SERBIAN ACADEMY OF
NONLINEAR SCIENCES

1st CONFERENCE ON NONLINEARITY

Editors

B. Dragovich, Ž. Čupić

Belgrade, 2020
SERBIA

P r o c e e d i n g s

of the

1st CONFERENCE ON NONLINEARITY

(October 11 – 12, 2019, Belgrade, Serbia)

Editors

B. Dragovich, Ž. Čupić

Publisher

Serbian Academy of Nonlinear Sciences

Belgrade, 2020

SERBIA

Naslov: 1st CONFERENCE ON NONLINEARITY
(Prva konferencija o nelinearnosti)

Autor: Grupa autora

Izdavač: Srpska akademija nelinearnih nauka, Beograd, Srbija

Godina izdanja: 2020

Štampa: Skripta Internacional, Beograd

Tiraž: 150
ISBN: 978-86-905633-4-0

Urednici: Dragović Branko, Čupić Željko
Nelinearne nauke-Zbornici

1st Conference on Nonlinearity

Belgrade, October 11 - 12, 2019

Organizer

- **Serbian Academy of Nonlinear Sciences (SANS), Belgrade**

Coorganizers

- **Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade**
- **SEENET-MTP (Southeastern European Network in Mathematical and Theoretical Physics, Niš)**

Sponsor

- **Ministry of Education, Science and Technological Development, Republic of Serbia**

Program Committee

Milivoj Belić (Texas A & M University in Qatar, Doha, Qatar)
Milan Dimitrijević (Belgrade Observatory, and SANS, Serbia)
Branko Dragovich (Institute of Physics, MI SASA, and SANS, Serbia)
Vladan Djordjević (Serbian Academy of Sciences and Arts, and SANS, Serbia)
Branislav Sazdović (Institute of Physics, and SANS, Serbia)
Branko Urošević, (Faculty of Economics, University of Belgrade, and SANS, Serbia)

Organizing Committee

Branko Dragovich, Chairman (Belgrade, Serbia)
Željko Čupić, Vice Chairman (Belgrade, Serbia)
Ćemal Dolićanin (Novi Pazar, Serbia)
Goran Djordjević (Niš, Serbia)
Božidar Jovanović (Belgrade, Serbia)
Nikolaj Ostrovski (Subotica, Serbia)

PREFACE

This book contains some papers related to the talks presented at the first *Conference on Nonlinearity*, held at the Mathematical Institute of the Serbian Academy of Sciences and Arts, on October 11–12, 2019, Belgrade, Serbia. This conference is organized by the Serbian Academy of Nonlinear Sciences (SANS) in cooperation with the Mathematical Institute SASA and the Southeastern European Network in Mathematical and Theoretical Physics (SEENET-MTP).

Nonlinear phenomena are present everywhere in nature – from interactions between elementary particles to the dynamics of the universe as a whole. Nonlinear methods are used in research of all sciences – from fundamental to applied. Contemporary comfortable human life largely depends on technological achievements based on nonlinear processes.

Serbian Academy of Nonlinear Sciences is a scientific society whose members are scientists that significantly contributed to developments of nonlinear sciences in Serbia. The main goal of SANS is a strong fruitful suport to versatile developments of nonlinear sciences, particularly in Serbia. SANS has its origin in *South-Slavenian Academy of Nonlinear Sciences* founded on May 20, 1998 in Belgrade, as a branch of the *Academy of Nonlinear Sciences* in Moscow. Presently SANS has 30 members: 16 full (academicians), 7 corresponding, 2 honorary, and 5 foreign members. Organization of scientific meetings – colloquiums and conferences on nonlinearity – are among principal activities of SANS.

This *Conference on Nonlinearity* was organized on the occasion of the 110th anniversary since the birth of Nikolay Nikolaevich Bogolyubov, an outstanding world known Soviet and Russian scientist, who made significant contributions in many parts of theoretical and mathematical physics. This volume of proceedings contains a review

article on the life and scientific activities of N. N. Bogolyubov, written by A. G. Zagorodny – academician of the National Academy of Sciences of Ukraine and director of the *Bogolyubov Institute for Theoretical Physics* of the National Academy of Sciences of Ukraine, Kiev. It is worth noting that Nikolay N. Bogolyubov published his first scientific paper at the age of 15, and obtained a Candidate of Sciences degree and Doctor of Sciences degree at the ages of 19 and 21, respectively. Among other activities, N. N. Bogolyubov was the director of the Joint Institute for Nuclear Research (JINR), Dubna, Russia, fo many years. One of us (B. Dragovich) spent one year (1976) and after that he visited JINR many times. N. N. Bogolyubov visited the Institute of Physics, Belgrade, 1978 and signed a document on scientific collaboration between JINR and the Institute of Physics. At that time, B. Dragovich had the opportunity to spend a week with N. N. Bogolyubov and enjoy his very pleasant communication and scientific influence.

