.

References

1. M. Aaboud et al. [ATLAS], Phys. Lett. B **784**, 173–191 (2018)
2. A.M. Sirunyan et al. (CMS Collaboration), Phys. Rev. Lett. **120**, 23180 (2018)
3. [ATLAS], Measurement of Higgs boson production in association with a $t\bar{t}$ pair in the diphoton decay channel using 139 fb^{-1} of LHC data collected at $\sqrt{s} = 13$ TeV by the ATLAS experiment, ATLAS-CONF-2019-004 (March 2019)
4. A.M. Sirunyan et al. [CMS], Measurements of $\text{t}\bar{\text{t}}\text{H}$ production and the CP structure of the Yukawa interaction between the Higgs boson and top quark in the diphoton decay channel. arXiv:2003.10866 [hep-ex] (March 2020)
5. D. Choudhury, T.M.P. Tait, C.E.M. Wagner, Phys. Rev. D **65**, 053002 (2002); D.E. Morrissey, C.E.M. Wagner, Phys. Rev. D **69**, 053001 (2004)
6. G. Couture, M. Frank, C. Hamzaoui, M. Toharia, Phys. Rev. D **95**(9), 095038 (2017)
7. M. Aaboud et al. [ATLAS], Phys. Rev. D **98**(11), 112010 (2018)
8. A. Azatov, M. Toharia, L. Zhu, Phys. Rev. D **80**, 035016 (2009)
9. J.F. Gunion, H.E. Haber, G.L. Kane, S. Dawson, Front. Phys. **80**, 1 (2000)
10. L. Randall, R. Sundrum, Phys. Rev. Lett. **83**, 3370 (1999); ibid., Phys. Rev. Lett. **83**, 4690 (1999)

Part VII
Quantum Information

Mirror Dirac Leptogenesis

Kevin Earl, Chee Sheng Fong, Thomas Gregoire, and Alberto Tonero

Abstract We consider a mirror world scenario, in which light *Dirac neutrinos* are generated from a seesaw mechanism and leptogenesis occurs at high scale without violating lepton number. After leptogenesis, the conservation laws of the theory imply the visible baryon-minus-lepton asymmetry to be equal to the mirror baryon-minus-lepton asymmetry. We extend previous work by presenting a detailed study of this Dirac leptogenesis mechanism by constructing the full set of Boltzmann Equations (BEs) for both cases of unflavored and flavored regimes. We show that Z_2 breaking and lepton/mirror lepton flavor effects can be exploited to enhance the final baryon-minus-lepton asymmetry in our world by several orders of magnitude.

Keywords Leptogenesis · Dirac neutrinos · Mirror world models

1 The Model

We consider a model that, in addition to the SM sector, is characterized by the presence of a mirror sector with the same structure and field content as the SM. In addition to SM fields and their mirror copies, we add heavy singlet neutrinos to both sectors with a Dirac mass term which serves as a portal between the two sectors [1]:

$$\begin{aligned}\mathcal{L} = & \, i\bar{N}_{Ri}\not{\partial}N_{Ri} + i\bar{N}'_{Ri}\not{\partial}N'_{Ri} - M_i\bar{N}_{Ri}N'^{c}_{Ri} \\ & - y_{\alpha j}\bar{l}_{L\alpha}\tilde{\Phi}N_{Rj} - y'_{\alpha j}\bar{l}'_{L\alpha}\tilde{\Phi}'N'_{Rj} + \text{h.c.}\end{aligned} \tag{1}$$

The fields $l_{L\alpha}$ and Φ are SM lepton and Higgs doublets charged under the SM EW $SU(2)_L \times U(1)_Y$ and $\tilde{\Phi} = i\sigma_2\Phi^*$, while $l'_{L\alpha}$ and Φ' are mirror lepton and

K. Earl · T. Gregoire · A. Tonero (✉)
Ottawa-Carleton, Institute for Physics, Carleton University, Ottawa, ON, Canada
e-mail: kearl@physics.carleton.ca; gregoire@physics.carleton.ca

C. S. Fong
Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, SP, Brazil

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_57

Higgs doublets that transform under the mirror EW group $SU(2)'_L \times U(1)'_Y$ and $\tilde{\Phi}' = i\sigma_2 \Phi'^*$. The fields N_{Ri} and N'_{Ri} are heavy fermions which are singlets under both the SM and the mirror EW group. The number of generations in the mirror sector is not fixed, though a minimum of two generations of mirror fermions as well as N_{Ri} and N'_{Ri} are required for leptogenesis and to explain the two observed neutrino squared mass differences. In Eq. (1), the Dirac mass term is responsible for connecting the SM and the mirror sector. In addition, we have chosen, without loss of generality, the basis where the Dirac mass matrix M, the charged lepton and mirror lepton Yukawa couplings (not shown above) are real and diagonal. Other portal interactions might exist, such as a Higgs portal or photon-mirror photon mixing, but are more model dependent.

The structure of Eq. (1) can be obtained by imposing a global $U(1)$ symmetry that can be identified with the total lepton number $L_{\text{tot}} = L - L'$ which is an extended lepton number defined in terms of both SM lepton number (L) and mirror lepton number (L'). We have

$$L_{\text{tot}}(l_{L\alpha}) = L_{\text{tot}}(N_{Ri}) = -L_{\text{tot}}(l'_{L\alpha}) = -L_{\text{tot}}(N'_{Ri}), \qquad L_{\text{tot}}(\Phi) = L_{\text{tot}}(\Phi') = 0. \tag{2}$$

In this case, the SM and twin right-handed neutrinos combine to form heavy Dirac states

$$N = N_R + (N'_R)^c. \tag{3}$$

With this definition, we can rewrite the model Lagrangian as follows:

$$\mathcal{L} = i\bar{N}_i \partial\!\!\!/ N_i - M_i \bar{N}_i N_i - y_{\alpha j} \bar{l}_\alpha \tilde{\Phi} P_R N_j - y'_{\alpha j} \bar{l}'_\alpha \tilde{\Phi}' P_R N_j^c + \text{h.c.} \tag{4}$$

1.1 Heavy N Decay and CP Violation

In the model considered in this study we can have the following decay processes involving the heavy neutrinos $N_i \to l_\alpha \Phi$, $N_i \to \bar{l}'_\alpha \bar{\Phi}'$, and its antiparticle $\bar{N}_i \to \bar{l}_\alpha \bar{\Phi}$, $\bar{N}_i \to l'_\alpha \Phi'$. For generic complex Yukawa couplings y and y' in Eq. (4) we can have CP violation in the decays of the heavy neutrinos and this will imply non-zero $\Delta\Gamma(N_i)_\alpha \equiv \Gamma(N_i \to l_\alpha \Phi) - \Gamma(\bar{N}_i \to \bar{l}_\alpha \bar{\Phi})$ in the visible sector and non-zero $\Delta\Gamma'(\bar{N}_i)_\alpha \equiv \Gamma(\bar{N}_i \to l'_\alpha \Phi') - \Gamma(N_i \to \bar{l}'_\alpha \bar{\Phi}')$ in the mirror sector. Let us define $\Gamma(N_i) \equiv \sum_\alpha [\Gamma(N_i \to l_\alpha \Phi) + \Gamma(N_i \to \bar{l}'_\alpha \bar{\Phi}')]$ and $\Gamma(\bar{N}_i) \equiv \sum_\alpha [\Gamma(\bar{N}_i \to \bar{l}_\alpha \bar{\Phi}) + \Gamma(\bar{N}_i \to l'_\alpha \Phi')]$. CPT conservation implies that

$$\Gamma(N_i) = \Gamma(\bar{N}_i) \equiv \Gamma_{N_i}, \tag{5}$$

where

$$\Gamma_{N_i} = \frac{M_i}{16\pi} [(y^\dagger y)_{ii} + (y'^\dagger y')_{ii}]. \tag{6}$$

It follows that

$$\sum_{\alpha} \Delta\Gamma(N_i)_{\alpha} = \sum_{\alpha} \Delta\Gamma'(\bar{N}_i)_{\alpha}. \tag{7}$$

Furthermore, we can define the CP violating parameters in the SM and mirror sector as follows:

$$\epsilon_{i\alpha} = \frac{\Delta\Gamma(N_i)_{\alpha}}{2\Gamma_{N_i}}, \tag{8}$$

$$\epsilon'_{i\alpha} = \frac{\Delta\Gamma'(\bar{N}_i)_{\alpha}}{2\Gamma_{N_i}}. \tag{9}$$

The relation in Eq. (7) shows that the total CP violation in the visible and hidden sector is the same, namely

$$\epsilon_i = \epsilon'_i, \tag{10}$$

where $\epsilon_i \equiv \sum_{\alpha} \epsilon_{i\alpha}$ and $\epsilon'_i \equiv \sum_{\alpha} \epsilon'_{i\alpha}$. The relation above is ensured by the presence of a global $B - B' - L_{\text{tot}}$ symmetry as discussed in the previous section. The explicit computation of $\epsilon_{i\alpha}$ gives

$$\epsilon_{i\alpha} = \frac{1}{8\pi} \frac{1}{(y^{\dagger}y)_{ii} + (y'^{\dagger}y')_{ii}} \sum_{k} \left[\frac{1}{1-x_k} \text{Im}[(y^{\dagger}y)_{ki} y_{\alpha k} y^*_{\alpha i}] + \frac{\sqrt{x_k}}{1-x_k} \text{Im}[(y'^{\dagger}y')_{ik} y_{\alpha k} y^*_{\alpha i}] \right], \tag{11}$$

where $x_k = M_k^2/M_i^2$. This parameter is obtained from the interference between the tree- and loop-level diagrams in the top row of Fig. 1. Interestingly, the right diagram on the top row of Fig. 1 involves mirror particles in the loop and in performing the computation, "the propagator of the internal N_k picks up a mass term instead of the momentum" and is the only diagram that contributes in the unflavored case.

By a similar computation for the decay to the mirror sector (bottom row of Fig. 1), we obtain $\epsilon'_{i\alpha}$

$$\epsilon'_{i\alpha} = \frac{1}{8\pi} \frac{1}{(y^{\dagger}y)_{ii} + (y'^{\dagger}y')_{ii}} \sum_{k} \left[\frac{1}{1-x_k} \text{Im}[(y'^{\dagger}y')_{ki} y'_{\alpha k} y'^*_{\alpha i}] + \frac{\sqrt{x_k}}{1-x_k} \text{Im}[(y^{\dagger}y)_{ik} y'_{\alpha k} y'^*_{\alpha i}] \right]. \tag{12}$$

Notice that in general the two CP parameter $\epsilon_{i\alpha}$ and $\epsilon'_{i\alpha}$ are different; however, they become equal in the Z_2 symmetric limit where $y = y'$.

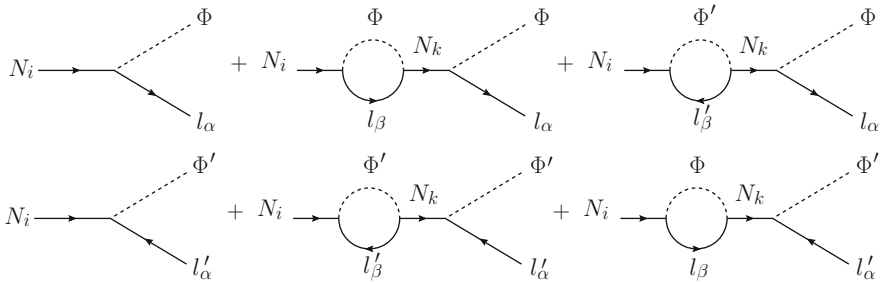


Fig. 1 Diagrams responsible for the decay of the heavy neutrino N into SM particles (top row) and the mirror particles (bottom row), contributing to the calculation of the CP violating parameters $\epsilon_{i\alpha}$ and $\epsilon'_{i\alpha}$

The middle diagrams of Fig. 1 involving the same type of particles in the loop as in the final states are relevant only in the flavored case. They give rise respectively to the first terms in the square brackets of Eqs. (11) and (12) which vanish only if one sums over α.

Summing over α, one can verify that Eq. (10) holds with

$$\epsilon_i = \epsilon'_i = \frac{1}{8\pi} \frac{1}{(y^\dagger y)_{ii} + (y'^\dagger y')_{ii}} \sum_k \frac{\sqrt{x_k}}{1 - x_k} \mathrm{Im}[(y'^\dagger y')_{ik} (y^\dagger y)_{ik}]. \tag{13}$$

This parameter measures the total amount of CP violation induced by the decay of the heavy neutrino N_i. Compared to the standard leptogenesis result [2, 3], there is no triangle or vertex diagram contribution and the CP violation comes only from the interference between the tree-level and the one-loop self-energy diagrams [1].

2 Flavor Enhancement

Here we will discuss a novel enhancement effect that can be achieved in this model [4]. Assuming that $|y'| \gg |y|$, from Eqs. (11) and (12), we have the following parametric dependence:

$$\epsilon_{i\alpha} \sim a(y^4/y'^2) + b(y^2), \tag{14}$$

$$\epsilon'_{i\alpha} \sim a(y'^2) + b(y^2), \tag{15}$$

where $a(x)$ represents the "purely flavor terms" i.e. the first terms in the square brackets of Eqs. (11) and (12) while $b(x)$ represents the second terms in the square brackets of Eqs. (11) and (12). Notice that the purely flavored term in the ϵ' parameter is enhanced by a factor of $\sim y'^2/y^2 \sim P^{-1}$ with P the branching ratio

for singlet decay to the SM sector.[1] Hence if N_i decays mostly to the mirror sector, the flavored CP parameters in the mirror sector can be enhanced accordingly. In this case, even if the flavored CP parameters in the SM are not enhanced, due to conservation of $(B-L)-(B'-L')$ charge, the enhanced production of mirror $(B'-L')$ asymmetry will be fed back to the SM sector, resulting in an overall enhancement of asymmetry production. This is an appealing feature of the model.

As a concrete example, we study N_1 leptogenesis by choosing the following parameters:

$$(P_{1e}, P_{1\mu}, P_{1\tau}) = 10^{-3}(1/3, 1/3, 1/3), \tag{16}$$

$$(P'_{1e}, P'_{1\mu}, P'_{1\tau}) = 0.999(8\times 10^{-3}, 2\times 10^{-3}, 0.99), \tag{17}$$

$$(\epsilon_{1e}, \epsilon_{1\mu}, \epsilon_{1\tau}) = -(1/3, 1/3, 1/3)\epsilon_1^{\max}, \tag{18}$$

$$(\epsilon'_{1e}, \epsilon'_{1\mu}, \epsilon'_{1\tau}) = (1000, 990, -1991)\epsilon_1^{\max}, \tag{19}$$

where the total CP parameter is $\epsilon_1 = \epsilon'_1 = -\epsilon_1^{\max}$. By setting $M_1 = 8\times 10^8$ GeV, $f = 500$ GeV, $m_3 + m_1 = 0.1$. The final baryon asymmetry Y_B obtained from solving the BEs as a function of K_1

$$K_1 \equiv \frac{\Gamma_{N_1}}{H(T=M_1)} \tag{20}$$

is plotted in Fig. 2. The solid red and purple dashed lines refer respectively to the case with zero and thermal initial N_1 abundances while the dotted horizontal line is the observed baryon asymmetry $Y_B = 8.7\times 10^{-11}$.

