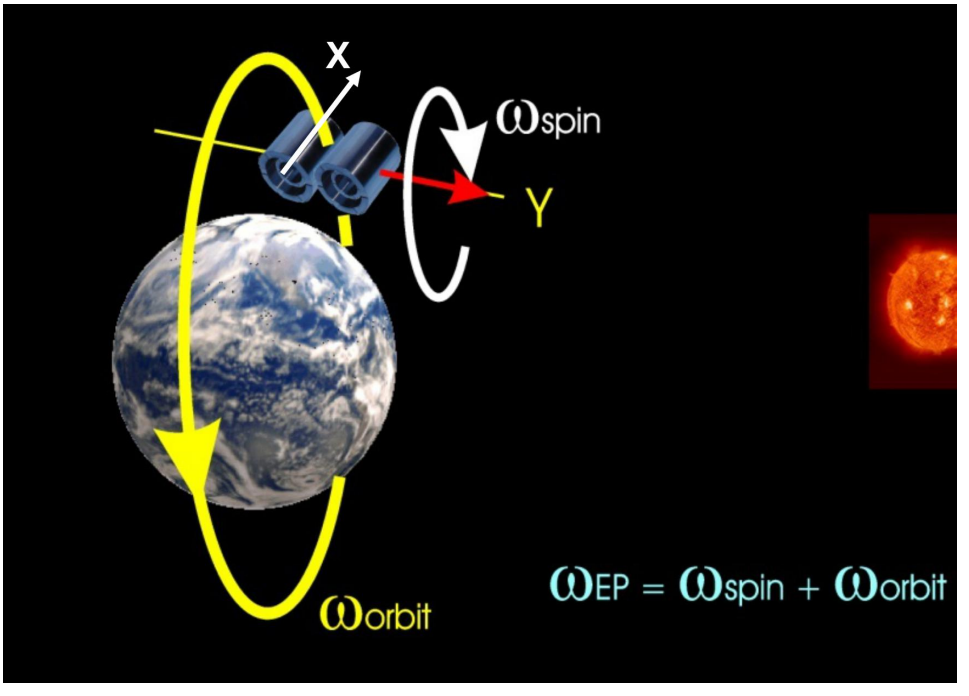
Comptes Rendus
de l'Académie des sciences

1878-1535 (electronic)

Physique



Volume 21, Special Issue 2, juin 2020

Special issue / *Numéro thématique*

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Académie des sciences — Paris

INSTITUT DE FRANCE
Académie des sciences

CENTRE MERSENNE

Comptes Rendus

Physique

Objective of the journal

Comptes Rendus Physique is a peer-reviewed electronic journal of international standing, covering all fields of physics and astrophysics. It publishes mainly thematic issues, but also original research articles, preliminary announcements, review articles, historical perspectives, pedagogical texts or conference proceedings, without length limit, in English or in French. It also publishes special issues devoted to certain recent and/or significant aspects of the discipline, whose authors are chosen from among the most active researchers on the subject and whose coordination is assured by guest editors.

Comptes Rendus Physique is published according to a virtuous policy of diamond open access, free for authors (no publication fees) as well as for readers (immediate and permanent open access).

Editorial director: Étienne Ghys

Editors-in-Chief: D. Gratias, J. Villain

Editorial Board: Jacqueline Bloch, Christian Bordé, Hélène Bouchiat, Alexandre Bouzdine, Yves Bréchet, Françoise Combes, Jean Dalibard, Michel Davier, Daniel Estève, Stéphan Fauve, Pierre Fayet, Frédérique de Fornel, Maurice Goldman, Guy Laval, Chaouqi Misbah, Jean-Yves Ollitrault, Nathalie Palanque-Delabrouille

Editorial secretary: Julien Desmarets

About the journal

All journal's information, including the text of published articles, which is fully open access, is available from the journal website at https://comptes-rendus.academie-sciences.fr/physique/.

Author enquiries

For enquiries relating to the submission of articles, please visit this journal's homepage at https://comptes-rendus.academie-sciences.fr/physique/.

Contact

Académie des sciences
23, quai de Conti, 75006 Paris, France
Tel: (+33) (0)1 44 41 43 72
CR-Physique@academie-sciences.fr

The articles in this journal are published under the license Creative Commons Attribution 4.0 International (CC-BY 4.0) https://creativecommons.org/licenses/by/4.0/deed.en

Comptes Rendus
Physique
2020, 21, nº 2

Contents / *Sommaire*

Jacques Villain
Foreword 135-137

Pierre Touboul, Gilles Métris, Manuel Rodrigues, Yves André, Alain Robert
The MICROSCOPE space mission to test the Equivalence Principle 139-150

Michael Le Bars, Louis-Alexandre Couston, Benjamin Favier, Pierre Léard, Daniel Lecoanet, Patrice Le Gal
Fluid dynamics of a mixed convective/stably stratified system—A review of some recent works 151-164

Pierre Delplace, Antoine Venaille
From the geometry of Foucault pendulum to the topology of planetary waves 165-175

Maria J. Esteban
Mathematical questions about the computation of eigenvalues of Dirac operators with critical potentials in atomic and molecular physics 177-183

Slava Rychkov
3D Ising model: a view from the conformal bootstrap island 185-198

Yves Colin de Verdière
Des bruits et des ondes: mathématiciens et physiciens en interaction 199-202

Comptes Rendus
Physique
2020, 21, nº 2, p. 135-137
https://doi.org/10.5802/crphys.26

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Foreword

Jacques Villain [a]

[a] Theory Group, Institut Laue Langevin, F-38054 Grenoble Cedex 9, France
E-mail: jvillain@infonie.fr

This issue brings together invited contributions, mainly from the winners of prizes of the Academy of Sciences. Their purpose was to explain their research to non-specialists.

Most of these non-specialists are convinced of the equivalence of gravitational mass and inertial mass, in other words it is the same mass m which intervenes in the expression of the force $f = mm'/r^2$ between two masses m and m' at distance r, and in the relation $f = ma$ between the force f and the acceleration a. Yet this is an important question in the very current problem of the quantification of gravity and general relativity. Pierre Touboul, Manuel Rodrigues, Gilles Métris and Yves André were able to demonstrate this equivalence experimentally with a hitherto unattained precision of 10^{-14} and for this reason received the Servant prize in 2019. This determination requires, as they explain in their article co-authored with Alain Robert, the use of a satellite and a very precise analysis of the various forces involved.

Long before artificial satellites circled the earth, there were winds that did so too. Notably stratospheric winds which, curiously, change direction about once a year (more precisely the period is about 28 months). This phenomenon was reproduced by numerical simulations of Michael LeBars, who was awarded the Leconte prize for this reason. In his article, with his co-authors he explains the phenomenon, which is due to the interaction of the stratosphere with the turbulent troposphere below. As they write, "the rapid small-scale turbulence excites waves at intermediate scale, which propagate and interact non-linearly to generate large-scale circulations." The authors have also experimentally investigated in the laboratory some properties of such two-layer stratified media.

Atmospheric waves (but also oceanic and even seismic) are also central to the article by Pierre Delplace and Antoine Venaille. Their concern, however, is more mathematical; they show how geometric concepts commonly used in condensed matter physics (Berry phase, Chern number) can find application on the scale of the earth. The authors have not yet received a prize from the Académie des Sciences, but Pierre Delplace had given a very nice invited talk at this Academy, and we thought it would be useful to have a written version.

Maria Esteban's article is devoted to solving the Dirac equation, for example for an electron in an atom. She managed to circumvent one of the difficulties encountered in some methods, namely phantom (or spurious, as she denotes them) solutions that lack physical significance. For her work, the Jacques-Louis Lions Prize was awarded to Maria Esteban. It is a Mathematics award and the winner is a mathematician. We are happy that mathematicians are interested in physics!

Slava Rychkov's article is devoted to critical exponents of three-dimensional Ising model. Unlike those of the two-dimensional Ising model, they are not known exactly, but the author (Mergier-Bourdeix Prize 2019) describes a method that allows them to be determined with better precision than all the other methods. Older physicists will appreciate that Slava Rychkov gives, at the beginning of his article, a current definition of the word "emerging", which did not exist (to my knowledge) 40 years ago, and which is now very fashionable.

The last article is also the work of a mathematician who is interested in physics: Yves Colin de Verdière (Emile Picard medal 2018) briefly describes his collaborations with physicists; more precisely a geophysicist specializing in earthquakes, and specialists in fluid mechanics.

We thank the authors for agreeing to contribute to this issue, and also for submitting their contribution early. Other laureates sent their contribution later, or the referees' reports were a bit late, so their articles will appear in a later issue of C.R. Physique.

Avant-propos

Ce fascicule réunit principalement des contributions invitées des lauréats de prix de l'Académie des Sciences.[1] Leur tâche est d'expliquer leurs recherches à des non-spécialistes.

La plupart de ces non-spécialistes sont persuadés de l'équivalence de la masse gravitationnelle et de la masse inertielle, autrement dit c'est la même masse m qui intervient dans l'expression de la force $f = mm'/r^2$ entre deux masses m et m' à distance r, et dans la relation $f = ma$ entre la force f et l'accélération a. C'est pourtant une question importante dans le problème très actuel de la quantification de la gravitation et de la relativité générale. Pierre Touboul, Manuel Rodrigues, Gilles Métris et Yves André Pierre ont su démontrer expérimentalement cette équivalence avec une précision de 10^{-14} jamais atteinte encore et ont reçu pour cette raison le prix Servant 2019. Cette détermination nécessite, comme ils l'expliquent dans leur article dont Alain Robert est coauteur, l'utilisation d'un satellite et une analyse très précise des diverses forces qui interviennent.

Bien avant que des satellites artificiels tournent autour de la terre, il y avait des vents qui tournent aussi. Notamment des vents stratosphériques qui ont la curieuse propriété de changer de sens à peu près une fois par an (plus précisément la période est d'environ 28 mois). Ce phénomène a été reproduit par des simulations numériques de Michael LeBars, auquel pour cette raison le prix Leconte a été décerné. Avec ses coauteurs marseillais, dans son article il explique le phénomène, qui est dû à l'interaction de la stratosphère avec la troposphère turbulente qui se trouve au dessous. Comme ils l'écrivent, « la turbulence rapide à petite échelle excite des ondes à moyenne échelle, qui se propagent et interagissent non linéairement pour générer des circulations à grande échelle ». Les auteurs ont également étudié expérimentalement en laboratoire certaines propriétés de tels milieux stratifiés à deux couches.

Les ondes atmosphériques (mais aussi océaniques et même sismiques) sont aussi au centre de l'article de Pierre Delplace and Antoine Venaille. Leur préoccupation est cependant plus mathématique ; ils montrent comment des concepts géométriques habituellement utilisés en physique de la matière condensée (phase de Berry, nombre de Chern) peuvent trouver leur application à l'échelle de la terre. Les auteurs n'ont pas encore reçu de prix de l'Académie des Sciences, mais Pierre Delplace avait fait un très beau séminaire invité devant cette Académie, et nous avons pensé qu'il serait utile d'en avoir une version écrite.

L'article de Maria Esteban est consacré à la résolution de l'équation de Dirac, par exemple pour un électron dans un atome. L'une des difficultés, qu'elle parvient à éviter, est la présence dans

[1] https://www.academie-sciences.fr/fr/Ceremonies/premiere-remise-prix-2019.html
https://www.academie-sciences.fr/fr/Laureats/laureats-2019-des-prix-thematiques-premiere-ceremonie.html

certaines méthodes de résolution, de solutions fantômes sans existence physique (que l'autrice appelle *spurieuses*). Pour ses travaux, le prix Jacques-Louis Lions a été décerné à Maria Esteban. C'est un prix de Mathématiques et la lauréate est une mathématicienne. Réjouissons-nous que les mathématicien-ne-s s'intéressent à la physique !

L'article de Slava Rychkov est consacré aux exposants critiques du modèle d'Ising tridimensionnel. Contrairement à ceux du modèle d'Ising bidimensionnel ils ne sont pas connus exactement, mais l'auteur (prix Mergier-Bourdeix 2019) décrit une méthode qui permet de les déterminer avec une précision meilleure que toutes les autres méthodes. Les vieux physiciens apprécieront que Slava Rychkov donne, au début de son article, une définition du mot « émergent », qui dans le sens qu'il donne n'existait pas (à ma connaissance) il y a 40 ans, et qui est actuellement très à la mode.

Le dernier article est encore l'œuvre d'un mathématicien qui s'intéresse à la physique : Yves Colin de Verdière (médaille Emile Picard 2018) y décrit brièvement ses collaborations avec des physicien-ne-s ; plus précisément un géophysicien spécialiste des tremblements de terre, et des spécialistes de mécanique des fluides.

Nous remercions les auteurs d'avoir accepté de contribuer à ce numéro, et aussi d'avoir envoyé leur contribution tôt. D'autres lauréats nous ont envoyé une contribution plus tardive, ou les expertises ont un peu tardé, de sorte que leurs articles paraîtront dans un numéro ultérieur des C.R. Physique.

Jacques Villain
Editor-in-Chief
France
jvillain@infonie.fr

Comptes Rendus
Physique
2020, 21, nº 2, p. 139-150
https://doi.org/10.5802/crphys.24

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

The MICROSCOPE space mission to test the Equivalence Principle

La mission spatiale MICROSCOPE pour le test du Principe d'Equivalence

Pierre Touboul [a], Gilles Métris [*, b], Manuel Rodrigues [a], Yves André [c] and Alain Robert [d]

[a] ONERA, Université Paris Saclay, Chemin de la Hunière, BP 80100, F-91123 Palaiseau Cedex, France

[b] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur, 250 avenue Albert Einstein, F-06560 Valbonne, France

[c] CNES, 18 avenue Edouard Belin, F-31401 Toulouse, France

E-mails: pierre.touboul@onera.fr (P. Touboul), gilles.metris@oca.eu (G. Métris), manuel.rodrigues@onera.fr (M. Rodrigues), yves.andre@cnes.fr (Y. André), AlainJM.Robert@cnes.fr (A. Robert)

Abstract. The MICROSCOPE space experiment aimed to test the Equivalence Principle with a much better accuracy than ever before. Its principle is to compare the free fall of concentric test masses embedded in a space accelerometer onboard a satellite orbiting the Earth. The effect of non-gravitational forces on the motion of the satellite is strongly reduced thanks to the so-called drag-free system. MICROSCOPE ran from April 2017 until October 2019. The analysis of the first series of measurements leads to an improvement of about an order of magnitude on the accuracy of the test of the Equivalence Principle. No violation has been detected for the pair of masses in platinum and titanium at the level of 10^{-14}.

MICROSCOPE, proposed by ONERA and OCA as science leaders and developed by CNES as project manager, is the first European space mission dedicated to fundamental physics on low Earth orbit. ZARM, PTB and ESA are the main European contributors.

Résumé. La mission spatiale MICROSCOPE avait pour objectif de tester le Principe d'équivalence (PE) avec une précision bien meilleure que ce qui avait été fait jusqu'alors. Ce type de test a un enjeu important car, tandis que le PE est un pilier de la relativité générale, il n'est pas imposé par la plupart des théories alternatives visant à étendre la gravitation pour l'unifier avec les autres interactions de la physique. Fondamentalement l'expérience consiste à comparer les chutes libres de différentes masses. Pour des raisons de mise en œuvre, le mouvement des masses n'est pas libre mais contrôlé par un accéléromètre (la charge utile du satellite) et c'est la force électrostatique nécessaire à maintenir les masses au repos qui constitue la mesure. Plus précisément, on compare les forces par unité de masse exercées sur des masses concentriques et on recherche dans leur différence la signature d'une différence de comportement vis-à-vis de la gravité terrestre. L'avantage d'un

* Corresponding author.

test dans l'espace est de permettre une chute quasi-infinie et de minimiser de nombreuses perturbations environnementales. Le satellite est équipé de micro-propulseurs dont les poussées sont asservies pour d'une part contrebalancer les forces non-gravitationnelles et d'autre part maintenir une loi d'attitude très stable.

Le satellite MICROSCOPE a été lancé en avril 2016 et a fonctionné avec succès jusqu'en octobre 2018, date à laquelle sa passivation a été réalisée. De nombreuses sessions de mesures ont été réalisées, non seulement pour faire le test du PE dans différentes conditions mais aussi pour caractériser l'expérience et étalonner les instruments. Les analyses de l'ensemble des données est en cours de finalisation, mais les résultats obtenus à partir des toutes premières sessions de mesure apportent déjà un progrès important par rapport à l'état de l'art : une seule session de mesure a permis d'améliorer la précision du test d'un ordre de grandeur, concluant à l'absence de violation du PE pour le couple de matériaux platine-titane, au niveau de 10^{-14}. Une autre session de mesure comparant 2 masses de mêmes compositions (platine) a permis de vérifier l'absence de systématismes importants dans l'expérience puisque : aucune violation supérieure à 10^{-14} n'a été détectée.

Keywords. Equivalence principle, Space experiment, Satellite, Accelerometers, Drag free, Inertial mass, Gravitationnal mass.

Mots-clés. Principe d'équivalence, Expérience spatiale, Satellite, Accéléromètres, Compensation de trainée, Masse inertielle, Masse gravitationnelle.

2020 Mathematics Subject Classification. 83B05.

1. Introduction

At the beginning of the 17th century, as part of his work on falling bodies based on the observation of the descent of different balls on inclined planes, Galileo noted that these movements were identical regardless of the size and material composition of these bodies [1]. This result, in apparent contradiction with daily experience, was obtained thanks to the virtual elimination of the main non-gravitational disturbance constituted by dragging by the atmosphere and which affects bodies differently according to their cross section and their mass. Half a century later, Newton confirmed this universality of free fall by comparing the beatings of pendulums with various test masses and presented this universality as a consequence of the law of gravitation and Newton's second law [2]: the first states that the force of gravity which attracts bodies to each other is proportional to their mass (called gravitational in this case) m_G; the second also asserts that the resistance of a body to a modification of its movement (acceleration) by a given force is proportional to its inertial mass m_I. Therefore the acceleration resulting from the gravitational attraction is proportional to the ratio m_G/m_I:

$$F = G\frac{M_G m_G}{r^2} \quad \text{and} \quad a = \frac{F}{m_I} \Longrightarrow a = \frac{GM_G}{r^2}\frac{m_G}{m_I}. \tag{1}$$

Newton was well aware that these two concepts of mass were very different but the experimental verification of the universality of free fall led to the conclusion that the ratio m_G/m_I is the same for all bodies: the gravitational and inertial masses are proportional or equivalent. This equivalence is generally quantified by the Eötvös parameter:

$$\eta(A,B) = \frac{\dfrac{m_{G,A}}{m_{I,A}} - \dfrac{m_{G,B}}{m_{I,B}}}{\frac{1}{2}\left(\dfrac{m_{G,A}}{m_{I,A}} + \dfrac{m_{G,B}}{m_{I,B}}\right)} = \frac{a_A - a_B}{\frac{1}{2}(a_A + a_B)}. \tag{2}$$

The equivalence was checked more and more precisely in the following centuries to reach 10^{-9} at the turn of the 20th century, thanks to Eötvös' experiments using a torsion balance [3]. An experiment with much more efficient technology is currently running at Washington University in Seattle; it recorded the most precise determination, before MICROSCOPE, of the Eötvös parameter for laboratory bodies: they demonstrated that the value of this parameter was compatible

with 0 for the couples (Beryllium, Titanium) and (Beryllium, Aluminium) with order of accuracy of 2×10^{-13} [4]. Another type of experiment consists in comparing the movements (very close to free falls) of different celestial bodies. The most precise experiment of this type consists in comparing the movement of the Moon and the Earth in the gravitational field of the Sun, thanks to the very precise measurement of the Earth–Moon distance by means of laser telemetry [5]. For such massive bodies, another focus, in addition to a possible violation due to different compositions of the bodies, is the search for a violation related to the self-gravitational energy of the bodies (Nordvedt effect) [6]. The best result was obtained by [7]. The two types of experience are therefore highly complementary.

In 1907 Albert Einstein defined the Equivalence Principle (EP) [8] which is one of the pillars of his theory of gravitation, General Relativity (GR) [9]. The equivalence of the gravitational and inertial masses as well as the universality of free fall are consequences of the EP. GR has led to many new predictions which have all been verified since, such as for example the existence of black holes and gravitational waves confirmed in 2016 [10]. It remains, however, to unify the theory of gravitation, relevant to large scales, with quantum mechanics relevant to small scales. There are several theories seeking to fulfil this objective and, unlike GR, they most often allow a violation of the EP [11]. Thus, experiments that are able to detect whether or not the EP has been violated with a greater degree of precision, bring important landmarks for these theories.

Ground based experiments as cited above are limited by the disturbing environment or by the duration of the experiment. In space far from Earth vibrations and day/night temperature variations, and with much more measurement time, it is possible to compensate for the external disturbing forces and torques in order to provide the best environment ever in a low Earth orbit.

The first idea of an EP test in space was first proposed in the 1970's [12] and studied at the end of the 20th century at Stanford University [13] with the STEP mission. STEP comprised cylindrical accelerometers with electrostatic control for the start-up and SQUID detectors for the fine test-mass position measurement. STEP was a complex cryogenic mission proposal with an expected accuracy of 10^{-18} on the EP test. Taking advantage of the Myriade line of CNES micro-satellite, MICROSCOPE appeared at the beginning of the 21st century to be a simpler and quicker mission with an objective of testing the EP at the level of 10^{-15} [14,15], already a breakthrough with respect to current results.

In this paper we present the first space experiment which aims at testing the EP. The principle of the mission is described in Section 2 while brief descriptions of the satellite and of the instrument are presented in Sections 3 and 4. The very first results of the mission are summarised in Section 5.

2. The MICROSCOPE mission

MICROSCOPE aimed to test the Equivalence Principle with an unprecedented precision of 10^{-15}. The T-SAGE (Twin Space Accelerometers for Gravitation Experiment) scientific payload, provided by ONERA, was integrated within a CNES micro-satellite. It was launched and injected into a 710 km altitude, circular orbit, by a Soyouz launcher from Kourou on April 25, 2016. The orbit is sun-synchronous, dawn-dusk (i.e. the ascending node stays at 18h mean solar time) in order to have long eclipse-free periods (eclipses are defined as periods within the Earth's shadow and happen only between May and July).

In the spirit, the experiment aims to compare the free fall of several test masses orbiting the Earth. But, for practical reasons, the implementation is slightly more sophisticated and relies on two nested control loops.

The first loop is inside the payload T-SAGE constituted by 4 test masses grouped by pairs in two differential accelerometers. Each test mass is placed between pairs of electrodes and its motion

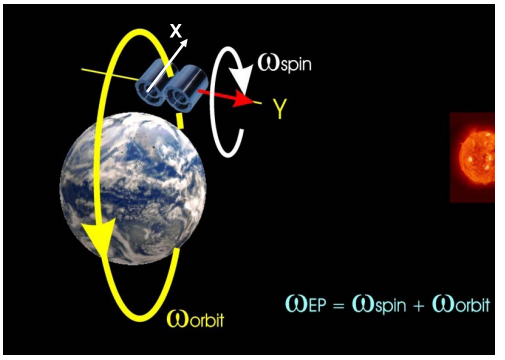


Figure 1. Configuration of the experiment: the most sensitive axis x (along the axis of the cylinders) of the accelerometers is maintained parallel to the orbital plane and rotates around the axis normal to the orbit.

with respect to its cage fixed to the satellite is monitored by capacitive sensors. This motion can then be controlled at rest by applying the appropriate electrostatic force calculated by a PID (Proportional Integral Derivative). This means that this electrostatic force compensates for all other forces. In that way, knowledge of the applied electrostatic potential allows us to measure the acceleration which would affect the test mass with respect to the satellite in the absence of the electrostatic force. That is why, in the following, we will use the terminology "measured acceleration" even if the masses are motionless with respect to the satellite.

The other major control loop in the MICROSCOPE experiment is included in the Drag-Free and Attitude Control System (DFACS) which applies accelerations on the satellite in order to cancel (or at least to considerably reduce), the level of the common mode measured acceleration. This is achieved by means of cold gas thrusters. This system also ensures a very accurate control of the pointing as well as the angular velocity and acceleration based on the measurements of angular position delivered by the stellar sensors and of the angular acceleration delivered by T-SAGE.

Even if T-SAGE measures the linear acceleration along the 3 axes, the measurement along the x-axis, which is also the axis of the cylindrical test masses, is the most accurate. This axis is controlled, thanks to the DFACS, parallel to the orbital plane and rotates with a frequency $f_{\rm spin}$ around the y-axis orthogonal to the orbital plane (Figure 1). In these conditions the component g_x of the Earth' gravity, and then the searched EP signal $\eta\, g_x$, varies with a very stable frequency $f_{\rm EP} = f_{\rm orb} + f_{\rm spin}$ where $f_{\rm orb}$ is the mean orbital frequency.

At first view, the main perturbation comes from the Earth's gravity gradient, due to the fact that, despite our best efforts in terms of manufacturing, the centres of mass of the test masses are not located in exactly the same place, but are separated by an off-centering $\vec{\Delta}$ of a few tens of micrometers. At the altitude of MICROSCOPE this leads to a differential acceleration of a few 10^{-11} m·s^{-2}, much larger than the accuracy of $10^{-15} \times (7.9\ \text{m·s}^{-2}) = 7.9 \times 10^{-15}$ m·s^{-2} targeted for the EP signal. But thanks to the careful design of the experiment (orbit close to circular, sensitive axis maintained in the orbital plane) this gravity gradient signal is mainly concentrated at DC and 2 $f_{\rm EP}$ frequencies and well decorralated from the EP signal. Moreover, the component Δ_y has a fully negligible impact whereas the components Δ_x and Δ_z can be accurately estimated in flight and the corresponding terms of the gradient can be corrected [16].

Up to negligible correcting terms, the model of the measured acceleration takes the simple form:

$$\Gamma^{(d)}_{x,\rm corr} = b^{(d)}_x + \delta_x g_x + \Delta_x S_{xx} + \Delta_z S_{xz} + k^{(d)}_x \Gamma^{(c)}_x + \theta^{(d)}_z \Gamma^{(c)}_y + \theta^{(d)}_y \Gamma^{(c)}_z + n^{(d)}_x, \tag{3}$$

where

- $\Gamma^{(d)}_{x,\mathrm{corr}}$ is the difference of the accelerations measured for the two quasi-concentric test masses along the x-axis,
- $b^{(d)}_x$ is the differential bias, mainly constant but also potentially including low frequency thermal effects,
- δ_x is very close to the Eötvös parameter η,
- g_x is the gravity acceleration projected along the x-axis,
- Δ_x and Δ_z are the components along x and z respectively, of the vector separating the two test masses,
- S_{xx} and S_{xz} are components of the gradient of acceleration along x due to differences of position along x and z respectively: $S_{xx} = T_{xx} + \Omega_y^2 + \Omega_z^2$ and $S_{xz} = T_{xz} - \Omega_x\Omega_z$ where $[T]$ is the gravity gradient tensor (the space derivative of the gravity acceleration) and $\vec{\Omega}$ is the angular velocity of the satellite,
- $\vec{\Gamma}^{(c)}$ is the common mode applied acceleration residue when the DFACS is operating,
- $k^{(d)}_x$ is the difference of scale factor related to the measurement process for the two test masses,
- θ_y and θ_z are the test mass relative misalignment around y and z,
- $n^{(d)}_x$ is the differential noise.

g_x, S_{xx} and S_{xz} can be computed very accurately from the known position and attitude (i.e. the orientation in space) of the satellite, whereas $b^{(d)}_x$, δ_x, Δ_x and Δ_z are parameters estimated during the data analysis process [17]. The common mode acceleration effect is also corrected thanks to calibration sessions which allow us to estimate $k^{(d)}_x$, θ_y and θ_z.

3. The satellite and the acceleration and attitude control system

3.1. *Satellite design*

The MICROSCOPE satellite was designed and developed by CNES as a space laboratory devoted to test the Equivalence Principle. The satellite points the instrument accurately along 3 directions, protects it against non-gravitational forces, and ensures an ultra-stable thermal and gravitational environment, with a very low level of microphonic or micro-acceleration disturbances. It is in close interaction with its payload, the science instrument developed by ONERA described in Section 4.

The satellite was derived from the Myriade micro-satellite product line and its architecture was based on DEMETER [18] and PARASOL, the first satellites of the Myriade family [19]. But the design of MICROSCOPE was extensively adapted for this mission (Figure 2).

The first design driver was the reduction of the mean level of applied accelerations on the instrument introducing a new function: the acceleration control (currently named drag-free control) [20]. This function uses the payload as an inertial sensor to measure the external forces.

The acceleration control is performed by the cold gas micro propulsion system (provided by ESA), which allows us to counteract the perturbations (atmospheric drag, radiation pressure, electromagnetic forces) at a very small level of tens of μN. The combination of the six degrees of freedom (angular and linear acceleration) enables the control of torques and forces in a single subsystem, called DFACS.

The thermal and mechanical architecture design was driven by the need to centre the instrument as close as possible to the satellite centre of mass and by the high thermal stability requirement around f_{EP} : better than 1 mK at the sensor unit interface and 10 mK at the associated analog electronics interface. Therefore, these most thermally sensitive payload units have been integrated in a specific cocoon at the core of the satellite structure: the BCU, the payload box.