On behalf of all involved in the organization of this conference, we would like to express our gratitude to the Ministry of Education, Science and Technological Development of the Republic of Serbia for a very useful financial support, including the publication of these Proceedings. We are also thankful to all the speakers at the conference and the authors of contributions to the Proceedings. We hope that this collection of papers will be useful not only to participants of this conference but also to all others who are interested in nonlinearity.

SANS plans to continue with the organization of *Conference on Nonlinearity* in the future – one international conference with a period of two years. We will be happy to see all participants of the first conference again and welcome new ones.

Belgrade, June 2020

E d i t o r s

Branko Dragovich
(President of SANS)

Zeljko Cupic
(Scientific secretary of SANS)

CONTENTS

Founder of modern theoretical and mathematical physics

A. G. Zagorodny*
Bogolyubov Institute for Theoretical Physics,
Nat. Acad. of Sci. of Ukraine, Kyiv, Ukraine

Abstract

The main periods of life and scientific activity of Academician Bogolyubov are presented. His role in the progress of theoretical and mathematical physics, the formation of scientific schools and the foundation of new departments, scientific laboratories and institutes are discussed.

"Science is the main and only goal in my life."
Mykola Bogolyubov

The Conference on Nonlinearity organized on the occasion of the 110-th anniversary of Mykola Bogolyubov provided a good opportunity to recollect the scientific achievements of this brilliant scientist and to discuss once again his invaluable contribution to the development of various fields of theoretical and mathematical physics – nonlinear mechanics, nuclear physics, quantum field theory, high energy physics, condensed matter physics, etc. The departments, research groups, and even whole institutes founded by him continue to work nowadays. The Bogolyubov scientific school, which has grown on his ideas, is successfully developing already in its fourth generation.

Mykola Bogolyubov was born on August 21, 1909 in Nizhny Novgorod. That very year the Bogolyubov family moved from Nizhny Novgorod to Nizhyn, Chernihiv province, where Mykola's father – by that time already

* e-mail address: azagorodny@bitp.kiev.ua

known theologist Mykola Mykhailovych Bogolyubov – got the position of professor of scripture at the Prince Bezborodko Historical and Philological Institute. It should be noted that this institute had a long educational traditions and a high reputation. Mykola Hohol, Yevhen Hrebinka, Leonid Hlibov, and many other outstanding figures had been its students. Four years later, Mykola Mykhailovych became a professor of theology at St. Vladimir' University and the family moved to Kiev.

In 1917, when Bogolyubov was eight, he entered the preparatory class of the First Alexander Classic Kiev Gymnasium, but he studied there for less than two years. In 1920, Soviet power was finally established in Kiev, the Department of Theology was closed, and Mykola's father was forced to take a parish in the village of Velyka Krucha, Poltava province. Mykola began to attend the Velykokruchans'ka seven-year school and graduated in 1922. When recollecting this school, Bogolyubov said that it's pedagogical team would be honor to the best schools of the capital. By the way, the certificate on the graduation of the seven-year school was the only document on education that Bogolyubov received in his life, and the words about his officially received education are "I became a scientist in Velika Krycha". Due to school and home education, at the age of 13 Mykola Bogolyubov had knowledge at the graduate level of the faculty of physics and mathematics of the university.

In 1922, the Bogolyubov family returned to Kiev. Bogolyubov's father asked the famous mathematician academician Dmytro Grave for advise concerning the further education of his elder son. Professor Grave, after acquaintance with young Bogolyubov, told his father that attending lectures at any university would make no sense for the young man, but advised to continue education individually. Since then Bogolyubov began to participate in the seminars of Academician Grave. In the spring of 1923 young Bogolyubov met his teacher, mentor, and future colleague Academician Mykola Krylov, who began to give special classes in mathematics and mechanics for Bogolyubov. When Mykola Bogolyubov was 15 years old, he published his first scientific paper, and on June 1, 1925, a special decree was adopted by the Ukrhlavnauka, which stated: "In view of phenomenal gift for mathematics, to consider M. Bogolyubov as a postgraduate student of the Department of Mathematics since July 18, 1925. Add him to the payroll list".

In 1930, Bogolyubov received his first recognition – he got the Award of the Academy of Sciences of Bologna (Italy). The same year, namely, on April 6, 1930, at the General Meeting of the Department of Physics and Mathematics of the All-Ukrainian Academy of Sciences (VUAN), on the recommendation of D. Grave and M. Krylov, the degree of Doctor of Sciences was awarded to Bogolyubov without thesis. Academic title of Professor on the Department of Theory of Functions was conferred on Bogolyubov in 1936 after he had began teaching at Kiev University. Later, in 1939, M. Bogolyubov was elected a corresponding member, in 1948 – academician of the Academy of Sciences of the Ukrainian SSR. In 1947 he

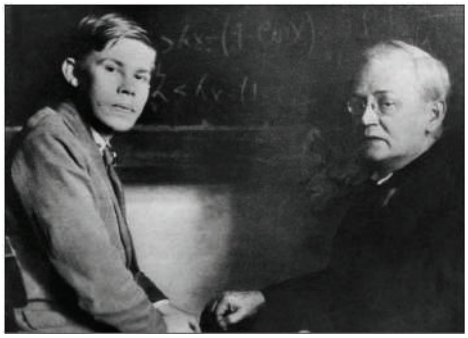


Figure 1: M.M. Bogolyubov and Academician M.M. Krylov

became a corresponding member of the Academy of Sciences of the USSR, and in 1953 he became its full member.