3 Notes and Comments

Regarding the flavor effects, we have the following comments:

- In both cases of thermal and zero initial abundance we have that flavor effects are impotent in the weak washout regime. In this case the washout terms are negligible during decays of N_1, allowing one to sum over the source term and the final asymmetry will be proportional to the total CP parameter ϵ_1 which is too small in the three-flavor regime. For the parameters specified in the previous section, as shown in Fig. 2, a sufficient baryon asymmetry can be generated for $K_1 \gtrsim 15$ and $K_1 \gtrsim 55$ respectively for thermal and zero initial N_1 abundance.

[1] For N_1 leptogenesis, the flavored terms will have an additional $M_1/M_{k>1}$ suppression while for N_2 leptogenesis, there can be an enhancement of M_2/M_1. Here will assume $M_1/M_{k>1}$ is of a factor of a few and focus on N_1 leptogenesis.

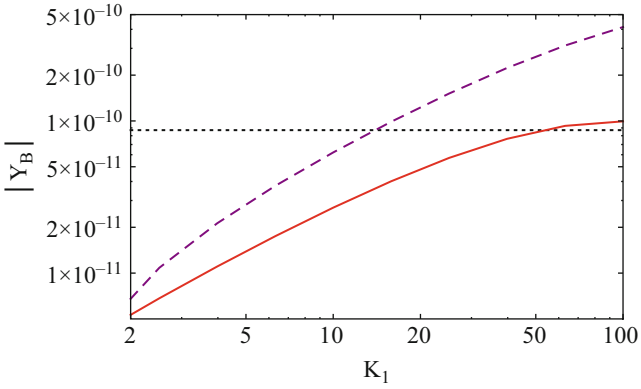


Fig. 2 The final baryon asymmetry Y_B as a function of K_1 with $M_1 = 8 \times 10^8$ GeV in the three-flavor regime assuming the parameters specified in Eqs. (16)–(19) and below them. The red solid and purple dashed lines represent respectively the baryon asymmetry obtained with zero and thermal initial N_1 abundance. The dotted horizontal line represents the value of the observed baryon asymmetry $Y_B = 8.7 \times 10^{-11}$. For the parameters specified above, the baryon asymmetry with initial thermal N_1 abundance has the wrong sign which could be changed by flipping the sign of CP parameters in Eqs. (18) and (19)

- For the zero initial N_1 abundance, the largest asymmetry is induced in the flavor α for which $P'_{1\alpha} K_1 \sim \mathcal{O}(1)$ in order to have a significant washout of the initial "wrong" sign baryon asymmetry generated during N_1 production. For our choice of parameters, the largest asymmetry is generated in the τ sector.
- On the other hand, for thermal initial N_1 abundance, the largest asymmetry is induced in the flavor α for which $P'_{1\alpha} K_1$ is the smallest, i.e. the washout is the smallest. For our choice of parameters, we have that the largest asymmetry is generated in the μ sector. Moreover, for the parameters specified above, the final baryon asymmetry obtained with initial thermal N_1 abundance has the wrong sign compared to that with zero initial N_1 abundance. The correct sign can be obtained by flipping the signs of CP parameters in Eqs. (18) and (19).
- Since the model is symmetric under the exchange of $y \leftrightarrow y'$, we can also achieve the same enhancement by having $|y| \gg |y'|$.

References

1. P.H. Gu, Nucl. Phys. B **872**, 38 (2013). arXiv:1209.4579 [hep-ph]. https://doi.org/10.1016/j.nuclphysb.2013.03.014
2. M. Fukugita, T. Yanagida, Phys. Lett. B **174**, 45 (1986). https://doi.org/10.1016/0370-2693(86)91126-3
3. L. Covi, E. Roulet, F. Vissani, Phys. Lett. B **384**, 169 (1996) [hep-ph/9605319]. https://doi.org/10.1016/0370-2693(96)00817-9
4. K. Earl, C.S. Fong, T. Gregoire, A. Tonero, JCAP. **03**, 036 (2020). arXiv:1903.12192 [hep-ph]. https://doi.org/10.1088/1475-7516/2020/03/036

Fast Tests for Probing the Causal Structure of Quantum Processes

Giulio Chiribella and Swati

Abstract The identification of causal relations is a cornerstone of the scientific method. Traditional approaches to this task are based on classical statistics. However, such classical approaches do not apply in the quantum domain, where a broader spectrum of causal relations becomes accessible. New approaches to quantum causal inference have been developed in recent years, and promising new features have been discovered. In this paper, we review and partly expand the framework and results of reference (Chiribella and Ebler, Nat. Commun. 10(1):1472, 2019), which demonstrated quantum speedups in the identification of various types of causal relations induced by reversible processes.

Keywords Causal inference · Quantum estimation theory · Indefinite causal order

1 Introduction

Identifying cause-effect relations is a fundamental problem in science and engineering [1–3]. In its simplest form, the problem can be described as follows: an

G. Chiribella (✉)
QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong

Department of Computer Science, University of Oxford, Oxford, UK

Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada

HKU Shenzhen Institute of Research and Innovation, Shenzhen, China
e-mail: giulio@cs.hku.hk

Swati
QICI Quantum Information and Computation Initiative, Department of Computer Science, The University of Hong Kong, Pokfulam, Hong Kong

Institute for Quantum Science and Engineering, Department of Physics, Southern University of Science and Technology, Shenzhen, China

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_58

experimenter has collected some amount of raw data about the values of a set of variables, and wants to determine whether a certain variable X influences another variable Y in the set. Since the data are typically subject to noise and fluctuations, the problem of identifying causal relations is ultimately a statistical inference problem.

Traditionally, causal inference methods have been designed for classical variables. Recent advances in quantum science and engineering motivate an extension of these methods to the quantum domain. Experimental techniques can now address individual quantum systems, initialize them in a given quantum state, and subject them to a variety of quantum measurements. In this context, the presence of causal dependencies between two quantum systems acquire a concrete practical relevance. For example, if the interaction of two quantum spins induces a causal dependence between them, then the state of one spin can be controlled by operating on the other spin, to some degree that is determined by the strength of the causal relation.

For quantum systems, classical methods of causal inference turn out to be inadequate. The reason is that such methods assume that randomness can always be reduced to ignorance about the initial conditions of some additional, latent variable. This assumption is at odds with the violation of Bell inequalities, which rules out local realistic models for quantum correlations. For this reason, classical causal models cannot be applied to the Bell scenario [4]. A similar conclusion also holds for more general scenarios, including more than two quantum systems, and/or timelike separations [5, 6].

The breakdown of classical causal inference methods calls for a new, genuinely quantum formulation of the notions of cause and effect. Several frameworks have been proposed to date, with different features and sometimes different purposes. The framework of *quantum combs*, introduced by Chiribella, D'Ariano, and Perinotti in a series of works [7–9], describes networks of quantum processes connected with one another according to a given causal structure. In this framework, one can express the fact that a given process induces a causal dependence between a quantum system and another. More broadly, this notion can also be generalized to probabilistic theories beyond quantum mechanics [10], and to an even broader class of theories described by symmetric monoidal categories [11, 12]. Frameworks for describing causal relations in quantum theory and beyond have been developed in [13–15]. More recently, a quantum framework for describing causal relations between quantum systems has been developed by Allen et al. [16]. This framework, known as *quantum causal models*, can be regarded as an enrichment of the framework of quantum combs, with new conditions that allow one to express the *conditional independences* among quantum variables.

Given a framework for describing causal relationships among quantum variables, one can develop strategies for identifying such relations. Two interesting examples were presented in Refs. [17, 18]. In these works, the authors considered the problem of determining whether two quantum systems, say two photons, are in a spacelike or timelike configuration. Equivalently, this amounts to determining whether the states of the two systems are marginals of a bipartite quantum state (spacelike configuration, corresponding to a past common cause in the joint preparation of the two systems), or whether the state of one system is obtained from the state of the

other by a quantum process (timelike configuration, corresponding to a cause-effect relation directed from one system to the other). Remarkably, the authors found out that certain quantum correlations can distinguish between these two situations, thus defying the classical motto "correlation does not imply causation."

The fact that quantum correlations can be witnesses of causal relationships suggests that quantum measurements could offer more powerful ways to identify causal relations compared to their classical counterpart. In a recent work [19], Chiribella and Ebler showed that quantum features like entanglement and superposition can lead to speedups in the identification of various types of causal relations. In particular, they addressed the following question: given a variable, which variable out of a list of candidate variables, is the effect of it? For simplicity, they focussed on the basic scenario where the cause-effect relation is induced by a reversible process. Classically, it turns out that the minimum probability of error is $p_{\mathrm{err}}^{\mathrm{C}}(N) = 1/[2d^{N-1}]$, where d is the dimension of the quantum systems associated with the given variables, and N is the number of times the variables are probed. This means that the error probability decays exponentially as $p_{\mathrm{err}}^{\mathrm{C}}(N) \approx 2^{-R_{\mathrm{C}}N}$, with decay rate $R_{\mathrm{C}} = \log_2 d$. In stark contrast, Chiribella and Ebler showed that the error probability for quantum strategies decays quadratically faster, with an exponential decay rate equal to $2\log_2 d$. Similar advantages arise in the task of identifying the presence of a causal link between two variables, and in the task of identifying the cause of a given effect.

This paper reviews the framework and the results of Ref. [19] in a way that is suitable for non-specialists. The paper is organized as follows: in Sect. 2, we provide preliminary notions that will be used later in the paper. In Sect. 3, we review the problem of identifying the effect of a given variable, and we provide the minimum error probability and its decay rate for classical strategies. Then, we provide quantum strategies that achieve a speedup over their classical counterpart (Sect. 4). In Sect. 5, we generalize the above results for the case of multiple hypotheses and provide some applications. Finally, we conclude in Sect. 6 by discussing directions of future research.

2 Preliminaries

In this section we review and expand the framework of Ref. [19], providing some additional definitions that help clarifying the different types of causal relations induced by quantum processes.

2.1 Notation

For a given Hilbert space $\mathcal{H}$, we denote by $B(\mathcal{H})$ the space of bounded operators on $\mathcal{H}$, by $T(\mathcal{H})$ the space of trace-class operators, and by $\mathsf{St}(\mathcal{H}) := \{\rho \in T(\mathcal{H}) \mid \rho \geqslant 0\,,\ \mathrm{tr}[\rho] = 1\}$ the convex set of density operators. In the context of causal inference,

quantum systems are often called *quantum variables*. We will denote quantum systems by Roman letters, such as $A, B, \ldots$, and the corresponding Hilbert spaces as $\mathcal{H}_A, \mathcal{H}_B, \ldots$, respectively. We will use the shorthand notation $\mathsf{St}(A) := \mathsf{St}(\mathcal{H}_A)$.

Let A and B be two quantum systems, and let $\mathcal{H}_A$ and $\mathcal{H}_B$ be the corresponding Hilbert spaces. We will denote by $A \otimes B$ the composite system consisting of subsystems A and B, corresponding to the tensor product Hilbert space $\mathcal{H}_{A\otimes B} = \mathcal{H}_A \otimes \mathcal{H}_B$. The partial trace over the Hilbert space $\mathcal{H}_A$ will be denoted as tr_A.

A quantum process with input A and output B is described by a linear, completely positive, trace-preserving map $\mathcal{C} : T(\mathcal{H}_A) \to T(\mathcal{H}_B)$, mapping input states $\rho \in \mathsf{St}(A)$ into output states $\mathcal{C}(\rho) \in \mathsf{St}(B)$. Linear, completely positive and trace-preserving maps will be sometimes abbreviated as CPTP maps. The set of CPTP maps from $T(\mathcal{H}_A)$ to $T(\mathcal{H}_B)$ will be denoted by $\mathsf{CPTP}(A \to B)$.

2.2 Cause-Effect Relations Induced by Quantum Processes

Let A and B be two quantum systems.

Definition 1 We say that a process $\mathcal{C} \in \mathsf{CPTP}(A \to B)$ *induces a cause-effect relation* from A to B if and only if $\mathcal{C}$ is non-constant. If this is the case, then we say that A is a cause for B, and that B is an effect of A.

In other words, a process induces a causal relation from A to B if and only if changing the state of system A can induce changes in the state of system B. The ability to induce changes serves as a witness of the fact that A is a cause for B.

Definition 1 can be generalized to processes involving multiple inputs and outputs.

Definition 2 We say that a bipartite process $\mathcal{D} \in \mathsf{CPTP}((A \otimes A') \to (B \otimes B'))$ *induces a cause-effect relation* from A to B if and only if there exists at least one state $\alpha \in \mathsf{St}\mathcal{H}_{(A')}$ such that the *reduced process*

$$\mathcal{D}_\alpha \in \mathsf{CPTP}(A \to B)\,, \qquad \rho \mapsto \mathrm{tr}_{B'}[\mathcal{D}(\rho \otimes \alpha)] \tag{1}$$

is non-constant. If this is the case, then we say that A is a *cause* for B, and that B is an *effect* of A.

Note that a process $\mathcal{C} \in \mathsf{CPTP}(A \to B)$ may not induce a cause-effect relation from A to B, but may still be the reduced process of some other process $\mathcal{D} \in \mathsf{CPTP}((A \otimes A') \to (B \otimes B'))$ that does induce a cause-effect relation from A to B. This observation shows that the presence of a cause-effect relation, as defined in Definitions 1 and 2, is a property of the process under consideration, and not just of the variables A and B. In other words, a cause-effect relation that is actually present may not be detected by inspecting the process from A to B alone.

The notion of "process inducing a cause-effect relation" provided in Definitions 1 and 2 is rather weak, because it allows the influence of the cause on the effect to be

arbitrarily small. A stronger notion arises when the cause-effect relation is *faithful*, in the following sense:

Definition 3 We say that a process $\mathcal{C} \in \mathsf{CPTP}(A \to B)$ *induces a faithful cause-effect relation* from A to B if and only if $\mathcal{C}$ is correctable, meaning that there exists another process $\mathcal{R} \in \mathsf{CPTP}(B \to A)$ such that

$$\mathcal{R} \circ \mathcal{C} = \mathcal{I}_A \,, \tag{2}$$

where $\mathcal{I}_A \in \mathsf{CPTP}(A \to A)$ is the identity process. When this is the case, we say that B is a *causal intermediary* of A.