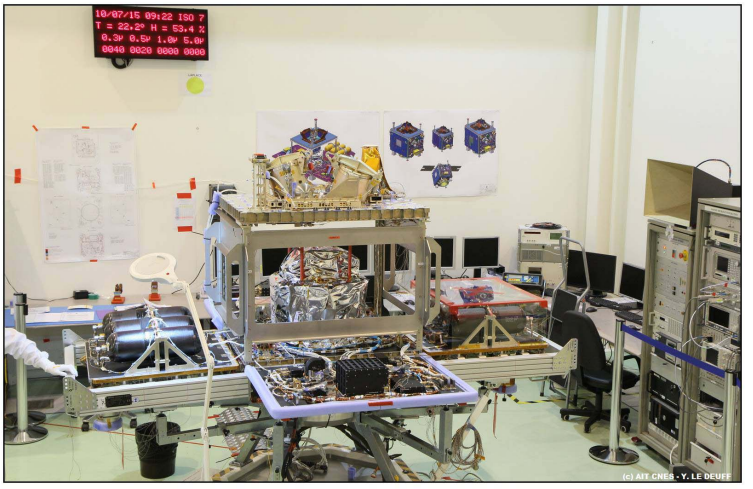


Figure 2. (© AIT-CNES Y. Le Deuff). The MICROSCOPE satellite being integrated. This is a space laboratory of about 300 kg. Once closed, the shape is of a cube measuring 1.4 m × 1 m × 1.5 m. We distinguish in particular the payload T-SAGE (at the centre, under the silver coating), and the micro-propulsion 2 × 3 tanks (on the left and right walls), carrying 16 kg of gas under a pressure of 400 bar at the beginning of life.

For the other Myriade satellites, the payload was located on top of the platform and the tank of the propulsion system was at the centre. The need to ensure an ultra-stable thermal and gravitational environment, with a very low level of microphonic or micro-acceleration disturbances, imposed a lot of constraints in terms of conception, component selection, manufacturing and integration.

3.2. *The attitude and acceleration control subsystem*

The diagram of the DFACS operation principle is shown in Figure 3. The satellite must protect the payload and thus the test masses from all non-gravitational forces perturbing the EP test experiment and so an active control of the acceleration and of the attitude of the satellite has been implemented. The performance of the overall mission was evaluated taking into consideration all subsystems. To specify the DFACS and the payload in close link was a real challenge.

The DFACS in orbit performance exceeded expectations. Some results are summarised here. The common acceleration of the spacecraft was reduced to much better than the specified 10^{-12} m·s^{-2} around the EP test frequency f_{EP}, see Figure 4. Around twice this frequency, the specification is relaxed by a factor 3 to 10, sufficient to estimate the Earth's gravity gradient effect and thus to calibrate the off-centring of the pair of concentric test masses. This control had to deal with non-gravitational accelerations (atmospheric braking and radiation pressure) leading to a mean common mode acceleration signal greater than 10^{-8} m·s^{-2}.

At f_{EP}, the angular pointing is controlled to less than 7 µrad with an angular velocity stability better than 10^{-9} rad·s^{-1} and the angular acceleration better than 10^{-11} rad·s^{-2}. This function needs a very sensitive 6-axis sensor and very fine actuators. For this reason, the accelerometers, which are also able to measure angular accelerations, are used as the main sensor of the DFACS control loop with a priori onboard correction of the scale factor to better than 5% and of the biases to better than a few percent. Scale factors matching is better estimated to an accuracy of a few

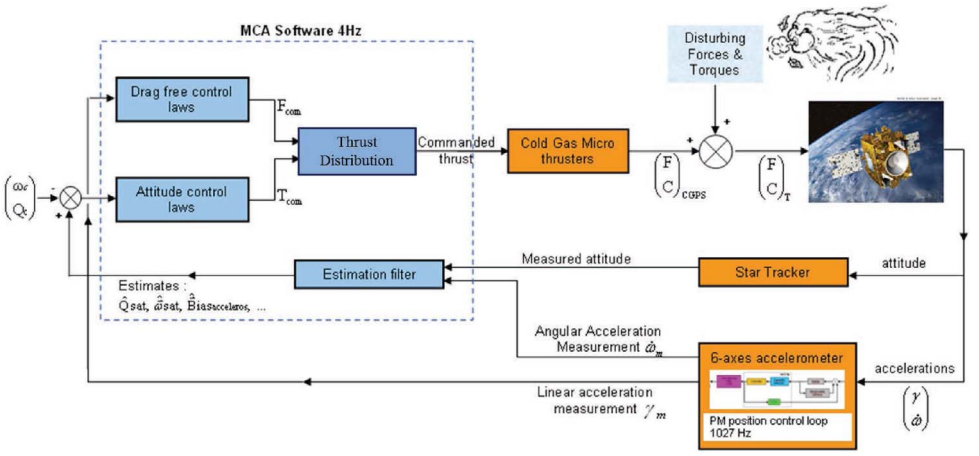


Figure 3. DFACS control loop schematic diagram.

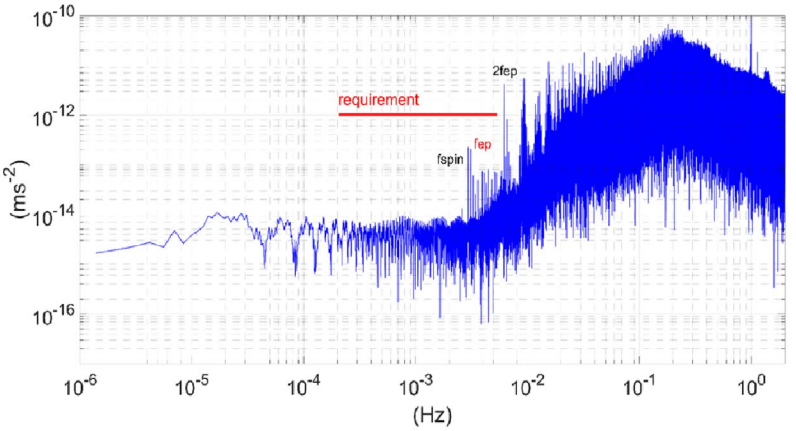


Figure 4. Discrete Fourier Transform of the measured acceleration on the drag-free sensor along X axis.

10^{-5} in dedicated calibration sessions to correct the measurements in the scientific data process: indeed the onboard DFACS needs only rough estimations. The choice of the test mass reference in the drag-free loop was defined by the scientific needs at each session. Estimated attitude is the result of the hybridisation between the attitude measurements provided by the Star Trackers and the angular accelerations measured by T-SAGE. The control laws for the acceleration and the attitude servo-loops define the total forces F_{com} and torques T_{com} (Figure 3) to be applied on the satellite to compensate for the external perturbations. The commanded forces and torques are transformed into 8 thrust orders sent to the Cold Gas Propulsion System (CGPS).

4. The science instrument

The satellite comprises only one payload: T-SAGE (Twin-Space Accelerometers for Gravity Experiment). T-SAGE was designed on the legacy of more than 40 years' experience of developing electrostatic accelerometers [21] which provided the best means of mapping the Earth's gravity

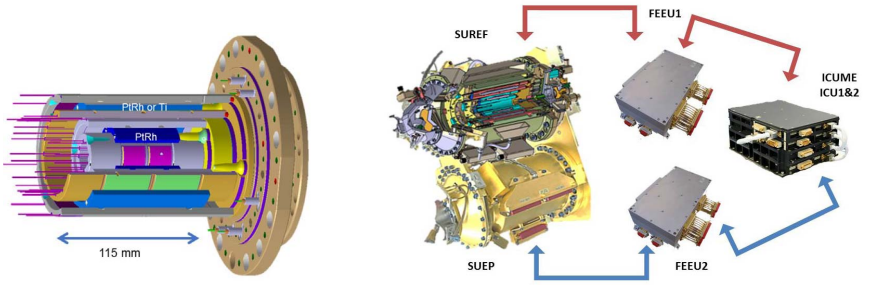


Figure 5. Diagram of T-SAGE. Left: cross-section of one SU with the pair of concentric test masses surrounded by a set of 4 cylinders carrying the electrodes. Right: the two SUs, the two FEEUs and the two ICUs stacked into one unit called ICUME.

field: GRACE [22], GOCE [23,24]. The payload is composed of two sensor units (SU), see Figure 5, each one associated with very accurate analog electronics (Front End Electronic Unit, FEEU) and a digital electronic unit for the test-mass servo-loop and communication with the satellite (Interface Control Unit, ICU). It operates at room temperature and is based on full electrostatic accelerometers.

Two concentric and cylindrical test masses in each SU define the accelerometer core. The SU using two test masses made of two different materials, platinum–rhodium alloy for one and titanium alloy for the other, is called SUEP and is used to test the WEP. The other SU with both test masses made of platinum–rhodium alloy serves as a reference to check the whole experiment and gives more weight to the SUEP results. Each test mass is surrounded by a set of electrodes (Figure 5) that provides a 100 kHz capacitive detection of its motion which is converted in the FEEU into a DC voltage proportional to the displacement at first order. The detection signal is entered at the input of a 40-bit digital controller in the ICU, based on a PID (Proportional Integral Derivative) servo-loop. The output of the PID is representative of the test mass degrees of freedom acceleration and transmitted to the spacecraft onboard computer (OBC).

The stability of the accelerometric measurement [17] is mainly obtained thanks to the very accurate reference voltage applied to the test-mass through a thin gold wire of 7 μm. This DC voltage, V_p (in Figure 6), determines at first order the scale factor of the instrument when opposite voltages are applied to opposite electrodes of one degree of freedom. However, the gold wire damping introduces a Nyquist noise [25], which constitutes the performance limit for this type of accelerometer.

In order to perform the EP test in the best accelerometric environment conditions, the OBC picks up the outer test mass measurements and calculates the necessary thrusts to be applied on each thruster to nullify the accelerometer outputs. The disturbing forces (air drag and solar pressure) and torques (magnetic and gravitational) felt by the satellite are thus taken into account and counteracted. The scientific data process calculates the difference of the measured acceleration in order to extract the Eötvös parameter. After a few months of operation, the differential acceleration noise was established to be quite stable and evaluated to less than 2×10^{-11} m·s^{-2}·Hz$^{-1/2}$ at $f_{\mathrm{EP}} = 0.9 \times 10^{-3}$ Hz for the SUREF and to less than 5×10^{-11} m·s^{-2}·Hz$^{-1/2}$ at $f_{\mathrm{EP}} = 3 \times 10^{-3}$ Hz for the SUEP.

5. First results

The MICROSCOPE mission is divided into different measurement sessions. Sessions represent a time span during which the satellite and the instrument keep the same configuration (spin, drag-

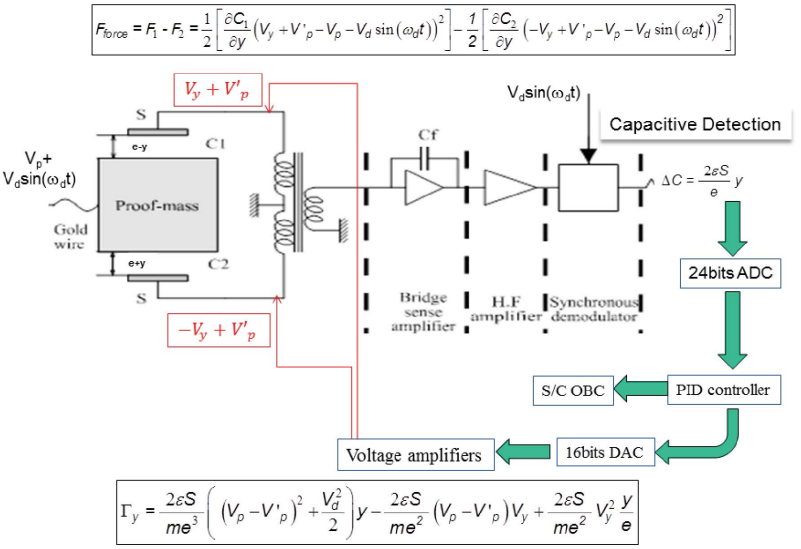


Figure 6. Schematic of one degree of freedom servo-loop control.

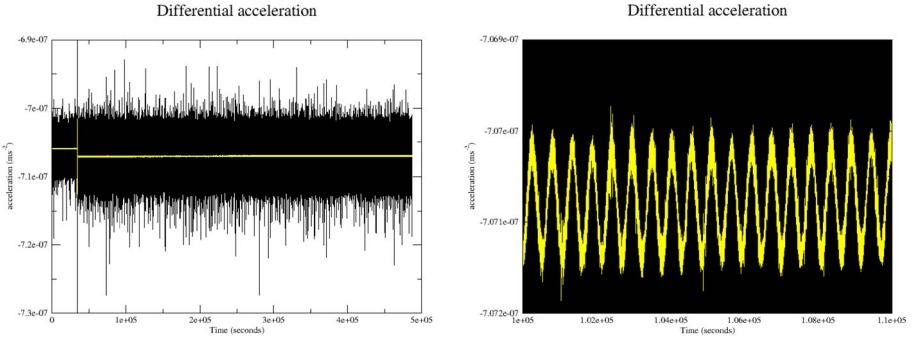


Figure 7. Differences in the accelerations measured along the X-axis, between the two test masses of the SUREF instrument. Raw data (black) and after a running average over 240 points (yellow). The zoom on the averaged data (right panel) highlights the periodic signal due to the gravity gradient.

free control law etc.). Some of these sessions are directly devoted to the EP test while others are used to calibrate or characterise the experiment. EP sessions are the longest, most of them lasting 120 orbital periods (about 8 days), while calibration sessions typically last a few orbits. Figures 7 and 8 show the differential acceleration measured during 2 EP sessions that were among some of the first to take place at the end of the commissioning phase: one session using the instrument SUEP and one session using the instrument SUREF.

The SUREF session had a total duration of 82 orbits, but the precise attitude is available only for the last 62 orbits which limits the analysis to this period. For this session, the spin frequency of the satellite was $f_{\text{spin2}} = 7.568 \times 10^{-4}$ Hz leading to an EP frequency (the frequency at which the Earth apparently rotates with respect to a frame fixed to the satellite) $f_{\text{EP}} = 9.249 \times 10^{-4}$ Hz. The

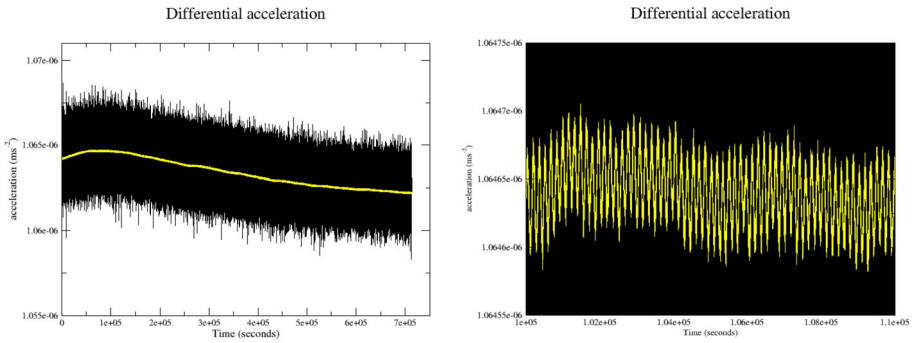


Figure 8. Differences in the accelerations measured along the X-axis, between the two test masses of the SUEP instrument. Raw data (black) and after a running average over 240 points (yellow). The zoom on the averaged data (right panel) highlights the periodic signal due to the gravity gradient.

estimated values of the parameters are [17, 26]:

$$\begin{aligned} \delta &= (4 \pm 4) \times 10^{-15}, \\ \Delta_x &= (-35.39 \pm 0.02)\ \mu\text{m}, \\ \Delta_z &= (5.55 \pm 0.02)\ \mu\text{m}. \end{aligned} \tag{4}$$

Note that the above errors are only statistical errors at 1 sigma. A very preliminary and conservative assessment has been conducted for the Eötvös parameter. This leads to [17]

$$\delta(\text{Pt, Pt}) = [+4 \pm 4(\text{stat}) \pm 8(\text{syst})] \times 10^{-15} \quad (1\sigma \text{ statistical uncertainty}). \tag{5}$$

The SUEP session lasts 120 orbits and the associated spin frequency is $f_{\text{spin3}} = 2.943 \times 10^{-3}$ Hz leading to the EP frequency $f_{\text{EP}} = 3.111 \times 10^{-3}$ Hz. The estimated parameters are

$$\begin{aligned} \delta &= (-1 \pm 9) \times 10^{-15}, \\ \Delta_x &= (20.14 \pm 0.05)\ \mu\text{m}, \\ \Delta_z &= (-5.55 \pm 0.05)\ \mu\text{m}. \end{aligned} \tag{6}$$

Including the systematic errors we get

$$\delta(\text{Ti, Pt}) = [-1 \pm 9(\text{stat}) \pm 9(\text{syst})] \times 10^{-15} \quad (1\sigma \text{ statistical uncertainty}). \tag{7}$$

6. Conclusion and prospects

Using only one session representing slightly more than eight days of measurement, we have already obtained an accuracy about ten times better that the state of the art before MICROSCOPE. The results are fully compatible with the Equivalence Principle: the free fall of the test masses in platinum and titanium are identical at the 10^{-14} level. Confidence in our results has been strengthened by a double check: first, a very conservative evaluation of the systematic errors leads to an assessment better than 10^{-14} and second, the same experiment conducted with identical test masses in platinum provides a null result with an accuracy better than 10^{-14} [17]. A by-product of this experiment, which is not fundamental but gives a good idea of its level of sensitivity, is the estimation of the distance between the test masses with a precision of a few hundredths of μm using the induced gravity gradient.

Since then, ten times more sessions have been acquired, both to test the EP and to characterise the whole experiment. This will allow us to improve the statistical error and also the assessment

of systematic errors (articles in preparation). The MICROSCOPE satellite was deactivated in October 2018 and is slowly desorbiting as predicted with the two wings deployed to increase the air braking.

References

[1] G. Galilei, *Discours et démonstrations mathématiques concernant deux sciences nouvelles*, Presses Universitaires de France, 1995, French translation by M. Clavelin.

[2] I. Newton, *Principes mathématiques de la philosophie naturelle*, Librairie scientifique et technique Albert Blanchard, 1966, French translation by la Marquise du Chastellet.

[3] L. Eötvös, D. Pekár, E. Fekete, "Beiträge zum Gesetz der Proportionalität von Trägheit and Gravität", *Ann. Phys.* **68** (1922), p. 11, English translation in Annales Universitatis Scientiarium Budapestiensis de Rolando Eötvös Nominate, Sectio Geologica, 7, 111, 1963.

[4] S. Schlamminger, K.-Y. Choi, T. A. Wagner, J. H. Gundlach, E. G. Adelberger, "Test of the equivalence principle using a rotating torsion balance", *Phys. Rev. Lett.* **100** (2008), no. 4, article ID 041101.

[5] J. Chabé, C. Courde, J.-M. Torre, S. Bouquillon, A. Bourgoin, M. Aimar, D. Albanèse, B. Chauvineau, H. Mariey, G. Martinot-Lagarde, N. Maurice, D.-H. Phung, E. Samain, H. Viot, "Recent progress in lunar laser ranging at grasse laser ranging station", *Earth Space Sci.* **7** (2020), no. 3, article ID e2019EA000785.

[6] K. Nordtvedt, "Testing relativity with laser ranging to the moon", *Phys. Rev.* **170** (1968), p. 1186-1187.

[7] V. Viswanathan, A. Fienga, O. Minazzoli, L. Bernus, J. Laskar, M. Gastineau, "The new lunar ephemeris INPOP17a and its application to fundamental physics", *Mon. Not. R. Astron. Soc.* **476** (2018), no. 2, p. 1877-1888.

[8] A. Einstein, "Über das Relativitätsprinzip und die aus demselben gezogene Folgerunge", *Jahrb. Radioakt Elektronik* **4** (1907), p. 411, English translation in The Collected Papers of Albert Einstein Volume 2, A. Beck and P. Havas (Princeton University Press, 1989) doc.47.

[9] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie", *Ann. Phys.* **354** (1916), p. 769-822.

[10] B. P. Abbott *et al.*, "Observation of gravitational waves from a binary black hole merger", *Phys. Rev. Lett.* **116** (2016), article ID 061102.

[11] T. Damour, "Theoretical aspects of the equivalence principle", *Class. Quantum Gravity* **29** (2012), no. 18, article ID 184001.

[12] P. K. Chapman, A. J. Hanson, "An Eötvös experiment in earth orbit", in *Proc. Conf. on Experimental Tests of Gravitation Theories* (R. W. Davies, ed.), vol. JPL TM 33-499, California Institute of Technology, Passadena, 1970, p. 228.

[13] C. Everitt, T. Damour, K. Nordtvedt, R. Reinhard, "Historical perspective on testing the equivalence principle", *Adv. Space Res.* **32** (2003), no. 7, p. 1297-1300.

[14] P. Touboul, M. Rodrigues, G. Métris, B. Tatry, "MICROSCOPE, testing the equivalence principle in space", *C. R. Acad. Sci., Paris IV* **2** (2001), no. 9, p. 1271-1286.

[15] P. Touboul, G. Métris, V. Lebat, A. Robert, "The MICROSCOPE experiment, ready for the in-orbit test of the equivalence principle", *Class. Quantum Gravity* **29** (2012), no. 18, article ID 184010.

[16] E. Hardy, A. Levy, M. Rodrigues, P. Touboul, G. Métris, "Validation of the in-flight calibration procedures for the MICROSCOPE space mission", *Adv. Space Res.* **52** (2013), no. 9, p. 1634-1646.

[17] P. Touboul, G. Métris, M. Rodrigues, Y. André, Q. Baghi, J. Bergé, D. Boulanger, S. Bremer, R. Chhun, B. Christophe, V. Cipolla, T. Damour, P. Danto, H. Dittus, P. Fayet, B. Foulon, P.-Y. Guidotti, E. Hardy, P.-A. Huynh, C. Lämmerzahl, V. Lebat, F. Liorzou, M. List, I. Panet, S. Pires, B. Pouilloux, P. Prieur, S. Reynaud, B. Rievers, A. Robert, H. Selig, L. Serron, T. Sumner, P. Visser, "Space test of the equivalence principle: first results of the MICROSCOPE mission", *Class. Quantum Gravity* **36** (2019), no. 22, article ID 225006.

[18] C. Fallet, M. Le Du, C. Pittet, P. Prieur, A. Torres, "First in orbit results from DEMETER", in *28th Annual AAS Control and Guidance Conference*, Univelt, San Diego, CA, 2005.

[19] M. Le Du, J. Maureau, P. Prieur, "Myriade : an adaptative concept", in *5th ESA International Conference on Spacecraft Guidance, Navigation and Control Systems in Frascati*, ESA Publications Division, Oordwijk, 2002.

[20] P. Prieur, T. Lienart, M. Rodrigues, P. Touboul, T. Denver, J. L. Jorgensen, A. M. Bang, G. Metris, "MICROSCOPE mission: on-orbit assessment of the drag-free and attitude control system", in *Paper presented at 31st International Symposium on Space Technology and Science (ISTS 2017), Matsuyama-Ehime, Japan*, 2017.

[21] P. Touboul, B. Foulon, M. Rodrigues, J. P. Marque, "In orbit nano-g measurements, lessons for future space missions", *Aerosp. Sci. Technol.* **8** (2004), no. 5, p. 431-441.

[22] J. Flury, S. Bettadpur, B. D. Tapley, "Precise accelerometry onboard the grace gravity field satellite mission", *Adv. Space Res.* **42** (2008), no. 8, p. 1414-1423.

[23] J.-P. Marque, B. Christophe, B. Foulon, "Accelerometers of the GOCE mission: return of experience from one year of in-orbit", in *ESA Living Planet Symposium*, ESA Special Publication, vol. 686, 2010, p. 57.

[24] R. Rummel, W. Yi, C. Stummer, "GOCE gravitational gradiometry", *J. Geod.* **85** (2011), p. 777-790.

[25] E. Willemenot, P. Touboul, "On-ground investigation of space accelerometers noise with an electrostatic torsion pendulum", *Rev. Sci. Instrum.* **71** (2000), no. 1, p. 302-309.

[26] P. Touboul, G. Métris, M. Rodrigues, Y. André, Q. Baghi, J. Bergé, D. Boulanger, S. Bremer, P. Carle, R. Chhun, B. Christophe, V. Cipolla, T. Damour, P. Danto, H. Dittus, P. Fayet, B. Foulon, C. Gageant, P.-Y. Guidotti, D. Hagedorn, E. Hardy, P.-a. Huynh, H. Inchauspe, P. Kayser, S. Lala, C. Lämmerzahl, V. Lebat, P. Leseur, F. Liorzou, M. List, F. Löffler, I. Panet, B. Pouilloux, P. Prieur, A. Rebray, S. Reynaud, B. Rievers, A. Robert, H. Selig, L. Serron, T. Sumner, N. Tanguy, P. Visser, "MICROSCOPE mission: first results of a space test of the equivalence principle", *Phys. Rev. Lett.* **119** (2017), no. 23, p. 231101-1-231101-7.

Comptes Rendus
Physique
2020, 21, nº 2, p. 151-164
https://doi.org/10.5802/crphys.17

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Fluid dynamics of a mixed convective/stably stratified system—A review of some recent works

Dynamique des fluides d'un système mixte convectif / stablement stratifié — Une revue de quelques travaux récents

Michael Le Bars*, [a], Louis-Alexandre Couston [a], Benjamin Favier [a], Pierre Léard [a], Daniel Lecoanet [b] and Patrice Le Gal [a]

[a] CNRS, Aix Marseille Univ, Centrale Marseille, IRPHE, Marseille, France

[b] Princeton Center for Theoretical Science and Department of Astrophysical Sciences, Princeton, New Jersey 08544, USA

E-mails: lebars@irphe.univ-mrs.fr (M. Le Bars), lacouston@gmail.com (L.-A. Couston), favier.benjamin@gmail.com (B. Favier), pierre.leard@gmail.com (P. Léard), dlecoanet@gmail.com (D. Lecoanet), legal@irphe.univ-mrs.fr (P. Le Gal)

Abstract. Numerous fluid systems organise into a turbulent layer adjacent to a stably stratified one, for instance, planetary atmospheres and stellar interiors. Capturing the coupled dynamics of such systems and understanding the exchanges of energy and momentum at the interface between the two layers are challenging, because of the large range of involved time- and length-scales: indeed, the rapid small-scale turbulence excites waves at intermediate scale, which propagate and interact non-linearly to generate large-scale circulations, whose most famous example is the quasi-biennial oscillation of the Earth's atmosphere. We review here some recent progress on the wave characterisation and on the non-linear mean flow generation, based on the combined experimental and numerical study of a model laboratory system. Applications in climate and stellar modelling are also briefly discussed.

Résumé. De nombreux systèmes fluides s'organisent en une couche turbulente adjacente à une couche stratifiée stable, comme par exemple les atmosphères planétaires et les intérieurs stellaires. La compréhension des échanges d'énergie et de quantité de mouvement à l'interface entre ces deux couches, et l'appréhension de leur dynamique couplée sont difficiles, en raison de la grande gamme d'échelles de temps et de longueur impliquées : en effet, la turbulence rapide à petite échelle excite des ondes à moyenne échelle, qui se propagent et interagissent non linéairement pour générer des circulations à grande échelle, dont le plus célèbre exemple est l'oscillation quasibiennale de l'atmosphère terrestre. Dans cet article, nous passons en revue

* Corresponding author.

quelques progrès récents sur la caractérisation des ondes et sur la génération non-linéaire d'un écoulement moyen, obtenus par l'étude combinée, expérimentale et numérique, d'une configuration modèle au laboratoire. Les conséquences possibles de nos résultats pour la modélisation climatique et stellaire sont aussi brièvement discutées.

Keywords. Internal gravity waves, Convection, Wave—mean flow interactions, Quasi-biennial oscillation (QBO), Atmospheric and stellar dynamics.

Mots-clés. Ondes internes de gravité, Convection, Interactions ondes — écoulemement moyen, Oscillation quasi-biennale, Dynamique atmosphérique et stellaire.

2020 Mathematics Subject Classification. 76-XX.

1. Introduction

Numerous natural systems exhibit a specific organisation with a turbulent convective layer adjacent to a stably stratified one: examples include planetary atmospheres with their troposphere/stratosphere, and stellar interiors with their convective/radiative zones. The dynamics of such coupled, two-layer systems are quite complex and scatter over large ranges of timescales and length-scales. Indeed, motions in the convective layer excite internal gravity waves (IGWs) which propagate from the interface into the stratified region, sustained by gravity and the progressive decrease of the ambient density profile (for a full description of IGWs and their properties, see [1]). IGWs carry momentum and energy, and are thus to be accounted for closing the energy budget of such coupled systems, in order to make e.g. relevant mid- and long-term climate prediction. Waves are also of direct interest in e.g. asteroseismology to probe otherwise inaccessible stellar interiors [2]. Besides, waves can non-linearly generate large scale horizontal flows with global, long-term, dynamical consequences. Such a mechanism has for instance been invoked to explain the apparent misalignment of some exoplanets around hot stars [3]. But its most striking evidence is the quasi-biennial oscillation (QBO) of the Earth's atmosphere at altitudes ranging from about 16 to 50 km, corresponding to a nearly periodic reversal of the equatorial stratospheric winds between easterlies and westerlies with a mean period of 28 to 29 months [4]. Similar oscillations have also been reported in Jupiter's and Saturn's atmospheres [5, 6].

The QBO is classically explained by the specific "anti-diffusive" nature of IGWs (e.g. [7]): IGWs are more prone to lose momentum when they propagate in the same direction as the ambient flow, which in turn is reinforced by the deposition of wave momentum. The mechanism for energy dissipation and wave damping can be due to different phenomena such as radiation, wave breaking and interaction with critical levels in atmospheres and stars, viscous dissipation in experiments... In all cases, as illustrated in Figure 1 (left), starting from e.g. an eastward wind (horizontal mean flow $\bar{u}$) plus two IGWs emitted at the interface with the same frequency and amplitude but with opposite directions, the eastward-propagating wave rapidly deposits its energy and locally increases the ambient wind, while the westward-propagating wave rises higher up and finally damps while generating a westward wind at larger altitude. This appealing mechanism was theorised in [7–9] in a one-dimensionsional (1D) model solving only for the mean flow equation in the linearly stratified domain: there, the time derivative of the mean flow equals its viscous dissipation plus a source term coming from the momentum flux from damped fluctuations (i.e. the vertical gradient of their associated Reynolds stress, corresponding to the horizontal average of the product of the horizontal and vertical velocity fluctuations); it was evaluated analytically by

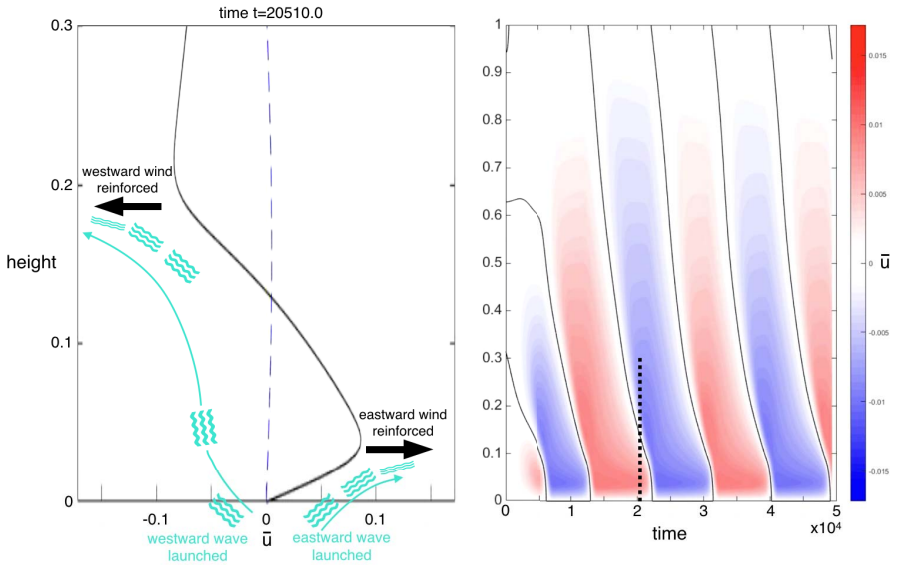


Figure 1. Sketch of the classical QBO model of Lindzen–Holton–Plumb [7–9] (left) and time evolution of the horizontal mean flow $\bar{u}$ as a function of depth computed from the associated 1D model, solving for the mean flow in the presence of a monochromatic, linear wave source in the WBK limit (right). On the left, the dashed line shows the initial mean flow profile and the solid line the profile at $t = 20510$, shown as a vertical dotted lined on the right. Time is adimensionalised by the buoyancy frequency N and lengths by the domain height H.

considering the weakly damped, Doppler shifted, linear internal gravity waves in the WKB limit (i.e. assuming scale separation between the waves and the mean flow) (see details in e.g. [10]). An example of obtained QBO is shown in Figure 1 (right). The monochromatic QBO mechanism was also demonstrated experimentally in the famous study by Plumb and McEwan [11], recently extended by Semin *et al.* [12]: they used oscillating membranes at the boundary of a linearly-stratified salty-water layer in order to force a standing wave pattern in a cylindrical shell container, mimicking the equatorial stratospheric band. However, in this classical model of the QBO and in its experimental realisation, the wave forcing remains steady and monochromatic, as opposed to the atmospheric configuration where it is due to turbulent tropospheric motions [4]. Besides, the excitation is driven by forced interface displacements and only the stratified layer is modelled, neglecting any coupling with the turbulent source. In Global Climate Models (GCMs) capable of spontaneously exhibiting a QBO, part of the waves responsible for its generation, including non-orographic IGWs excited by moist convection [13], are not resolved: they have to be parameterized, and the chosen parameterization scheme significantly affects the obtained results [14, 15]. It thus remains a challenge to observe and understand if/how/when a large-scale, reversing flow spontaneously emerges from a wide range of naturally excited IGWs, in a self-organising coupled two-layer system. And even before doing so, deciphering the mechanism of wave excitation in such a coupled convective/stably stratified system, as well as predicting the spectral characteristics of the associated wave field, are still debated. These are the tasks we have started to tackle over the last few years combining experiments and numerical simulations. This paper presents a rapid review of our recent contributions [16–21] and of some of the remaining open questions.

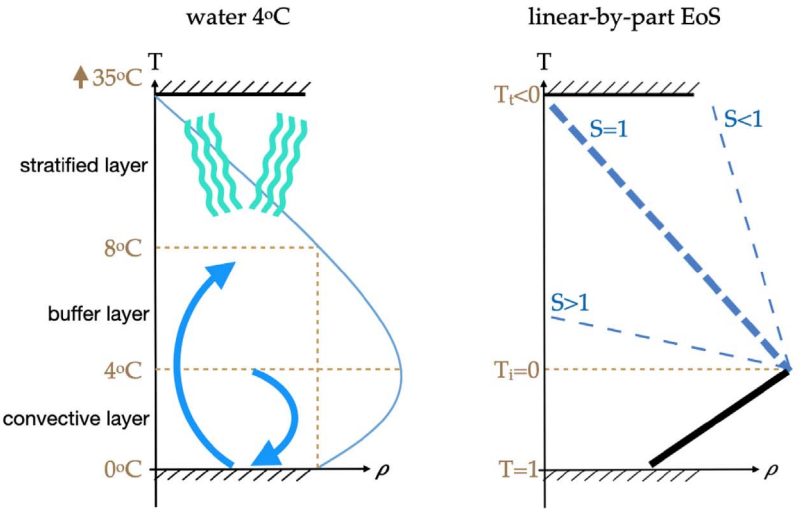


Figure 2. Density variation as a function of temperature for water around 4 °C (left) and for the generalised equation of state used in our numerical simulations (right).

2. Experimental and numerical investigation tools

2.1. *A self-consistent two-layer system in the laboratory: convection in water around* 4 °*C*

Previous studies of wave excitation by convection have used either a forced plume [22] or a transient Rayleigh–Bénard system, starting from a thermally stratified configuration and suddenly reversing the buoyancy profile from a boundary [23, 24]. But as first realized by Townsend [25], a self-organising, stationary, two-layer convective/stably stratified system can be relatively readily obtained in the laboratory using water, thanks to its specific property of having its maximum density at 4 °C with a nearly parabolic equation of state around it: as sketched in Figure 2 (left), in a simple reverse Rayleigh–Bénard configuration with cooling from below at 0 °C and heating from above at e.g. 35 °C, a two-layer system spontaneously emerges, with a turbulent convective layer below a stably-stratified one. Cold buoyant plumes rise from the bottom plate at 0 °C, cross the 4 °C isotherm, and theoretically equilibrate around the 8 °C isotherm, when neglecting diffusive effects; reciprocally, dense plumes at 4 °C detach from the maximum density interface and sink into the cooler, convective layer. IGWs propagate in the stably stratified layer above 8 °C. Note that because of the dissymetry between rising and sinking convective structures, the region between 4 °C and 8 °C is very specific: it is called the buffer layer [26].

Figure 3 shows our experimental realisation of this system [21], following an earlier, less evolved version of the set-up [16]. The tank is made of 2 cm thick acrylic sides, a temperature controlled bottom copper plate, and a transparent, temperature controlled, electric heater as a top boundary. Inner dimensions are 32×32 cm^2 in horizontal and $H = 20$ cm in height. A cylinder of outer diameter 29 cm and thickness 0.4 cm might be centred inside this tank to obtain an axisymmetric geometry prone to the development of large-scale horizontal flows, as shown by the historical work of Plumb and McEwan [11]. Velocity measurements are performed in a vertical central plane using Particle Imaging Velocimetry (PIV), to characterize both convective motions and propagating IGWs (Figure 3 bottom left). Additionally, wave dynamics and the possible presence of large-scale horizontal flows are assessed by performing horizontal PIV (Figure 3 bottom right) and scanning over the whole depth of the tank. Further details on the experimental set-up can be found in [21].

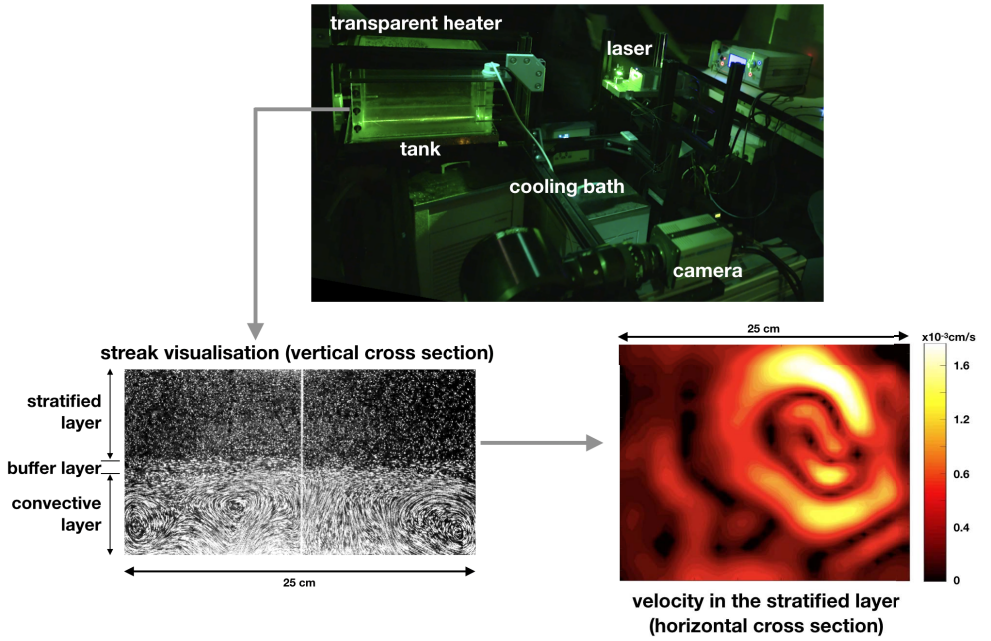


Figure 3. Picture of the experimental set-up (top) and illustration of velocity measurements in a vertical plane (bottom left) and in an horizontal plane (bottom right). The vertical cross-section shows a streak pattern obtained by superimposing 15 images (i.e. duration of 7.5 s) from the PIV acquisition movie. The horizontal cross-section shows an instantaneous PIV field in the stratified domain, at a distance ∼1 cm above the interface with the convective zone.

This experimental system is fully characterised by 3 dimensionless parameters, defined as

- the Prandtl number $Pr = \nu/\kappa$, equal to the ratio of kinematic viscosity ν to thermal diffusivity κ averaged over the whole domain,
- the Rayleigh number Ra based on the convection-driving density difference $\Delta\rho = \rho(4\,^\circ C) - \rho(0\,^\circ C)$, $Ra = g\Delta\rho H^3/\rho_0\kappa\nu$, where g is the gravity and ρ_0 the mean density,
- the top temperature anomaly relative to the "inversion" 4 °C temperature, non-dimensionalised by the convection-driving temperature difference (0 °C to 4 °C here), named T_t.

We also define the mean buoyancy frequency N from the total density contrast over the stratified layer depth. In our experiment, $Pr = 7$, $Ra = 7 \times 10^6$, $T_t = -7.75$, and $N = 0.135$ Hz. As illustrated in Figure 3, the flow in the convective region is chaotic, with turbulent plumes advected by large-scale circulation and typical velocities around 1 mm/s. The buffer layer is clearly apparent, subject to a strong horizontal shear discussed in details in [21]. The stratified domain sustains IGWs with typical velocities around 10 μm/s: they appear in the horizontal cross-section as concentric rings, similar to waves propagating from an impact point at the surface of a lake.

2.2. *Generalisation in direct numerical simulations*

While convection in water around 4 °C allows a nice experimental realisation of a self-consistent, two-layer configuration, it is also intrinsically limited in terms of parameter space exploration. Indeed, considering a fixed total domain size, we have two adjusting parameters: the top and bot-

tom temperatures. We aim at maximising the Rayleigh number to reach a chaotic state: we thus use the maximum possible driving temperature contrast, hence a bottom temperature of 0 °C. Adjusting the top temperature anomaly allows changing the relative depth of the two layers. However, at steady-state, heat flux conservation between the convective and the stratified domains intrinsically fixes the value of the buoyancy frequency: indeed, according to the Howard's historical scaling law [27], the convective heat flux does not depend on the convective layer depth, but only on its driving temperature difference, which is here fixed at 4 °C; this fixes the temperature gradient in the diffusive, stratified layer, which determines the buoyancy frequency. Besides, we aim for a deep enough stratified layer to allow for wave propagation, and lateral heat losses render the system highly non-linear (see e.g. discussion in [17]): this significantly limits any change in the temperature anomaly. Finally, using water also fixes the value of the Prandtl number to 7, which is limiting since as we show in the following, Pr has a tremendous influence on the long-term dynamics. Hence, to further explore the dynamics of our self-organising two-layer system, we have also used numerical simulations, expanding upon the experimental model.

We have first solved the non-Oberbeck Boussinesq Navier–Stokes and temperature equations using the approximate parabolic equation of state for water, hence closely reproducing the experiments. To do so, we have used either the open spectral solver Dedalus [28] in two dimension (2D), with periodic horizontal boundary conditions [17], or the open spectral element solver NEK5000 [29] in three dimension (3D), with perfectly insulating, rigid vertical boundaries [21]. As will be detailed in the following, this has allowed us to investigate the mechanism of wave excitation [17], to assess the experimental uncertainties issued from e.g. non-perfect thermal boundary conditions, and to explore the influence of the Prandtl number on the dynamics [21].

Then, to provide a more systematic exploration of a larger parameter space, we have also considered an equation of state with a constant thermal expansion coefficient in each layer, but changing sign and value around a chosen inversion temperature (Figure 2 right). Non-Oberbeck Boussinesq Navier–Stokes and temperature equations are then solved using the open spectral solver Dedalus [28] with periodic boundary conditions in the horizontal direction, both in 2D [18, 19] and in 3D (see [20] and Figure 4 left). The ratio S of the thermal expansion coefficients in the stratified vs. convective layers determines the stiffness of the interface, and reveals three different regimes (see Figure 4 right and [18]): a whole-layer convective regime at small stiffness, where the interface is destroyed by rising plumes; a two-layer regime at large stiffness, where the interface remains flat but gravity internal modes are excited by Reynolds stress fluctuations from the convective layer; and an intermediate regime in between, with a deformable interface and propagating IGWs, actually corresponding to the experimental, water configuration. In the following, we will focus on this last case only.

3. Wave properties

3.1. *Mechanism for wave excitation*

A good knowledge of the physical mechanism for wave excitation is fundamental for correct IGW parameterization in climate models and valid interpretation of asteroseismology observations (see e.g. [30]): as reviewed for instance in [22], it is thus the subject of a long-standing debate, with two main possible models sketched in Figure 5. In the mechanical oscillator model, convective updrafts rise up and deflect the interface with the stratified zone, hence locally initiating propagating IGWs. On the contrary, the deep forcing model assumes excitation all over the convective domain from the Reynolds stress associated with turbulent fluctuations: generated IGWs are first evanescent in the convective domain where no global stratification exists, and turn into propagating IGWs if/when they reach the interface and the stratified domain.

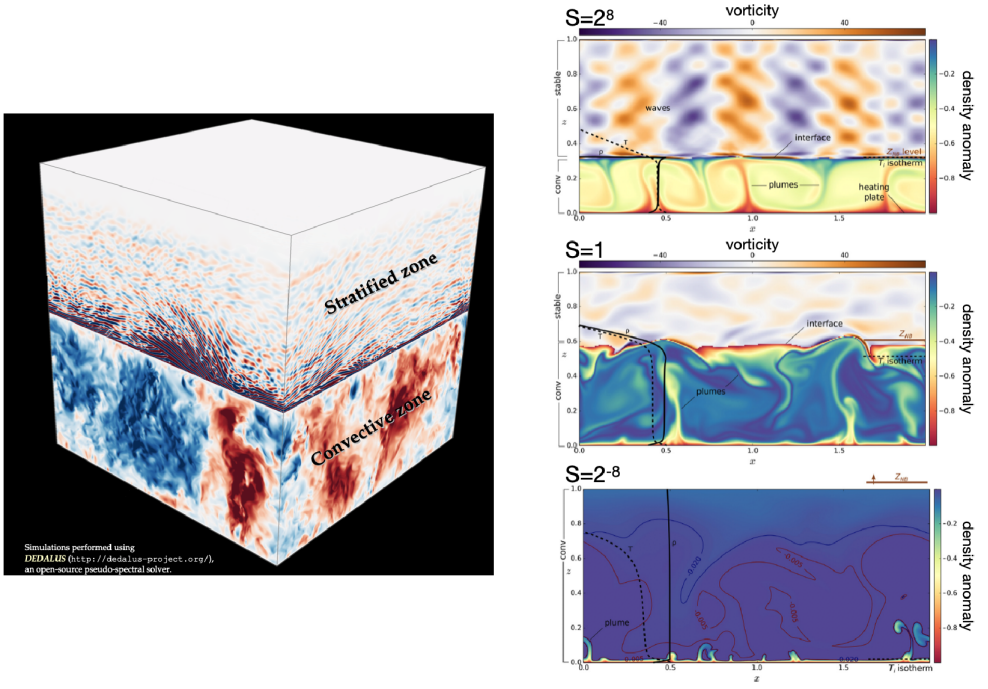


Figure 4. Snapshots from numerical simulations at thermal equilibrium using a piecewise linear equation of state. Left: vertical velocity in a 3D simulation for $Ra = 2 \times 10^8$, $Pr = 1$, $T_t = -52$ and $S = 400$, with red corresponding to upward motions and blue to downward motions. Note that the color scale is adjusted in each zone separately for better visualization. Right: 2D simulations for a given $Ra = 8 \times 10^7$, $Pr = 1$, $T_t = -20$, and 3 different stiffnesses S representative of the 3 different regimes of the system's dynamics, reproduced from [18]. Colors show the density anomaly in the convection zone below the neutral buoyancy height noted Z_{NB}, and the vorticity in the stratified zone above it (except for the smaller value of S where Z_{NB} is outside of the domain). We also show the mean temperature and mean density profiles as dashed and solid lines.

To quantitatively assess which model is the most relevant for our configuration, we use a 2D full simulation of the water experiment, together with 2 models of the simulation where we only solve for the linear wave equation in terms of vertical displacement, together with an adhoc source of excitation. The flexibility of the Dedalus solver [28] is especially suited for this type of approach. The first model of the simulation corresponds to the deep forcing, where we follow the approach of Lighthill [31], adapted to our configuration: we first compute from the full simulation the Reynolds stress all over the convective domain, and we then use it as the excitation term in the wave equation, with a buoyancy frequency equal to zero in the convective domain and to its horizontal and temporal average in the stratified domain. The second model of the simulation corresponds to the mechanical oscillator mechanism: we calculate from the full simulation the position of a chosen isotherm as a function of time. We then use these position fluctuations as the bottom boundary condition for the wave displacement in the model, solving wave propagation in the stratified domain above this isotherm only. We have considered 2 different isotherms encompassing the effective interface location: the 5 °C which is very close to the density maximum, and the 8 °C which corresponds to the maximum height of rising plumes at 0 °C in the absence of dissipation (since the equation of state is parabolic with a maximum at 4 °C).

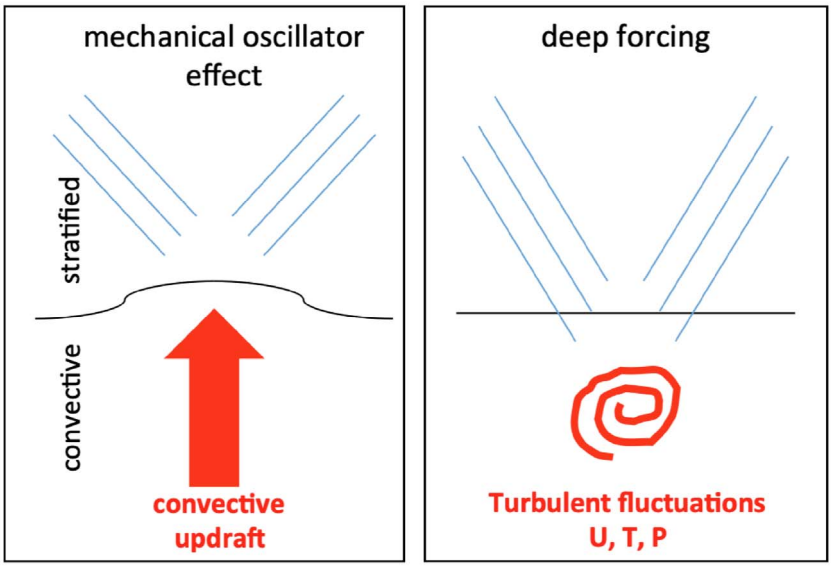


Figure 5. Sketch of the two possible mechanisms for IGW excitation from convection.

Comparison between the full simulation and the 2 models is shown in Figure 6, considering the power spectrum density of the vertical velocity as a function of depth in the stratified region. The deep forcing model agrees remarkably well with the full simulation, while the mechanical oscillator model exhibits high frequency IGWs with overestimated amplitude. The interface forcing considers plumes hammering on the interface, hence produces impulsive excitations that translate into high-frequency waves; it does not correctly account for the regularisation/smoothing of the complete flow, like e.g. the sweeping motions along the interface of the thermal uplifts advected by the large-scale convective motions shown in Figure 3. Comparing Figures 6(c) and (d), it is clear that the forcing by the 8 °C isotherm fluctuations does a much better job, as one would expect from the fact that by then, wave amplitudes are small and the dynamics is much more linear. Neverthless, the high-frequency signature in the spectrum is still very apparent, contrary to both the full simulation and the bulk forcing model. In conclusion, and even if in visualizations we clearly see strong, but intermittent evidences of the mechanical oscillator in the form of wave clusters emerging from impinging rising plumes (see e.g. Figure 4(left)), the wave energy distribution is clearly dominated by the Reynolds stress coupling with the convective layer. It is thus better described by a deep forcing model, at least in the explored range of parameters.

3.2. *Wave flux*

Acknowledging that deep forcing by Reynolds stress is the predominant source of IGWs, it is possible to compute the full temporal and spatial spectrum of linear waves in the stratified region from an adhoc modelling of the turbulent region [32]. For instance, Lecoanet and Quataert [33] describe the flow in the convective region as a Kolmogorov turbulent cascade from an injection scale corresponding to the large-scale circulation. They then predict that for weakly damped waves, the energy flux spectrum scales like $k_\perp^4 f^{-13/2}$, where $k_\perp$ and f are the wave horizontal wavenumber and frequency, respectively; the total wave flux decreases as a power law of the distance from the interface, with an exponent $-13/8$. Despite the simplicity of the underlying mechanistic model, this analytical prediction shows remarkable agreement with our

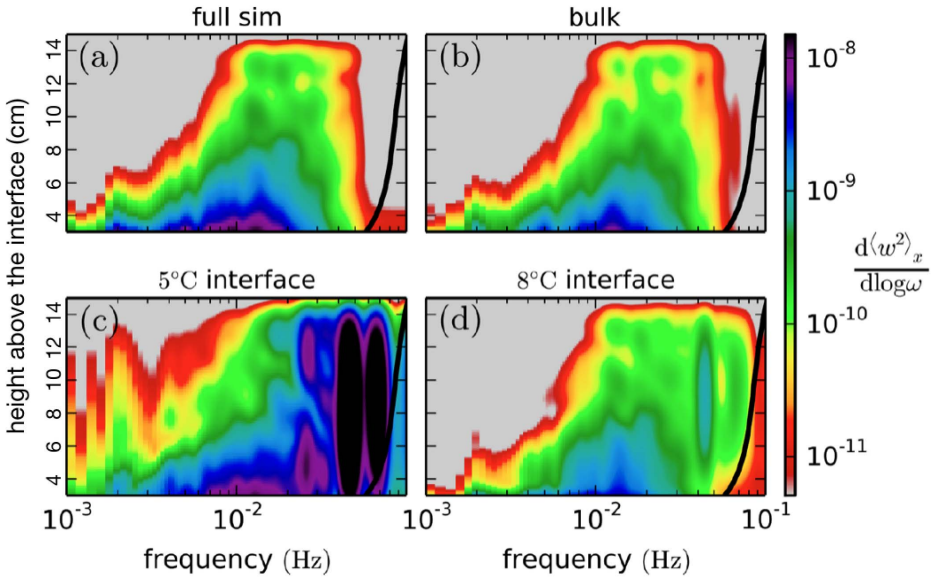


Figure 6. Spectrograms of the vertical velocity squared in the full simulation of the 4 °C experiment (a), in the deep forcing model (b) and in the mechanical oscillator model using the 5 °C (c) and the 8 °C (d) isotherms. The black line shows the horizontally and temporally averaged buoyancy frequency profile. Reproduced from [17].

3D simulations, as shown in [20] and illustrated in Figure 7. These scalings can thus be used to assess possible wave signatures revealed by asteroseismology. For instance, Bowman *et al.* [34] recently reported low-frequency photometric variability in a large number of hot massive stars, which they interpret as the surface signature of IGWs excited by the deep convective core [35]. If it was so, the analysis of those waves would provide a unique probe inside the otherwise inaccessible depths of those mysterious objects. However, the observed spectral signature does not match with our validated model, and we rather interpret it as the trace of some subsurface convection [30]. This issue is currently debated [36].

4. Mean flow generation and reversals in the stratified layer

4.1. *Influence of the Prandtl number*

Beyond IGWs characterization, we have also assessed the generation of a mean flow in the stratified region of our 4 °C experiment by systematically measuring the azimuthal mean of the azimuthal velocity as a function of depth and time: results are shown in Figure 8 (left). While the experimental velocity field exhibits reversals on a typical time much longer than the wave periods (7.4–250 s typically), the observed signal cannot be related to a QBO process, because the phase propagation of reversals goes slightly upward, as opposed to the clear downward signature observed both in atmospheric data [4] and in the ideal, 1D, monochromatic QBO model (Figure 1 right). By reproducing our experiment using 3D numerical simulation, we have checked that this signal is not due to any improper boundary condition, like e.g. lateral heat losses which could have induced unwanted natural convection [21]. Actually, an estimate of the viscous propagation of a velocity perturbation from the interface shows a quantative agreement with the experimental signal (see dotted lines in Figure 8 left): the stratified layer is viscously coupled to the interface

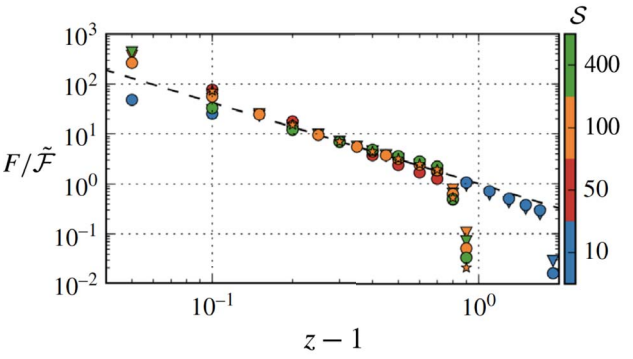


Figure 7. Total flux carried by IGWs in the stratified domain (normalised by its theoretical evaluation, see details in [20]), as a function of height above the interface located at $z = 1$. Triangles, circles and stars show results from our 3D simulations for $Ra = 4\times10^7, 2\times10^8, 10^9$, and the dashed line the theoretical scaling $(z-1)^{-13/8}$. Color shows the stiffness S, $Pr = 1$ in all cases, and T_t is adjusted so as to conserve the same stratified layer depth. Deviations at small z are due to non-wave flows present around the interface region. Departure at large z comes from imposed boundary conditions in the simulations. Note that the simulations with $S = 10$ (i.e. blue color) were performed in a deeper computational domain, hence depart close to $z-1 = 2$ instead of $z-1 = 1$. Reproduced from [20].

region. In water, viscosity is indeed the dominant diffusive effect, as quantified by its Prandtl number $Pr = 7$. Reproducing the same two-layer configuration in a numerical simulation with $Pr = 0.1$ actually shows the vanishing of this viscous coupling as well as some tenuous signature of a QBO-like pattern [21].

The Prandtl number thus has a fundamental influence on the generated mean flow that we want to address systematically. However, such a study is extremely costly from a numerical point of view: it requires numerous and long computations with an highly performant solver, and remains barely feasible today in 3D. We have thus started this systematic study using our 2D Dedalus model with periodic boundary conditions and a piecewise linear equation of state [19]. Three illustrative results are shown in Figure 8 (right). At $Pr = 0.3$, a QBO is clearly observed; at $Pr = 1$, a QBO is obtained but is barely visible; and at $Pr = 3$, the QBO signature completely disappears and the mean flow has an upward phase suggestive of a viscous coupling with the convective region. The thresholds between these different regimes deserve a more detailled, dedicated study, and surely depend on the level of turbulence in the convective domain. Nevertheless, our first results here highlight that in numerical modelling of the longterm dynamics of an atmosphere ($Pr = 0.7$) or a star ($Pr = 10^{-5}$), the Prandtl number based on molecular viscosity and thermal diffusivity should not be fixed at 1 for numerical convenience, as commonly done.