Intuitively, variable B is a causal intermediary for variable A if all the possible causal influences of A can be reconstructed from B. In other words, every process from A to a third variable B' must factorize into the process from A to B, followed by some process from B to B'. This intuition is formalized by the following proposition:

Proposition 1 *The process $\mathcal{C} \in \mathsf{CPTP}(A \to B)$ induces a faithful cause-effect relation from A to B if and only if, for every quantum system B', and for every process $\mathcal{E} \in \mathsf{CPTP}(A \to B')$ there exists a process $\mathcal{D} \in \mathsf{CPTP}(B \to B')$ such that*

$$\mathcal{E} = \mathcal{D} \circ \mathcal{C} \,. \tag{3}$$

Proof Suppose that $\mathcal{C}$ induces a faithful cause-effect relation. Then, let $\mathcal{R} \in \mathsf{CPTP}(B \to A)$ be the process that inverts $\mathcal{C}$, namely $\mathcal{R} \circ \mathcal{C} = \mathcal{I}_A$. Then, for every system B' and every process $\mathcal{E} \in \mathsf{CPTP}(A \to B')$, one has $\mathcal{E} = \mathcal{D} \circ \mathcal{C}$, with $\mathcal{D} := \mathcal{E} \circ \mathcal{R}$.

Conversely, suppose that, for every quantum system B' and for every process $\mathcal{E} \in \mathsf{CPTP}(A \to B')$, there exists a process $\mathcal{D} \in \mathsf{CPTP}(B \to B')$ such that $\mathcal{E} = \mathcal{D} \circ \mathcal{C}$. In particular, one can pick $B' \equiv A$ and $\mathcal{E} \equiv \mathcal{I}_A$, in which case the condition $\mathcal{E} = \mathcal{D} \circ \mathcal{C}$ implies that $\mathcal{C}$ is correctable, and therefore induces a faithful cause-effect relation. □

In general, the presence of a faithful cause-effect relation from A to B does not imply that the causal influences of A propagate *exclusively* through B. For example, quantum secret sharing protocols, such as those presented in Ref. [20], provide examples of processes where a given cause can have multiple causal intermediaries. The situation is much simpler when the cause-effect relation is *reversible*, in the following sense:

Definition 4 We say that a process $\mathcal{C} \in \mathsf{CPTP}(A \to B)$ *induces a reversible cause-effect relation* between A and B if and only if $\mathcal{C}$ is reversible, meaning that there exists another process $\mathcal{R} \in \mathsf{CPTP}(B \to A)$ such that

$$\mathcal{R} \circ \mathcal{C} = \mathcal{I}_A \qquad \text{and} \qquad \mathcal{C} \circ \mathcal{R} = \mathcal{I}_A \,, \tag{4}$$

where $\mathcal{I}_X \in \mathsf{CPTP}(X \to X)$ is the identity process on system $X \in \{A, B\}$.

In this paper, we will focus on situations where the cause-effect relation between two variables is reversible. In this case, the presence of a cause-effect relation between variables A and B rules out the possibility of any cause-effect relation from A to any other variable B' that is independent of B. More precisely, if a process $\mathcal{D} \in \mathsf{CPTP}(A \to B \otimes B')$ is such that its reduced process $\mathcal{C} = (\mathcal{I}_B \otimes \mathrm{tr}_{B'}) \circ \mathcal{D}$ is reversible, then $\mathcal{D}$ must be of the form $\mathcal{D} = \mathcal{C} \otimes \beta$, where $\beta \in \mathsf{St}(B')$ is some fixed state of system B'.

2.3 Discrimination of Causal Hypotheses

Consider the situation where an experimenter has access to a black box, implementing a quantum process with a given set of input systems and a given set of output systems. The goal of the experimenter is to figure out the causal relations among the systems involved in the process. For example, the black box could implement a process with input system A and output system B, and the experimenter may want to figure out whether the process induces a cause-effect relation.

In general, we will assume that the black box is guaranteed to satisfy one, and only one, of k possible hypotheses $(\mathsf{H}_i)_{i=1}^k$ about the cause-effect relations occurring between its inputs and outputs. No further information about the process is available to the experimenter. To figure out which hypothesis is correct, the experimenter will set up a test designed to probe the causal relations.

We will use the term *causal hypotheses* for hypotheses on the causal relations induced by a given process. The problem of distinguishing between alternative causal hypotheses will be called *discrimination of causal hypotheses*. To illustrate the problem, we will focus on one basic instance: discover which of two variables B and C is the effect of a given variable A. More specifically, we consider the following alternative hypotheses:

H_1: B is a causal intermediary for A, and C is uniformly random,
H_2: C is a causal intermediary for A, and B is uniformly random.

For simplicity, we will assume that systems A, B, and C have the same dimension, equal to $d < \infty$. In terms of the process $\mathcal{C} \in \mathsf{CPTP}(A \to (B \otimes C))$ the two hypotheses correspond to the following statements:

H_1: $\mathcal{C}$ has the form $\mathcal{C}(\cdot) = U \cdot U^\dagger \otimes I/d$, for some unitary operator $U \in B(\mathbb{C}^d)$,
H_2: $\mathcal{C}$ has the form $\mathcal{C}(\cdot) = I/d \otimes V \cdot V^\dagger$, for some unitary operator $V \in B(\mathbb{C}^d)$.

In either case, the unitary operators U and V are unknown to the experimenter.

In general, every hypothesis on the causal relations between inputs and outputs is in one-to-one correspondence with a subset of CPTP maps. In the above example, the two hypotheses H_1 and H_2 correspond to the sets

$$\mathsf{H}_1 = \left\{ U \cdot U^\dagger \otimes I/d \mid U \in B(\mathbb{C}^d)\,,\; U^\dagger U = U U^\dagger = I \right\} \tag{5}$$

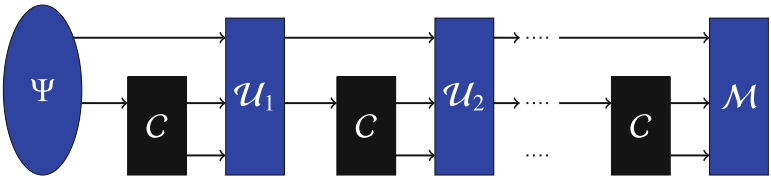


Fig. 1 General setup for testing causal hypotheses on a process $\mathcal{C}$ with one input variable and two output variables [7–9]. The experimenter prepares an initial state Ψ, possibly entangled with an auxiliary system, and probes the process for N times. Between one execution of the process $\mathcal{C}$ and the next, the experimenter can intervene by performing a unitary gate $\mathcal{U}_i$ on the output systems of $\mathcal{C}$ and on some auxiliary system in the laboratory. Finally, a measurement $\mathcal{M}$ is performed and the outcome is used to produce a guess of the correct causal hypothesis

$$\mathsf{H}_2 = \left\{ I/d \otimes V \cdot V^\dagger \mid V \in B(\mathbb{C}^d)\,,\, V^\dagger V = V V^\dagger = I \right\}. \tag{6}$$

The problem is to determine whether the given process $\mathcal{C}$ belongs to H_1 or to H_2.

For the discrimination between these two hypotheses we will consider setups that probe the unknown process $\mathcal{C}$ for N times, by inserting them in a sequential quantum circuit, as illustrated in Fig. 1. As a figure of merit, we will consider the minimization of the probability of error, in the worst case over all possible processes $\mathcal{C}$ that are compatible with the causal hypotheses (5) and (6).

In general, the error probability in distinguishing two known quantum processes decays exponentially with the number of interrogations [21]. Informally, this means that the error probability scales asymptotically as $p_{\rm err}(N) \approx 2^{-N R}$ for some exponent $R > 0$, called the *decay rate*. More formally, the decay rate is defined as $R := \liminf_{N\to\infty} -\log[p_{\rm err}(N)]/N$, where we assume base 2 for the logarithm here and in the rest of the paper. In the following, we will use the decay rate to assess the performance of various strategies for distinguishing causal hypotheses.

3 Optimal Classical Strategy for Finding a Causal Intermediary

Let us consider the classical version of the problem in Sect. 2.3. The classical version involves three classical random variables A, B, and C, each with sample space of cardinality $d < \infty$. The two causal hypotheses are:

H_1: B is a permutation of A, and C is uniformly random,
H_2: C is a permutation of A, and B is uniformly random.

In either case, the exact form of the permutation is unknown to the experimenter.

For parallel strategies where the unknown process is probed on N independent inputs, the optimal error probability $p^{\rm C}_{\rm err}(N)$ is [19]

$$p_{\text{err}}^{\text{C}}(N) = \frac{1}{2d^{N-1}}. \tag{7}$$

This implies the optimal decay rate is

$$R_{\text{C}} = \log d. \tag{8}$$

A result by Hayashi [22] implies that the decay rate cannot be improved by considering more general sequential strategies such as those in Fig. 1.

4 Quantum Advantages

Classical random variables can be regarded as quantum systems that have lost coherence with respect to a privileged orthonormal basis of "classical states." But what if the coherence is preserved? In the following we will show that the possibility of probing processes through coherent superpositions of classical inputs can enhance our ability to identify the correct causal hypothesis.

4.1 A Benefit of Quantum Coherence

The three classical random variables A, B, and C considered in the previous section can be considered as the decohered versions of three quantum systems A, B, and C with d-dimensional Hilbert spaces. Similarly, a permutation of the values of the random variable A according to an element π of the symmetric group $\mathsf{S}(d)$ can be regarded as the decohered version of the unitary permutation operator $U_\pi = \sum_{i=1}^{d} |\pi(i)\rangle\langle i|$, where $\{|i\rangle\}_{i=1}^{d}$ is the orthonormal basis representing the classical states.

In this scenario, the two "classical" causal hypotheses become

H_1: $\mathcal{C}$ has the form $\mathcal{C}(\cdot) = U_\pi \cdot U_\pi^\dagger \otimes I/d$, for some permutation $\pi \in \mathsf{S}(d)$,
H_2: $\mathcal{C}$ has the form $\mathcal{C}(\cdot) = I/d \otimes U_\sigma \cdot U_\sigma^\dagger$, for some permutation $\sigma \in \mathsf{S}(d)$.

To distinguish between these two hypotheses, the experimenter could prepare N probes in the superposition state

$$|e_0\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle \ , \tag{9}$$

which is invariant under permutations. Thus, the unknown process $\mathcal{C}$ yields either $\left(|e_0\rangle\langle e_0| \otimes \frac{\mathbb{I}}{d}\right)^{\otimes N}$ or $\left(\frac{\mathbb{I}}{d} \otimes |e_0\rangle\langle e_0|\right)^{\otimes N}$ depending on which causal hypothesis is correct.

The probability of error can be computed by taking advantage of the symmetry of the problem, which implies that the worst case error probability is equal to the average error probability when the prior probability for the two hypotheses H_1 and H_2 is uniform. In this case, Helstrom's theorem [23] states that the minimum error probability in distinguishing between the states ρ_1 and ρ_2, given with uniform *a priori* probability, is

$$p_{\text{err,ave}}(\rho_1, \rho_2) = \frac{1 - \frac{1}{2}\|\rho_1 - \rho_2\|_1}{2}, \tag{10}$$

where $\|G\|_1 := \text{tr}\sqrt{G^\dagger G}$ denotes the trace norm of a generic trace-class operator $G \in T(\mathcal{H})$. Applied to our problem, Helstrom's theorem yields the error probability

$$p_{\text{err}}^{\text{coh}} = \frac{1}{2d^N}. \tag{11}$$

By comparison with the classical error probability (7), we can see that coherence provides a reduction of the error probability by a factor d. This is a relatively minor improvement, as it does not affect the decay rate. Still, it is an interesting first illustration of how quantum effects can affect the discrimination of causal hypotheses.

4.2 General Quantum Scenario

Let us move on now to the general quantum scenario, where the relation between cause and effect is described by an arbitrary unitary operator (not necessarily a permutation operator, as in the previous subsection).

At first, this problem appears to be much more challenging for our experimenter, since the dependence between cause and effect can be any arbitrary element of the special unitary group $\mathsf{SU}(d)$. Surprisingly, however, the *same* error probability as in Eq. (11) can be achieved.

Let us consider the case where N is an integer multiple of d, say $N = d\,t$ for some non-negative integer t. In this case, a universal quantum strategy with error probability (11) is to divide the N probes into groups of d, and to prepare each group of probes in $\mathsf{SU}(d)$ singlet state

$$|S_d\rangle = \frac{1}{\sqrt{d!}} \sum_{k_1,\cdots,k_d} \epsilon_{k_1\cdots k_d} |k_1, \cdots, k_d\rangle, \tag{12}$$

where $\epsilon_{k_1\cdots k_d}$ represents the totally antisymmetric tensor and the summation extends over all the vectors in the computational basis.

Each of the d particles in each group is then fed into one use of the unknown process. Since the singlet state is invariant under unitary transformations, the

problem becomes to distinguish between the state $|S_d\rangle\langle S_d|^{\otimes t}\otimes(I/d)^{\otimes N}$ (hypothesis H_1) and $(I/d)^{\otimes N}\otimes|S_d\rangle\langle S_d|^{\otimes t}$ (hypothesis H_2). In this case, Helstrom's theorem again gives error probability $1/(2d^N)$, which is equal to the value $p_{\rm err}^{\rm coh}$ obtained in the previous subsection.

If N is not a multiple of d, a rough strategy is to use only the first $\widetilde{N} := d\,\lfloor N/d\rfloor$ probes, and to apply the above procedure. In this way, one obtains error probability $1/(2d^{\widetilde{N}})$, which is suboptimal but still has the same decay rate as the coherent strategy we saw in the previous subsection.

4.3 Parallel Strategies with an External Reference System

We now show that the decay rate of the minimal error probability can be doubled if the experimenter uses an additional reference system entangled with N input probes in the protocol. For simplicity, we will assume that N is a multiple of d.

The basic idea for constructing the quantum-enhanced strategy is the following. In the absence of a reference system, the strategy was to divide the N probes into N/d subgroups of size d, and to put the particles in each group in the singlet state. However, there are many ways of dividing N particles into groups of d. Each of these ways leads to the error probability $p_{\rm err}^{\rm coh} = 1/[2d^N]$. What about trying all possible ways in a coherent superposition?

Consider an external reference system with an orthonormal basis $\{|i\rangle\}_{i=1}^{G_{N,d}}$, where the index i labels the possible ways in which N particles can be divided into groups of d, and $G_{N,d}$ denotes the number of such ways. Then, one can construct the superposition state

$$|\psi\rangle = \frac{1}{\sqrt{G_{N,d}}}\sum_{i=1}^{G_{N,d}}\left(|S_d\rangle^{\otimes N/d}\right)_i\otimes|i\rangle\ , \tag{13}$$

where $\left(|S_d\rangle^{\otimes N/d}\right)_i$ denotes the product of N/d singlet states distributed in the i-th way.