4.2. *Parameterisation of IGWs and the resulting QBO*

Actually, computing the full dynamics of a two-layer system, including all the time- and length-scales of convection, of waves, and of their non-linear long-term interactions, remains limited to idealised or local configurations. In Global Climate Models (GCMs) for instance, physical variables are typically evaluated every ten minutes on a grid with 100 km resolution in the horizontal and a few hundred meters in the vertical (e.g. [37]); phenomena at smaller scales—like non-orographic IGWs—are not resolved but appear as parameterizations. Some GCMs are capable of

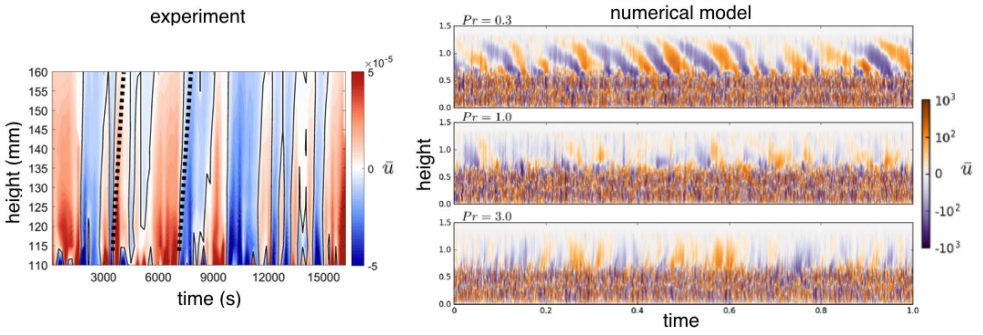


Figure 8. Evolution as a function of time and height of the azimuthal average of the azimuthal velocity in our 4 °C experiment where $Pr = 7$ (left), and of the horizontal average of the horizontal velocity for three 2D simulations with the piecewise linear equation of state considering $Pr = 0.3, 1, 3$ respectively (right). In numerical simulations, $T_t = -43$ and the other parameters are slightly adjusted to maintain the same depth and buoyancy frequency in the stratified region, i.e. $Ra = 8 \times 10^7, 5.6 \times 10^7, 4.4 \times 10^7$ and $S = 0.33, 0.14, 0.06$, respectively. Adapted from [21] and [19]. Note that the experimental results show the stratified domain only. The dotted lines show the typical viscous propagation of a perturbation from the interface.

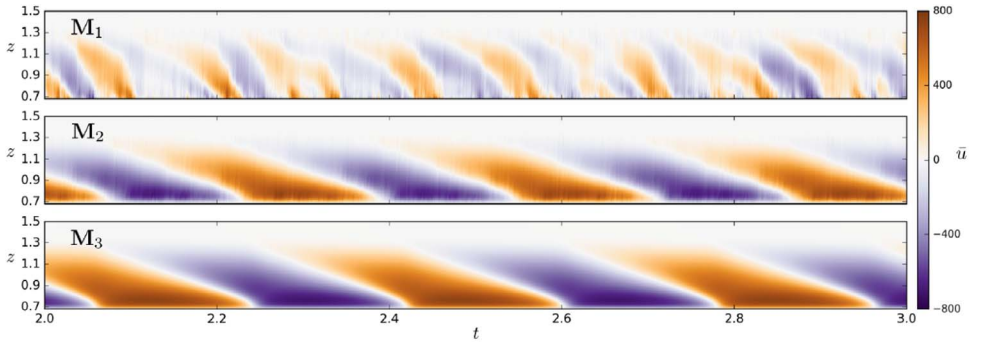


Figure 9. QBO-like flows produced in the stratified domain of a full 2D simulation of the coupled system (M1, same parameters as Figure 8 top right); in a 2D model of the stratified layer only, using as bottom boundary conditions the forcing extracted from the full simulation at the depth of neutral buoyancy (M2); and in a 1D Lindzen–Holton–Plumb-like model sustained with the energy spectrum measured at the interface of the full simulation (M3). Reproduced from [19].

producing realistic QBO (see e.g. [15] and references therein), whose signature provides a reliable test for proving the validity of the model [38]. However, different parameterization schemes lead to different predictions. In addition to very interesting, recent initiatives in assessing QBO modelling uncertainties by performing coordinated numerical benchmarks with GCMs [14, 15], our simplified model could offer an unique opportunity to assess the minimum necessary ingredients for a relevant treatment of IGWs and QBO.

Figure 9 (top) shows again the mean flow in the stratified layer from the full simulation at $Pr = 0.3$ introduced in Figure 8 (top right). The QBO signature is clearly visible. We then consider 2 models of the simulation. In M2 (Figure 9 middle), we solve the full Boussinesq Navier–Stokes

equations in the stratified layer only, using as bottom boundary conditions the forcing in velocity and temperature measured in the full simulation at the interface (i.e. depth of neutral buoyancy). Finally, bottom Figure (M3) shows results from a 1D Lindzen–Holton–Plumb-like model (i.e. solving only the mean flow equation with a forcing term computed from the weakly damped, Doppler shifted, linear internal gravity waves in the WKB limit), using as a forcing term the linear superimposition (with no cross-correlation) of a large range of wave contributions, each wave being excited with an amplitude given by the energy spectrum measured at the interface of the full simulation. Comparing M2 and M3 shows that the 1D Lindzen–Holton–Plumb-like model actually does a decent job in reasonably reproducing, at much lower numerical cost, the signal obtained from a given energy input at the interface, despite all underlying approximations (WKB, weakly damped IGWs, etc.). However, both models fail in predicting the relevant period and amplitude of flow reversals in the full simulation shown in M1, which reinforces the conclusion already raised in Section 3.1 on the source of IGWs: considering wave excited by interface fluctuations only is not sufficient for producing the relevant wave spectrum over the whole stratified domain, hence for correctly modelling their long-term non-linear effects. One must actually consider the whole Reynolds stress generated in the convective region. This is clearly not feasible for GCMs parameterization. But one should at least consider, beyond energy spectrum, higher order statistical description of the wave interface fluctuations, in order to better account for the properties of the convective turbulent source, including in particular intermittency and wave packet production: this is done for instance by [13].

5. Conclusion and open questions

In conclusion, by combining laboratory experiments and numerical simulations, we have successfully characterised the mechanism and characteristics of IGWs excitation in a self-organising two-layer convective/stably stratified system. Our model has also demonstrated that beyond the historical 1D, monochromatic model of Lindzen, Holton and Plumb [7–9], and in complement to GCMs where part of the relevant waves still have to be parameterized [15], slowly reversing mean flows may spontaneously emerge from a stochastic convective excitation, provided the Prandtl number is low enough. Various challenges now remain to be tackled. First from an experimental point of view, main challenges are: (i) to produce QBO-like reversals in a set-up with a stochastic excitation, and (ii) to explain why the only successful experimental QBO up-to-now has been obtained in salty water, i.e. with an equivalent Prandtl number of 700, which seems at odds with our previous conclusion. Then from a numerical point of view, main challenges are: (i) to obtain QBO-like reversals in 3D direct numerical simulations of the full coupled system, and (ii) to extend our results on wave excitation and propagation to more realistic configurations, including in particular compressibility and rotation effects [39], as well as a spherical geometry. Finally, beyond atmospheric and stellar applications, it would be of great interest to evaluate the consequences of waves in other natural two-layer systems, like e.g. the Earth's iron core, where the presence of a convective domain is the prevalent model for explaining the generation of the Earth's magnetic field, but where the presence of a stratified layer has recently been proposed [40]: as in stars and atmospheres, no doubt that excited waves and associated mean flow in this stratified layer would have a strong signature, here imprinted in the magnetic field [41].

Acknowledgements

The authors are supported by the European Research Council under the European Union's Horizon 2020 research and innovation program through Grant No. 681835-FLUDYCO-ERC-2015-CoG.

References

[1] B. R. Sutherland, *Internal Gravity Waves*, Cambridge University Press, Cambridge, UK, 2010.

[2] C. Aerts, J. Christensen-Dalsgaard, D. W. Kurtz, *Asteroseismology*, Springer Science & Business Media, Netherlands, 2010.

[3] T. Rogers, D. N. Lin, H. H. B. Lau, "Internal gravity waves modulate the apparent misalignment of exoplanets around hot stars", *Astrophys. J. Lett.* **758** (2012), no. 1, p. L6.

[4] M. Baldwin, L. Gray, T. Dunkerton, K. Hamilton, P. Haynes, W. Randel, J. Holton, M. Alexander, I. Hirota, T. Horinouchi *et al.*, "The quasi-biennial oscillation", *Rev. Geophys.* **39** (2001), no. 2, p. 179-229.

[5] C. B. Leovy, A. J. Friedson, G. S. Orton, "The quasiquadrennial oscillation of Jupiter's equatorial stratosphere", *Nature* **354** (1991), no. 6352, p. 380.

[6] T. Fouchet, S. Guerlet, D. Strobel, A. Simon-Miller, B. Bézard, F. Flasar, "An equatorial oscillation in Saturn's middle atmosphere", *Nature* **453** (2008), no. 7192, p. 200.

[7] R. S. Lindzen, J. R. Holton, "A theory of the quasi-biennial oscillation", *J. Atmos. Sci.* **25** (1968), no. 6, p. 1095-1107.

[8] J. R. Holton, R. S. Lindzen, "An updated theory for the quasi-biennial cycle of the tropical stratosphere", *J. Atmos. Sci.* **29** (1972), no. 6, p. 1076-1080.

[9] R. Plumb, "The interaction of two internal waves with the mean flow: Implications for the theory of the quasi-biennial oscillation", *J. Atmos. Sci.* **34** (1977), no. 12, p. 1847-1858.

[10] A. Renaud, L.-P. Nadeau, A. Venaille, "Periodicity disruption of a model quasibiennial oscillation of equatorial winds", *Phys. Rev. Lett.* **122** (2019), no. 21, article ID 214504.

[11] R. Plumb, A. McEwan, "The instability of a forced standing wave in a viscous stratified fluid: A laboratory analogue of the quasi-biennial oscillation", *J. Atmos. Sci.* **35** (1978), no. 10, p. 1827-1839.

[12] B. Semin, N. Garroum, F. Pétrélis, S. Fauve, "Nonlinear saturation of the large scale flow in a laboratory model of the quasibiennial oscillation", *Phys. Rev. Lett.* **121** (2018), no. 13, article ID 134502.

[13] F. Lott, L. Guez, "A stochastic parameterization of the gravity waves due to convection and its impact on the equatorial stratosphere", *J. Geophys. Res.* **118** (2013), no. 16, p. 8897-8909.

[14] N. Butchart, J. Anstey, K. Hamilton, S. Osprey, C. McLandress, A. Bushell, Y. Kawatani, Y.-H. Kim, F. Lott, J. Scinocca *et al.*, "Overview of experiment design and comparison of models participating in phase 1 of the SPARC Quasi-Biennial Oscillation initiative (QBOi)", *Geosci. Model Dev.* **11** (2018), no. 3, p. 1009-1032.

[15] A. Bushell, J. Anstey, N. Butchart, Y. Kawatani, S. Osprey, J. Richter, F. Serva, P. Braesicke, C. Cagnazzo, C.-C. Chen *et al.*, "Evaluation of the Quasi-Biennial Oscillation in global climate models for the SPARC QBO-initiative", *Q. J. R. Meteorol. Soc.* (2020), p. 1-31.

[16] M. Le Bars, D. Lecoanet, S. Perrard, A. Ribeiro, L. Rodet, J. M. Aurnou, P. Le Gal, "Experimental study of internal wave generation by convection in water", *Fluid Dyn. Res.* **47** (2015), no. 4, article ID 045502.

[17] D. Lecoanet, M. Le Bars, K. J. Burns, G. M. Vasil, B. P. Brown, E. Quataert, J. S. Oishi, "Numerical simulations of internal wave generation by convection in water", *Phys. Rev. E* **91** (2015), no. 6, article ID 063016.

[18] L.-A. Couston, D. Lecoanet, B. Favier, M. Le Bars, "Dynamics of mixed convective–stably-stratified fluids", *Phys. Rev. Fluids* **2** (2017), no. 9, article ID 094804.

[19] L.-A. Couston, D. Lecoanet, B. Favier, M. Le Bars, "Order out of chaos: slowly reversing mean flows emerge from turbulently generated internal waves", *Phys. Rev. Lett.* **120** (2018), no. 24, article ID 244505.

[20] L.-A. Couston, D. Lecoanet, B. Favier, M. Le Bars, "The energy flux spectrum of internal waves generated by turbulent convection", *J. Fluid Mech.* **854** (2018), article ID R3.

[21] P. Léard, B. Favier, P. Le Gal, M. Le Bars, "Coupled convection and internal gravity waves excited in water around its density maximum at 4 °C", *Phys. Rev. Fluids* **5** (2020), no. 2, article ID 024801.

[22] J. K. Ansong, B. R. Sutherland, "Internal gravity waves generated by convective plumes", *J. Fluid Mech.* **648** (2010), p. 405-434.

[23] J. W. Deardorff, G. E. Willis, D. K. Lilly, "Laboratory investigation of non-steady penetrative convection", *J. Fluid Mech.* **35** (1969), no. 1, p. 7-31.

[24] M. Michaelian, T. Maxworthy, L. Redekopp, "The coupling between turbulent, penetrative convection and internal waves", *Eur. J. Mech. B* **21** (2002), no. 1, p. 1-28.

[25] A. Townsend, "Natural convection in water over an ice surface", *Q. J. R. Meteorol. Soc.* **90** (1964), no. 385, p. 248-259.

[26] S. Perrard, M. Le Bars, P. Le Gal, "Experimental and numerical investigation of internal gravity waves excited by turbulent penetrative convection in water around its density maximum", in *Studying Stellar Rotation and Convection*, Springer, Berlin, Heidelberg, Germany, 2013, p. 239-257.

[27] L. N. Howard, "Convection at high Rayleigh number", in *Applied Mechanics*, Springer, New York, USA, 1966, p. 1109-1115.

[28] K. J. Burns, G. M. Vasil, J. S. Oishi, D. Lecoanet, B. P. Brown, "Dedalus: A flexible framework for numerical simulations with spectral methods", *Phys. Rev. Res.* **2** (2020), no. 2, article ID 023068.

[29] P. Fisher, J. Lottes, S. Kerkemeier, "Nek5000 v17.0", http://nek5000.mcs.anl.gov (2017).

[30] D. Lecoanet, M. Cantiello, E. Quataert, L.-A. Couston, K. J. Burns, B. J. Pope, A. S. Jermyn, B. Favier, M. Le Bars, "Low-frequency variability in massive stars: Core generation or surface phenomenon?", *Astrophys. J. Lett.* **886** (2019), no. 1, p. L15.
[31] M. J. Lighthill, *Waves in Fluids*, Cambridge University Press, Cambridge, UK, 2001.
[32] P. Goldreich, P. Kumar, "Wave generation by turbulent convection", *Astrophys. J.* **363** (1990), no. 2, p. 694-704.
[33] D. Lecoanet, E. Quataert, "Internal gravity wave excitation by turbulent convection", *Mon. Not. R. Astron. Soc.* **430** (2013), no. 3, p. 2363-2376.
[34] D. M. Bowman, S. Burssens, M. G. Pedersen, C. Johnston, C. Aerts, B. Buysschaert, M. Michielsen, A. Tkachenko, T. M. Rogers, P. V. Edelmann *et al.*, "Low-frequency gravity waves in blue supergiants revealed by high-precision space photometry", *Nat. Astron.* **3** (2019), no. 8, p. 760-765.
[35] P. Edelmann, R. Ratnasingam, M. Pedersen, D. Bowman, V. Prat, T. Rogers, "Three-dimensional simulations of massive stars. I. Wave generation and propagation", *Astrophys. J.* **876** (2019), no. 1, p. 4.
[36] D. Bowman, "What physics is missing in theoretical models of high-mass stars: new insights from asteroseismology", *preprint*, arXiv:1912.12653 (2019).
[37] T. Crueger, M. A. Giorgetta, R. Brokopf, M. Esch, S. Fiedler, C. Hohenegger, L. Kornblueh, T. Mauritsen, C. Nam, A. K. Naumann *et al.*, "ICON-A, the atmosphere component of the ICON earth system model: II. Model evaluation", *J. Adv. Model. Earth Syst.* **10** (2018), no. 7, p. 1638-1662.
[38] J. A. Anstey, J. F. Scinocca, M. Keller, "Simulating the QBO in an atmospheric general circulation model: Sensitivity to resolved and parameterized forcing", *J. Atmos. Sci.* **73** (2016), no. 4, p. 1649-1665.
[39] L.-A. Couston, D. Lecoanet, B. Favier, M. Le Bars, "Shape and size of large-scale vortices: A generic fluid pattern in geophysical fluid dynamics", *Phys. Rev. Res.* **2** (2020), no. 2, article ID 023143.
[40] S. Labrosse, "Thermal evolution of the core with a high thermal conductivity", *Phys. Earth Planet. Inter.* **247** (2015), p. 36-55.
[41] E. Jaupart, B. Buffett, "Generation of MAC waves by convection in Earth's core", *Geophys. J. Int.* **209** (2017), no. 2, p. 1326-1336.

Comptes Rendus
Physique
2020, 21, nº 2, p. 165-175
https://doi.org/10.5802/crphys.28

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

From the geometry of Foucault pendulum to the topology of planetary waves

De la géométrie du pendule de Foucault à la topologie des ondes planétaires

Pierre Delplace[*, a] **and Antoine Venaille**[a]

[a] Univ Lyon, Ens de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, F-69342 Lyon, France

E-mails: pierre.delplace@ens-lyon.fr (P. Delplace), antoine.venaille@ens-lyon.fr (A. Venaille)

Abstract. The physics of topological insulators makes it possible to understand and predict the existence of unidirectional waves trapped along an edge or an interface. In this review, we describe how these ideas can be adapted to geophysical and astrophysical waves. We deal in particular with the case of planetary equatorial waves, which highlights the key interplay between rotation and sphericity of the planet, to explain the emergence of waves which propagate their energy only towards the East. These minimal ingredients are precisely those put forward in the geometric interpretation of the Foucault pendulum. We discuss this classic example of mechanics to introduce the concepts of holonomy and vector bundle which we then use to calculate the topological properties of equatorial shallow water waves.

Résumé. La physique des isolants topologiques permet de comprendre et prédire l'existence d'ondes unidirectionnelles piégées le long d'un bord ou d'une interface. Nous décrivons dans cette revue comment ces idées peuvent être adaptées aux ondes géophysiques et astrophysiques. Nous traitons en particulier le cas des ondes équatoriales planétaires, qui met en lumière les rôles clés combinés de la rotation et de la sphéricité de la planète pour expliquer l'émergence d'ondes qui ne propagent leur énergie que vers l'est. Ces ingrédients minimaux sont précisément ceux mis en avant dans l'interprétation géométrique du pendule de Foucault. Nous discutons cet exemple classique de mécanique pour introduire les concepts d'holonomie et de fibré vectoriel que nous utilisons ensuite pour le calcul des propriétés topologiques des ondes équatoriales en eau peu profonde.

Keywords. Waves, Coriolis force, Chern numbers, Geometric phase, Astrophysical and geophysical flows.

Mots-clés. Ondes, Force de Coriolis, Nombre de Chern, Phase géométrique, Flots géophysiques et astrophysiques.

* Corresponding author.

1. The renewal of topological waves

As recalled by Michael Berry [1], the investigation of the topological properties of waves started during the "miraculous 1830s" with the discovery of their singularities: the singularity of the intensity explaining the emergence rainbows; the singularity of the phase, as amphidromic points[1] discovered at that time in the North sea; and the singularity of the polarization whose theoretical prediction led to the observation of the conical refraction in optics. Topology of waves was enriched in the late seventies, for instance with the discovery of wavefront dislocations of water waves that emerge when scattered by a vortex, thus providing a classical analog to the quantum Aharonov–Bohm effect [2]. This example anticipates how topological properties of quantum wave functions may inspire the search for novel topological properties of classical waves. Indeed, for the last ten years, it was realized that topological properties similar to that of the integer quantum Hall effect and of the recently discovered topological insulators could be engineered in metamaterials with classical waves of various kinds, from optics [3] to mechanics [4, 5] and acoustics [6, 7]. These topological properties are related to phase singularities of the complex eigenstates of the system in a parameter (or reciprocal) space, and translate in real space as the existence of trapped boundary modes that can be used to guide energy, through the celebrated *bulk-boundary correspondence* [8]. These confined states are often referred to as *topological modes*.

Coincidentally, it was during this same "miraculous" 1830 decade that Gaspard-Gustave Coriolis formalised the celebrated inertial force that nowadays bears his name. A spectacular manifestation of this force is revealed by the slow deviation of the Foucault pendulum, an effect that can precisely be apprehended with a geometrical approach [9]. As we shall see below, the Coriolis force is also involved in a singularity of geophysical fluid waves, associated to the twisting of eigenmodes around degeneracy points in their dispersion relation. This topological property is closely related to the geometric interpretation of the Foucault pendulum. It manifests itself through the existence of peculiar equatorially trapped eastward oceanic and atmospheric waves, which bear strong formal similarities with boundary states of a topological material. It was shown over the last few years that topological waves are indeed ubiquitous in natural systems, with application to equatorial dynamics [10], astroseismology [11], plasma [12], or active matter [13, 14].

We review here the recent input of topological tools inherited from topological insulators to these geophysical and astrophysical waves. We put emphasis on equatorial waves, which highlight the crucial role of Earth rotation and curvature. We propose to use the Foucault pendulum as a starting point to introduce key notions of geometrical properties induced by a rotating planet, and then use these tools to address the topology of equatorial waves.

2. Coriolis force-induced geometrical effects

2.1. *Coriolis force on Earth*

Coriolis force is an inertial force perceived by an object of velocity $\mathbf{v}$ in a rotating frame of reference. Its effect is to deviate the object's trajectory in a direction perpendicular to both $\mathbf{v}$ and the rotation vector $\mathbf{\Omega}$, i.e. it reads $\mathbf{F}_c = 2M\mathbf{v}\times\mathbf{\Omega}$ where M is the mass of the object. Owing to its rotation, Earth naturally induces a Coriolis force for an observer at the surface of the planet. However, its effect is clearly negligible at the scale of a human being: to experience the Coriolis force of only 1% amplitude of our weight, we should run at about 2500 km/h.[2] This is due to

[1] Amphidromic points are phase singularity where tidal amplitude must vanish.

[2] $2v\Omega/g = 10^{-2}$ with $g = 10\ \mathrm{m\cdot s^{-2}}$ and $\Omega = 2\pi/(24\times 60\times 60)\ \mathrm{s^{-1}}$.

the small value of the angular velocity of our planet. One can experience this force in inertial carousels where the visitors are invited to pass each others a ball; they fail because of the strong deviation induced by the fast rotation of the carousel.

The Coriolis force cannot be neglected anymore when the motion occurs over a time scale T comparable to the period of rotation, i.e. when $1/(\Omega T)$ is of order one or smaller. This is well illustrated by the celebrated Foucault pendulum, whose vertical oscillation plane significantly deviates from its initial orientation when oscillating long enough. Similarly, slow motion of typical velocity U in ocean currents and atmospheric winds are strongly affected by the Coriolis force at the scale L of a planet, that is again when $U/(\Omega L)$ is of order one or smaller. In the case $U/(\Omega L) \ll 1$, a striking manifestation of Earth rotation is that currents or winds blow along pressure lines, as horizontal momentum equations are dominated by a balance between Coriolis and pressure forces. This property is used to draw midlatitude weather maps, where cyclones and anticyclones are visualized by using isobars. The long time behavior of the Foucault pendulum and the slow motion of winds or oceanic currents at the scale of a planet thus both reveal the influence of Earth rotation. As we explain below, geometrical tools are also in both cases particularly useful to understand central aspects of their dynamics. It is necessary for that purpose to introduce a second key ingredient: Earth's curvature.

Because of Earth's roundness, the effect of the Coriolis force depends on the latitude. Dynamics at the poles is very much like in the carousel picture, since the plane tangent to the sphere is in that case perpendicular to the angular rotation vector $\mathbf{\Omega}$. The situation is quite different at the Equator, as the local tangent planes contain the planet rotation vector $\mathbf{\Omega}$. Standard descriptions of the Foucault pendulum and planetary waves rely on two assumptions: the motion it nearly horizontal, and in-plane component of the angular rotation $\mathbf{\Omega}$ are neglected. In the Foucault pendulum case, both assumptions are justified in a small amplitude limit. In the geophysical case, this can be justified by considering the limit of vanishing aspect ratio between the vertical fluid layer thickness and horizontal scale of motion [15]. In both cases, only the horizontal components of the Coriolis force matters, and the effect of planet rotation is encoded into the *Coriolis parameter*

$$f = 2\Omega \sin\theta, \tag{1}$$

with θ the latitude. This parameter is twice the projection of the planet rotation vector on the local vertical axis ($2\mathbf{\Omega}\cdot\hat{\mathbf{n}}$). The Coriolis parameter increases from the South pole ($\theta = -\pi/2$) to the North pole ($\theta = \pi/2$), and vanishes at the Equator ($\theta = 0$). The latitudinal variations of the Coriolis parameter f is the second key ingredient to understand important aspects of Foucault pendulum dynamics and geophysical flows at planetary scale. In fact, it plays a central role in the geometrical arguments exposed in the following sections.

2.2. *A geometrical look at the Foucault pendulum*

In January 1851, Léon Foucault observes in his basement the slow but "irresistible" clockwise deviation of the oscillation plane of a two meter long pendulum hung from the ceiling, thus subtly revealing Earth's rotation [16]. Its demonstration was reproduced a month later at the Observatoire de Paris and the next month at the Pantheon in Paris, this time with a steel wire of 67 m and a globe of 28 kg, in order to make visible the Earth's rotation to everyone [17].

This slow deviation of the oscillatory plane can be directly inferred from a standard procedure, when taking into account the Coriolis force acting on the pendulum. In the small amplitude limit, one can safely neglect the vertical velocity of the pendulum, and the horizontal trajectory can then be written in the compact form $z(t) = x(t) + \mathrm{i}y(t)$. Integrating the Newton equations in the limit where the Coriolis parameter is small compared to the pulsation of the pendulum ($f \ll \omega$) then leads to $z(t) = \mathrm{e}^{-\mathrm{i}f/2t}(\alpha \mathrm{e}^{-\mathrm{i}\omega t} + \beta \mathrm{e}^{\mathrm{i}\omega t})$, where the constants α and β depend on the

initial conditions, which implies a slow clockwise deviation of the vertical oscillation plane of the pendulum in the Northern hemisphere ($f > 0$).

The deviation of the Foucault pendulum can also be apprehended from elegant geometrical considerations in an inertial frame of reference [9]. This approach does not require any knowledge of the Coriolis force, and is motivated by the slow motion of Earth rotation relative to that of the pendulum oscillations. This scale separation is referred to as the *adiabatic*[3] limit. Locally, the pendulum trajectory is described in a tangent plane to Earth, as sketched in Figure 1(a). The pendulum inertia tends to simply translate the pendulum trajectory parallel to itself in space. Of course, the pendulum is retained on Earth because of its weight and of the tension of the wire it is attached to. Those constraints guarantee that the pendulum trajectory always remains in a plane tangent to Earth. After one period of pendulum oscillation, Earth has *adiabatically* rotated: the pendulum trajectory has barely changed, but it does not belong any longer to the same tangent plane to the sphere in the inertial frame of reference.

We are then left with the question of how its projected trajectory, that we shall represent with a tangent vector to Earth's surface, can remain *as constant as possible* when continuously moved from a tangent plane to another, that are arranged along a curve on the surface of Earth, more specifically at fixed latitude? This question is known in Mathematics as the one of *parallel transport* of vectors, and was in particular worked through by Tullio Levi-Civita at the beginning of the XXth century in parallel to Einstein's general relativity.

In the pendulum case, parallel transport gives a procedure to displace a vector in $\mathbb{R}^3$ from a tangent plane T_xS^2 to the sphere S^2 at a point x to another tangent plane $T_{x+\mathrm{d}x}S^2$. This procedure tells that, when parallel transporting a vector along a curve, the rate of change of this vector must be normal to the surface, as it is so in the Euclidian planar geometry. For instance, the blue vector in the tangent plane to the North pole in Figure 1(a) is parallel transported along a red curve (black vector) so that their difference (green vector) is normal to the surface.

Vector fields of parallel transported tangent vectors are illustrated in Figure 1 along different curves. These are intuitively obtained when the curve corresponds to a geodesics, e.g. along the Equator or along a Meridian (see Figure 1(a)). Perhaps less intuitively, a vector parallel-transported along an arbitrary cycle generally does not return identical to itself, but makes an angle with the initial vector.

As for the Foucault pendulum, we only care about parallel transport of vectors along a longitude. The particular case of the Equator ($\theta = 0$), that is a geodesic, follows from the discussion above: in that case, a local observer contemplating the pendulum in the tangent planes does not see any deviation of its oscillations. It is thus worth stressing that the deviation of the Foucault pendulum is a consequence of the interplay between Earth rotation and Earth local curvature: for instance, there would not be any deflection on a rotating cylinder. The deflection angle ϕ_F that the pendulum makes between its initial and final orientations after one day depends on the latitude θ. It can be easily obtained geometrically, by noting that the collection of planes tangent to the sphere along a longitude are also tangent to a cone (see Figure 1(b)). By unfolding and flattening this cone, the parallel transport of vectors reverts to the usual and intuitive one of Euclidian geometry. Crucially, this operation requires to cut the cone, and we choose the cut position to coincide with the initial and final position of the pendulum (blue line in Figure 1(b)). It becomes clear that, to be parallel transported along a longitude, a tangent vector in the Northern Hemisphere has to rotate clockwise, as observed by Foucault. Moreover, it follows from an elementary geometrical analysis that it acquires an angle $2\pi\sin\theta$ after a full cycle. After one day of oscillations, the pendulum thus makes a deflection angle $\phi_F = 2\pi(1-\sin\theta)$ with its initial orientation.

[3] This adiabatic limit is not related to heat transfers in the system.

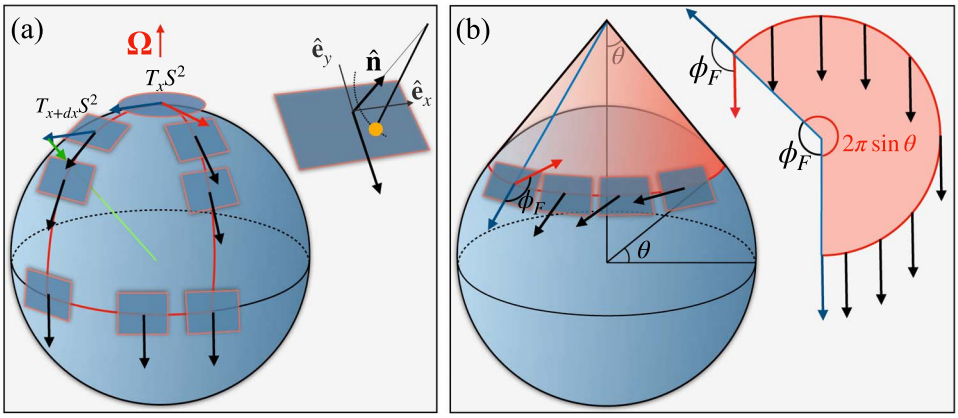


Figure 1. (a) Parallel transport of a tangent vector along geodesics. After a close path, the transported vector (red) differs from the initial one (blue) by an angle. Inset: Foucault pendulum; its trajectory (dashed), defines the black arrow once projected onto the tangent plane. (b) Parallel transport of a tangent vector along a close path (red) at fixed latitude θ, that continuously rotates in its tangent plane. Inset: unfold and flattened red cone.

To summarize, the mismatch angle ϕ_F of the Foucault pendulum follows from two ingredients: (1) parallel transport along a longitude, that is due to the adiabatic rotation of Earth, and (2) the curvature of the surface. More generally, such a phase mismatch of a parallel transported vector over a loop on a surface reads

$$\phi_F = \int_\Sigma \kappa \, \mathrm{d}S \tag{2}$$

where κ is the gaussian curvature of the surface and Σ is the surface enclosed by the loop. Considering Earth as a perfect sphere of radius R, this curvature is simply $\kappa = 1/R^2$, and one recovers the result aforementioned which is nothing but the solid angle delimited by the close path along a circle of latitude.

It has been found that the polarization of seismic shear waves travelling over the Earth slowly rotates similarly to the Foucault pendulum [18]. In the following we present another consequence of Earth rotation and sphericity on planetary waves that is related to a topological number.

2.3. *From geometrical phases to topological numbers for waves*

The phase accumulated by a system over a cycle is a generic property of vector bundles, called holonomy. A vector bundle can be seen as a continuous collection of vector spaces parametrized over a close manifold. In Foucault pendulum case, the collection of vector space are tangent planes $\mathbb{R}^2$, and the base space is the sphere S^2. Vector bundles are mathematical objects that appear in physics for instance when a gauge freedom is involved. In quantum mechanics, the local choice of the phase of the wavefunction is an important example of gauge freedom that gives rise to $U(1)$-vector bundles. There, quantum eigenstates can acquire a geometrical phase, known as the Berry phase [19, 20], when a periodic modulation of the system is performed adiabatically. There is thus a conceptual common root between the quantum holonomy, and that of the Foucault pendulum. However, the nature of the fiber bundle involved is different, and the curvatures used to describe local geometrical properties of these bundles are also different. In the Foucault pendulum case, the Gaussian curvature was used in (2). In the quantum case, a different (two-form) curvature $\mathscr{F}^{(n)}(\boldsymbol{\lambda})$ is employed, called Berry curvature, which is a property of parametrized

eigenstates $\psi_n(\boldsymbol{\lambda})$ with $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots)$ a point in parameter space. The Berry curvature is a physical observable that was measured in of quantum [21] and classical [22] systems. In particular, it appears in the semiclassical equations of motion of electronic wavepackets by yielding a correction to the group velocity to that predicted from the dispersion relation alone. A similar effect is currently investigated in geophysical ray tracing [23].

Similarly to the phase mismatch (2), the Berry phase results from the integration of Berry curvature, which formally reads

$$\phi_B^{(n)} = \int_\Sigma \mathscr{F}^{(n)} \quad \mathscr{F}^{(n)} = \sum_{i,j} \mathrm{i} \left(\frac{\partial \psi_n^\dagger}{\partial \lambda_i} \frac{\partial \psi_n}{\partial \lambda_j} - \frac{\partial \psi_n^\dagger}{\partial \lambda_j} \frac{\partial \psi_n}{\partial \lambda_i} \right) \mathrm{d}\lambda_i \wedge \mathrm{d}\lambda_j. \tag{3}$$

A pedagogical model introduced by Berry to illustrate this geometrical phase consists of a quantum spin $\hat{\mathbf{S}}$ coupled to a slowly varying classical magnetic field $\mathbf{B}$ [19]. The dynamics is encoded in the Hamiltonian $\hat{H} = \mu \mathbf{B} \cdot \hat{\mathbf{S}}$ where μ is a constant involving the gyromagnetic ratio. Spin eigenstates are denoted by $|m\rangle$, with $m = \{-S, -S+1, \ldots, S\}$. When the orientation of $\mathbf{B}$ is varied *adiabatically* along a close path, a spin eigenstate $|m\rangle$ acquires a Berry phase $\phi_B^{(m)} = -m\Omega_s$ where Ω_s is the solid angle drawn by $\mathbf{B}$, by analogy with the deflection angle of the Foucault pendulum in Figure 1(b).

When Σ consists in the entire close manifold base space, the holonomy along its boundary is meaningless. Still, the integration of the Berry curvature is meaningful, and is actually an integer-valued topological index that counts the number of singularities of the $U(1)$-vector bundle. For instance, in (2), the integration of the Gaussian curvature over an orientable close surface Σ is the Euler–Poincaré integer number $(1/2\pi) \int_\Sigma \kappa \, \mathrm{d}S = 2(1-g)$ that only depends on the genus g of the surface. For the sphere ($g = 0$), the value 2 then obtained is the number of vortices where any tangent vector field necessarily vanishes. The integration of the Berry curvature over the base space has a quite similar meaning; it corresponds to the number of phase singularities of the complex eigenstate. This topological number, called the (first) Chern number, was introduced in physics in 1982 to explain the unexpected remarkable quantization of the transverse conductivity of the quantum Hall effect [24] and was later found to predict the number of unidirectional modes propagating without dissipation along the edge of the sample [8].

Although the topological quantization of the conductivity is a specific property of quantum electronic wavefunctions, the existence of topologically protected unidirectional edge modes was later realized to be an ubiquitous property of wave dynamics across all fields of physics, provided that time-reversal symmetry is broken. This requirement is satisfied by the presence of a perpendicular magnetic field in the Hall effect. It must be engineered differently when dealing with classical neutral waves. The formal similarity between Lorentz and Coriolis forces fruitfully suggests that rotating the system is a natural alternative. A nice illustration of the effect of rotation on mechanical waves was realized with a lattice of gyroscopes coupled by springs [4]. A regime of couplings can be found such that when a gyroscope is excited at the edge of the lattice, it generates a wave that propagates clockwise along the boundary, without spreading into the bulk, despite the abruptness of the irregular edge (Figure 2). Another possibility is to rotate the entire system itself rather than its constituants. In the case of a thin fluid layer in a rotating tank, this leads to the emergence of unidirectional trapped modes that propagate along the tank's wall (Figure 2). Such waves are nothing but a lab implementation of coastal Kelvin waves, that propagate along natural coats of lakes and continents due to the Coriolis force induced by Earth's rotation [15]. Note that contrary to the case of gyroscopes, and to other macroscopic experiments designed to exhibit mechanical or fluid topological waves [5, 7], there is no underlying lattice in the rotating fluid case, which is not without causing important conceptual issues concerning the topological nature of these boundary waves [25–27].

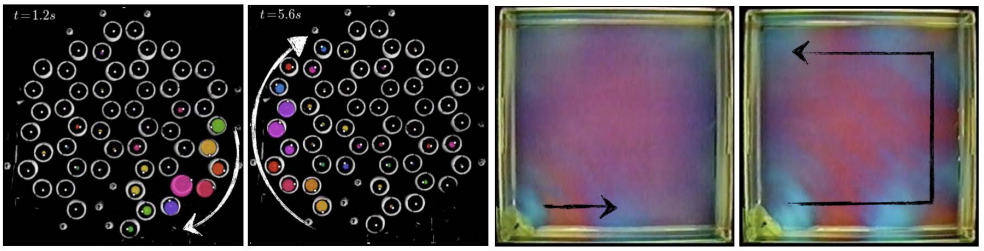


Figure 2. Left: clockwise mechanical edge mode propagating in an array of coupled gyroscopes, reproduced from [4]. Colors encode the phase of the gyroscopes. Right: laboratory realization of a coastal Kelvin wave in a rotating tank, reproduced from [28]. Colors encodes variations in the fluid layer thickness. Snapshots are taken at two successive times to visualize wave propatation.

2.4. *Topological equatorial waves*

Equatorial atmospheric and oceanic waves are another emblematic example of unidirectional trapped modes in geophysics. The Equator around which these waves are trapped plays a role analogous to that of a coast for the Kelvin waves, except that the Equator is not a boundary of the fluid. Nevertheless, it can somehow be interpreted as a kind of interface between two Hemispheres where the Coriolis parameter f defined in (1) has opposite sign. Because time-reversal symmetry is broken by the Coriolis parameter f, one may expect peculiar topological waves at the interface. Indeed, he problem of a continuous system with such a smooth interface turns out to be a convenient framework to show the topological origin of confined unidirectional waves, in the sense that their number is precisely given by the Chern invariant.

Equatorial waves at planetary scale are well described by the linearized shallow water model (see Figure 3(a)). This textbook model describes an incompressible fluid whose thickness is much smaller than the horizontal length scale of motion. In this limit, hydrostatic balance holds in the vertical direction, and the horizontal velocity field $\mathbf{v}(\mathbf{r})$ is depth independent. The dynamics then follows from mass and horizontal momentum conservation. Shallow water waves are solutions of this model linearized around a state of rest. In the case where the dynamics takes place on a flat surface, it can be written as the following Hermitian eigenvalue problem, quite similarly to a quantum mechanical problem

$$\begin{pmatrix} 0 & -\mathrm{i}f(y) & k_x \\ \mathrm{i}f(y) & 0 & \mathrm{i}\partial_y \\ k_x & \mathrm{i}\partial_y & 0 \end{pmatrix} \begin{pmatrix} \tilde{v}_x \\ \tilde{v}_y \\ \tilde{\eta} \end{pmatrix} = \tilde{\omega}(k_x) \begin{pmatrix} \tilde{v}_x \\ \tilde{v}_y \\ \tilde{\eta} \end{pmatrix}. \tag{4}$$

The two horizontal velocity components $(\tilde{v}_x, \tilde{v}_y)$ are coupled to the free surface elevation $\tilde{\eta}$. Time unit has been chosen so that the intrinsic celerity $c = \sqrt{gH}$ has been set to unity, with g the standard gravity and H the averaged layer thickness (see Figure 3(a)). In (4), k_x is the eastward wave vector component while y refers to the direction pointing to the North pole. The function $f(y)$ accounts for the variations with latitude of the Coriolis parameter.

Eigenmodes of (4) are tricky to derive analytically for an arbitrary profile $f(y)$, and becomes even more complicated when curvature effects are taken into account. A useful and standard simplification called the β-plane approximation consists in linearizing the Coriolis parameter around the Equator, i.e. $f(y) = \beta y$. This approximation was fruitfully used by Taroh Matsuno in 1966 to derive the frequency spectrum shown in Figure 3(b) [29]. This spectrum, besides being

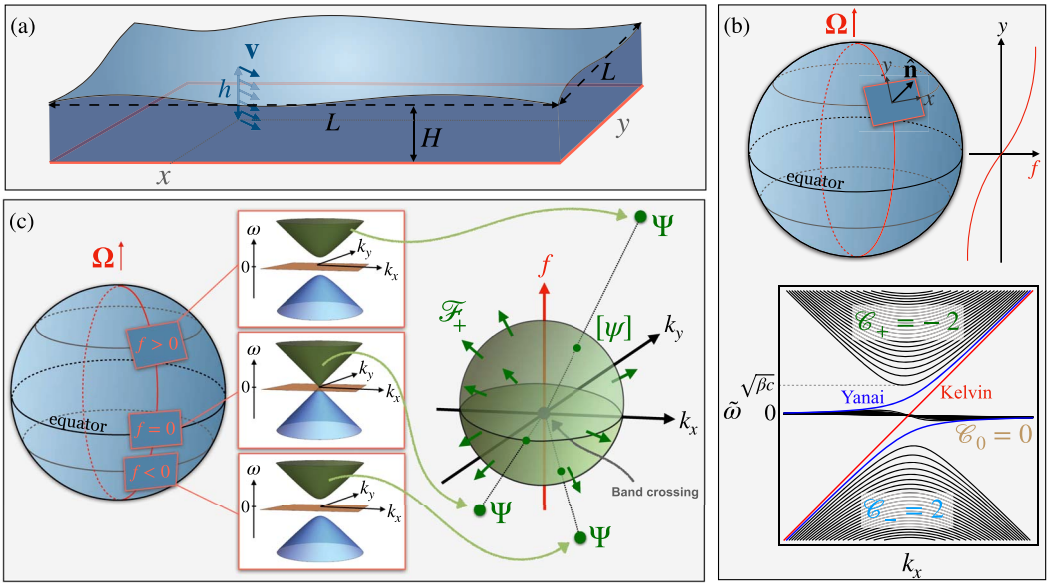


Figure 3. (a) Shallow water model. Thickness H is much smaller than horizontal length scale L. (b) Upper panel: variation of the Coriolis parameter with latitude on the sphere. Lower panel: dispersion relation of shallow water waves computed by T. Matsuno in 1966 under the beta plane approximation $f = \beta y$. Notice the presence of two unidirectional modes filling the gap between low frequency Rossby waves and high frequency Poincaré waves. Those are the equatorial Kelvin and Yanai waves. (c) Dispersion relation of shallow water waves with constant f, for different tangent planes to the rotating sphere. The gap vanishes when $f = 0$. In that case, a degeneracy point occurs at the origin in (k_x, k_y) plane. The right panel is a schematic representation of the eigenmode bundles in parameter space (k_x, k_y, f). The base space in a sphere surrounding the degeneracy point. The topological charge of this degeneracy point explains the emergence of equatorial Kelvin and Yanai waves.

discrete, shows that the modes gather in two groups separated by a gap of amplitude $\sqrt{\beta c}$: the low frequency Rossby waves and the high frequency Poincaré waves. Negative eigenvalues are also shown, for reasons that will become clear in the topological analysis below. These modes are simply related to the positive ones by a symmetry $\tilde{\omega}(k_x) \to -\tilde{\omega}(-k_x)$ which is analogous to the particle-hole symmetry in quantum mechanics.

In addition to Rossby and Poincaré wave modes, Matsuno found two additional waves whose dispersions connect the two previous branches by continuously bridging the frequency gap. These two modes, called equatorial Kelvin and Yanai waves, share the remarkable property to have a positive group velocity at any wave vector, so that they always propagate energy eastward, in contrast to Rossby and Poincaré waves. Moreover, there is a finite range of frequency in which these modes are the only one to be excited. This phenomenology is precisely that of the edge modes of a two-dimensional topological material where time-reversal symmetry is broken. In topological insulators, the topological properties emerge in the bulk, that is in a simplified edgeless problem where translation symmetry is restored. The situation is a bit different here, since the equatorial problem is already edgeless. However, a "bulk Hamiltonian" can also be assigned to the situation where translation symmetry would be restored in the y direction, by

considering the Coriolis function varying with latitude as a (constant) parameter f, and then Fourier transform with respect to the y coordinate. This leads to the distinct parametrized eigenvalue problem

$$\begin{pmatrix} 0 & -\mathrm{i}f & k_x \\ \mathrm{i}f & 0 & k_y \\ k_x & k_y & 0 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ \eta \end{pmatrix} = \omega(k_x, k_y, f) \begin{pmatrix} v_x \\ v_y \\ \eta \end{pmatrix} \tag{5}$$

where the eigenfrequencies simply consist in three bands $\omega_\pm = \pm\sqrt{k_x^2 + k_y^2 + f^2}$ and $\omega_0 = 0$. Dispersion relations are shown for different values of f in Figure 3(c). The modes $\omega_\pm$ are separated to the flat one ω_0 by a frequency gap of amplitude $|f|$. In other words, at fixed f, the system behaves as a two-dimensional insulator with broken time-reversal symmetry. In a condensed matter context, the standard next move would be to compute the Chern number of each of the bands, by integrating the Berry curvature—obtained from the eigenstates of (5)—over the close manifold parameter space, in that case, the two-dimensional Brillouin zone span by k_x and k_y, that is equivalent to a torus. However, for geophysical waves, and more generally for waves in continuous media, the absence of a Brillouin zone makes this procedure inappropriate. Instead of considering the waves for a single fixed value of f, one can look on the continuous family of solutions in parameter space (k_x, k_y, f).

Three eigenfrequencies and their corresponding eigenvectors, that we shall note $\Psi_\pm$ and Ψ_0, are assigned to each point of the parameter space. Once normalized, the eigenstates do not depend on their distance from the origin in parameter space, but only on the orientation. In other words, they live on a sphere that surrounds the origin. Note that the origin is a very particular point of the problem, since there the three bands touch. Bands touching points are known to be a source of Berry curvature which, once integrated over the sphere that surrounds it, yields a quantized Chern number $\mathscr{C} = (1/2\pi)\int_{S^2} \mathscr{F} \in \mathbb{Z}$. Surprisingly, this shallow water problem is formally analog to that of the quantum spin coupled to a classical magnetic field briefly introduced in Section 2.3. Indeed, the Hermitian matrix in (5) playing the role of the quantum Hamiltonian takes the form $\mathbf{B}\cdot\hat{\mathbf{S}}$ where $\mathbf{B} = (k_x, k_y, f)$ and $\hat{\mathbf{S}}$ is a spin-1 operator. The Chern numbers can then be inferred directly from the Berry phase when the solid angle is that of the entire sphere, i.e. $\mathscr{C}_m = (1/2\pi)(-m)4\pi$ that gives $(2, 0, -2)$ for the three bands [10].

Importantly, these Chern numbers are directly related to a spectral property of the equatorial frequency spectrum $\tilde{\omega}(k_x)$: a band n of this spectrum gains $-\mathscr{C}_n$ modes when increasing k_x, as emphasized in Figure 3(b). It follows that modes have to transit from band to band; this is referred to as a *spectral flow*. The set of Chern numbers found in the parameterized problem (5) thus imposes that two modes of positive group velocity have to fill the gap, hence the topological origin of the Kelvin and the Yanai waves [30]. This topological information does not depend on the shape of the Coriolis function $f(y)$, which can be continuously modified without changing the spectral flow. The existence of these two unidirectional modes is thus guaranteed beyond the β plane approximation.

First observations of equatorial Kelvin and Yanai waves in Earth atmosphere and oceans came almost simultaneously to their theoretical predictions, in the late sixties and the seventies. Their discovery has been central to our modern understanding of tropical dynamics. Indeed, they are involved in the most important internal modes of climate variability in the equatorial area. For instance, equatorial Kelvin waves travelling across the Pacific ocean are precursors of El Niño events, atmospheric equatorial Kelvin waves are often invoked in the explanation for superrotation phenomenon reported in other planets, including exoplanets, and the Quasi-Biennial Oscillations of stratospheric winds are partly driven by atmospheric Yanai and equatorial Kelvin waves in the middle atmosphere, see e.g. [15].

3. Conclusion and prospects

Unidirectional equatorial waves can be predicted from topological considerations without solving the rotating shallow water model on the sphere, just as the deviation of the Foucault pendulum could be inferred from geometrical considerations only, without deriving the equations of motion. Besides, it highlights the important role of broken symmetries in a system at hand, and of interfaces induced by the symmetry breaking parameter. This topological approach is not restricted to the shallow water model, nor to systems subject to the Coriolis force. Similar arguments can be used to predict in a simple way the existence of remarkable waves in various complicated continuous systems such as astrophysical and geophysical flows. The strategy consists in searching for band crossing points in dispersion relations, identifying the three-dimensional parameter space in which these degeneracies occur, and computing the topological charge describing eigenmodes twisting around these degeneracy points in parameter space. It is then possible to use these information to predict the emergence of trapped modes along an interface in physical space. Beyond the equatorial case, this method has already been used to predict the emergence of Lamb-like waves in compressible-stratified fluids [11], with possible applications in astroseismology, or to peculiar plasma waves [12], with possible experimental realizations.The discovery of other topological waves in seismology, astrophysics, and geophysical fluid dynamics is at hand.

References

[1] M. Berry, “Making waves in physics”, *Nature* **403** (2000), p. 21.
[2] M. V. Berry, R. G. Chambers, M. D. Large, C. Upstill, J. C. Walmsley, “Wavefront dislocations in the Aharonov–Bohm effect and its water wave analogue”, *Eur. J. Phys.* **1** (1980), p. 154.
[3] T. Ozawa, H. M. Price, A. Amo, N. Goldman, M. Hafezi, L. Lu, M. C. Rechtsman, D. Schuster, J. Simon, O. Zilberberg, I. Carusotto, “Topological photonics”, *Rev. Mod. Phys.* **91** (2019), article ID 015006.
[4] L. M. Nash, D. Kleckner, A. Read, V. Vitelli, A. M. Turner, W. Irvine, “Topological mechanics of gyroscopic metamaterials”, *Proc. Natl Acad. Sci. USA* **112** (2015), no. 47, p. 14495-14500.
[5] R. Süsstrunk, S. D. Huber, “Observation of phononic helical edge states in a mechanical topological insulator”, *Science* **349** (2015), no. 6243, p. 47-50.
[6] X. Zhang, M. Xiao, Y. Cheng, M.-H. Lu, J. Christensen, “Topological sound”, *Commun. Phys.* **1** (2018), no. 1, p. 97.
[7] A. Souslov, B. C. van Zuiden, D. Bartolo, V. Vitelli, “Topological sound in active-liquid metamaterials”, *Nat. Phys.* **13** (2017), p. 1091.
[8] Y. Hatsugai, “Chern number and edge states in the integer quantum Hall effect”, *Phys. Rev. Lett.* **71** (1993), p. 3697-3700.
[9] M. Berry, “The quantum phase, five years after”, in *Geometric Phases on Physics* (A. Shapere, F. Wilczek, eds.), World Scientific, Singapore, 1989, p. 7.
[10] P. Delplace, J. Marston, A. Venaille, “Topological origin of equatorial waves”, *Science* **358** (2017), no. 6366, p. 1075-1077.
[11] M. Perrot, P. Delplace, A. Venaille, “Topological transition in stratified fluids”, *Nat. Phys.* **15** (2019), no. 8, p. 781-784.
[12] J. B. Parker, J. Marston, S. M. Tobias, Z. Zhu, “Topological gaseous plasmon polariton in realistic plasma”, *Phys. Rev. Lett.* **124** (2020), no. 19, article ID 195001.
[13] S. Shankar, M. Bowick, M. C. Marchetti, “Topological sound and flocking on curved surfaces”, *Phys. Rev. X* **7** (2017), no. 3, article ID 031039.
[14] A. Souslov, K. Dasbiswas, M. Fruchart, S. Vaikuntanathan, V. Vitelli, “Topological waves in fluids with odd viscosity”, *Phys. Rev. Lett.* **122** (2019), no. 12, article ID 128001.
[15] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation*, 2 ed., Cambridge University Press, Cambridge, UK, 2017.
[16] L. Foucault, “Démonstration physique du mouvement de rotation de la Terre au moyen du pendule”, *C. R. Séances Acad. Sci.* **32** (1851), p. 157.
[17] J. Sommeria, “Foucault and the rotation of the Earth”, *C. R. Phys.* **18** (2017), no. 9–10, p. 520-525.
[18] R. Snieder, C. Sens-Schönfelder, E. Ruigrok, K. Shiomi, “Seismic shear waves as Foucault pendulum”, *Geophys. Res. Lett.* **43** (2016), no. 6, p. 2576-2581.
[19] M. V. Berry, “Quantal phase factors accompanying adiabatic changes”, *Proc. R. Soc. Lond. A* **392** (1984), p. 45.

[20] B. Simon, "Holonomy, the quantum adiabatic theorem, and Berry's phase", *Phys. Rev. Lett.* **51** (1983), p. 2167-2170.
[21] G. Jotzu, M. Messer, R. Desbuquois, M. Lebrat, T. Uehlinger, D. Greif, T. Esslinger, "Experimental realization of the topological Haldane model with ultracold fermions", *Nature* **515** (2014), p. 237.
[22] M. Wimmer, H. M. Price, I. Carusotto, U. Peschel, "Experimental measurement of the Berry curvature from anomalous transport", *Nat. Phys.* **13** (2017), no. 6, p. 545.
[23] N. Perez, P. Delplace, A. Venaille, "Manifestation of Berry curvature in geophysical ray tracing", *preprint*, arXiv:2010.05575 (2020).
[24] D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, "Quantized hall conductance in a two-dimensional periodic potential", *Phys. Rev. Lett.* **49** (1982), p. 405-408.
[25] C. Tauber, P. Delplace, A. Venaille, "A bulk-interface correspondence for equatorial waves", *J. Fluid Mech.* **868** (2019), article ID R2.
[26] C. Tauber, P. Delplace, A. Venaille, "Anomalous bulk-edge correspondence in continuous media", *Phys. Rev. Res.* **2** (2020), no. 1, article ID 013147.
[27] G. M. Graf, C. Jud, H. Tauber, "Topology in shallow-water waves: a violation of bulk-edge correspondence", *preprint*, arXiv:2001.00439 (2020).
[28] S. Sakai, I. Lizawa, E. Aramaki, "GFD Online ressources", https://www.gfd-dennou.org/.
[29] T. Matsuno, "Quasi-geostrophic motions in the equatorial area", *J. Meteorol. Soc. Japan. Ser. II* **44** (1966), no. 1, p. 25-43.
[30] F. Faure, "Manifestation of the topological index formula in quantum waves and geophysical waves", *preprint*, arXiv:1901.10592 (2019).

Comptes Rendus
Physique
2020, 21, nº 2, p. 177-183
https://doi.org/10.5802/crphys.16

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Mathematical questions about the computation of eigenvalues of Dirac operators with critical potentials in atomic and molecular physics

Quelques questions mathématiques sur le calcul des valeurs propres des opérateurs de Dirac avec potentiels critiques en physique atomique et moléculaire

Maria J. Esteban [a]

[a] CEREMADE (CNRS UMR no 7534), PSL university, Université Paris-Dauphine, Place de Lattre de Tassigny, F-75775 Paris 16, France
E-mail: esteban@ceremade.dauphine.fr

Abstract. This Note describes various analytical and computational results concerning the calculation of Dirac eigenvalues, or more generally, of operators with gaps. An algorithm based on an abstract theorem characterizing the eigenvalues in gaps was found years ago, but it is only recently that a delicate analysis to identify and study the domains of those operators has allowed to put that algorithm on a firm basis concerning the choice of approximation basis sets, and this both for light and for heavy atoms. The works described here concern joint papers with several collaborators: J. Dolbeault, M. Lewin, M. Loss, E. Séré and M. Vanbreugel.

Résumé. Cette Note présente divers résultats analytiques et numériques concernant le calcul des valeurs propres de l'opérateur de Dirac, ou plus généralement, des opérateurs avec des "gaps" spectraux. Un algorithme basé sur un théorème abstrait caractérisant les valeurs propres dans les écarts a été trouvé il y a des années, mais ce n'est que récemment qu'une analyse délicate pour identifier et étudier les domaines de ces opérateurs a permis de mettre cet algorithme sur une base ferme concernant le choix des approximations, et ceci aussi bien pour les atomes légers que pour les atomes lourds. Les travaux décrits ici concernent des travaux réalisés avec plusieurs collaborateurs : J. Dolbeault, M. Lewin, M. Loss, E. Séré et M. Vanbreugel.

Keywords. Dirac operator, Eigenvalue, Operator domain, Eigenvalue computation, Self-adjoint operator.

Mots-clés. Opérateur de Dirac, Valeur propre, Domaine de l'opérateur, Calcul des valeurs propres, Opérateur auto-adjoint.

2020 Mathematics Subject Classification. Primary 49S05; Secondary 35Q75, 47-08, 65D07, 70-08, 81-08, 81Q05, 81Q10, 81V45, 81V55.

Version française abrégée

Depuis 1997, en collaborant avec un groupe de chercheurs (J. Dolbeault, M. Lewin, M. Loss, E. Séré and M. Vanbreugel), nous nous sommes intéressés à l'étude des propriétés et des valeurs du spectre ponctuel d'opérateurs *à gaps*, et en particulier, d'opérateurs de Dirac. Nous cherchions également à proposer un algorithme de calcul libre des problèmes liés à la pollution spectrale. En effet, il est bien connu que selon le choix des bases de calcul et selon les méthodes utilisées, des états spurieux peuvent apparaître dans les calculs. Pour des références dans la littérature physique et chimique voir par exemple [1–7]. Des analyses et propositions intéressantes peuvent être trouvées dans [8–15].

Après avoir analysé le problème de manière abstraite, nous avons obtenu des applications particulièrement intéressantes pour l'opérateur de Dirac perturbé par un potentiel extérieur électromagnétique. Utilisant une décomposition de l'espace des spineurs proposée par Talman [16], l'application du théorème abstrait nous a permis de trouver un algorithme qui est très performant pour le calcul des valeurs propres d'opérateurs de Dirac, et ceci sans avoir à faire face à l'apparition de valeurs spurieuses. De nombreux résultats dans la littérature proposent des choix de bases finies particulières pour éviter la pollution spectrale. Dans un des derniers travaux présentés dans cette Note, nous montrons que dans notre approche il y a une énorme liberté dans le choix des bases finies, et que dans tous les cas, il n'y aura pas de valeur spurieuse, et que l'algorithme convergera vers les valeurs propres. Des résultats récents s'intéressant à l'étude détaillée des domaines des opérateurs nous ont donc permis de donner une base solide à cet algorithme pas seulement pour les calculs impliquant des atomes légers, mais aussi, et surtout, dans le cas des atomes lourds, qui sont les plus intéressants dans une théorie relativiste.