While a classical randomization over all the possible ways to group the N particles is useless, the quantum superposition (13) turns out to be *very* useful. After some algebra [19], it is possible to show that the optimal setup using the state (13) has error probability

$$p_{\rm err}^{\rm Q}(N) = \frac{m(N,d)}{2d^N}\left(1-\sqrt{1-\frac{1}{m(N,d)^2}}\right), \tag{14}$$

where

$$m(N,d) := N! \prod_{i=1}^{d} \frac{(d-i)!}{(N/d+d-1)!} \tag{15}$$

is the multiplicity of the trivial representation of $\mathsf{SU}(d)$ in the N-fold tensor representation $U \mapsto U^{\otimes N}$. For large N, the above expression can be approximated as

$$p_{\rm err}^{\rm Q}(N) \approx \frac{1}{4m(N,d)\, d^N}\,. \tag{16}$$

Using the explicit expression (15), we then obtain the decay rate

$$R_{\rm Q} = -\lim_{N\to\infty} \frac{\log[p_{\rm err}^{\rm Q}(N)]}{N} = 2\log(d) = 2R_{\rm C}\,. \tag{17}$$

Thus, the presence of entanglement between the probes and the external reference system allows one to double the decay rate of the error probability.

The asymptotic limit can be achieved for sufficiently small number of interrogations of the unknown process. For example, an error probability less than 10^{-6} can be achieved by using 12 interrogations for discrimination of a causal relation between two quantum bits, while 20 interrogations are necessary for classical binary variables in order to achieve the same error threshold.

4.4 The Ultimate Quantum Limit

In the previous Sects. 4.2 and 4.3, we considered strategies where the unknown process was applied for N times in parallel on N input systems. These strategies are a special case of the sequential strategies shown in Fig. 1. Can our experimenter further reduce the error probability by using these more general strategies?

A useful tool to address this question is the notion of *fidelity divergence* between two processes, introduced in Ref. [19], and later generalized to a broader set of channel divergences in Ref. [24]. The fidelity divergence between processes $\mathcal{C}_1$ and $\mathcal{C}_2$ is defined as

$$\partial F(\mathcal{C}_1,\mathcal{C}_2) := \inf_{R} \inf_{\rho_1,\rho_2} \frac{F\left(\rho_1',\rho_2'\right)}{F(\rho_1,\rho_2)}, \tag{18}$$

where ρ_1 and ρ_2 are states of the process input and of a reference system R, $\rho_1' := (\mathcal{C}_1\otimes\mathcal{I}_R)(\rho_1)$, and $\rho_2' := (\mathcal{C}_2\otimes\mathcal{I}_R)(\rho_2)$. In order for the above expression to be well defined, the infimum is taken over all the states (ρ_1,ρ_2) such that $F(\rho_1,\rho_2)\neq 0$.

Let us denote by $p_{\rm err}^{\rm seq}(\mathcal{C}_1,\mathcal{C}_2;N)$ the error probability in distinguishing between the two processes $\mathcal{C}_1$ and $\mathcal{C}_2$ using a sequence of N interrogations, as in Fig. 1. It can be shown that the error probability satisfies the bound [19]

$$p_{\rm err}^{\rm seq}(\mathcal{C}_1,\mathcal{C}_2;N)\geqslant\frac{\partial F(\mathcal{C}_1,\mathcal{C}_2)^N}{4}. \tag{19}$$

In the special case where $\mathcal{C}_1=\mathcal{U}\otimes\mathbb{I}/d$ and $\mathcal{C}_2=\mathbb{I}/d\otimes\mathcal{V}$ with $\mathcal{U}(\cdot)=U\cdot U^\dagger$ and $\mathcal{V}(\cdot)=V\cdot V^\dagger$ being fixed unitary processes, the fidelity divergence is $1/d^2$, and the error probability is lower bounded as

$$p_{\rm err}^{\rm seq}(\mathcal{C}_1,\mathcal{C}_2;N)\geqslant\frac{1}{4d^{2N}}. \tag{20}$$

Hence, the decay rate cannot be larger than $2\log d$, *even if the unitaries U and V are known!* If the unitaries U and V are unknown, as in the causal discrimination scenario, then the decay rate can only be $2\log d$ or smaller. This observation proves that the decay rate $2\log d$, achievable with a parallel strategy, is optimal among all decay rates achievable by arbitrary sequential strategies.

As a further curiosity, one may ask whether the rate could be improved in some exotic scenario where the order of the N interrogations is indefinite, unlike in the scenario of Fig. 1, where the N interrogations happen in a well-defined sequential order. In principle, quantum probability theory is logically compatible with scenarios where quantum processes are combined in an indefinite causal order [25–27]. Physically, these scenarios may arise in exotic quantum gravity regimes, although research on such realizations is still in its infancy (and, of course, no complete theory of quantum gravity has been formulated yet). Still, as a theoretical possibility, one can already investigate the question of whether the ability to test a process for N times in an indefinite causal order could help identifying the causal relations occurring between its inputs and outputs.

For the identification of the causal intermediary, the answer turns out to be negative. The proof strategy is to bound the discrimination error for two simple processes, namely, $\mathcal{C}_1=\mathcal{I}\otimes\mathbb{I}/d$ and $\mathcal{C}_2=\mathbb{I}/d\otimes\mathcal{I}$ using arbitrary setups with indefinite order. Using semidefinite programming, Ref. [19] showed that the minimum error probability over all setups that place N uses of the unknown process in an indefinite order is

$$p_{\rm err}^{\rm ind}(\mathcal{C}_1,\mathcal{C}_2;N)\geq\frac{1}{2}\left(1-\sqrt{1-\frac{1}{d^{2N}}}\right). \tag{21}$$

This result establishes the decay rate $2\log d$ as the ultimate limit placed by quantum mechanics to the identification of a causal intermediary. In addition, the strong duality of semidefinite programming guarantees that there exists a suitable setup (possibly requiring indefinite causal order) that achieves the above error probability exactly.

5 Other Examples of Speedups in the Identification of Causal Hypotheses

5.1 Multiple Candidates for the Causal Intermediary

In the previous sections, we have seen how to identify the causal intermediary of a variable A among two possible candidates B and C. What about more than two candidates? For $k > 2$ candidates, the derivation is technically more involved, but the main results remain unchanged:

- In the classical case, when causal relations are described by arbitrary permutations, the minimal error probability is given by

$$p^{\mathrm{C}}_{\mathrm{err},k} = \frac{(k-1)}{2d^{N-1}} + O(d^{-2N}) , \tag{22}$$

  and the decay rate is $\log d$.
- In the quantum case, parallel strategies without a reference system achieve error probability

$$p^{\mathrm{Q}}_{\mathrm{err},k} = \frac{(k-1)}{2d^{N}} + O(d^{-2N}) , \tag{23}$$

  and the decay rate is still $\log d$. In contrast, parallel strategies using entanglement with an external reference system can achieve a doubled decay rate $2\log d$.

5.2 Detection of Causal Link Between Two Variables

A basic example of identification of causal hypotheses is to determine whether there is a causal link between two variables A and B. In this case, the task is to determine whether B is a causal intermediary for A, or whether B fluctuates at random independently of A. As it turns out, entanglement with a reference system can once again double the decay rate of the error probability: the classical error probability decays with rate $\log d$, while the quantum error probability with reference systems decays with doubled rate $2\log d$.

5.3 Identification of the Cause of a Variable

Another interesting problem is to identify which variable in a given set $\{A_1, \cdots A_m\}$ is the cause for a given variable B (assuming that one and only one variable can be the cause). Again, we assume that causal relation is induced by a reversible process

(a permutation in the classical case, or a unitary gate in the quantum case) and that all the variables have sample space of cardinality d in the classical case, or Hilbert space of dimension d in the quantum case.

Classically, the problem is to find the random variable A_i such that B is a permutation of A_i, with $i \in \{1, \cdots, m\}$. In the simplest case, when the permutation is known, the cause can be determined without error by interrogating the unknown process $\lceil \log_d(m) \rceil$ times.

In the quantum case, Ref. [19] showed that, if the unitary operator inducing the causal relation is known, then there exists a test that achieves error probability

$$p_{\text{err}}(N) = \frac{m-1}{d^{2N}+m-1}. \tag{24}$$

This means that one can get an error probability smaller than any desired $\epsilon > 0$ by using $N = \lceil (1+\epsilon)(\log_d m)/2 \rceil$ interrogations, which is approximately half of the number of classical interrogations.

In the case where the dependency between cause and effect is unknown (arbitrary permutation in the classical case, or arbitrary unitary operator in the quantum case), the analysis is more complex. However, it is still possible to show that quantum strategies can identify the cause using only $N = \lceil (1+\epsilon)(\log_d m)/2 \rceil$ interrogations.

6 Discussion and Conclusions

In this paper, we reviewed the framework and the results of Ref. [19], which showed that quantum features such as coherence and entanglement offer advantages in detecting cause-effect relations induced by reversible processes.

In the problem of identifying the effect of a given variable, it was shown that entanglement between the probes and an external reference system can double the rate at which the error probability decays. For classical random variables with d possible values, the decay rate is $\log d$. For quantum systems of dimension d, the decay rate is $2 \log d$. As it turns out, the value $2 \log d$ is the ultimate limit posed by quantum theory to the problem of identifying the causal intermediary of a given variable.

Interestingly, both the classical and quantum decay rates can be expressed as $\log \dim \mathsf{St}_{\mathbb{R}}(A)$ where $\mathsf{St}_{\mathbb{R}}(A)$ is the vector space spanned by linear combinations of states of system A. For classical systems, the states are probability distributions over the sample space, and the dimension of the corresponding vector space is d. For quantum systems, the states are density matrices, and the dimension of the vector space is d^2. It would be interesting to study the problem of causal hypothesis discrimination in toy theories with higher dimensional state spaces, such as quantum theory on quaternionic Hilbert spaces [28] or the quartic toy theory proposed by Życzkowski in Ref. [29].

On a more practical side, it is important to extend the analysis from the idealized scenario where the cause-effect dependencies are induced by reversible processes, to the more realistic scenario where they are induced by general noisy processes. Preliminary results in Ref. [19] indicate that quantum advantages may still persist for sufficiently low noise levels. However, a fully general treatment of noise is still lacking and will be important for future applications.

Given the success of causal discovery algorithms in classical statistics and machine learning, it is natural to expect that the development of quantum causal discovery algorithms may have applications to the burgeoning field of quantum machine learning [30–32]. This connection is largely unexplored and represents an exciting direction of future research.

Acknowledgments GC is grateful to R Spekkens, D Schmid, and M T Quintino for stimulating discussions on the notion of causality for quantum processes. This work is supported by the National Science Foundation of China through Grant No. 11675136, by Hong Kong Research Grant Council through Grants No. 17326616 and 17300918, by the Croucher Foundation, and by the John Templeton Foundation through grant 61466, The Quantum Information Structure of Spacetime (qiss.fr). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. Research at the Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

References

1. P. Spirtes, C. Glymour, R. Scheines (eds.) *Causation, Prediction, and Search* (MIT Press, Cambridge, 2000)
2. J. Pearl, *Causality* (Cambridge University Press, Cambridge, 2009)
3. J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers Inc., San Francisco, CA, 1988)
4. C.J. Wood, R.W. Spekkens, The lesson of causal discovery algorithms for quantum correlations: causal explanations of bell-inequality violations require fine-tuning. New J. Phys. **17**(3), 033002 (2015)
5. R. Chaves, G. Carvacho, I. Agresti, V. Di Giulio, L. Aolita, S. Giacomini, F. Sciarrino, Quantum violation of an instrumental test. Nat. Phys. **14**(3), 291–296 (2018)
6. T. Van Himbeeck, J. Bohr Brask, S. Pironio, R. Ramanathan, A.B. Sainz, E. Wolfe, Quantum violations in the instrumental scenario and their relations to the bell scenario. Quantum **3**, 186 (2019)
7. G. Chiribella, G.M. D'Ariano, P. Perinotti, Quantum circuit architecture. Phys. Rev. Lett. **101**(6), 060401 (2008)
8. G. Chiribella, G.M. D'Ariano, P. Perinotti, Memory effects in quantum channel discrimination. Phys. Rev. Lett. **101**(18), 180501 (2008)
9. G. Chiribella, G.M. D'Ariano, P. Perinotti, Theoretical framework for quantum networks. Phys. Rev. A **80**(2), 022339 (2009)
10. G. Chiribella, G.M. D'Ariano, P. Perinotti, Probabilistic theories with purification. Phys. Rev. A **81**(6), 062348 (2010)
11. B. Coecke, R.W. Spekkens, Picturing classical and quantum Bayesian inference. Synthese **186**(3), 651–696 (2011)

12. B. Coecke, Terminality implies non-signalling. Electron. Proc. Theor. Comput. Sci. **172**, 27–35 (2014)
13. J. Henson, R. Lal, M.F. Pusey, Theory-independent limits on correlations from generalized Bayesian networks. New J. Phys. **16**(11), 113043 (2014)
14. J. Pienaar, Č. Brukner, A graph-separation theorem for quantum causal models. New J. Phys. **17**(7), 073020 (2015)
15. F. Costa, S. Shrapnel, Quantum causal modelling. New J. Phys. **18**(6), 063032 (2016)
16. J.-M.A. Allen, J. Barrett, D.C. Horsman, C.M. Lee, R.W. Spekkens, Quantum common causes and quantum causal models. Phys. Rev. X **7**, 031021 (2017)
17. J.F. Fitzsimons, J.A. Jones, V. Vedral, Quantum correlations which imply causation. Sci. Rep. **5**(1), 18281 (2015)
18. K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R.W. Spekkens, K.J. Resch, A quantum advantage for inferring causal structure. Nat. Phys. **11**(5), 414–420 (2015)
19. G. Chiribella, D. Ebler, Quantum speedup in the identification of cause–effect relations. Nat. Commun. **10**(1), 1472 (2019)
20. R. Cleve, D. Gottesman, H.-K. Lo, How to share a quantum secret. Phys. Rev. Lett. **83**(3), 648 (1999)
21. N. Yu, L. Zhou, Chernoff bound for quantum operations is faithful (2017). Preprint. arXiv:1705.01642
22. M. Hayashi, Discrimination of two channels by adaptive methods and its application to quantum system. IEEE Trans. Inf. Theor. **55**(8), 3807–3820 (2009)
23. C.W. Helstrom, Quantum detection and estimation theory. J. Stat. Phys. **1**(2), 231–252 (1969)
24. M. Berta, C. Hirche, E. Kaur, M.M Wilde, Amortized channel divergence for asymptotic quantum channel discrimination (2018). Preprint. arXiv:1808.01498
25. G. Chiribella, G.M. D'Ariano, P. Perinotti, B. Valiron, Beyond quantum computers (2009). Preprint. arXiv:0912.0195
26. O. Oreshkov, F. Costa, Č. Brukner, Quantum correlations with no causal order. Nat. Commun. **3**(1), 1092 (2012)
27. G. Chiribella, G.M. D'Ariano, P. Perinotti, B. Valiron, Quantum computations without definite causal structure. Phys. Rev. A **88**(2), 022318 (2013)
28. H. Barnum, M.A. Graydon, A. Wilce, Some nearly quantum theories. Electron. Proc. Theor. Comput. Sci. **195**, 59–70 (2015)
29. K. Życzkowski, Quartic quantum theory: an extension of the standard quantum mechanics. J. Phys. A: Math. Theor. **41**(35), 355302 (2008)
30. M. Schuld, I. Sinayskiy, F. Petruccione, An introduction to quantum machine learning. Contemp. Phys. **56**(2), 172–185 (2015)
31. J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, S. Lloyd, Quantum machine learning. Nature **549**(7671), 195–202 (2017)
32. V. Dunjko, H.J. Briegel, Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Rep. Progr. Phys. **81**(7), 074001 (2018)

Qubits as Edge State Detectors: Illustration Using the SSH Model

Meri Zaimi, Christian Boudreault, Nouédyn Baspin, Hichem Eleuch, Richard MacKenzie, and Michael Hilke

Abstract As is well known, qubits are the fundamental building blocks of quantum computers, and more generally, of quantum information. A major challenge in the development of quantum devices arises because the information content in any quantum state is rather fragile, as no system is completely isolated from its environment. Generally, such interactions degrade the quantum state, resulting in a loss of information.