1. Introduction

In 1928 [17] Paul Dirac derived an operator for quantum electrodynamics, starting from the usual classical expression of the energy of a free relativistic particle of momentum $p \in \mathbb{R}^3$ and mass m,

$$E^2 = c^2|p|^2 + m^2c^4. \tag{1}$$

His aim was to propose a local differential operator of first order with respect to $p = -\mathrm{i}\hbar\nabla$:

$$D_{m,c,\hbar} = -\mathrm{i}c\hbar\boldsymbol{\alpha}\cdot\nabla + mc^2\beta = -\mathrm{i}c\hbar\sum_{k=1}^{3}\boldsymbol{\alpha}_k\partial_k + mc^2\boldsymbol{\beta}, \tag{2}$$

where α_1, α_2, α_3 and β are Hermitian matrices which have to satisfy the following anticommutation relations:

$$\begin{cases} \alpha_k\alpha_\ell + \alpha_\ell\alpha_k = 2\delta_{k\ell}I_4, \\ \alpha_k\beta + \beta\alpha_k = 0, \\ \beta^2 = I_4. \end{cases} \tag{3}$$

It can be proved [18] that the smallest dimension in which (3) can take place is 4 (i.e. α_1, α_2, α_3 and β should be 4×4 Hermitian matrices), meaning that D_c has to act on $L^2(\mathbb{R}^3,\mathbb{C}^4)$. The usual representation in 2×2 blocks is given by

$$\beta = \begin{pmatrix} I_2 & 0 \\ 0 & -I_2 \end{pmatrix}, \quad \alpha_k = \begin{pmatrix} 0 & \sigma_k \\ \sigma_k & 0 \end{pmatrix} \quad (k = 1,2,3),$$

where the Pauli matrices are defined as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In which follows we will always work in the framework of atomic units, where we can assume that the constants $\hbar, c$ and m are equal to 1. For this particular choice of units, the unperturbed Dirac operator will be denoted by H_0:

$$H_0 := -\mathrm{i}\boldsymbol{\alpha}\cdot\nabla + \beta.$$

It is well known that the Dirac operator is essentially self-adjoint and that its spectrum is all continuous spectrum and equal to

$$(-\infty,-1]\cup[1,+\infty).$$

Of course the main interesting situation for atomic or molecular physics is to consider the Dirac operator perturbed by an external electromagnetic potential, corresponding to the interaction of the electron(s) with other charged particles and its evolution under the action of an external magnetic field. Without magnetic field, one is thus interested in looking at operators of the form H_0+V, and with magnetic field, one has to replace the usual gradient operator ∇ by the magnetic gradient $\nabla_{\mathbf{A}} = \nabla + \mathrm{i}\mathbf{A}$, where $\mathbf{A}$ is a potential related to the magnetic field $\mathbf{B}$ by $\mathbf{B} = \operatorname{curl}\mathbf{A}$. In this case, the magnetic Dirac operator is denoted by $H_{\mathbf{A}}$:

$$H_{\mathbf{A}} := -\mathrm{i}\boldsymbol{\alpha}\cdot\nabla_{\mathbf{A}} + \beta.$$

If now we consider the perturbed operator $H_{\mathbf{A}}+V$ and the potential V is for instance Coulomb-like, eigenvalues of $H_{\mathbf{A}}+V$ appear in the gap of the essential spectrum and those eigenvalues correspond to discrete electronic states in the atom or the molecule. Computing eigenvalues of operators in gaps is notoriously difficult. For instance, the ground state, or better said, the minimal electronic eigenvalue, cannot be found by a simple minimization procedure.

Another serious issue is that depending on the choice of the approximating basis sets and on the computing algorithm, spurious eigenvalues not converging towards exact eigenvalues can appear. For examples and discussions on this issue see for instance [1–7] in the Physics and Chemistry literature. Interesting analysis and propositions to deal with this issue can be found in [8–15].

In [19–22], under adequate assumptions, variational min–max formulas were provided for the eigenvalues in gaps of self-adjoint operators A. These formulas are based on a decomposition $\mathscr{H} = \Lambda^+\mathscr{H}\oplus\Lambda^-\mathscr{H}$ given by two orthogonal projectors $\Lambda^\pm$ of the ambient Hilbert space $\mathscr{H}$, and take the general form

$$\lambda^{(k)} = \inf_{\substack{W\subset F^+\\ \dim(W)=k}} \sup_{\psi\in W\oplus F^-} \frac{\langle\psi, A\psi\rangle}{\|\psi\|^2}. \tag{4}$$

Here, $F^\pm = \Lambda^\pm F$, with F a dense subspace of the domain of $\mathscr{H}$, such that the quadratic form $\langle\psi, A\psi\rangle$ is well-defined on $F^+\oplus F^-$. See also the recent articles [23–25]. Based on a simple and very useful orthogonal decomposition proposed by Talman [16], it was proved by Dolbeault, Esteban and Séré in [22] for the case without magnetic field, and later in [26, 27] by Dolbeault, Esteban and Loss for the case of an external constant magnetic field, that the above abstract result implies that for electrostatic potentials having at worst singularities of the Coulomb type, $-\nu/|x|$, with $0<\nu\le 1$, the eigenvalues of the operator $H_{\mathbf{A}}+V$ can be found by the following simple and computable procedure: for functions $\varphi\in L^2(\mathbb{R}^3,\mathbb{C}^2)$, consider the quadratic form

$$Q_{\mathbf{A},V,\lambda}(\varphi) := \int_{\mathbb{R}^3}\left(\frac{|\boldsymbol{\sigma}\cdot\nabla_{\mathbf{A}}\varphi|^2\,\mathrm{d}x}{1+\lambda-V} + (1-\lambda+V)|\varphi|^2\right)\mathrm{d}x$$

which is decreasing in λ. If the electrostatic potential is not very large, and in particular, if the singularities of V are not worse than $-\nu/|x|$, with $0<\nu\le 1$, condition which is necessary to prove that the operator $H_{\mathbf{A}}+V$ can be defined as a self-adjoint operator in a physically meaningful manner (see Esteban–Loss [26, 28]), then, the quadratic form $Q_{\mathbf{A},V,\lambda}$ is positive for λ in some interval $[-\infty, b)$, $b\in(-1,1)$. Moreover the first (smallest) eigenvalue of $H_{\mathbf{A}}+V$ will be the smallest

λ for which there is a function φ satisfying $Q_{\mathbf{A},V,\lambda}(\varphi) = 0$. More concretely, and equivalently, let $T(\lambda)$ be the operator defined via the quadratic form which acts on 2-spinors:

$$(\varphi, T(\lambda)\varphi) := \int_{\mathbb{R}^3} \left(\frac{|\boldsymbol{\sigma} \cdot \nabla_{\mathbf{A}} \varphi|^2}{\lambda + 1 - V} + (1 - \lambda + V)|\varphi|^2 \right) \mathrm{d}x$$

and consider its first eigenvalue, $\mu_1(\lambda)$. Since $T(\lambda)$ is monotone decreasing with respect to λ, there exists a unique λ_1 such that $\mu_1(\lambda_1) = 0$. Then λ_1 is the smallest eigenvalue of $H_{\mathbf{A}} + V$ in the gap $(-1, 1)$. Furthermore, for every positive integer k, if λ_k is the unique root of the equation $\mu_k(\lambda) = 0$, then λ_k is the k-th eigenvalue of $H_{\mathbf{A}} + V$ in the gap $(-1, 1)$, counted with multiplicity. All these results were proved in [22] in all cases where the operator $H_{\mathbf{A}} + V$ is self-adjoint and when $\mathbf{A}$ corresponds to a constant magnetic field and when V is not too large (the concrete conditions are stated in the theorems contained in [22]).

It is quite simple to propose now an algorithm to compute the eigenvalues of $H_{\mathbf{A}} + V$. For that purpose let us choose an n-dimensional space of functions from $\mathbb{R}^3$ to $\mathbb{C}^2$ and generated by $\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$. Let $T_n(\lambda)$ be the $n \times n$ matrix whose elements are given by

$$T_n^{i,j}(\lambda) = \int_{\mathbb{R}^3} \left(\frac{(\sigma \cdot \nabla_{\mathbf{A}} \varphi_i, \sigma \cdot \nabla_{\mathbf{A}} \varphi_j)}{\lambda + 1 - V} + (1 - \lambda + V)(\varphi_i, \varphi_j) \right) \mathrm{d}x.$$

Let $\mu_1^n(\lambda)$ the smallest eigenvalue of $A_n(\lambda)$. Then, the unique zero of the map $\lambda \mapsto \mu_1^n(\lambda)$, λ_1^n, is an approximation of the first eigenvalue of $H_{\mathbf{A}} + V$ in the gap $(-1, 1)$ if the set $\{\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots\}$ generates a space F like the one present in the above abstract theorem about eigenvalues in gaps. Or more generally, if the set $\{\varphi_1, \varphi_2, \ldots, \varphi_n, \ldots\}$ generates the domain of the operator $H_{\mathbf{A}} + V$. Using this algorithm, in [29], in collaboration with Dolbeault, Séré and Vanbreugel, we computed the ground state and the ground state energy for an electron in the electrostatic field created by light and heavy nuclei (H, He^+, Cr^{23+} and Th^{89+}). The algorithm converged beautifully, without the presence of spurious eigenvalues that are often present in Dirac eigenvalues computations. In this case we chose the functions φ_n as Hermite polynomials. In [30], with Dolbeault and Séré we performed the same kind of computations for a diatomic configuration, with both light and heavy atoms in two separate locations (H_2^+ and Th_2^{179+}). These computations were done choosing B-spline functions for the functions φ_n and in cylindrical coordinates. Again the computations ran perfectly, and the values obtained in these computations fitted perfectly experimental data and data obtained by using other algorithms. In [27], with Dolbeault and Loss we also made computations for the magnetic case, again for light and heavy nuclei, and once again the results fitted extremely well existing results obtained by other means. In particular we were able to produce new eigenvalue approximations for heavy atoms in cases that had not been dealt with before.

But, even if the above computations were excellent and the algorithm was robust and very efficient, there was a deep problem behind them. Indeed, for electrostatic potentials involving Coulomb-like singularities $-\nu/|x|$, with $0 < \nu < \sqrt{3}/2$, the domain of the Dirac–Coulomb operator $H_0 - \nu/|x|$ is equal to the Sobolev space $H^1(\mathbb{R}^3, \mathbb{C}^4)$ defined as the 4-spinors which are square integrable and such that all their first derivatives are also square integrable. But when $\nu > \sqrt{3}/2$ it is known that the domain is contained in $H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$ and contains $H^1(\mathbb{R}^3, \mathbb{C}^4)$, but it is not equal to any of these two spaces. In the concrete case of Coulomb potentials, the domain could be computed explicitly, but not for other potentials having the same singularities, but not being exactly Coulomb-like. It was therefore delicate to use the above abstract theorem to derive the algorithm and use it with basis sets that maybe were not generating the domain of the operator. This is why recently in [31–33]), with Lewin and Séré we have addressed this issue, describing in full detail the domains of the Dirac–Coulomb-like operators for large ν, that is, for $\sqrt{3}/2 \leq \nu \leq 1$ and trying to see which properties are necessary for the basis sets for the above algorithm to converge to the eigenvalues, and not to some numbers above them.

In [31] we described the domains of Dirac–Coulomb-like operators for $0 < \nu \leq 1$. In various papers written in the 70's and 80's we can find proposals of physically meaningful definitions of Dirac–Coulomb operators as self-adjoint operators for $0 < \nu < 1$ and in all cases the domain was shown to be a strict subspace of $H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$. In the end, all those proposals were shown to be equivalent, of course. For full details on all those developments see [18, 34–42]. The limit case $\nu = 1$ is harder, and this was dealt with by Esteban and Loss in [28] by using a novel method to prove self-adjointness for operators with gaps. This was later extended by Arrizabalaga *et al.* in [43, 44]. In the case $\nu = 1$ the domain is not a subspace of $H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$ anymore. This full description of the domain of Dirac–Coulomb-like operators was the first step towards the full understanding of how to interpret the results of the numerical computations done by the above algorithm. In [31] we found an additional result proving that the space $C_c^\infty(\mathbb{R}^3 \setminus \{0\}, \mathbb{C}^2)$ (compactly supported functions which are infinitely derivable in all points except at the origin) is dense in the space of the upper components (2-spinors) of the elements of the domain. This density argument is key in the proof of our main result which states that in the min–max characterization of the eigenvalues (4) of the operator $H_0 - \nu/|x|$, $0 < \nu \leq 1$, the space F can be any among those satisfying

$$C_c^\infty(\mathbb{R}^3 \setminus \{0\}, \mathbb{C}^4) \subset F \subset H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$$

The immediate consequence of this result is that when running the above algorithm we can take basis sets that span any space between $C_c^\infty(\mathbb{R}^3 \setminus \{0\}, \mathbb{C}^4)$ and $H^{1/2}(\mathbb{R}^3, \mathbb{C}^4)$! This creates an impressive flexibility in the choice of the basis sets, ensuring that whichever basis set we take in this class, will lead towards the eigenvalues of $H_0 - \nu/|x|$.

Of course, the above results would not be very useful if they were only applicable to Dirac–Coulomb operators. In [31] we describe the class of electrostatic potentials V to which the above results also apply, the main condition being that they are not too positive, so that they are mainly attractive, and also that if they are singular at some point, the singularity cannot exceed $-\nu/|x|$ for $0 < \nu \leq 1$.

The above recent results settle the question of when and how we can run the above algorithm with guarantee that what we will find in the end will be good approximations of the eigenvalues of perturbed Dirac operators.

More recently, with Lewin and Séré we have gone further and consider the case of several singularities, case that is important in molecular computations. In [32, 33], we have extended the results of [31] to this case. And not only to this case, but to the much more general case of Dirac operators perturbed by general singular measures. In this case, which goes beyond Coulomb singularities, more technically refined arguments are needed to describe the domains and prove the validity of the min–max arguments used to find the corresponding eigenvalues. In the above papers, we have also considered other interesting questions that we have not been able to answer completely concerning, for instance, the dependence of the eigenvalues on the geometric shape of the nuclear distribution. A very simple question is: if we consider two identical nuclei of charge $\nu \in (0, 1/2)$, at a distance $R > 0$, would the energy of an electron be above the energy of an electron in the electrostatic field created by a single nucleus of charge 2ν? In the non-relativistic case the answer is yes. And not only that, actually the ground state energy grows with the distance R. For Dirac operators such results do not exist. Only numerical results have been obtained for instance by Artemyev *et al.* [45] showing that the dependence of the ground state energy with respect to the distance between the two nuclei follows the same pattern in the relativistic and in the non-relativistic cases.

Acknowledgement

This Note has been prepared while the author was being partially supported by the project molQED, Project-ANR-17-CE29-0004 of the French National Research Agency (ANR).

References

[1] G. W. F. Drake, S. P. Goldman, "Application of discrete-basis-set methods to the Dirac equation", *Phys. Rev. A* **23** (1981), p. 2093-2098.

[2] I. P. Grant, "Conditions for convergence of variational solutions of Dirac's equation in a finite basis", *Phys. Rev. A* **25** (1982), p. 1230-1232.

[3] W. Kutzelnigg, "Basis set expansion of the Dirac operator without variational collapse", *Int. J. Quantum Chem.* **25** (1984), p. 107-129.

[4] R. E. Stanton, S. Havriliak, "Kinetic balance: A partial solution to the problem of variational safety in Dirac calculations", *J. Chem. Phys.* **81** (1984), p. 1910-1918.

[5] K. G. Dyall, K. Fægri, "Kinetic balance and variational bounds failure in the solution of the Dirac equation in a finite Gaussian basis set", *Chem. Phys. Lett.* **174** (1990), p. 25-32.

[6] G. Pestka, "Spurious roots in the algebraic Dirac equation", *Chem. Phys. Lett.* **376** (2003), p. 659-661.

[7] V. M. Shabaev, I. I. Tupitsyn, V. A. Yerokhin, G. Plunien, G. Soff, "Dual kinetic balance approach to basis-set expansions for the Dirac equation", *Phys. Rev. Lett.* **93** (2004), article ID 130405.

[8] M. Lewin, E. Séré, "Spectral pollution and how to avoid it", *Proc. Lond. Math. Soc.* **100** (2009), p. 864-900.

[9] M. Levitin, "Spectral pollution and second-order relative spectra for self-adjoint operators", *IMA J. Numer. Anal.* **24** (2004), p. 393-416.

[10] E. B. Davies, M. Plum, "Spectral pollution", *IMA J. Numer. Anal.* **24** (2004), p. 417-438.

[11] L. Boulton, M. Levitin, "On approximation of the eigenvalues of perturbed periodic Schrödinger operators", *J. Phys. A* **40** (2007), p. 9319-9329.

[12] A. C. Hansen, "On the approximation of spectra of linear operators on Hilbert spaces", *J. Funct. Anal.* **254** (2008), p. 2092-2126.

[13] J. Descloux, "Essential numerical range of an operator with respect to a coercive form and the approximation of its spectrum by the Galerkin method", *SIAM J. Numer. Anal.* **18** (1981), p. 1128-1133.

[14] A. Pokrzywa, "Method of orthogonal projections and approximation of the spectrum of a bounded operator", *Studia Math.* **65** (1979), p. 21-29.

[15] A. Pokrzywa, "Method of orthogonal projections and approximation of the spectrum of a bounded operator. II", *Studia Math.* **70** (1981), p. 1-9.

[16] J. D. Talman, "Minimax principle for the Dirac equation", *Phys. Rev. Lett.* **57** (1986), p. 1091-1094.

[17] P. Dirac, "The quantum theory of the electron", *Proc. R. Soc. Lond. A* **117** (1928), p. 610-624.

[18] B. Thaller, *The Dirac Equation*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1992.

[19] M. Esteban, E. Séré, "Existence and multiplicity of solutions for linear and nonlinear Dirac problems", in *Partial Differential Equations and their Applications (Toronto, ON, 1995)*, CRM Proc. Lecture Notes, vol. 12, Amer. Math. Soc., Providence, RI, 1997, p. 107-118.

[20] M. Griesemer, H. Siedentop, "A minimax principle for the eigenvalues in spectral gaps", *J. Lond. Math. Soc.* **60** (1999), p. 490-500.

[21] J. Dolbeault, M. J. Esteban, E. Séré, "Variational characterization for eigenvalues of Dirac operators", *Calc. Var. Partial Differential Equations* **10** (2000), p. 321-347.

[22] J. Dolbeault, M. J. Esteban, E. Séré, "On the eigenvalues of operators with gaps. Application to Dirac operators", *J. Funct. Anal.* **174** (2000), p. 208-226.

[23] S. Morozov, D. Müller, "On the minimax principle for Coulomb–Dirac operators", *Math. Z.* **280** (2015), p. 733-747.

[24] D. Müller, "Minimax principles, Hardy–Dirac inequalities, and operator cores for two and three dimensional Coulomb–Dirac operators", *Doc. Math.* **21** (2016), p. 1151-1169.

[25] L. Schimmer, J. P. Solovej, S. Tokus, "Friedrichs extension and min–max principle for operators with a gap", *Ann. Henri Poincaré* **21** (2020), p. 327-357.

[26] J. Dolbeault, M. J. Esteban, M. Loss, "Relativistic hydrogenic atoms in strong magnetic fields", *Ann. Henri Poincaré* **8** (2007), p. 749-779.

[27] J. Dolbeault, M. J. Esteban, M. Loss, "Characterization of the critical magnetic field in the Dirac–Coulomb equation", *J. Phys. A* **41** (2008), article ID 185303.

[28] M. J. Esteban, M. Loss, "Self-adjointness for Dirac operators via Hardy–Dirac inequalities", *J. Math. Phys.* **48** (2007), article ID 112107.

[29] J. D. Talman, "Minimax principle for the Dirac equation", *Phys. Rev. Lett.* **57** (1986), p. 1091-1094.

[30] J. Dolbeault, M. J. Esteban, E. Séré, "A variational method for relativistic computations in atomic and molecular physics", *Int. J. Quantum Chem.* **93** (2003), p. 149-155.

[31] M. Esteban, M. Lewin, E. Séré, "Domains for Dirac–Coulomb min-max levels", *Rev. Mat. Iberoam.* **35** (2019), p. 877-924.

[32] M. J. Esteban, M. Lewin, E. Séré, "Dirac-Coulomb operators with general charge distribution. I. Distinguished extension and min-max formulas", 2020, Working paper or preprint.

[33] M. J. Esteban, M. Lewin, E. Séré, "Dirac-Coulomb operators with general charge distribution. II. The lowest eigenvalue", 2020, Working paper or preprint.
[34] U.-W. Schmincke, "Distinguished selfadjoint extensions of Dirac operators", *Math. Z.* **129** (1972), p. 335-349.
[35] R. Wüst, "A convergence theorem for selfadjoint operators applicable to Dirac operators with cutoff potentials", *Math. Z.* **131** (1973), p. 339-349.
[36] R. Wüst, "Distinguished self-adjoint extensions of Dirac operators constructed by means of cut-off potentials", *Math. Z.* **141** (1975), p. 93-98.
[37] R. Wüst, "Dirac operations with strongly singular potentials", *Math. Z.* **152** (1977), p. 259-271.
[38] G. Nenciu, "Self-adjointness and invariance of the essential spectrum for Dirac operators defined as quadratic forms", *Comm. Math. Phys.* **48** (1976), p. 235-247.
[39] M. Klaus, R. Wüst, "Characterization and uniqueness of distinguished selfadjoint extensions of Dirac operators", *Comm. Math. Phys.* **64** (1978), p. 171-176.
[40] J. J. Landgren, P. A. Rejto, "An application of the maximum principle to the study of essential self-adjointness of Dirac operators. i", *J. Math. Phys.* **20** (1979), p. 2204-2211.
[41] J. J. Landgren, P. A. Rejto, M. Klaus, "An application of the maximum principle to the study of essential self-adjointness of Dirac operators. ii", *J. Math. Phys.* **21** (1980), p. 1210-1217.
[42] T. Kato, "Holomorphic families of Dirac operators", *Math. Z.* **183** (1983), p. 399-406.
[43] N. Arrizabalaga, "Distinguished self-adjoint extensions of Dirac operators via Hardy–Dirac inequalities", *J. Math. Phys.* **52** (2011), article ID 092301.
[44] N. Arrizabalaga, J. Duoandikoetxea, L. Vega, "Self-adjoint extensions of Dirac operators with Coulomb type singularity", *J. Math. Phys.* **54** (2013), article ID 041504.
[45] A. N. Artemyev, A. Surzhykov, P. Indelicato, G. Plunien, T. Stöhlker, "Finite basis set approach to the two-centre Dirac problem in Cassini coordinates", *J. Phys. B* **43** (2010), article ID 235207.

Comptes Rendus
Physique
2020, 21, nº 2, p. 185-198
https://doi.org/10.5802/crphys.23

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

3D Ising model: a view from the conformal bootstrap island

Modèle d'Ising en 3D, vu de l'île d'auto-amorçage conforme

Slava Rychkov [a, b]

[a] Institut des Hautes Études Scientifiques, Bures-sur-Yvette, France

[b] Laboratoire de Physique de l'Ecole normale supérieure, ENS, Université PSL, CNRS, Sorbonne Université, Université de Paris, F-75005 Paris, France

E-mail: slava.rychkov@gmail.com

Abstract. We explain how the axioms of Conformal Field Theory are used to make predictions about critical exponents of continuous phase transitions in three dimensions, via a procedure called the conformal bootstrap. The method assumes conformal invariance of correlation functions, and imposes some relations between correlation functions of different orders. Numerical analysis shows that these conditions are incompatible unless the critical exponents take particular values, or more precisely that they must belong to a small island in the parameter space.

Résumé. Nous expliquerons comment utiliser les axiomes de la théorie conforme des champs pour faire des prédictions sur les exposants critiques des transitions de phase continues en trois dimensions, par une procédure dite "auto-amorçage conforme". La méthode consiste à faire l'hypothèse que les fonctions de corrélation ont l'invariance conforme, puis à imposer certaines relations entre les fonctions de corrélation d'ordre différent. L'analyse numérique montre alors que ces conditions ne sont compatibles que si les exposants critiques ont des valeurs particulières, plus précisément s'ils se trouvent dans un îlot étroit de l'espace des paramètres.

Keywords. Phase transitions, Critical exponents, Conformal symmetry, Ising model, Emergence.

Mots-clés. Transitions de phase, Exposants critiques, Symétrie conforme, Modèle d'Ising, Émergence.

1. Introduction

Physics has many *emergent laws*, which follow in a non-obvious way from more fundamental microscopic laws. Whenever this happens, we have two separate goals: to understand how the emergent law arises, and to explore its consequences.

One example is the Gibbs distribution of equilibrium statistical mechanics: the probability for the system in thermal equilibrium at temperature T to be found in a state n of energy E_n is proportional to $\exp(-E_n/T)$. One may be interested in deriving this emergent law from microscopic models of thermalization, or in exploring the myriad of its physical consequences.

This text, based on a recent talk for an audience of mathematical physicists, is about the "conformal field theory" (CFT), a set of emergent laws governing critical phenomena in equilibrium statistical mechanics (such as the liquid–vapor critical point or the Curie point of ferromagnets). CFT makes certain assumptions about the state of the system at a critical point. These assumptions can be given reasonable physical explanations, but for the purposes of this talk we will view them as axioms.

CFT is an "emergent law of second degree" with respect to the Gibbs distribution, by itself emergent. It sidesteps the Gibbs distribution similarly to how the Gibbs distribution sidesteps a thermalization model. Future work should derive the CFT axioms with mathematical rigor from the Gibbs distribution. Our goal here will be to explain the axioms and how they lead to concrete predictions for observable quantities through a procedure called the "conformal bootstrap".

CFT/bootstrap approach to critical phenomena is an alternative to the better-known Wilson's renormalization group (RG) theory. The RG is more directly related to the Gibbs distribution than CFT, although it too is not fully mathematically justified. The RG will not be treated here except for a few comments.

We will not give many references, which can be found in the recent review [1]. See also lecture notes [2–4]. An excellent set of recorded lectures is [5].

2. The first two CFT axioms

We will describe the axioms of Conformal Field Theory (CFT) on $\mathbb{R}^d$, $d \geqslant 3$. These axioms are well established in the physics literature. We will present them in a form hopefully more accessible to mathematicians. In particular, we will try to avoid (or at least explain) excessive physics jargon. Similar axioms, with additional bells and whistles, hold in $d = 2$ dimensions [6].[1]

Suppose we are given a collection of real-valued functions

$$\mathcal{T} = \{G_{i_1,\dots,i_n}(x_1,\dots,x_n)\}, \tag{1}$$

defined for $x_p \in \mathbb{R}^d$, $x_p \neq x_q$ $(p,q = 1,\dots,n)$, *where* $n \geqslant 1$ *and the indices* i_p *are non-negative integers.*

Functions (1) are called "n-point correlators of fields $A_{i_1},\dots,A_{i_n}$" and are also denoted by

$$\langle A_{i_1}(x_1)A_{i_2}(x_2)\cdots A_{i_n}(x_n)\rangle. \tag{2}$$

The collection $\mathcal{T}$ is called a CFT if it satisfies certain axioms stated below. Different CFT's are just different collections of correlators satisfying those axioms.[2]

Note that the "field" A_i is just a label, a name, and (2) is just a notation for $G_{i_1,\dots,i_n}(x_1,\dots,x_n)$. The statistical average operation suggested by this notation does not have a direct meaning in the CFT axioms. It will be handy in the *interpretation* of the axioms (Section 4).

[1] Our axioms should be viewed as a sketch of future complete axiomatics, which has not yet been written up in the mathematics literature. A different approach to axiomatize CFTs in $d \geqslant 3$ (akin to Segal's 2d CFT axiom) is in [7], but it makes the connection to concrete calculations less explicit. A nicer starting point is the recent mathematics paper [8] which develops 2d CFT from the conformal bootstrap perspective. It would be interesting to generalize it to $d \geqslant 3$.

[2] One also uses the term "Conformal Field Theory" in a meta-sense, as the study of all possible CFTs.

Axiom 1 (Simple properties). *Correlators have the following properties:*

(a) *They are invariant under permutation of any two fields:*[3]

$$\langle A_{i_1}(x)A_{i_2}(y)\cdots\rangle = \langle A_{i_2}(y)A_{i_1}(x)\cdots\rangle, \text{ etc.} \tag{3}$$

(b) *Index $i=0$ is associated with the "unit field", replaced by 1 under the correlator sign:*

$$\langle A_0(x)\times \text{anything}\rangle = \langle \text{anything}\rangle. \tag{4}$$

(c) *The 1-point (1pt) correlators are given by*

$$\langle A_0(x)\rangle \equiv 1, \quad \langle A_i(x)\rangle \equiv 0 \quad (i\geqslant 1). \tag{5}$$

(d) *The 2pt correlators are given by*

$$\langle A_i(x)A_j(y)\rangle = \frac{\delta_{ij}}{|x-y|^{2\Delta_i}} \quad (x,y\in\mathbb{R}^d). \tag{6}$$

where δ_{ij} is the Kronecker symbol, and $\Delta_i \geqslant ((d-2)/2)$ ($i\geqslant 1$) is a real number called "scaling dimension of field A_i". For the unit field we have $\Delta_0 = 0$.