Topological edge states are promising in this regard because they are in ways more robust against noise and decoherence. But creating and detecting edge states can be challenging. We describe a composite system consisting of a two-level system (the qubit) interacting with a finite Su–Schrieffer–Heeger chain (a hopping model with alternating hopping parameters) attached to an infinite chain. In this model, the dynamics of the qubit changes dramatically depending on whether or not

M. Zaimi
Centre de Recherches Mathématiques, Université de Montréal, Montréal, QC, Canada
e-mail: meri.zaimi@umontreal.ca

C. Boudreault
Département des sciences de la nature, Collège militaire royal de Saint-Jean, Saint-Jean-sur-Richelieu, QC, Canada
e-mail: Christian.Boudreault@cmrsj-rmcsj.ca

N. Baspin · M. Hilke
Department of Physics, McGill University, Montréal, QC, Canada
e-mail: nouedyn.baspin@mail.mcgill.ca; hilke@physics.mcgill.ca

H. Eleuch
Department of Applied Sciences and Mathematics, College of Arts and Sciences, Abu Dhabi University, Abu Dhabi, United Arab Emirates

Institute for Quantum Science and Engineering, Texas A&M University, College Station, TX, USA

R. MacKenzie (✉)
Département de physique, Université de Montréal, Montréal, QC, Canada
e-mail: richard.mackenzie@umontreal.ca

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_59

an edge state exists. Thus, the qubit can be used to determine whether or not an edge state exists in this model.

Keywords Open quantum systems · Decoherence · Topological materials · Edge states

1 Introduction

The two-level system (TLS) is the simplest nontrivial quantum system. Its simplicity notwithstanding, many important systems are TLSs. Some familiar examples are: a spin-1/2 particle (two spin states), a photon (two polarizations), a two-level atom (the two levels), a quantum dot (empty/full), two-meson systems (K, $\bar{K}$), two-flavor neutrino oscillations ($\nu_{1,2}$). Some of the above play the role of qubits, the building blocks of quantum information systems (quantum computers, teleportation, etc.).

An isolated TLS, like any isolated quantum system, will evolve unitarily. This implies that pure states remain pure; assuming the two basis states are coupled, a system put in one state will oscillate back and forth between the two.

However, no system is perfectly isolated; in reality, a TLS interacts with its environment and becomes entangled with it. From the point of view of the TLS, entanglement with the environment is indistinguishable from a mixed state. We say that the pure state becomes impure, or it decoheres. In addition, in many TLSs (including the one we will study here) the interaction can permit a particle to escape from the system to the environment. In this case, from the point of view of the TLS probability is not conserved.

Decoherence and nonconservation of probability are almost always undesirable; in particular, decoherence results in a loss of information and also a loss of the potential advantage of quantum vs classical computing, quantum vs classical communication, etc. Thus, understanding (and, usually, minimizing) decoherence is critically important to the functioning of quantum devices. As an example, in [1] a tripartite system was studied: a TLS coupled to one end of a finite chain (or channel) whose other end was coupled to a semi-infinite chain; both chains were described by tight-binding Hamiltonians. The question addressed was: how can one reduce the decoherence of the TLS? It was found that adding noise to the channel did the trick, essentially due to Anderson localization: if excitations in the channel are localized, it becomes hard for a particle in the TLS to make its way to the far side of the channel and escape to infinity.

Here, we study a similar system with a very different goal in mind (Fig. 1). The main difference is that the channel is a Su–Schrieffer–Heeger (SSH) [2] chain (free of disorder) described by a hopping parameter with alternating hopping strengths. Such chains have topological edge states (for a review, see [3]), and rather than trying to minimize the decoherence of the TLS, we imagine measuring its decoherence rate to determine whether the system to which it is attached has edge

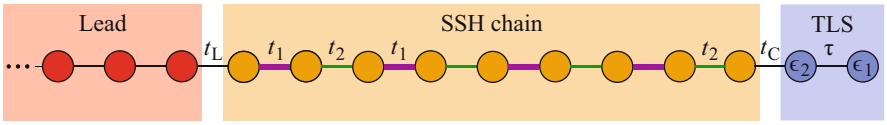


Fig. 1 Tripartite system geometry. Rightmost SSH chain hopping parameter is t_1 or t_2 depending on whether number of sites N is even or odd, respectively (odd case shown here)

states. As we will see, the presence of edge states greatly increases the decoherence rate.

2 Two-Level System: A Rapid Review

We review the isolated TLS, mostly to establish notation to be used in what follows. The TLS Hamiltonian is

$$H^{\mathrm{DD}} = \begin{pmatrix} \epsilon_2 & \tau \\ \tau & \epsilon_1 \end{pmatrix} \equiv \begin{pmatrix} \epsilon_0 - \delta_0/2 & \tau \\ \tau & \epsilon_0 + \delta/2 \end{pmatrix}. \tag{1}$$

The energies are $\lambda_\pm = \frac{1}{2}(\epsilon_1 + \epsilon_2 \pm \delta) = \epsilon_0 \pm \frac{\delta}{2}$, where $\delta = \sqrt{(\epsilon_1 - \epsilon_2)^2 + 4\tau^2}$.

The energy-dependent Green's function is defined by $G^{\mathrm{DD}}(E) = (E - H^{\mathrm{DD}})^{-1}$; its Fourier transform gives the time-dependent Green's function, which is a sum of oscillatory terms with frequencies given by the energies; for instance,

$$G^{\mathrm{DD}}_{12}(t) = -\frac{2\pi i \tau}{\delta}\left(e^{-i\lambda_+ t} - e^{-i\lambda_- t}\right). \tag{2}$$

When we couple the TLS to the rest of the system, it will decohere; this will be seen in the Green's function, which will exhibit decaying behavior [1].

3 Su–Schrieffer–Heeger Model and Edge States

The SSH Hamiltonian [2], proposed in the context of the polymer polyacetylene for reasons we will not go into here, is

$$H_{\mathrm{SSH}} = \begin{pmatrix} 0 & t_1 & & & & \\ t_1 & 0 & t_2 & & & \\ & t_2 & 0 & t_1 & & \\ & & t_1 & 0 & \ddots & \\ & & & \ddots & \ddots & t \\ & & & & t & 0 \end{pmatrix}, \tag{3}$$

where $t = t_1$ or t_2 for N even or odd, respectively. We will assume $t_1, t_2 > 0$ for simplicity, and for now we assume N is even and write $N = 2M$. Much of what follows is known [4–7]; we repeat it to establish notation and to focus on results to be used below.

To solve the Schrödinger equation, translational invariance (by two sites) suggests the following ansatz:

$$|\psi\rangle = \sum_{n=0}^{M-1} \left(A\,|2n+1\rangle + B\,|2n+2\rangle\right) e^{in2k}. \tag{4}$$

We can take k between $\pm\pi/2$ since $k \to k + \pi$ has no effect on $|\psi\rangle$. The middle components of the Schrödinger equation (all but the first and last) determine the dispersion relation and also the ratio A/B. The former is

$$E^2 = t_1^2 + t_2^2 + 2\,t_1 t_2 \cos 2k. \tag{5}$$

Assuming k is real, $(t_1 - t_2)^2 < E^2 < (t_1 + t_2)^2$ so there are two energy bands. For any allowed energy, (5) has two equal and opposite solutions for $\pm k$ where we assume $k > 0$. Thus the general solution to the middle equations is a linear combination of the solutions for $\pm k$.

The edge components of the Schrödinger equation (the first and last) determine the ratio of these two solutions, and also the energy eigenvalues. The latter are given by the solutions of the following equation for k, where $r = t_1/t_2$ and we have written $\mathrm{s}_j = \sin(jk)$.

$$r\,\mathrm{s}_{N+2} + \mathrm{s}_N = 0 \tag{6}$$

where $r = t_1/t_2$ and we have written $\mathrm{s}_j = \sin(jk)$.

This equation cannot be solved analytically; however, numerically or graphically (see Fig. 2) we find that there are N real solutions, as required, for $r > r_{\mathrm{C}}$ whereas there are two fewer real solutions for $r < r_{\mathrm{C}}$, where [4]

$$r_{\mathrm{C}} \equiv \frac{N}{N+2}. \tag{7}$$

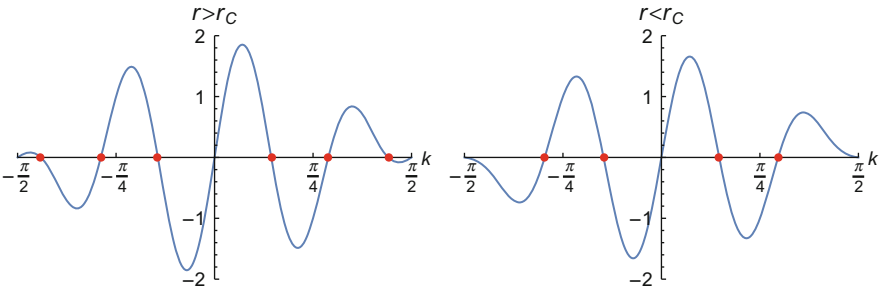


Fig. 2 Graphical solution of (6) for $N = 6$ ($r_C = 0.75$). Left panel: $r = 0.9$; six solutions. Right panel: $r = 0.7$; four solutions (Note that $k = 0, \pm\pi/2$, although solutions of (6), do not correspond to solutions to the SE)

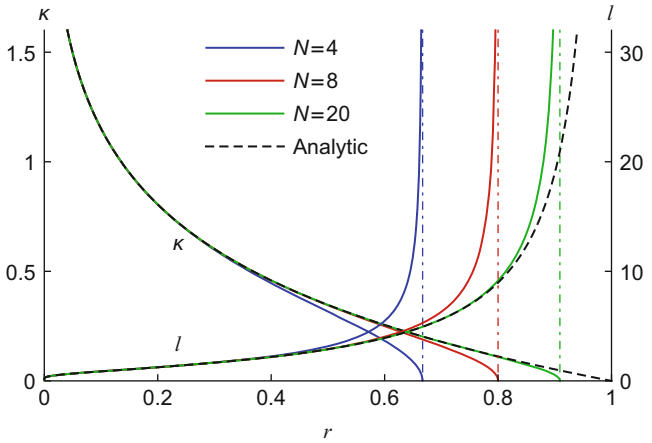


Fig. 3 Decay rate κ and decay length l of edge states for various values of N. Also displayed is an analytic solution to (8) for $N \to \infty$

Thus, for $r < r_C$ there are two missing solutions. These turn out to be solutions of complex wave number, $k = \pi/2 \pm i\kappa$, where κ is the positive solution of

$$\frac{\sinh(N\kappa)}{\sinh((N+2)\kappa)} = r \tag{8}$$

the solution of which is displayed in Fig. 3 for various values of N. These states, having complex k, are exponentially confined to the edges of the system: they are edge states. Also displayed is $l = 1/\kappa$, the penetration length of the edge states. As $r \to r_C$ from below, we see that the length scale goes to infinity; the "edginess" of the edge states becomes irrelevant if $l \gg N$.

We conclude with a brief discussion of the case N odd, which is in fact much simpler. It is easy to show that no matter the value of r, there is always exactly one

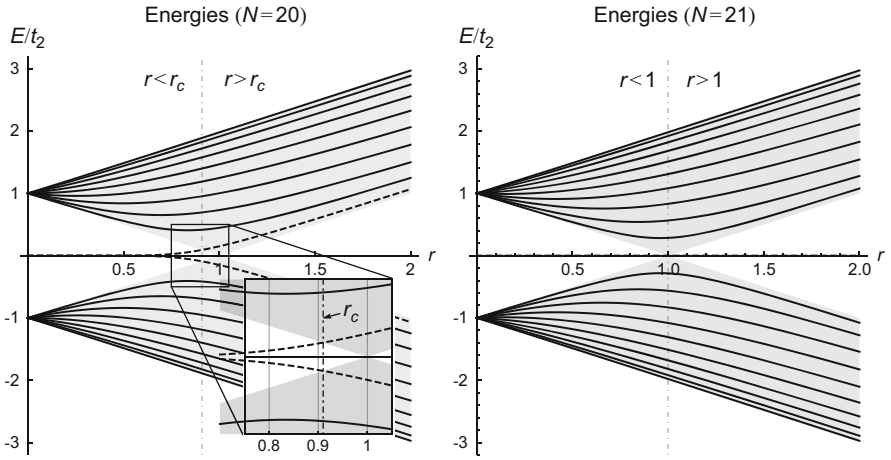


Fig. 4 Energy spectra for two values of N as a function of r. The shaded regions are the bands for $N = \infty$. Dotted lines outside the bands are edge states. The inset on the left focuses on the transition between an edge state (to the left of the vertical broken line) and a non-edge state (to the right, in the shaded region)

zero-energy edge state (the remainder of the spectrum being symmetric). This state is confined to the left (right) edge for $r < 1$ ($r > 1$) with decay length $l = 1/|\log r|$. Figure 4 displays the spectra for $N = 20$ and 21.