(e) *The set of scaling dimensions $\{\Delta_i\}$ is called the "spectrum". It is a discrete set without accumulation points (i.e. there are finitely many scaling dimensions below any $\Delta_* < \infty$).*

Axiom 2 (Conformal invariance). *Correlators are conformally invariant, in the sense that they satisfy the constraint*

$$G_{i_1,\ldots,i_n}(x_1,\ldots,x_n) = \left(\prod_{p=1}^{n}\lambda(x_p)^{\Delta_{i_p}}\right) G_{i_1,\ldots,i_n}(f(x_1),\ldots,f(x_n)), \tag{7}$$

or equivalently, using notation (2)*,*

$$\langle A_{i_1}(x_1)\cdots A_{i_n}(x_n)\rangle = \left(\prod_{p=1}^{n}\lambda(x_p)^{\Delta_{i_p}}\right)\langle A_{i_1}(f(x_1))\cdots A_{i_n}(f(x_n))\rangle, \tag{8}$$

where $f(x)$ is an arbitrary conformal transformation of $\mathbb{R}^d$ and $\lambda(x) = |\partial f/\partial x|^{1/d}$ is its scale factor.

Recall that conformal transformations satisfy the constraint $\partial f^\mu/\partial x^\nu = \lambda(x)R^\mu{}_\nu(x)$ where $R^\mu{}_\nu(x)\in \mathrm{SO}(d)$. For $d\geqslant 3$, these transformations form a group $\mathrm{SO}(d+1,1)$.

Remark 1. Conformal transformations of $\mathbb{R}^d$ may send points to infinity, and should be thought more properly as acting on $\mathbb{R}^d\cup\{\infty\}$, the d-dimensional analogue of the Riemann sphere. To treat the point at infinity on equal footing with the other points, one can put $\mathbb{R}^d\cup\{\infty\}$ in one-to-one correspondence with the d-dimensional unit sphere S^d via the stereographic projection. This subtlety will be glossed over here.

2.1. *Basic consequences of conformal invariance*

We will state without proof a few basic consequences of the above axioms. One can check that the 2pt correlators given in Axiom 1(d) are consistent with Axiom 2. Note that the same scaling dimension Δ_i has to appear in all n-point correlators involving the field A_i. The 3pt correlators are fixed by Axiom 2 up to an overall factor:

$$\langle A_i(x_1)A_j(x_2)A_k(x_3)\rangle = \frac{c_{ijk}}{x_{12}^{\Delta_i+\Delta_j-\Delta_k}x_{13}^{\Delta_i+\Delta_k-\Delta_j}x_{23}^{\Delta_j+\Delta_k-\Delta_i}}, \tag{9}$$

[3] One can also consider CFTs with fields having fermionic statistics, whose correlators change sign under permutations. Such CFTs are important e.g. for describing quantum critical points of many-electron systems. Here we only consider commuting fields for simplicity.

where c_{ijk} is totally symmetric by Axiom 1(a), and we denoted $x_{ij} = |x_i - x_j|$. For 4pt correlators Axiom 2 implies the following functional form:

$$\langle A_i(x_1)A_j(x_2)A_k(x_3)A_l(x_4)\rangle = \left(\frac{x_{24}}{x_{14}}\right)^{\Delta_i-\Delta_j}\left(\frac{x_{14}}{x_{13}}\right)^{\Delta_k-\Delta_l}\frac{g_{ijkl}(u,v)}{x_{12}^{\Delta_i+\Delta_j}x_{34}^{\Delta_k+\Delta_l}}, \tag{10}$$

where $g_{ijkl}(u,v)$ is a function of conformally invariant cross-ratios:

$$u = \frac{x_{12}^2x_{34}^2}{x_{13}^2x_{24}^2}, \quad v = u|_{1\leftrightarrow 3} = \frac{x_{23}^2x_{14}^2}{x_{13}^2x_{24}^2}. \tag{11}$$

By Axiom 1(a) functions g_{ijkl} with permuted indices are all related, e.g. permutation $1 \leftrightarrow 3$ generates the constraint:

$$u^{-\frac{\Delta_i+\Delta_j}{2}}g_{ijkl}(u,v) = v^{-\frac{\Delta_k+\Delta_j}{2}}g_{kjil}(v,u), \text{etc.} \tag{12}$$

2.2. *Primaries and descendants*

Group-theoretically, the transformation

$$A(x) \to \lambda(x)^{\Delta}A(f(x)) \tag{13}$$

is an irreducible representation π_Δ of the conformal group on scalar functions $A : \mathbb{R}^d \to \mathbb{R}$. Equation (7) means that the correlators $G_{i_1,\ldots,i_n}$ belong to the invariant subspace of the tensor product representation $\otimes_{p=1}^n \pi_{\Delta_{i_p}}$ (so they can be called "invariant tensors").

We formulated Axioms 1, 2, 3 for the fields transforming as (13), called "scalar fields". These axioms can and should be extended to allow for fields with tensor indices. First of all, we have to add fields $\partial^\alpha A_i(x)$ which are partial derivatives (of arbitrary order) of the fields A_i. Their correlators are defined as derivatives of the original ones:

$$\langle \partial^\alpha A_i(x)\cdots\rangle := \partial_x^\alpha\langle A_i(x)\cdots\rangle. \tag{14}$$

This is, in a sense, just a convenient notation. The basic fields $A_i(x)$ whose correlators transform as (13) are called "primaries", while their derivatives "descendants". Transformation rules for correlators of descendants can be obtained by differentiating (13).

The second extension is a bit less trivial. We should generalize (13), allowing for fields with values in a finite-dimensional vector space V, $\dim V > 1$, transforming under the conformal group via

$$A(x) \to \lambda(x)^{\Delta}\rho(R(x))A(f(x)), \tag{15}$$

where ρ is an irreducible representation of $\mathrm{SO}(d)$ acting in V. Such fields are called "primary spinning fields". One example is $V = \{\text{symmetric traceless rank-}l\text{ tensors}\}$. Correlators of spinning fields then take values in the tensor product $\otimes_{p=1}^n V_{i_p}$ and satisfy a conformal invariance constraint similar to (7) but with factors of $\rho_{i_p}(R(x_p))$ in the l.h.s. (Derivatives of spinning fields are then also added as in (14).) Adding spinning fields would complicate the notation a bit. We will neglect them here, although practical conformal bootstrap computations always allow for their presence.

3. The OPE axiom

The last "OPE axiom" will relate different correlators, and in particular correlators with different n. This is unlike the previous axioms which involved one n-point correlator at a time.[4]

[4] Except the rather trivial Axiom 1(b).

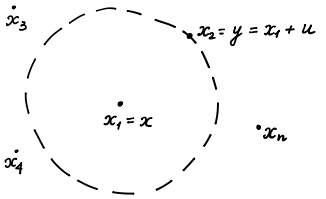


Figure 1. The OPE expansion applies when all points x_p ($p \geqslant 3$) lie further from x_1 than x_2.

Suppose we are given two collections of real numbers

$$\{\lambda_{ijk}\}, \{s^{(r)}_{ijk}, r \geqslant 1\}, \tag{16}$$

where i, j, k run over the field indices (non-negative integers). With these numbers as coefficients, "Operator Product Expansion" (OPE) is constructed as a set of formal equalities (one for each pair of fields A_i and A_j):

$$\begin{aligned} A_i(x)A_j(y) = \sum_{k=0}^{\infty} \frac{\lambda_{ijk}}{|u|^{\Delta_i+\Delta_j-\Delta_k}} \\ \times [A_k(x) + s^{(1)}_{ijk} u_\mu \partial^\mu_x A_k(x) + (s^{(2)}_{ijk} u_\mu u_\nu + s^{(3)}_{ijk} u^2 \delta_{\mu\nu}) \partial^\mu_x \partial^\nu_x A_k(x) + \cdots], \end{aligned} \tag{17}$$

where $u = y - x$. Using the OPE for the first pair of fields inside the n-point correlator (2) with $n \geqslant 2$, we get a set of candidate identities among correlators:

$$\langle A_i(x)A_j(y)\mathbf{\Pi}\rangle = \sum_{k=0}^{\infty} \frac{\lambda_{ijk}}{|u|^{\Delta_i+\Delta_j-\Delta_k}} [\langle A_k(x)\mathbf{\Pi}\rangle + s^{(1)}_{ijk} u_\mu \partial^\mu_x \langle A_k(x)\mathbf{\Pi}\rangle + \cdots], \tag{18}$$

where we denoted $i_1 = i$, $i_2 = j$, $x_1 = x$, $x_2 = y$, and $\mathbf{\Pi} = \Pi_{p=3}^n A_{i_p}(x_p)$ is the product of all other fields in the correlator. In the l.h.s. we have an n-point correlator, while in the r.h.s. we have an infinite series of $(n-1)$-point correlators and derivatives thereof.

The OPE axiom gives a condition for when the candidate identity (18) is a true identity.

Axiom 3 (OPE). *There exists a set of coefficients* (16)*, such that* (18) *holds as a true relation between correlators (the series in the r.h.s. converges absolutely to the l.h.s.) as long as* $|x_p - x| > |u|$ *for all* $p \geqslant 3$ *(see Figure 1).*

What can be said about coefficients (16) which make this axiom work? To see this, let us apply (18) to a 3pt correlator. Because the 2pt correlators vanish for non-identical fields (δ_{ij} in Axiom 1(d)), the sum in the r.h.s. collapses to the single k value, and we get:

$$\langle A_i(x)A_j(y)A_k(x_3)\rangle = \frac{\lambda_{ijk}}{|u|^{\Delta_i+\Delta_j-\Delta_k}} \left[\frac{1}{|x-x_3|^{2\Delta_k}} + s^{(1)}_{ijk} u_\mu \partial^\mu_x \frac{1}{|x-x_3|^{2\Delta_k}} + \cdots \right], \tag{19}$$

where we used that $\langle A_k(x)A_k(x_3)\rangle = 1/|x-x_3|^{2\Delta_k}$. On the other hand, we already know that the 3pt correlator in the l.h.s. has form (9) by Axiom 2. Let us then expand (9) for small y and match with (19). From the leading term we find $\lambda_{ijk} = c_{ijk}$.[5] Relative to this overall normalization, the subleading terms on the l.h.s. are fixed, and this allows to determine $s^{(r)}_{ijk}$ uniquely as rational functions of Δ_i, Δ_j, Δ_k and d. We conclude that all coefficients (16) can be uniquely determined by demanding that the OPE axiom works for the 3pt correlators. Furthermore, the axiom says that the same set of coefficients should then also work for any n-point correlators.

[5] In particular we learn that λ_{ijk} has to be symmetric, just as c_{ijk}. Note also that by putting $A_k = A_0 = 1$ in (9) and by using Axiom 1(d), we get $\lambda_{ij0} = c_{ij0} = \delta_{ij}$.

Definition 1. A conformal field theory (CFT) $\mathcal{T}$ in $d \geqslant 3$ dimensions is a collection of correlators (1) satisfying Axioms 1–3.

Remark 2. We included in our axioms the conditions $\Delta_i \geqslant (d-2)/2$ for $i \geqslant 1$ in Axiom 1(d), and $\lambda_{ijk} \in \mathbb{R}$ in (16). By this, we are restricting our discussion to a subclass of conformal field theories called reflection positive (or unitary). Many statistical physics systems at criticality (such as the Ising model or $O(N)$ models) are known to be described by unitary CFTs.

3.1. *CFT data*

The spectrum Δ_i and the OPE coefficients λ_{ijk} comprise the "dataset" of a CFT $\mathcal{T}$:

$$\text{Data}(\mathcal{T}) = \{\Delta_i, \lambda_{ijk}\}. \tag{20}$$

As discussed, Data($\mathcal{T}$) is in one-to-one correspondence with the 2pt and 3pt correlators of $\mathcal{T}$. Moreover, knowing Data($\mathcal{T}$) we can reconstruct all n-point correlators, for an arbitrarily high n. Indeed, from Data($\mathcal{T}$) we can construct the OPE (the coefficients $s^{(r)}_{ijk}$ are not included in Data($\mathcal{T}$) since they are uniquely determined by Δ's and d). Then, we can recursively reduce any n-point correlator to lower-point ones, until we get to the known 2pt and 3pt correlators.[6]

We thus see that the dataset Data($\mathcal{T}$) encodes full information about the CFT $\mathcal{T}$. Below we will describe a program of classifying CFTs by classifying their data sets. But first let us discuss the interpretation of the CFT axioms.

4. Interpretation

Notation $\langle A_{i_1}(x_1) A_{i_2}(x_2) \cdots A_{i_n}(x_n)\rangle$ for $G_{i_1,\ldots,i_n}(x_1,\ldots,x_n)$ acquires a meaning in the interpretation of the CFT axioms, as correlation functions of statistical systems at their critical points. CFT calculations are then interpreted as predictions for the critical exponents of statistical physics models. Although the CFT calculations based on the axioms are completely rigorous, the interpretation step is at present non-rigorous. Hopefully it will be justified in the future.

Let us discuss how this works for the 3d Ising model: a lattice model with the Hamiltonian $H = -\sum_{\langle xy\rangle} S_x S_y$ where $S_x = \pm 1$ are spins on a cubic lattice, with the nearest-neighbor ferromagnetic interaction.

The "3d Ising CFT" is a CFT in $d = 3$ describing the critical point of this model, and of any other model in the same universality class. Just as the lattice Ising model, this CFT has a global $\mathbb{Z}_2$ invariance with all fields divided into $\mathbb{Z}_2$-even and $\mathbb{Z}_2$-odd.[7] It contains a $\mathbb{Z}_2$-odd scalar primary field denoted $\sigma(x)$, whose correlators

$$\langle \sigma(x_1)\sigma(x_2)\cdots\sigma(x_n)\rangle \tag{21}$$

are interpreted as the 3d Ising model spin correlation functions

$$\langle S_{x_1} S_{x_2} \cdots S_{x_n}\rangle \tag{22}$$

[6] We should take care that the OPE is used for a pair of fields at positions x_1, x_2 verifying conditions of Axiom 3, so that it converges. This is the case if x_2 is the unique position with the minimal distance from x_1. There are degenerate configurations when such a pair cannot be found, because each point has two or more nearest neighbors at equal distance (e.g. the vertices of a regular polygon). It is then always possible to apply a small conformal transformation which moves points to a non-degenerate configuration. In the new configuration the OPE converges and we can compute the value of the correlator. We then conformal-transform back to the original configuration. This way we can compute correlators in any configuration of non-coincident points.

[7] We have not included the notion of global symmetry in the CFT axioms, but this extension is straightforward. It just means that all fields transform in finite-dimensional irreducible representations of a compact global symmetry group G, forming a direct product with the conformal group. All correlators are invariant tensors of G, and the OPE respects this additional symmetry.

computed at the critical temperature $T = T_c$, at distances $|x_p - x_q|$ much larger than the lattice spacing. While $\langle \cdots \rangle$ in (21) is just a notation, in (22) it is a true average with respect to the Gibbs distribution, in the thermodynamic limit. By Axiom 2, correlator (21) is conformally invariant, and thus in particular scale invariant, scale transformations being a part of the conformal group. This means (Δ_σ is the scaling dimension of σ):

$$\langle \sigma(\lambda x_1)\sigma(\lambda x_2)\cdots\sigma(\lambda x_n)\rangle = \lambda^{-n\Delta_\sigma}\langle \sigma(x_1)\sigma(x_2)\cdots\sigma(x_n)\rangle. \tag{23}$$

On the other hand, Equation (22) clearly does not have such an exact scale invariance, already because it is defined on a lattice. The precise statement of agreement at large distances is[8]

$$\lim_{|x_p - x_q|\to\infty} \frac{\langle S_{x_1} S_{x_2}\cdots S_{x_n}\rangle}{\langle \sigma(x_1)\sigma(x_2)\cdots\sigma(x_n)\rangle} = C^n, \tag{24}$$

where C is some constant, which is n-independent but non-universal (e.g. it would change if we add next-to-nearest interactions to the lattice model, which does not change the universality class).

Other 3d Ising CFT fields will correspond to other lattice-scale operators. E.g. we can consider the product of two nearby spins (separated in an arbitrary direction)

$$E_x = S_x S_{x+1} - \langle S_x S_{x+1}\rangle, \tag{25}$$

where $\langle S_x S_{x+1}\rangle$ is subtracted so that $\langle E_x\rangle = 0$. The 3d Ising CFT contains a $\mathbb{Z}_2$-even scalar primary $\varepsilon(x)$ whose correlators describe long-distance limits of the E_x correlators, similarly to (24).

More generally, we expect to have a CFT associated with every universality class of continuous phase transitions. This CFT will share global symmetry ($\mathbb{Z}_2$, $O(N)$, etc) with the universality class, and its scaling dimensions will determine the critical exponents. It has not been proven yet, starting from the lattice models or in any other way, that all these CFTs actually exist. This is the non-rigorous part of the CFT game.

CFT fields A_i and their scaling dimensions Δ_i also have counterparts in the RG approach to critical phenomena [9].[9] Namely, they correspond to the eigenvectors and the eigenvalues of RG transformation linearized near a fixed point describing a continuous phase transition. Fields of scaling dimension $\Delta_i < d$ ($\Delta_i > d$) correspond to the relevant (irrelevant) deformations of the fixed point. This dictionary is not needed for the actual CFT calculations, but only for interpreting the results.

We expect that the above-mentioned fields σ and ε are the only two relevant fields of the 3d Ising CFT. This follows from the experimental fact that the critical point of the 3d Ising model is in the same universality class as the liquid–vapor critical point, which is reached by tuning two parameters (pressure and temperature).

5. Conformal bootstrap program

5.1. *Consistency*

Conformal bootstrap program attempts to classify CFTs by classifying their datasets. That this may be possible was first suggested by Polyakov [10].

[8] Equivalently, one can consider a sequence of lattice models with a smaller and smaller lattice spacing a, and take the limit $a \to 0$ while keeping x_a fixed.

[9] On the other hand, the OPE coefficients λ_{ijk} do not feature prominently in the RG approach.

Figure 2. The 4pt correlator for this configuration of points can be reduced to 3pt correlators using the OPE (18) with (x,y) being one of the following pairs: (x_1,x_2), (x_2,x_3), (x_3,x_2), (x_4,x_3).

We call a dataset $\mathscr{D}=\{\Delta_i,\lambda_{ijk}\}$ "consistent" if it is a dataset of some CFT: $\mathscr{D}=\mathrm{Data}(\mathscr{T})$. Ideally, we would like to have a list of all consistent data sets:[10]

$$\{\mathrm{Data}(\mathscr{T}_1),\mathrm{Data}(\mathscr{T}_2),\ldots\}, \tag{26}$$

but it is not currently known how to generate such a list. The following question is less ambitious but still very interesting:

$$\textit{Q1: Given a trial dataset } \mathscr{D}\textit{, decide if it is inconsistent.} \tag{27}$$

It turns out that this has an algorithmic answer. This will allow progress on classification by ruling out inconsistent data sets (rather than by constructing consistent ones).

The idea is straightforward: given a trial dataset $\mathscr{D}=\{\Delta_i,\lambda_{ijk}\}$, we will try to construct all correlators, looking for some inconsistency with the axioms.

The first step is to construct the 2pt and 3pt correlators. These are simply given by explicit formulas from Axiom 1(d) and (9) with $c_{ijk}=\lambda_{ijk}$. So far no room for inconsistency.

Then we proceed to construct the 4pt correlators. For this we consider the OPE series reducing them to the 3pt correlators. All information needed to write down these series is contained in Δ_i and λ_{ijk}. But now we need to check a couple of things. First, do these series converge where Axiom 3 says they should? For this, the trial OPE coefficients λ_{ijk} should not grow too fast as a function of k for fixed i,j. The required growth condition can be shown to take a relatively simple form:

$$\sum_{k=0}^{\infty}(4\rho)^{\Delta_k}\lambda_{ijk}^2<\infty\quad\forall\rho<1. \tag{28}$$

Second, there are several ways to reduce a 4pt correlator to 3pt correlators via the OPE, and they all should agree in the overlapping regions of convergence. See Figure 2 for an example. This condition is called "crossing", and it is not automatically satisfied.[11] Assuming that it also holds, we can define the 4pt correlators as the sum of OPE series.

We then proceed to higher n-point correlators. Similarly to $n=4$, they are reduced to $(n-1)$-point correlators via the OPE, and we need to check convergence and crossing. It turns out that crossing for $n\geqslant 5$ is automatically satisfied once we impose crossing for all 4pt correlators. On the other hand, the general convergence condition is stronger than (28), and it can be expressed as follows. Consider an infinite matrix consisting of OPE coefficients λ_{ijk} with a fixed j and arbitrary i,k:

$$(M^{(j)})_{ik}=\lambda_{ijk}. \tag{29}$$

[10] We are not giving full details necessary to make this statement precise. One important subclass of CFTs are "local CFTs", which roughly correspond to critical points of lattice models with finite-range interactions. It is expected that most local CFTs are isolated. One exception are CFTs with "exactly marginal" fields of dimension $\Delta=d$, which form finite-dimensional continuous families. A folk conjecture says that exactly marginal fields in $d\geqslant 3$ require supersymmetry, which makes this exception non-generic.

[11] 4pt crossing constraints were first discussed in Refs. [10, 11]. The word "crossing" comes from an analogy with relativistic Quantum Field Theory. There, the $2\to 2$ scattering amplitude $\mathscr{M}(p_1,p_2\to p_3,p_4)$ is invariant under "crossing transformations", when one incoming particle is moved ("crosses") into the group of outgoing particles, while one outgoing particle crosses in the opposite direction.

Then any such $M^{(j)}$, viewed as an operator from i-indexed sequences to k-indexed sequences, should be bounded with respect to a certain weighted ℓ_2 norm (with Δ-depending weights).[12] Equation (28) follows from this general condition when we apply the operator to a sequence consisting of a single nonzero element, and demand that the result have a finite norm.

To summarize, consistent datasets are those which satisfy the general convergence and the 4pt crossing conditions. The convergence condition is the less interesting of the two. Below we will focus on the 4pt crossing, which will allow us to put constraints on the fields of low scaling dimension.

5.2. *Conformal blocks and 4pt crossing*

Here we will describe how to put 4pt crossing constraint into a more explicit form, by expanding 4pt correlators in a basis of special functions called conformal blocks.

Recall that conformally invariant 4pt correlators have form (10). When we compute the 4pt correlator in the r.h.s. of (10) using the OPE, we should get something consistent with this formula. Let us see how this happens. Applying the OPE to the first pair of fields, we get an expression of the form:

$$\langle A_i(x_1)A_j(x_2)A_k(x_3)A_l(x_4)\rangle = \sum_m \frac{\lambda_{ijm}}{x_{12}^{\Delta_i+\Delta_j-\Delta_m}}[\langle A_m(x_1)A_k(x_3)A_l(x_4)\rangle + \cdots], \tag{30}$$

where … denotes terms proportional to $s_{ijm}^{(r)}$ times derivatives acting on the 3pt correlator $\langle A_m A_k A_l\rangle$, which is in turn given by λ_{mkl} times an x-dependent function which can be read off from (9). It can be shown that by doing all derivatives and infinite sums over r, the r.h.s. of (30) takes the form:

$$\left(\frac{x_{24}}{x_{14}}\right)^{\Delta_i-\Delta_j}\left(\frac{x_{14}}{x_{13}}\right)^{\Delta_k-\Delta_l}\frac{1}{x_{12}^{\Delta_i+\Delta_j}x_{34}^{\Delta_k+\Delta_l}}\sum_m \lambda_{ijm}\lambda_{mkl}G_{\Delta_m}(u,v). \tag{31}$$

The functions $G_{\Delta_m}(u,v)$ appearing here are called "conformal block". These functions are fixed by conformal symmetry. They depends on the exchanged scaling dimension Δ_m, and on the space dimension d.[13] Notably, they do not depend on the OPE coefficients λ_{ijk} whose product appears as a prefactor in (31).

Theory of conformal blocks is huge and it's not possible to do it justice in this text. It has connections to representation theory, orthogonal polynomials, and integrable quantum mechanics. There are no fully general closed form expressions of conformal blocks in terms of the classical special functions. Fortunately, they admit rapidly convergent power series expansions which allow efficient numerical evaluation. This is what is used in practical applications.

The conformal block is simple only for the exchanged unit field: $A_m = A_0 = 1$, when we have:

$$G_0(u,v) = 1, \quad \lambda_{ij0} = \delta_{ij}, \quad \lambda_{0kl} = \delta_{kl}, \tag{32}$$

where we also gave the OPE coefficients for this case (see footnote 5).

Comparing (31) with (10) we see that they are consistent if we identify:

$$g_{ijkl}(u,v) = \sum_m \lambda_{ijm}\lambda_{mkl}G_{\Delta_m}(u,v). \tag{33}$$

[12] This is related to something called "radial quantization", which we do not describe in this text. This convergence condition for higher n-point corelators has not been discussed in detail in the literature.

[13] They also depend on the external dimension differences $\Delta_i - \Delta_j$, $\Delta_k - \Delta_l$ but we will omit this from the notation. In a full treatment involving spinning fields, the conformal blocks also depend on the spin of the fields.

This gives a compact formula to compute the 4pt correlators in terms of the CFT data. We can also obtain a compact expression for the 4pt crossing constraints, by substituting (33) into (12):

$$u^{-\frac{\Delta_i+\Delta_j}{2}}\sum_m \lambda_{ijm}\lambda_{mkl}G_{\Delta_m}(u,v) = v^{-\frac{\Delta_k+\Delta_j}{2}}\sum_m \lambda_{kjm}\lambda_{mil}G_{\Delta_m}(v,u). \tag{34}$$

As (12), this corresponds to the permutation $x_1 \leftrightarrow x_3$. Constraints corresponding to other permutations take a rather similar form. They should also be considered, although we will not discuss them here explicitly.

Now, we can test a trial dataset $\mathscr{D}$ for consistency, by checking (34) for all possible choices of i,j,k,l, in the region of overlapping convergence. This region is not empty. E.g., let us fix points $x_{1,3,4}$ so that x_4 is far away from $x_{1,3}$. Then the l.h.s. should converge within the set $\{x_2 : x_{12} < x_{13}\}$, and the r.h.s. in $\{x_2 : x_{23} < x_{13}\}$. These two balls have a nontrivial overlap.

5.3. *Partially specified datasets*

In Sections 5.1 and 5.2, we gave an answer to the consistency question (27). Unfortunately, the described procedure is not by itself practically useful, since it assumes that the trial dataset $\mathscr{D}$ is fully specified, which includes infinitely many parameters (all scaling dimensions and OPE coefficients). To correct for this, let us define the notion of a "partially specified trial dataset", which is a list $\mathscr{L}$ of finitely many assumptions on scaling dimensions and OPE coefficients. We say that $\mathscr{L}$ is consistent if there is at least one CFT $\mathscr{T}$ whose dataset Data($\mathscr{T}$) satisfies the assumptions. The following is then a more practical version of question (27):

Q2: Given a partially specified trial dataset $\mathscr{L}$, decide if it is inconsistent. (35)

Although this looks like a much harder question than (27), it turns out that this question can also be answered, based on (34), using numerical algorithms. This was first shown by Rattazzi, Tonni, Vichi and the author [12] and led to the rapid development of the numerical conformal bootstrap in the last 10 years. We will explain how this work on an example in the next section.

6. Example: constraining the 3d Ising CFT

Let us fix two real numbers Δ_1, Δ_2 in the interval $[1/2, 3]$, and consider the following list of assumptions $\mathscr{L} = \mathscr{L}(\Delta_1, \Delta_2)$ about a 3d CFT:

- $\mathbb{Z}_2$ global symmetry;
- there is one field which is $\mathbb{Z}_2$-odd, one which is $\mathbb{Z}_2$-even, and they have scaling dimensions Δ_1 and Δ_2;
- all other fields have scaling dimensions $\Delta_i \geqslant 3$ i.e. are irrelevant.

As discussed in Section 4, the 3d Ising CFT satisfies $\mathscr{L}(\Delta_\sigma, \Delta_\varepsilon)$. Our strategy will be to exclude a large part of the (Δ_1, Δ_2)-plane by showing that $\mathscr{L}(\Delta_1, \Delta_2)$ is inconsistent there. This will imply that the scaling dimensions of the 3d Ising CFT must belong to the remaining part of the plane.

6.1. *One crossing constraint*

Consider first the 4pt crossing for $\langle A_1 A_1 A_1 A_1\rangle$. Putting $i = j = k = l = 1$ in (34), we obtain:

$$u^{-\Delta_1}\sum_m p_m G_{\Delta_m}(u,v) = v^{-\Delta_1}\sum_m p_m G_{\Delta_m}(v,u), \quad p_m = \lambda_{11m}^2 \geqslant 0. \tag{36}$$

We know that $p_0 = G_0 = 1$ (see (32)), so isolating those terms we write this as

$$h(u,v) \equiv v^{-\Delta_1} - u^{-\Delta_1} = \sum_{m=2}^{\infty} p_m F_{\Delta_m}(u,v), \tag{37}$$
$$F_\Delta(u,v) := u^{-\Delta_1}G_\Delta(u,v) - v^{-\Delta_1}G_\Delta(v,u).$$

Note that F_Δ also depends on Δ_1. The sum starts from $m = 2$ because $\lambda_{111} = 0$ for the $\mathbb{Z}_2$-odd A_1.

Geometrically, Equation (37) means that h, viewed as a vector in a space of two-variable functions, belongs to a convex cone $\mathscr{C}$ generated by vectors F_{Δ_2} and F_Δ with $\Delta \geqslant 3$. We include all F_Δ with $\Delta \geqslant 3$ as generators of the cone since we don't know the exact values of Δ_m for $m \geqslant 3$, but only that $\Delta_m \geqslant 3$. Denote by $\mathscr{C}^*$ the dual convex cone, which is the set of all linear functionals α which are positive on all vectors generating the cone:

$$\alpha[F_{\Delta_2}] \geqslant 0, \quad \alpha[F_\Delta] \geqslant 0 \quad \forall \Delta \geqslant 3. \tag{38}$$

Suppose that there exists a functional $\alpha_0 \in \mathscr{C}^*$ such that

$$\alpha_0[h] < 0. \tag{39}$$

Then by acting with α_0 on (37) we get a contradiction. So, this equation cannot be satisfied for any nonnegative p_m. This is how one shows that the assumption $\mathscr{L}(\Delta_1, \Delta_2)$ is inconsistent: by exhibiting a functional α_0 which satisfies (38) and (39).