4 Tripartite System: TLS-SSH-Chain

We now study the tripartite system displayed in Fig. 1. Although it is an infinite-dimensional system, the effects of the SSH chain and semi-infinite chain on the TLS can be nicely incorporated into a 2×2 effective Hamiltonian for the TLS; these effects are simply given by a term added to the (1, 1) component of the Hamiltonian [7, 8]:

$$\epsilon_2 \to \epsilon_2 + \Sigma_{\mathrm{SSH},\infty} \equiv \epsilon_2'. \tag{9}$$

Here $\Sigma_{\mathrm{SSH},\infty}$ is proportional to the surface Green's function of the combined SSH chain and semi-infinite chain. This can be calculated analytically, although it is fairly nasty. The result is [7]

$$\Sigma_{\mathrm{SSH},\infty} = \begin{cases} {t_{\mathrm{C}}}^2 \dfrac{E t_2 s_N - \Sigma_\infty (t_1 s_{N-2} + t_2 s_N)}{t_2^2 (t_1 s_{N+2} + t_2 s_N) - E t_2 \Sigma_\infty s_N} & (N \text{ even}) \\ {t_{\mathrm{C}}}^2 \dfrac{t_2 (t_2 s_{N-1} + t_1 s_{N+1}) - E \Sigma_\infty s_{N-1}}{t_1 t_2 E s_{N+1} - t_1 \Sigma_\infty (t_2 s_{N+1} + t_1 s_{N-1})} & (N \text{ odd}) \end{cases} \tag{10}$$

where

$$\Sigma_\infty = \frac{{t_{\mathrm{L}}}^2}{2} \left(E - i\sqrt{4 - E^2} \right). \tag{11}$$

Note that ϵ_2' is complex, so the effective Hamiltonian is no longer Hermitian. This is related to the open nature of the TLS: being non-Hermitian, time evolution preserves neither probability nor purity, reflecting the fact that the particle can escape to its environment, and that the TLS and environment become entangled.

Defining $\lambda_\pm'$ and δ' as the quantities defined in Sect. 2 with the substitution (9), we can use these substitutions in the definition of $G^{\mathrm{DD}}(E)$ given earlier to get the new energy-dependent Green's function, $G^{\mathrm{DD}}_{\mathrm{SSH},\infty}(E)$. It is tempting to suppose that these substitutions also work for the time-dependent Green's function. This is not quite correct, since $\lambda_\pm'$ depend in a highly nontrivial way on E so the Fourier transform cannot be evaluated exactly. An analytical approximation which is justified in the weak-coupling limit ($t_{\mathrm{C}}\lambda 1$) [1] indicates that to a good approximation the new (complex) frequencies $\lambda_\pm'$ can be evaluated at the old frequencies: the time-dependent Green's function has, according to this approximation, frequencies $\lambda_\pm'(\lambda_\pm)$. According to this analytic approximation, the decay rates are given by the imaginary part of the frequencies, and we conclude that the decoherence time τ_ϕ is given by

$$\left(\tau_\phi\right)^{-1} \approx \min \left(-\frac{1}{2} \operatorname{Im} \left\{ \Sigma_{\mathrm{SSH},\infty}(\lambda_\pm) \pm \delta'(\lambda_\pm) \right\} \right). \tag{12}$$

This analytical approximation can be justified post hoc by comparing (12) with a numerical evaluation of the decoherence rate. Both are displayed in Fig. 5. The figure, which encapsulates our main result, merits some discussion. There are four cases to consider, two for each graph.

The graph on the left corresponds to $r = 1.21$, for which there are no edge states if N is even, while there is a right edge state if N is odd. If N is even, both TLS states lie in the gap, so there are no SSH states with which they can hybridize. Thus, the SSH chain represents a sort of potential barrier impeding escape of the particle to the semi-infinite chain. As a result, the tunneling rate decreases exponentially with N beyond about 20. If N is odd (blue, upper curve), the right-hand edge state couples strongly to the TLS forming a pair of hybridized wave functions which penetrate the SSH chain. This penetration facilitates decoherence, and the rate remains large as N increases. Since the hybridized wave function itself drops off exponentially away from the right edge of the SSH chain, its effect on decoherence drops off as

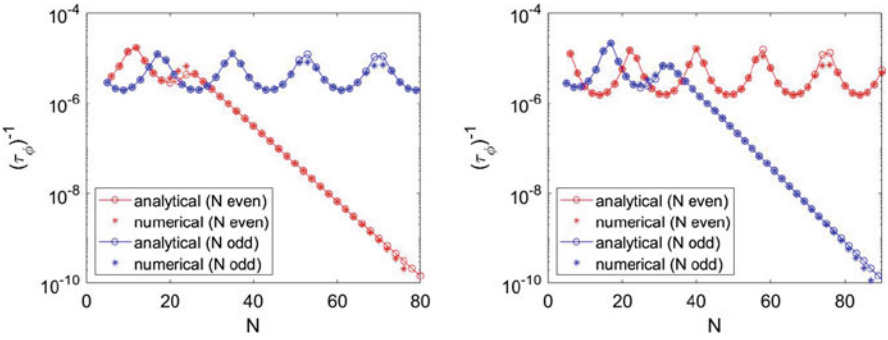


Fig. 5 Decoherence rate as a function of N for $t_1 = 1/t_2 = 1.1$ (left), $t_2 = 1/t_1 = 1.1$ (right). For both figures, $(\epsilon_1, \epsilon_2, \tau, t_C, t_L) = (.4022, .0022, .03, .035, .65)$. The values for $\epsilon_{1,2}$ were chosen so that the isolated TLS has a zero eigenvalue, corresponding exactly (N odd) and approximately (N even) to the edge state energy. The energy of the other TLS state lies in the SSH gap

N increases; although this is not apparent in Fig. 5, if we continue the graph beyond about $N = 120$, this effect is clearly seen [7]: eventually the blue curve on the left drops much like the red one does.

The graph on the right corresponds to $r \simeq 0.83$, for which there are two edge states if N is greater than 10 and even, while there is a left edge state if N is odd. Again, the behavior is dramatically different for even vs. odd parity. If N is even (red, upper curve), the presence of edge states results in hybridized wave functions which facilitate decoherence. As was discussed above ($r = 1.21$, N even), the decoherence is relatively independent of N until around $N = 120$, after which it drops exponentially. If N is odd (blue, lower curve), the absence of an edge state on the TLS side of the SSH chain impedes hybridization and giving rise to exponential decoherence suppression as N increases starting around $N = 30$.

While we believe the presence or absence of edge states explains in general terms the behavior exhibited in Fig. 5, one unresolved issue is why the dropoff in decoherence begins where it does. It is easy to show that for the parameters used the edge states have a characteristic width of about ten sites, so it is puzzling why the enhanced decoherence illustrated by the upper curves persists until beyond $N = 100$.

5 Conclusions

The interaction between a TLS and its environment can have a strong effect on the dynamics of the TLS. Here, we argued that coupling to one end of an SSH chain (which is coupled at the other end to an undimerized infinite chain) can have a very strong effect on the decoherence of the TLS. The effect is dramatically different depending on whether there is or is not an edge state at the TLS end of the SSH

chain: an edge state causes decoherence to remain high independent of chain length, whereas in the absence of an edge state decoherence decreases exponentially with chain length. This suggests using a TLS as a sort of edge state detector.

Acknowledgments This work was supported in part by the Natural Science and Engineering Research Council of Canada and by the Fonds de Recherche Nature et Technologies du Québec via the INTRIQ strategic cluster grant. RM is grateful for the hospitality of Perimeter Institute where part of this work was carried out. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development and by the Province of Ontario through the Ministry of Research, Innovation and Science.

References

1. H. Eleuch, M. Hilke, R. MacKenzie, Probing Anderson localization using the dynamics of a qubit. Phys. Rev. A **95**, 062114 (2017)
2. W.P. Su, J.R. Schrieffer, A.J. Heeger, Solitons in polyacetylene. Phys. Rev. Lett. **42**, 1698 (1979)
3. M.Z. Hasan, C.L. Kane, Colloquium: topological insulators. Rev. Mod. Phys. **82**, 3045 (2010)
4. P. Delplace, D. Ullmo, G. Montambaux, Phys. Rev. B **84**, 195452 (2011)
5. J.K. Asbóth, L. Oroszlány, A. Pályi, *A Short Course on Topological Insulators*. Lecture Notes in Physics, vol. 919 (Springer International Publishing, Cham, 2016)
6. C. Duncan, P. Öhberg, M. Valiente, Phys. Rev. B **97**, 195439 (2018)
7. M. Zaimi, C. Boudreault, N. Baspin, H. Eleuch, R. MacKenzie, M. Hilke (2019). arXiv:1910.09926 [cond-mat.mes-hall]
8. S. Datta, *Quantum Transport: Atom to Transistor* (Cambridge University Press, Cambridge, 2005)

RepLAB: A Computational/Numerical Approach to Representation Theory

Denis Rosset, Felipe Montealegre-Mora, and Jean-Daniel Bancal

Abstract We present a MATLAB/Octave toolbox to decompose finite dimensional representations of compact groups. Surprisingly, little information about the group and the representation is needed to perform that task. We discuss applications to semidefinite programming.

Keywords Representation theory · Compact groups

Early in the development of quantum formalism, some saw group theory as a mere nuisance. The label *Gruppenpest* is attributed to Pauli's talent for derision [1]. Today, the usefulness of group theory is no longer debated, usually a handful of groups are employed: symmetric, cyclic, Pauli and Clifford groups, and the (special) unitary group. It turns out that a larger variety of groups are present in quantum information computations. Thousands of Bell inequalities [2] have been discovered [3]: most with some form of symmetry [4]. And even a well understood problem (Schur-Weyl duality) required recent extensive work [5–9] to cater for partial transposition. In parallel, quantum information has seen the rise of numerical methods, especially those based on semidefinite positive (SDP) optimization. Density operators correspond naturally to SDP matrices, and so do a variety of quantum objects [10–12]; see, for example, SDP relaxations to test entanglement [13–16]. Moment-based (SDP) relaxations bound the set of quantum correlations [17–21]. Such SDP relaxations inherit the symmetries of the objects studied. We thus need an approach to decompose representations of a large class of symmetry groups. This paper presents the first compact, self-contained

D. Rosset (✉)
Perimeter Institute for Theoretical Physics, Waterloo, ON, Canada

F. Montealegre-Mora
Institute for Theoretical Physics, Universität zu Köln, Cologne, Germany
e-mail: fmonteal@thp.uni-koeln.de

J.-D. Bancal
Département de Physique Appliquée, Université de Genève, Genève, Switzerland

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_60

software to automate symmetry reduction of SDP optimization problems, with direct application to quantum information problems.[1]

An invariant SDP problem corresponds to a matrix invariant under a group representation. Following [24], let G be a compact group, with finite dimensional representation ρ acting on the representation space $V = \mathbb{C}^n$:

$$\rho : G \to \mathcal{U}(n), g \mapsto \rho_g, \tag{1}$$

where $\mathcal{U}(n)$ is the group of unitary[2] $n \times n$ matrices. If a SDP problem is ρ-invariant, it has a Hermitian matrix X with $X = \rho_g X \rho_g^\dagger$ for all $g \in G$, so that X is in the *commutant* [25, Sec 1.7] C_ρ of the algebra generated by ρ:

$$C_\rho = \{X \in \mathbb{C}^{n\times n} : [X, \rho_g] = 0, \forall g \in G\}, \qquad [X, \rho_g] = \rho_g X - X\rho_g \ . \tag{2}$$

The decomposition of ρ into irreducible subrepresentations (irreps) corresponds to

$$V = V^1 \oplus \ldots \oplus V^N, \qquad V^i = W^{i,1} \oplus \ldots \oplus W^{i,M_i}, \tag{3}$$

where the *isotypic components* V^i regroup the subspaces $W^{i,j}$ corresponding to identical irreps. We write $D_i = \dim W^{i,j}$ the *dimension* of the i-th irrep in the decomposition and M_i its *multiplicity*. There is a unitary U such that $\forall g \in G$:

$$\hat{\rho}_g = U\rho_g U^\dagger = \begin{pmatrix} \hat{\rho}_g^1 & & \\ & \ddots & \\ & & \hat{\rho}_g^I \end{pmatrix}, \qquad \hat{\rho}_g^i = \begin{pmatrix} \hat{\rho}_g^{i,1} & & \\ & \ddots & \\ & & \hat{\rho}_g^{i,M_i} \end{pmatrix} = \mathbb{1}_{M_i} \otimes \hat{\rho}_g^{i,1}, \tag{4}$$

as $\hat{\rho}_g^{i,j} = \hat{\rho}_g^{i,k}, \forall j, k$. By Schur's lemma [26], the matrix X has the form:

$$\hat{X} = UXU^\dagger = \begin{pmatrix} \hat{X}^1 & & \\ & \ddots & \\ & & \hat{X}^N \end{pmatrix}, \qquad \hat{X}^i = \Xi^i \otimes \mathbb{1}_{D_i}, \tag{5}$$

where Ξ^i is a $M_i \times M_i$ Hermitian matrix, so $X \succeq 0$ is equivalent to $\{\Xi^i \succeq 0\}_i$.[3]

[1] Noting that a similar approach is currently pursued for conformal bootstrap [22, 23].

[2] RepLAB works with nonunitary representations of compact groups too as they can have numerical advantages. We keep this presentation simple by assuming unitarity.

[3] This saves on memory and CPU time requirements. Consider a SDP problem $\min_X \text{tr}[CX]$ such that X is SDP and $\text{tr}[A_i X] = b_i$ for $i = 1, \ldots, m$. Assume X has dimension $n \times n$ with blocks of size n_i so that $n = n_1 + \ldots + n_I$. The complexity of standard interior point primal-dual methods is as follows. For CPU time [27]: when $m \gg n$, the factoring of the Schur complement matrix dominates in $\mathcal{O}(m^3)$. When $m \lambda n$ (which we observe is the common case), Cholesky factorizations

The *block-diagonalization* of an invariant SDP rests on the following.

Problem 1 Given a description of a compact group G and its representation ρ, return the change of basis matrix U, the dimensions $\{D_i\}$, and the multiplicities $\{M_i\}$ such that (4) and (5) are satisfied.

RepLAB solves this problem as follows. The compact group G is given as an oracle that samples from the Haar measure on G; for finite groups, a permutation representation can be provided for faster computations. The representation ρ is provided as an image function. Our algorithm then returns a floating-point approximation of the change of basis matrix U along with $\{D_i\}$, $\{M_i\}$. Additional structure is discovered and returned for real representations (see next section).

1 A Tour of RepLAB Features

Let us consider a small toy problem, resulting from a symmetrized moment-based relaxation [28]. We compute the quantum upper bound on the CHSH expression $I = \left\langle \psi \middle| \sum_{xy} (-1)^{xy} A_x B_y \middle| \psi \right\rangle$, where $|\psi\rangle$ is a bipartite quantum state and $\{A_x\}$, $\{B_y\}$ are ± 1-valued Hermitian operators acting on the first and second subsystem, respectively, for $x, y \in \{0, 1\}$.

The SDP program corresponds to

$$\begin{array}{l} \max 4y \\ \text{s.t. } X = C + Ay \succeq 0 \end{array}, \qquad C = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & & 1 & 1 \\ & 0 & 1 & -1 \\ 1 & 1 & 0 & \\ 1 & -1 & & 0 \end{pmatrix}, \tag{6}$$

where $y \in \mathbb{R}$, X is Hermitian and the constraint $X \succeq 0$ means that X is SDP. This SDP is invariant under the signed permutation matrices ρ_{g_1} and ρ_{g_2}:

$$\rho_{g_1} = \begin{pmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 1 & \\ & & & & -1 \end{pmatrix}, \qquad \rho_{g_2} = \begin{pmatrix} 1 & & & & \\ & & & -1 & 0 \\ & & & 0 & -1 \\ & -1 & 0 & & \\ & 0 & -1 & & \end{pmatrix}, \tag{7}$$

corresponding to the action of the signed permutations [29] $g_1 = [1, 3, 2, 4, -5]$ and $g_2 = [1, -4, -5, -2, 3]$ where we write a signed permutation using the image on its domain $g = [g(1), g(2), g(3), g(4), g(5)]$. The domain is $i \in \{1, \ldots, 5\}$; by repeated composition, following the convention $(g\dot{h})(i) = g(h(i))$, these generate the group $G = \langle g_1, g_2 \rangle$. We now show how to block diagonalize ρ using RepLAB

and eigenvalue computations usually dominate, in $\mathcal{O}((n_1)^3 + \ldots (n_I)^3)$. For memory: the problem data scales in $\mathcal{O}(mn^2)$ in the worst-case, but often less due to sparsity. The Schur complement matrix requires $\mathcal{O}(m^2)$ storage, and the matrices X and χ require storage in $\mathcal{O}((n_1)^2 + \ldots + (n_I)^2)$. When using our technique, the block-diagonalization of a SDP of size $n \times n$ produces a SDP with blocks of size $n_i' = M_i$.

running under either the MATLAB or Octave environment (the latest version is always available at https://github.com/replab/replab). Part of the transcript is truncated due to page limits, and // indicates a newline.

```
>> g1 = [1 3 2 4 -5]; g2 = [1 -4 -5 -2 -3];
>> G = replab.signed.Permutations(5).subgroup({g1 g2})
replab.signed.PermutationSubgroup
  domainSize: 5
  generator(1): [1, 3, 2, 4, -5]   //   generator(2): [1, -4, -5, -2, -3]
```

The *natural representation* represents G using signed permutation matrices:

```
>> rep = G.naturalRep
Orthogonal representation by images
  dimension: 5 // field: 'R' // group: replab.signed.PermutationSubgroup
  images_internal{1}: [1, 0, 0, 0, 0; 0, 0, 1, 0, 0; ...
  images_internal{2}: [1, 0, 0, 0, 0; 0, 0, 0, -1, 0; ...
```

and we ask RepLAB to decompose that representation

```
>> dec = rep.decomposition
Orthogonal similar representation (irreducible decomposition)
    dimension: 5 // field: 'R' // group: replab.signed.PermutationSubgroup
  basis.(:,1): [1, 0, 0, 0, 0].'
  basis.(:,2): [0, 0.5, 0.5, 0.70711, -1.5403e-37].'
  basis.(:,3): [0, 0.5, -0.5, 1.6492e-36, 0.70711].'
  basis.(:,4): [0, 0.5, -0.5, 1.3109e-36, -0.70711].'
  basis.(:,5): [0, 0.5, 0.5, -0.70711, -4.9235e-37].'
 component(1): Isotypic component R(1) (trivial)
 component(2): Isotypic component R(2)
 component(3): Isotypic component R(2)
```

The change of basis matrix is $U = \text{dec.basis}^\dagger$, it diagonalizes A and thus X:

$$U = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & \sqrt{2} & 0 \\ 0 & 1 & -1 & 0 & \sqrt{2} \\ 0 & 1 & -1 & 0 & -\sqrt{2} \\ 0 & 1 & 1 & -\sqrt{2} & 0 \end{pmatrix} /2, \qquad UAU^\dagger = \begin{pmatrix} 0 & & & & \\ & \sqrt{2} & & & \\ & & \sqrt{2} & & \\ & & & -\sqrt{2} & \\ & & & & -\sqrt{2} \end{pmatrix}, \tag{8}$$

with $D_1 = 1$, $D_2 = D_3 = 2$, and $M_{1,2,3} = 1$. The SDP problem (6) becomes $\max 4y$, such that $1 \pm \sqrt{2}y \geqslant 0$. This is easily solved for the maximum $4y^\star = 2\sqrt{2}$.

The transcript above emphasizes the choices we made in building RepLAB: we reuse plain types when possible (e.g., integer vectors to represent permutations), the library is object-oriented (so operations are discovered easily), and results are returned in floating-point arithmetic.

1.1 Unique Features

RepLAB supports compact groups. We decompose the representation $\rho : u \mapsto \rho_u = u \otimes u^*$ of the unitary group $\mathcal{U}(2)$ (u^* is the complex conjugate). The Bell state $\left|\Phi^+\right\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ spans the trivial representation, and its orthogonal subspace is another subrepresentation.

```
>> U = replab.U(2);         % creating the unitary group
>> defRep = U.definingRep; % instance of its defining representation
>> rep = kron(defRep, conj(defRep)) % representations can be manipulated
Unitary tensor representation
```

```
>> dec = rep.decomposition
Complex similar representation (irreducible decomposition)
     dimension: 4 // field: 'C' // group: 2 x 2 unitary matrices
   basis.(:,1): [-0.49064+0.50919i, 5.3749e-17+1.4048e-17i, ...
...
   basis.(:,4): [-0.69962+0.10261i, 1.8576e-17-1.6639e-17i, ...
  component(1): Isotypic component C(1) (trivial)
  component(2): Isotypic component C(3)

>> dec.nice % we ask for rational basis recovery
Complex similar representation (irreducible decomposition)
   ...
   basis.(:,1): [1, 0, 0, 1].'
   basis.(:,2): [0, 0, 1, 0].'
   basis.(:,3): [0, 1, 0, 0].'
   basis.(:,4): [1, 0, 0, -1].'
```

Most SDP solvers encode complex-valued problems using real matrices with overhead. It is thus more efficient to block diagonalize real SDPs over the real field. RepLAB identifies the three types of division algebras that appear in irreducible real representations (real-type `'R'`, complex-type `'C'`, quaternion-type `'H'`). Below we demonstrate a cyclic group and the quaternion group.

```
>> C3 = replab.S(3).cyclicSubgroup; decC3 = C3.naturalRep.decomposition
Orthogonal similar representation (irreducible decomposition)
     dimension: 3 // field: 'R' // ...
  component(1): Isotypic component R(1) (trivial)
  component(2): Isotypic component (harmonized) C(2)

>>> qminus = [-1 -2 -3 -4]; qi = [2 -1 4 -3]; qj = [3 -4 -1 2];
>>> Q = replab.signed.Permutations(4).subgroup({qminus qi qj});
>>> decQ = Q.naturalRep.decomposition
Orthogonal irreducible similar representation (irreducible decomposition)
     dimension: 4 // field: 'R' // ...
  component(1): Isotypic component (harmonized) H(4)
```

RepLAB standardizes the matrix encodings of the complex and quaternion division algebras: for example, `decC3.component(2).sample` returns the representation of a random group element, of the form `[a -b; b a]` for some `a` and `b`.

RepLAB contains out-of-the-box support for a variety of groups. Those groups can be combined using standard constructions (direct, semidirect, and wreath products, which are relevant in nonlocality [30]). RepLAB is compatible with the free MATLAB clone Octave [31]. It extends the toolbox YALMIP [32] by providing invariant SDP variables and constraints. It can also block diagonalize data in an extension of the SeDuMi SDP solver format [33].

2 Algorithm

The decomposition of ρ implies four subtasks. First, we sample generic elements from the commutant C_ρ. Second, we compute the eigendecomposition of a generic Hermitian matrix from C_ρ to obtain invariant subspaces. Third, we find which irreducible subrepresentations are equivalent and harmonize their bases. Fourth, only for real subrepresentations, we identify the representation type and express its division algebra in a standard encoding.

2.1 Sampling from the Commutant Algebra

We first obtain a generic element of the commutant algebra. Compared to other approaches [34–36], we sample first a generic matrix X from the Gaussian Unitary Ensemble [37]. This ensures that the distribution is invariant under unitary changes of basis, and provides guarantees on eigenvalue separation. We then project X on the commutant subspace. We compute $\overline{X} = \int_G \rho_g X \rho_g^{-1} \mathrm{d}\mu(g)$, where μ is the Haar measure of G. This simplifies to $\overline{X} = \frac{1}{|G|} \sum_{g \in G} \rho_g X \rho_g^{-1}$ for finite groups. We thus reduced the problem of sampling from C_ρ to the problem of projecting on C_ρ.

To average X over a compact group, we work with an oracle that samples elements from the Haar measure μ. Given an integer ν, we sample sets $T_1, \ldots, T_\nu$ from μ, typically with $|T_i| = 3$. We then define $\overline{X}_\nu = \Sigma_{\rho,T_\nu}[\Sigma_{\rho,T_{\nu-1}}[\ldots \Sigma_{\rho,T_1}[X]]]$, and observe that $\overline{X}_\nu \to \overline{X}$ when $\nu \to \infty$. We monitor convergence by the Frobenius norm $\Delta_i = \left| |\overline{X}_{i+1} - \overline{X}_i \right|_{\mathrm{FRO}}$, and fit $\Delta_i \approx \nu + \alpha \exp(-\beta i)$ (exponential convergence on top of numerical noise floor ν). We thus reduced the subtask of sampling from C_ρ to the task of sampling from G itself.

For finite groups, we use a trick from [38] where fixed sets $\{T_i\}$ are computed such that $\overline{X}_\nu = \overline{X}$ exactly for finite ν. Of course, one can always set $\nu = 1$ and $T_1 = G$, but in practice much better decompositions are obtained.[4]

2.2 Computing the Eigendecomposition of a Hermitian Matrix

We decompose numerically $\overline{X}' = \overline{X} + \overline{X}^\dagger$ into $\hat{X} = U \overline{X}' U^\dagger$, getting a change of basis matrix U such that $\hat{X}$ is diagonal. After eigenvalue reordering, $\hat{X}$ has the form (5) with fully diagonal blocks Ξ^i. Due to genericity, no eigenvalue repeats inside each Ξ^i and across them. We thus group equal eigenvalues and denote the basis of the corresponding eigenspaces by the matrices $U^1, \ldots, U^M$. Then

$$\sigma^i : g \mapsto \sigma^i_g = U^i \rho_g (U^i)^\dagger \tag{9}$$

are irreducible subrepresentations of ρ. The computational cost of this step is $\mathcal{O}(n^3)$ if n is the dimension of ρ.

[4] For example, every element of the symmetric group S_D can be written uniquely as a product of powers of the cycles $(1, 2)$, $(1, 2, 3)$, $\ldots$, $(1, \ldots, D)$, reducing the computational effort from $\mathcal{O}(|G|) = \mathcal{O}(D!)$ to $\mathcal{O}(D^2)$ image computations.

2.3 Grouping Equivalent Representations

We take another sample $\overline{X}''$ from C_ρ and apply the following.

Proposition 1 *Given two basis matrices U^i and U^j, let σ^i and σ_j be two subrepresentations of ρ according to (9). Let $\overline{X}''$ be sampled from C_ρ, and $F = U^i \overline{X}'' (U^j)^\dagger$. If $F = 0$, then (almost always) σ^i and σ^j are inequivalent. Otherwise, σ^i and σ^j are equivalent, $\sigma^i_g = F \sigma^j_g F^{-1}$, with αF unitary for some $\alpha \in \mathbb{C}$.*

Proof Noting that $(U^i)^\dagger U^i$ is a projector on the corresponding invariant subspace, we verify that F is an equivariant map: $\sigma^i_g F = F \sigma^j_g$ for all $g \in G$. We now use Schur's lemma [26, Prop. 4]. By the assumption of genericity, $F = 0$ happens only when F has to be zero, and σ^i is inequivalent to σ^j. Otherwise, there is a unitary [39] matrix A such that $\sigma^i_g = A \sigma^j_g A^{-1}$, and thus $\sigma^j_g (A^{-1} F) = (A^{-1} F) \sigma^j_g$. By Schur's lemma $A^{-1} F = \alpha \mathbb{1}$, and we have necessarily $F = \alpha A$ with $\alpha \in \mathbb{C}$.

Checking all pairs (i, j), we group the bases U^i into isotypic components and harmonize bases with A. We get the dimensions $\{D_i\}$ and multiplicities $\{M_i\}$.[5]

2.4 Post-processing Real Representations

Let σ be a real irreducible representation of G of unknown division algebra type. We sample a generic non-Hermitian element $\overline{X}$ of C_σ.[6] We compute the complex matrix $Z = (\overline{X} + \overline{X}^\top)/2 + i(\overline{X} - \overline{X}^\top)/2$, a generic Hermitian element of the commutant of the complexification of σ. If Z has a single eigenvalue with multiplicities, then σ is of real-type: it does not decompose further over $\mathbb{C}$. Otherwise, Z is either of complex-type or quaternion-type. We will describe in a future publication how we recognize quaternion-type representations and put the basis of the division algebra in a standard form.

3 Other Approaches to Solve Problem 1

Algebraic Approaches The GAP [40] package AREP [41] returns exact algebraic solutions when ρ is a monomial representation and G a finite solvable group, using

[5] Note that the same proposition can be adapted to decide whether two irreducible representations of G, $\sigma^1 : G \to \mathcal{U}(n_1)$ and $\sigma^2 : G \to \mathcal{U}(n_2)$ are equivalent, and computing the change of basis matrix between them. This problem, considered in [39] for finite groups only, can be solved for compact groups by applying the proposition above to $\rho : g \mapsto \rho_g = \sigma^1_g \oplus \sigma^2_g$.

[6] Usually, σ is a subrepresentation of some ρ, so that we sample from C_ρ and restrict.

algebraic techniques. However, equivalent irreps are not always grouped nor put in the same basis, and AREP works only over the complex numbers.

Character Table Approaches It is possible to use the character table [42–44] of a finite group to decompose a generic representation ρ: by expanding the character $\chi : g \mapsto \mathrm{tr}[\rho_g]$ as a linear combination of irreducible characters, one can construct a block-diagonal representation ρ' similar to ρ. However, this approach does not provide directly the change of basis matrix. GAP itself can compute the character table and irreducible representations, but does not provide the change of basis matrix that reveals the irreducible decomposition.[7] These first two approaches have the advantage of providing exact algebraic solutions. However, they do not apply to compact groups, and depend on a large library of algorithms and data tables. In comparison, RepLAB is lightweight and readily integrates with the MATLAB/Octave ecosystem related to semidefinite optimization.

Approaches Based on $*$-Algebra Decomposition By observing that the commutant algebra C_ρ is a $*$-algebra [36], it can be decomposed using the Artin–Wedderburn theorem; this has been discussed several times in the context of invariant SDP programs [35, 36, 45]. We found an incomplete implementation of the algorithm of [35] in the software package NCSOSTools [46]. However, the paper [35] prescribes the use of exotic matrix decomposition techniques, some of which do not have implementations available.