Numerically, one works with a finite dimensional space of functionals $\mathscr{A}(\Lambda)$ which are finite sums of partial derivatives at a particular point:

$$\alpha[f] = \sum_{m+n \leqslant \Lambda} \alpha_{m,n} \partial_u^m \partial_v^n f(u_0, v_0), \tag{40}$$

where Λ is a parameter, to be taken as large as possible to have the maximal constraining powers (within the available computer resources). One then minimizes $\alpha[h]$ over all $\alpha \in \mathscr{C}^* \cap \mathscr{A}(\Lambda)$, looking for a functional satisfying (39). This is a convex optimization problem (continuous linear programming), which can be solved by efficient numerical algorithms. If the minimum is negative, then we ruled out $\mathscr{L}(\Delta_1, \Delta_2)$. If it is positive, and cannot be made negative by increasing Λ, this would mean that $\mathscr{L}(\Delta_1, \Delta_2)$ is consistent with crossing for $\langle A_1 A_1 A_1 A_1 \rangle$.

With this procedure, Ref. [13] showed that the constraint $\mathscr{L}(\Delta_1, \Delta_2)$ is inconsistent in a significant portion of parameter space. Invoking an extra and so far unproven assumption, that the 3d Ising CFT lies at a singular boundary point of the consistent region (the so called "kink"), Refs. [13, 14] gave the first conformal bootstrap determination of $\Delta_\sigma, \Delta_\varepsilon$. Subsequent work has shown that the kink assumption is unnecessary, provided that one includes crossing constraints for the 4pt correlators $\langle A_1 A_1 A_2 A_2 \rangle$, $\langle A_2 A_2 A_2 A_2 \rangle$. We will now explain briefly how this was done.

6.2. *Several crossing constraints*

To increase the constraining power, a natural idea is to include crossing constraints for the other 4pt correlators of fields A_1 and A_2. While $\langle A_2 A_2 A_2 A_2 \rangle$ is completely analogous to $\langle A_1 A_1 A_1 A_1 \rangle$, one encounters a crucial difference when analyzing $\langle A_1 A_1 A_2 A_2 \rangle$. Namely, its conformal block expansion involves products of two different OPE coefficients $\lambda_{11m} \lambda_{22m}$. These products are not necessarily positive, because λ_{ijk} may have either sign. On the other hand, positivity of the coefficients $p_m = \lambda_{11m}^2$ played a crucial role in making the minimization problem of Section 6.1 convex. To overcome this obstacle, one analyzes all three correlators together, and considers the matrix

$$P_m = \begin{pmatrix} \lambda_{11m}^2 & \lambda_{11m}\lambda_{22m} \\ \lambda_{11m}\lambda_{22m} & \lambda_{22m}^2 \end{pmatrix}. \tag{41}$$

Crucially, this matrix is positive semidefinite: $P_m \succcurlyeq 0$. This condition is convex, and provides a good substitute for the simple positivity in the bootstrap problems involving multiple correlators. The resulting problem is that of continuous semidefinite programming, and it can still be attacked by efficient numerical algorithms. This was realized and carried out in Refs. [15–17] which found a consistent "island" near $\Delta_1 \approx 0.5181489(10)$, $\Delta_2 \approx 1.412625(10)$. The 3d Ising CFT point $(\Delta_\sigma, \Delta_\varepsilon)$ must live somewhere in this tiny island (Figure 3).

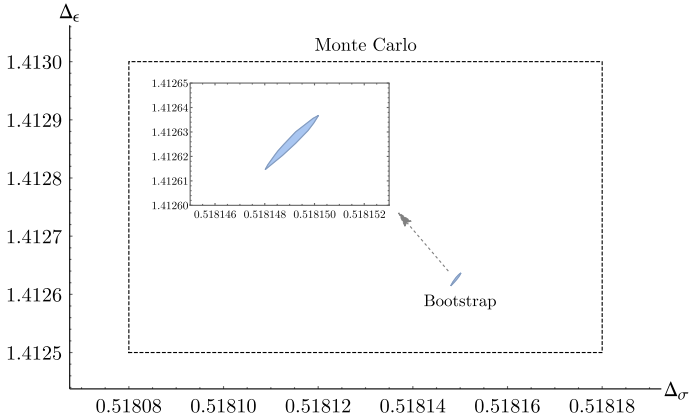


Figure 3. The bootstrap island to which the 3d Ising CFT must belong [17]. Also shown is a Monte Carlo result [18] for the same scaling dimensions. This plot used $\Lambda = 43$ in (40).

The scaling dimensions $\Delta_\sigma, \Delta_\varepsilon$ determine the main critical exponents of the 3d Ising model $\alpha, \beta, \gamma, \delta, \eta, \nu$. In what follows we will focus on η and ν, given by

$$\eta = 2\Delta_\sigma - 1, \tag{42}$$

$$\nu = 1/(d - \Delta_\varepsilon). \tag{43}$$

Equation (43) deserves a comment, because it expresses an off-critical quantity (the exponent ν describing behavior of the correlation length close to the critical point) via a critical theory parameter Δ_ε. This is an example of how CFT can make predictions about small deviations from the critical theory, which arise at short distances from relevant perturbations, and at large distances from the irrelevant ones. Such predictions are done via a technique called "conformal perturbation theory," which we have not explained. Equation (44) below is another simple example.

Another important quantity is the "correction to scaling" exponent ω. It appears in the rate $\sim 1/r^\omega$ at which the limit in (24) is achieved, assuming that all distances $|x_p - x_q| \sim r$ are of the same order. It also appears in the subleading singularities of all quantities exhibiting powerlaw behavior near the critical point (e.g. the specific heat). While describing deviations from criticality, ω like ν can be expressed in terms of a purely critical parameter:

$$\omega = \Delta_3 - d, \tag{44}$$

where Δ_3 is the scaling dimension of the leading irrelevant $\mathbb{Z}_2$-even scalar operator. The conformal bootstrap determines Δ_3 (and hence ω) by scanning the island in Figure 3 and reconstructing the spectrum which provides a solution to the 4pt crossing [19].

In Table 1 we report the values of the critical exponents ν, η, ω according to the conformal bootstrap, Monte Carlo simulations and RG calculations. We also include some experimental measurements of ν and η.[14] The conformal bootstrap predictions are the most precise, and they are in a good agreement with the Monte Carlo and RG. There is also reasonable agreement between the theory and the experiment, although the experimental accuracy is not amazing.

[14] Since ω parametrizes subleading powers, it is harder to measure, and we are not aware of any published result.

Table 1. Some representative theoretical and experimental determinations of the 3d Ising critical exponents. See [20], Section 3.2, for more references

Ref	Year	Method/Experiment	ν	η	ω
[17, 19]	2016	Conformal bootstrap	0.629971(4)	0.036298(2)	0.82968(23)
[18]	2010	Monte Carlo	0.63002(10)	0.03627(10)	0.832(6)
[21]	1998	RG	0.6304(13)	0.0335(25)	0.799(11)
[22]	1989	Binary fluid	0.628(8)	0.0300(15)	
[23]	2009	Binary fluid	0.629(3)	0.032(13)	
[24]	1994	Binary mixture	0.623(13)	0.039(4)	
[25]	2000	Liquid–vapor	0.62(3)		
[26]	1998	Liquid–vapor		0.042(6)	
[27]	1987	Uniaxial antiferromagnet	0.64(1)		

7. Conclusions

Conformal bootstrap calculations provide predictions for observable physical quantities from the CFT axioms. Agreement of these predictions with alternative theoretical determinations and the experiment increase our belief in the validity of the axioms.

Feynman [28] called the Gibbs distribution the "summit of statistical mechanics", the entire subject being either the "climb-up" to derive it, or the "slide-down" when it is applied. Echoing Feynman, we may call the CFT a summit of the theory of critical phenomena, the conformal bootstrap being the way to slide down. To climb up would be to prove the validity of the interpretation of the CFT axioms described in Section 4. Unfortunately, relatively little rigorous work has been done in the way of climbing up.[15] One should also not forget a second major peak in the same mountain range: the Renormalization Group.

Acknowledgements

This article is based on the talk at the mathematical physics workshop "Inhomogeneous Random Systems" (Institut Curie, Paris, January 28, 2020). I am grateful Ellen Saada, François Dunlop and Alessandro Giuliani for the organization and the invitation to speak. I am also grateful to Jacques Villain for the invitation to write this article, careful reading of the draft, and many suggestions on how to improve the presentation. SR is partly supported by the Simons Foundation grant 488655 (Simons Collaboration on the Nonperturbative Bootstrap), and by Mitsubishi Heavy Industries as an ENS-MHI Chair holder.

References

[1] D. Poland, S. Rychkov, A. Vichi, "The conformal bootstrap: theory, numerical techniques, and applications", *Rev. Mod. Phys.* **91** (2019), article ID 015002.

[2] S. Rychkov, *EPFL Lectures on Conformal Field Theory in $D \geq 3$ Dimensions*, SpringerBriefs in Physics, Springer, 2016.

[3] D. Simmons-Duffin, "The conformal bootstrap", in *Proceedings, Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings (TASI 2015): Boulder, CO, USA, June 1–26, 2015*, 2017, p. 1-74.

[15] The last 20 years, starting with Smirnov [29], have seen significant progress in showing rigorously conformal invariance of specific 2d models. This program is still far from establishing conformal invariance of a generic critical theory, let alone the full scope of the CFT axioms, such as the existence of a complete set of local operators and the OPE. We are also not aware of any relevant mathematical work in $d \geqslant 3$.

[4] S. M. Chester, "Weizmann Lectures on the Numerical Conformal Bootstrap", *arXiv: High Energy Physics - Theory* (2019), p. 93, http://arxiv.org/abs/1907.05147.
[5] P. Kravchuk, "Introduction to Conformal Blocks and Linear Programming. Lectures 1-4". PhD School "Bootstrap 2020", https://projects.iq.harvard.edu/bootstrap2020/videos.
[6] P. Di Francesco, P. Mathieu, D. Senechal, *Conformal Field Theory*, Springer-Verlag, New York, 1997.
[7] A. Schwarz, "Axiomatic conformal theory in dimensions ≥2 and AdS/CT correspondence", *Lett. Math. Phys.* **106** (2016), no. 9, p. 1181-1197.
[8] Y. Moriwaki, "Full vertex algebra and bootstrap – consistency of four point functions in 2d CFT". 2020, https://arxiv.org/abs/2006.15859.
[9] K. G. Wilson, J. B. Kogut, "The Renormalization group and the epsilon expansion", *Phys. Rep.* **12** (1974), p. 75-199.
[10] A. M. Polyakov, "Nonhamiltonian approach to conformal quantum field theory", *Zh. Eksp. Teor. Fiz.* **66** (1974), p. 23-42.
[11] S. Ferrara, A. F. Grillo, R. Gatto, "Tensor representations of conformal algebra and conformally covariant operator product expansion", *Ann. Phys.* **76** (1973), p. 161-188.
[12] R. Rattazzi, V. S. Rychkov, E. Tonni, A. Vichi, "Bounding scalar operator dimensions in 4D CFT", *JHEP* **12** (2008), p. 031.
[13] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, A. Vichi, "Solving the 3D ising model with the conformal bootstrap", *Phys. Rev.* **D86** (2012), article ID 025022.
[14] S. El-Showk, M. F. Paulos, D. Poland, S. Rychkov, D. Simmons-Duffin, A. Vichi, "Solving the 3d ising model with the conformal bootstrap II. c-Minimization and precise critical exponents", *J. Stat. Phys.* **157** (2014), p. 869.
[15] F. Kos, D. Poland, D. Simmons-Duffin, "Bootstrapping mixed correlators in the 3D ising model", *JHEP* **11** (2014), p. 109.
[16] D. Simmons-Duffin, "A semidefinite program solver for the conformal bootstrap", *JHEP* **06** (2015), p. 174.
[17] F. Kos, D. Poland, D. Simmons-Duffin, A. Vichi, "Precision islands in the ising and $O(N)$ models", *JHEP* **08** (2016), p. 036.
[18] M. Hasenbusch, "Finite size scaling study of lattice models in the three-dimensional ising universality class", *Phys. Rev. B* **82** (2010), article ID 174433.
[19] D. Simmons-Duffin, "The lightcone bootstrap and the spectrum of the 3d ising CFT", *JHEP* **03** (2017), p. 93.
[20] A. Pelissetto, E. Vicari, "Critical phenomena and renormalization-group theory", *Phys. Rep.* **368** (2002), p. 549-727.
[21] R. Guida, J. Zinn-Justin, "Critical exponents of the N vector model", *J. Phys.* **A31** (1998), p. 8103-8121.
[22] P. Damay, F. Leclercq, P. Chieux, "Critical scattering function in a binary fluid mixture: a study of sodium-deuteroammonia solution at the critical concentration by small-angle neutron scattering", *Phys. Rev. B* **40** (1989), p. 4696-4708.
[23] J. V. Sengers, J. G. Shanks, "Experimental critical-exponent values for fluids", *J. Stat. Phys.* **137** (2009), no. 5–6, p. 857-877.
[24] J. Schmitz, L. Belkoura, D. Woermann, "Light scattering and small angle neutron scattering experiments with an aggregating binary liquid mixture of critical composition", *Ann. Phys.* **506** (1994), no. 1, p. 1-12.
[25] D. M. Sullivan, G. W. Neilson, H. E. Fischer, A. R. Rennie, "Small angle neutron scattering from D_2O in the critical region", *J. Phys.: Condens. Matter* **12** (2000), no. 15, p. 3531-3542.
[26] P. Damay, F. Leclercq, R. Magli, F. Formisano, P. Lindner, "Universal critical-scattering function: an experimental approach", *Phys. Rev. B* **58** (1998), p. 12038-12043.
[27] D. P. Belanger, H. Yoshizawa, "Neutron scattering and the critical behavior of the three-dimensional ising antiferromagnet FEF_2", *Phys. Rev. B* **35** (1987), p. 4823-4830.
[28] R. Feynman, *Statistical Mechanics*, 7th ed., Frontiers in Physics, W. A. Benjamin, 1981.
[29] S. Smirnov, "Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits", *C. R. Math.* **333** (2001), no. 3, p. 239-244.

Comptes Rendus
Physique
2020, 21, nº 2, p. 199-202
https://doi.org/10.5802/crphys.25

Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Des bruits et des ondes : mathématiciens et physiciens en interaction

Noises and waves : mathematicians and physicists interacting

Yves Colin de Verdière [a]

[a] Université Grenoble-Alpes, Institut Fourier, Unité mixte de recherche CNRS-UGA 5582, BP 74, 38402-Saint Martin d'Hères Cedex, France
Courriel : yves.colin-de-verdiere@univ-grenoble-alpes.fr

Résumé. Nous décrivons la collaboration de l'auteur mathématicien avec deux groupes de physiciens : des sismologues et des physiciens des ondes dans les fluides.

Abstract. We describe the collaboration of the author, a mathematician, with two groups of physicists: seismologists and physicists of waves in fluids.

Mots-clés. Ondes internes, Bruit sismique.

Keywords. Internal waves, Seismic noise.

Depuis longtemps, les mathématiciens trouvent dans la physique de nombreuses sources d'inspiration. Dans cette petite note, je voudrais raconter deux expériences personnelles de cette interaction : l'utilisation du bruit sismique en imagerie découverte par mon collègue grenoblois Michel Campillo et la nature mathématique précise des attracteurs pour les ondes internes observés dans l'équipe de Thierry Dauxois à l'ENS de Lyon.

1. Bruit sismique et imagerie avec Michel Campillo (2003–2010)

Un beau jour de 2003, mon collègue physicien Bart van Tiggelen me dit à peu près ceci : « Tu sais, Michel Campillo fait de l'imagerie à partir du bruit sismique ». Je suis évidemment très intrigué, car tout le monde sait que le bruit est plutôt nuisible dans les expériences de physique. En tout cas, je n'en ai jamais entendu parler comme une source d'information. Me voilà donc parti rendre visite à Michel dont le laboratoire ISterre (Institut des sciences de la terre) est à deux pas de l'Institut Fourier. Michel m'explique entre autres que le grand géophysicien japonais Keiiti

Aki a posé depuis longtemps la question d'utiliser la coda sismique[1] dans l'imagerie. Il écrit une première publication (avec Anne Paul) [1] sur ce sujet avant de réaliser qu'on peut tirer parti du bruit enregistré en permanence par les sismomètres. Son idée est de considérer la corrélation temporelle donnée par

$$C_{A,B}(t) := \lim_{T\to+\infty} \frac{1}{T}\int_0^T \langle u(A,\tau)|u(B,t-\tau)\rangle \mathrm{d}\tau$$

où $u(x,t)$ est le champ de vitesse, i.e. le bruit au point x enregistré par un sismomètre. L'observation cruciale de Michel Campillo est que la corrélation $C_{A,B}(t)$ est très semblable au signal qui arrive en B s'il y a un tremblement de terre ou une explosion au point A. Cette remarquable idée ouvre la porte à de nouvelles possibilités d'imagerie sismique qui ont au moins deux avantages essentiels : la possibilité d'étudier des régions non sismiques et la possibilité de suivre les déformations au cours du temps, par exemple celle des cônes volcaniques.

Un enjeu est alors de concevoir des modèles mathématiques reflétant cette similitude. C'est ce que nous avons fait. Il y a deux difficultés principales :

1. Utiliser des sources de bruits localisées et non un bruit blanc, car on s'attend à ce que, par exemple, l'interaction de l'océan avec la croûte terrestre[2] soit une telle source localisée.
2. Décrire la propagation des ondes élastiques au sens de l'optique géométrique.

Il se trouve que ces deux difficultés trouvent leurs solutions, dans un contexte largement idéalisé, grâce à la boîte à outils développée depuis les années 70 par les mathématiciens pour étudier les équations aux dérivées partielles, soit les opérateurs pseudo-différentiels et les opérateurs intégraux de Fourier.

Nous avons alors décidé de démarrer un petit groupe de travail avec Michel et son équipe. Cette expérience m'a beaucoup marqué. Nous avons eu du mal à nous comprendre au début. Ensuite, à mon avis, le plus intéressant fut que nous avons réussi à échanger sur nos représentations mentales de ces phénomènes et cela fut source de progrès pour les deux parties. Je me souviens en particulier de leur avoir expliqué que la théorie de la diffusion n'est pas seulement une théorie de la propagation des ondes en présence d'un obstacle, mais qu'elle donne aussi une décomposition spectrale de l'Hamiltonien. Ce qui a contribué dans [2] à rendre très général un calcul compliqué fait sur un exemple « intégrable ».

Pour le lecteur intéressé, ma contribution mathématique est expliquée dans les articles [3–6], alors que la physique est expliquée dans [7].

2. Attracteurs pour les ondes internes observés par l'équipe de Thierry Dauxois (2016–?)

Il s'agit de travaux récents et en cours avec la mathématicienne Laure Saint-Raymond. Lors de son arrivée en 2016 à l'ENS de Lyon, Laure est allée voir les physiciens locaux et en particulier les expériences menées par Thierry Dauxois et ses élèves sur les ondes internes, voir par exemple la thèse de Christophe Brouzet [8]. Il s'agit d'une très belle expérience : dans un « aquarium », essentiellement bi-dimensionnel, dont les dimensions sont de l'ordre du mètre, de forme trapézoïdale, rempli d'eau salée stratifiée, on observe les ondes forcées par un oscillateur périodique. Pour un choix générique de la fréquence du forçage et de la géométrie du trapèze, l'onde se concentre sur une courbe fermée, sorte de trajectoire de billard. Comme j'avais jadis travaillé sur les fonctions propres du laplacien et leur éventuelle localisation sur les trajectoires périodiques du billard

[1] La coda sismique est la fin du signal sismique que l'on voit après les signaux nets dont on peut interprêter la nature géométrique.

[2] C'est ainsi que les géophysiciens Strasbourgeois voyaient la venue d'une tempête atlantique sur leurs sismomètres.

associé (les géodésiques périodiques), Laure eut l'idée de me proposer de collaborer en vue de comprendre la nature de ces attracteurs. Sous des hypothèses très simplifiées, nous avons décrit de tels attracteurs. L'équation d'onde forcée à laquelle on arrive est de la forme simplifiée

$$\frac{\mathrm{d}u}{\mathrm{d}t} + Hu - \nu\Delta u = f\mathrm{e}^{\mathrm{i}\omega t}, \quad u(t=0) = 0$$

où

- Δ est le laplacien, ici en dimension 2,
- $u(t)$ représente le champ de vitesse du fluide au temps t et donc u est aussi une fonction de la position x,
- H est un opérateur antisymétrique, non local et borné, qui résulte de l'élimination de la pression dans l'équation de Navier–Stokes. L'antisymmétrie provient de la conservation de l'énergie pour l'équation non forcée et à viscosité nulle,
- le membre de droite, où f est donnée, représente le forçage périodique.

On s'intéresse dans un premier temps, dans le cas où la viscosité ν est nulle, au comportement asymptotique de la solution $u(t)$ quand $t \to +\infty$. Ce que nous avons montré est que, pour un milieu générique, $u(t)$ converge, quand $t \to +\infty$, vers une distribution de Schwartz dont le support singulier est formé d'une (ou plusieurs) courbe(s) fermée(s). C'est cette courbe qui est l'attracteur. Ici l'opérateur H est borné dans l'espace de Hilbert naturel et son spectre près de la fréquence de forçage est absolument continu. On est donc en présence d'un spectre continu dans un domaine borné, ce qui n'a rien à voir avec le spectre discret d'un laplacien dans un tel domaine. La théorie spectrale de tels opérateurs n'avait jamais été étudiée faute de motivation et même le phénomène de résonance en présence de spectre continu n'avait pas été décrit. Nos travaux font appel à des concepts de géométrie différentielle (théorie des feuilletages génériques sur les surfaces : les feuilletages de Morse-Smale) et d'analyse (théorie de Mourre, opérateurs pseudo-différentiels) et font l'objet des publications [9, 10].

Le contexte de notre étude est un peu différent de celui des expériences : nous regardons ce qui se passe en l'absence de frontière, sur une surface fermée comme un tore. Les mêmes méthodes s'appliquent aussi aux ondes inertielles qui sont des solutions des équations de Navier–Stokes en présence de la force de Coriolis induite par la rotation. Ce comportement étrange (observé à la fois pour des fluides stratifiés ou en rotation) est lié à une brisure d'isotropie dans la propagation des ondes. Notre modèle montre que c'est alors un comportement générique de développer des attracteurs. Ces premiers travaux ouvrent un grand chantier : traiter le cas à bord, le cas de viscosité petite, la nature du spectre dans le complémentaire du spectre continu que nous avons décrit. Toutes ces questions donnent lieu à des échanges très constructifs avec la communauté des physiciens, et en particulier avec Michel Rieutord, astrophysicien de Toulouse qui s'intéresse au spectre des étoiles, voir [11].

3. Conclusions

Quitte à simplifier drastiquement la réalité physique très complexe, il est souvent possible d'en obtenir une description mathématique reproduisant bien, même si ce n'est que de façon qualitative, la phénoménologie physique. Ces descriptions ne sont finalement pas beaucoup plus éloignées de la nature que les expériences de laboratoire. Elles peuvent aider à une meilleure compréhension de la physique sous-jacente et donc à de futurs progrès. En tout cas, ces travaux permettent de prendre conscience de la grande unité de la science, au-delà de notre séparation moderne en des myriades de sous-disciplines. J'ajouterai que c'est un grand plaisir de travailler avec des chercheurs, mathématiciens ou physiciens, ayant une culture très différente de la sienne, surtout si l'on arrive à faire progresser ainsi les connaissances.

Références

[1] M. Campillo, A. Paul, « Long range correlations of seismic codas », *Science* **299** (2003), p. 547-549.
[2] Y. Colin de Verdière, « Scattering and correlations », *preprint*, arXiv :1103.4450 (2011).
[3] Y. Colin de Verdière, « Mathematical models for passive imaging I : general background », *preprint*, arXiv :math-ph/0610043 (2006).
[4] Y. Colin de Verdière, « Mathematical models for passive imaging II : Effective Hamiltonians associated to surface waves », *preprint*, arXiv :math-ph/0610044 (2006).
[5] Y. Colin de Verdière, « Semiclassical analysis and passive imaging », *Nonlinearity* **22** (2009), p. 45-75, Article le plus cité de ce journal pour l'année 2009.
[6] Y. Colin de Verdière, « A Semi-classical calculus of correlations », *C. R. Géosci.* **343** (2011), nº 8–9, p. 496-501, Thematic issue « Imaging and Monitoring with Seismic Noise ».
[7] P. Gouédard, L. Stéhly, F. Brenguier, M. Campillo, Y. Colin de Verdière, E. Larose, L. Margerin, P. Roux, F. J. Sànchez-Sesma, N. Shapiro, R. Weaver, « Cross-correlation of random fields : mathematical approach and applications », *Geophys. Prospect.* **56** (2008), p. 375-393, Loránd Eötvös Award 2009 de l'association des géophysiciens Européens (EAGE).
[8] C. Brouzet, « Internal wave attractors : from geometrical focusing to non-linear energy cascade and mixing », PhD thesis, ENS Lyon, 2016, https://hal.archives-ouvertes.fr/tel-01361201v1.
[9] Y. Colin de Verdière, L. Saint-Raymond, « Attractors for two dimensional waves with homogeneous Hamiltonians of degree 0 », *Comm. Pure Appl. Math.* **73** (2020), nº 2, p. 421-462.
[10] Y. Colin de Verdière, « Spectral theory of pseudo-differential operators of degree 0 and application to forced linear waves », *Anal. PDE* **13** (2020), nº 5, p. 1521-1537.
[11] M. Rieutord, L. Valdettaro, « Axisymmetric inertial modes in a spherical shell at low Ekman numbers », *J. Fluid Mech.* **844** (2018), p. 597-834.

Comptes Rendus

Physique

Objet de la revue

Les *Comptes Rendus Physique* sont une revue électronique évaluée par les pairs de niveau international, qui couvre l'ensemble des domaines de la physique et de l'astrophysique. Ils publient principalement des numéros thématiques, mais également des articles originaux de recherche, des annonces préliminaires, des articles de revue, des mises en perspective historiques, des textes à visée pédagogique ou encore des actes de colloque, sans limite de longueur, en anglais ou en français. Ils proposent également des numéros spéciaux consacrés à certains aspects récents et/ou significatifs de la discipline, dont les auteurs sont choisis parmi les chercheurs les plus actifs sur le sujet et dont la coordination est assurée par des rédacteurs en chef invités.

Les *Comptes Rendus Physique* sont diffusés selon une politique vertueuse de libre accès diamant, gratuit pour les auteurs (pas de frais de publications) comme pour les lecteurs (libre accès immédiat et pérenne).

Directeur de la publication : Étienne Ghys

Rédacteurs en chef : D. Gratias, J. Villain

Comité éditorial : Jacqueline Bloch, Christian Bordé, Hélène Bouchiat, Alexandre Bouzdine, Yves Bréchet, Françoise Combes, Jean Dalibard, Michel Davier, Daniel Estève, Stéphan Fauve, Pierre Fayet, Frédérique de Fornel, Maurice Goldman, Guy Laval, Chaouqi Misbah, Jean-Yves Ollitrault, Nathalie Palanque-Delabrouille

Secrétaire éditorial : Julien Desmarets

À propos de la revue

Toutes les informations concernant la revue, y compris le texte des articles publiés qui est en accès libre intégral, figurent sur le site https://comptes-rendus.academie-sciences.fr/physique/.

Informations à l'attention des auteurs

Pour toute question relative à la soumission des articles, les auteurs peuvent consulter le site https://comptes-rendus.academie-sciences.fr/physique/.

Contact

Académie des sciences
23, quai de Conti, 75006 Paris, France
Tél. : (+33) (0)1 44 41 43 72
CR-Physique@academie-sciences.fr

Les articles de cette revue sont mis à disposition sous la licence Creative Commons Attribution 4.0 International (CC-BY 4.0) https://creativecommons.org/licenses/by/4.0/deed.fr

Comptes Rendus
de l'Académie des sciences

Physique

Volume 21, nº 2, juin 2020

Special issue / *Numéro thématique*
Prizes of the French Academy of Sciences 2019 / *Prix 2019 de l'Académie des sciences*

Cover illustration / *Illustration de couverture*
© Manuel Rodrigues

Jacques Villain
Foreword 135-137

Pierre Touboul, Gilles Métris, Manuel Rodrigues, Yves André, Alain Robert
The MICROSCOPE space mission to test the Equivalence Principle 139-150

Michael Le Bars, Louis-Alexandre Couston, Benjamin Favier, Pierre Léard, Daniel Lecoanet, Patrice Le Gal
Fluid dynamics of a mixed convective/stably stratified system—A review of some recent works 151-164

Pierre Delplace, Antoine Venaille
From the geometry of Foucault pendulum to the topology of planetary waves 165-175

Maria J. Esteban
Mathematical questions about the computation of eigenvalues of Dirac operators with critical potentials in atomic and molecular physics 177-183

Slava Rychkov
3D Ising model: a view from the conformal bootstrap island 185-198

Yves Colin de Verdière
Des bruits et des ondes: mathématiciens et physiciens en interaction 199-202

1878-1535 (electronic)

Académie des sciences — Paris

COMPTES RENDUS
DE L'ACADÉMIE DES SCIENCES

Physique

Volume 21, nº 2, juin 2020