The approach used in RepLAB directly inspired by this last class of approaches; we will describe our technical improvements in a future publication. Note that RepLAB does not aim to replace a Computer Algebra System such as GAP System; rather, it focuses on the numerical symmetrization of optimization problems.

4 Conclusion

We presented a toolbox to numerically decompose arbitrary finite dimensional representation of compact groups. Surprisingly, the user needs to provide only little information about the group and its representation, and there is no need to compute much structure to accomplish our task. The current unoptimized code is able to deal with representations of size up to $\sim$10,000. We foresee RepLAB having impact in two ways. The first one is reducing the computational cost of solving SDPs, thus expanding the applicability of a wide variety of quantum information methods. The second one is pedagogical: by delegating all computations to the software, a hands-on approach to representation theory can be taught, focusing

[7] Note that for groups of small order, the approach discussed in [26] could give the change of basis matrix. But note that the authors of [23] remarked that the exact algorithms of GAP were sometimes slow and restricted their decompositions to groups of order <100 in their symmetrization of conformal bootstrap.

on the physics by working on concrete examples right at the start. Along the same line, RepLAB can be used to quickly check whether an algebraic analysis of the symmetries of a problem is worthwhile: while the bases returned are approximate, the dimensions and multiplicities of the irreducible subrepresentations are themselves not approximate.

Still, this manuscript presents an early version of our package, and it presents several areas for improvement. Can we obtain better scaling factors for the algorithm of Sect. 2? Currently, our running time is dominated by the eigendecomposition step, which has running time similar to a single solver iteration on the original SDP. Another route is to use the black-box algorithms above in the last resort, and exploit structural information: RepLAB has built in support for a dozen standard group and representation constructions, but is not exploiting that structure currently. Also, RepLAB works in hardware double floating-point precision, and uses a tolerance of $\varepsilon = 10^{-10}$ in all convergence tests. We are currently analyzing error propagation in all the steps of our algorithm, and will replace those tolerances by proper bounds. Finally, RepLAB performs limited exact solution recovery: can we extend those recovery methods?

Acknowledgments We acknowledge useful discussions with David Gross, Elie Wolfe, and Markus Heinrich. This research was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. FMM was funded by the DFG project number 4334.

Note Added During the review of this manuscript, we became aware of the publication of the package RepnDecomp [47], integrated in the release 4.11.0 of GAP (March 2020).

References

1. A. Szanton, in *The Recollections of Eugene P. Wigner*, ed. by A. Szanton (Springer US, Boston, MA, 1992), pp. 115–125. https://doi.org/10.1007/978-1-4899-6313-0_8
2. D. Rosset, J.D. Bancal, N. Gisin, J. Phys. A: Math. Theor. **47**(42), 424022 (2014). https://doi.org/10.1088/1751-8113/47/42/424022
3. N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, Rev. Mod. Phys. **86**(2), 419 (2014). https://doi.org/10.1103/RevModPhys.86.419
4. C. Śliwa, Phys. Lett. A **317**(3–4), 165 (2003). https://doi.org/10.1016/S0375-9601(03)01115-0
5. M. Mozrzymas, M. Horodecki, M. Studziński, J. Math. Phys. **55**(3), 032202 (2014). https://doi.org/10.1063/1.4869027
6. M. Studziński, M. Horodecki, M. Mozrzymas, J. Phys. A: Math. Theor. **46**(39), 395303 (2013). https://doi.org/10.1088/1751-8113/46/39/395303
7. M. Mozrzymas, M. Studziński, M. Horodecki, J. Phys. A: Math. Theor. **51**(12), 125202 (2018). https://doi.org/10.1088/1751-8121/aaad15

8. M. Mozrzymas, M. Horodecki, M. Studziński, J. Math. Phys. **55**(3), 032202 (2014). https://doi.org/10.1063/1.4869027
9. M. Studziński, P. Ćwikliński, M. Horodecki, M. Mozrzymas, Phys. Rev. A **89**(5), 052322 (2014). https://doi.org/10.1103/PhysRevA.89.052322
10. D. Rosset, D. Schmid, F. Buscemi (in preparation)
11. M.D. Choi, Linear Algebra Appl. **10**(3), 285 (1975). https://doi.org/10.1016/0024-3795(75)90075-0
12. A. Jamiołkowski, Rep. Math. Phys. **3**(4), 275 (1972). https://doi.org/10.1016/0034-4877(72)90011-0
13. A.C. Doherty, P.A. Parrilo, F.M. Spedalieri, Phys. Rev. Lett. **88**(18), 187904 (2002). https://doi.org/10.1103/PhysRevLett.88.187904
14. A.C. Doherty, P.A. Parrilo, F.M. Spedalieri, Phys. Rev. A **69**(2), 022308 (2004). https://doi.org/10.1103/PhysRevA.69.022308
15. A.C. Doherty, P.A. Parrilo, F.M. Spedalieri, Phys. Rev. A **71**(3), 032333 (2005). https://doi.org/10.1103/PhysRevA.71.032333
16. M. Navascués, M. Owari, M.B. Plenio, Phys. Rev. Lett. **103**(16), 160404 (2009). https://doi.org/10.1103/PhysRevLett.103.160404
17. M. Navascués, S. Pironio, A. Acín, New J. Phys. **10**(7), 073013 (2008). https://doi.org/10.1088/1367-2630/10/7/073013
18. M. Navascués, S. Pironio, A. Acín, Phys. Rev. Lett. **98**(1), 010401 (2007). https://doi.org/10.1103/PhysRevLett.98.010401
19. M. Navascués, S. Pironio, A. Acín, in *Handbook on Semidefinite, Conic and Polynomial Optimization*, ed. by M.F. Anjos, J.B. Lasserre. International Series in Operations Research & Management Science, vol. 166 (Springer US, New York, 2012), pp. 601–634. https://doi.org/10.1007/978-1-4614-0769-0_21
20. M. Navascués, A. Feix, M. Araújo, T. Vértesi, Phys. Rev. A **92**(4), 042117 (2015). https://doi.org/10.1103/PhysRevA.92.042117
21. E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, M. Navascues (2019). arXiv:1909.10519 [quant-ph]
22. D. Poland, S. Rychkov, A. Vichi, Rev. Mod. Phys. **91**(1), 015002 (2019). https://doi.org/10.1103/RevModPhys.91.015002
23. M. Go, Y. Tachikawa, J. High Energ. Phys. **2019**(6), 84 (2019). https://doi.org/10.1007/JHEP06(2019)084
24. K. Gatermann, P.A. Parrilo, J. Pure Appl. Algebra **192**(1–3), 95 (2004). https://doi.org/10.1016/j.jpaa.2003.12.011
25. B. Sagan, *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*, 2nd edn. Graduate Texts in Mathematics (Springer, New York, 2001). https://doi.org/10.1007/978-1-4757-6804-6
26. J.P. Serre, *Linear Representations of Finite Groups*. Graduate Texts in Mathematics (Springer, New York, 1977)
27. B. Borchers, J.G. Young, Comput. Optim. Appl. **37**(3), 355 (2007). https://doi.org/10.1007/s10589-007-9030-3
28. D. Rosset, arXiv:1808.09598 [quant-ph] (2018)
29. A. Young, Proc. Lond. Math. Soc. **2**(1), 255 (1928)
30. M.O. Renou, D. Rosset, A. Martin, N. Gisin, J. Phys. A: Math. Theor. **50**(25), 255301 (2017)
31. B. Röthlisberger, J. Lehmann, D. Loss, Comput. Phys. Commun. **183**(1), 155 (2002). https://doi.org/10.1016/j.cpc.2011.08.012
32. J. Lofberg, in *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)* (2004), pp. 284–289. https://doi.org/10.1109/CACSD.2004.1393890
33. J.F. Sturm, Optim. Methods Softw. **11**(1–4), 625 (1999). https://doi.org/10.1080/10556789908805766
34. A. Ibort, A. López-Yela, J. Moro (2016). arXiv:1610.01054 [math-ph]
35. T. Maehara, K. Murota, Jpn. J. Ind. Appl. Math. **27**(2), 263 (2010). https://doi.org/10.1007/s13160-010-0007-8

36. K. Murota, Y. Kanno, M. Kojima, S. Kojima, Jpn. J. Ind. Appl. Math. **27**(1), 125 (2010). https://doi.org/10.1007/s13160-010-0006-9
37. G.W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, 1st edn. (Cambridge University Press, New York, 2009)
38. L. Babai, K. Friedl, in *Proceedings 32nd Annual Symposium of Foundations of Computer Science* (1991), pp. 733–742. https://doi.org/10.1109/SFCS.1991.185442
39. M. Mozrzymas, M. Studziński, M. Horodecki, J. Phys. A: Math. Theor. **47**(50), 505203 (2014). https://doi.org/10.1088/1751-8113/47/50/505203
40. *GAP – Groups, Algorithms, and Programming, Version 4.7.8* (The GAP Group, 2015)
41. S. Egner, M. Püschel
42. G.J.A. Schneider, J. Symb. Comput. **9**(5), 601 (1990). https://doi.org/10.1016/S0747-7171(08)80077-6
43. S.B. Conlon, J. Symb. Comput. **9**(5), 535 (1990). https://doi.org/10.1016/S0747-7171(08)80072-7
44. U. Baum, M. Clausen, Math. Comput. **63**(207), 351 (1994). https://doi.org/10.2307/2153580
45. F. Vallentin, Linear Algebra Appl. **430**(1), 360 (2009). https://doi.org/10.1016/j.laa.2008.07.025
46. K. Cafuta, I. Klep, J. Povh, Optim. Methods Softw. **26**(3), 363 (2011). https://doi.org/10.1080/10556788.2010.544312
47. K. Hymabaccus, RepnDecomp: A GAP package for decomposing linear representations of finite groups. https://joss.theoj.org/papers/f260c31c9befafe72f990ced039c092a

Correction to: Leptophobic Z' in Supersymmetry and Where to Find Them

Correction to:
Chapter 52 in: M. B. Paranjape et al. (eds.),
***Quantum Theory and Symmetries*,**
CRM Series in Mathematical Physics,
https://doi.org/10.1007/978-3-030-55777-5_52

The original version of this chapter was inadvertently published with wrong figures (Fig. 1 to Fig. 4). Now, the correct figures (Fig. 1, Fig. 2, Fig. 3 and Fig. 4) have been updated in this chapter.
Also, the correct figures are given below:

The updated online version of this chapter can be found at
https://doi.org/10.1007/978-3-030-55777-5_52

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5_61

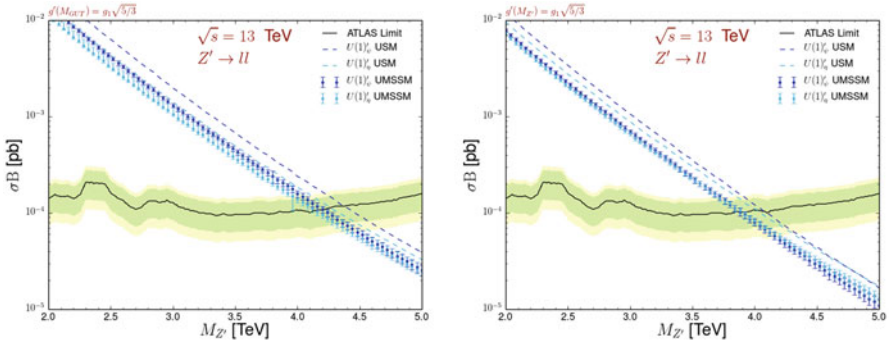


Fig. 1 The effect of symmetry breaking scale in both inclusive (UMSSM) and noninclusive (USM) supersymmetric particle cases is presented. The left panel shows Z' mass versus its corresponding cross section times branching ratio to dilepton final states where SUSY breaking occurring at GUT scale. The dotted line represents UMSSM and dashed line USM realisations, and the error bars show the variation caused by theoretical uncertainties. The right panel shows the same for SUSY breaking occurring at Z' mass scale. These results are presented in [11]

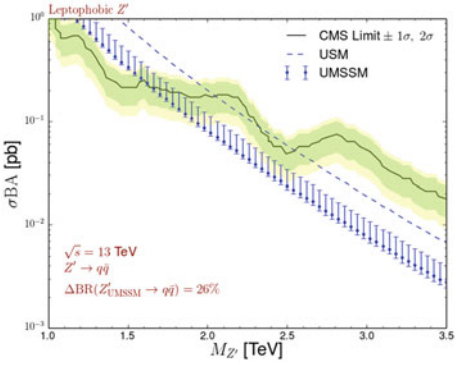


Fig. 2 Plot shows leptophobic Z' mass versus production cross section and dijet branching ratio with corresponding acceptance rate. The dashed line shows the USM sample, which assumes that Z' can only decay through the SM particles, and the dotted line shows UMSSM, including all supersymmetric particles. Error bars include scale and PDF variations as well as the variation in the dijet decay rate per sample. The CMS limit is taken from [3]

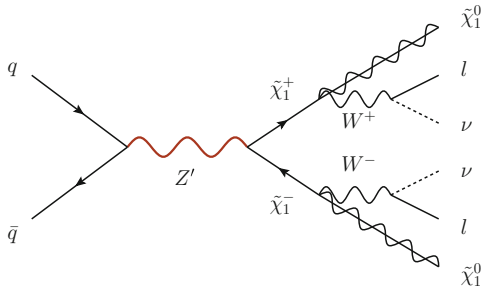


Fig. 3 Chosen analysis signal where Z' production and its decay through two lightest charginos. The figure has been produced with the help of the JaxoDraw package [17]

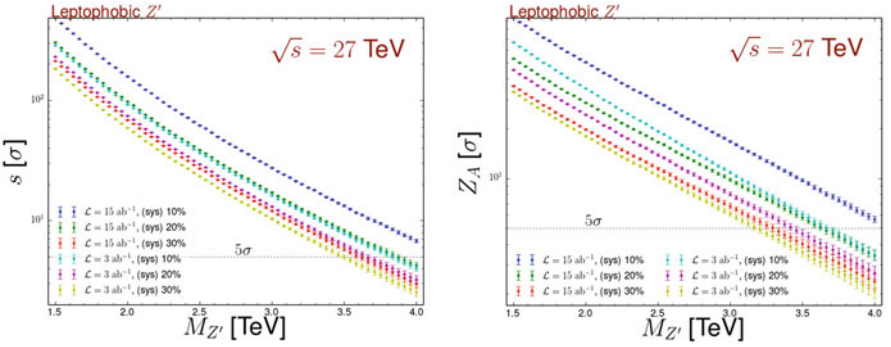


Fig. 4 Left panel shows the standard significance against Z' mass plotted for various luminosity and systematic uncertainty values. Right panel shows the same figure for Asimov significance, Z_A. These results are presented in [27]

Index

M. B. Paranjape et al. (eds.), *Quantum Theory and Symmetries*, CRM Series in Mathematical Physics, https://doi.org/10.1007/978-3-030-55777-5

R