The topmost results of Bogolyubov in 1932–37 include the foundation, together with his teacher, of a new section of mathematical physics – the theory of nonlinear oscillations, that later would be called nonlinear mechanics. In particular, they have developed new methods for integrating nonlinear differential equations describing vibration processes. These results have been summarized in many joint monographs by Bogolyubov and Krylov of this period. Among them are "On some formal decompositions of nonlinear mechanics", "New methods of nonlinear mechanics", "Application of methods of nonlinear mechanics to the theory of stationary oscillations", "Introduction to nonlinear mechanics". In 1955 a fundamental monograph by Bogolyubov and Mitropolsky "Asymptotic Methods in the Theory of Nonlinear Oscillations" was published.

In 1935–1936, Bogolyubov represents the Department of Mathematical Physics of VUAN abroad. He gives lectures on the theory of nonlinear oscillations at Henri Poincaré Institute in France, Belgian Mathematical Society, and Belgian Research Institute. In 1940, after the unification of the Northern Bukovyna with Ukraine Bogolyubov participated in the formation of the mathematical department at the Physics and Mathematics Faculty of Chernivtsi University.

In summer of 1941, M. Bogolyubov was evacuated to Ufa and then sent to Moscow. At this time, as Mykola Mykolayovych writes in his autobiography, he, while continuing theoretical studies in nonlinear mechanics, was mainly concerned with defense topics. Bogolyubov returned to Kiev in early 1944.

One of the most fruitful periods of creativity of Bogolyubov is associated with Kiev. Just here M. Bogolyubov initiated new fields of theoretical and mathematical physics, in particular, wrote his classical works on modern statistical theory.Particularly, in 1946 M. Bogolyubov published the world-famous book "Problems of the Dynamic Theory in Statistical Physics". The results given in this work have initiated a new stage in the

Figure 2: Opening of a new building of the Institute for Theoretical Physics

progress of statistical mechanics after the achievements related to Maxwell, Boltzmann, Gibbs. Bogolyubov proposed a dynamic approach to the formulation of the kinetic theory based on the chain of equations for equilibrium and nonequilibrium many-particle distribution functions – the chain of Bogolyubov-Borne-Green-Kirkwood-Yvon equations (here it should be noted that Bogolyubov gave the most general and mathematically rigorous chain derivation). Using a small-scale expansion of this chain and applying the assumption of the existence of a hierarchy of time scales (known in the world literature as the hierarchy of Bogolyubov's characteristic times), Bogolyubov obtained closed kinetic equations for one-particle distribution functions not only for neutral gas but also for plasma. The latter equation today is called the Bogolyubov-Balescu-Lennard kinetic equation. Instead of Boltzmann's hypothesis of molecular chaos, he proposed the principle of complete weakening of initial correlations (Bogolyubov's principle), which made it possible to calculate collision integrals on the basis of a reduced chain of equations for distribution functions. To describe the next stage in the evolution of the system, Bogolyubov obtained the equation of hydrodynamics.

1947 – another brilliant result: the microscopic theory of superfluidity. The article in which this theory was formulated has for many years remained one of the most cited works of our time. In this work Bogolyubov for the first time applied a new mathematical technique known today as Bogolyubov's canonical transformation. On the example of a weakly idealized Bose gas, Bogolyubov explained from the first principles the formation of the excitation spectrum of a superfluid helium and thus the nature of this macroscopic quantum phenomenon. He later summarized his mathematical

Figure 3: Academicians O.S. Davydov, O.G. Sitenko, and M.M. Bogolyubov after the opening of a new building of the Institute (1970)

formalism for the foundation of a microscopic theory of superconductivity. Bogolyubov perfectly studied the methods of secondary quantization for quantum statistical systems. His "Lectures on Quantum Statistics", published in 1949, could be a good illustration of effective application of this method to quantum statistics. This contributed to his interest in the problems of quantum field theory, where he also managed to obtain a number of outstanding results. A brilliant example is the development of a method for eliminating divergencies in the quantum field theory based on the use of the subtraction procedure, and proving one of the central theorems of the renormalization theory, known as the Bogolyubov-Parasyuk theorem. The discovery of the general form of the subtraction procedure and its justification were of great importance for the further development of high-energy physics. It made it possible, in particular, to prove the renormalizability of a unified theory of electroweak interactions, as well as of supersymmetric theories, to obtain operator expansions at short distances, to study phase transitions, and so on.

In 1951–1953, Mykola Bogolyubov worked at the top-secret object of the Soviet Union – "Arzamas-16" (Sarov), as well as at the Institute of Atomic Energy (now "Kurchatov Institute" in Moscow), where in parallel with the mathematical studies of the problems related to the creation of hydrogen weapons he worked on the problems concerning the magnetic fusion reactor. It should be noted that the results obtained then by Bogolyubov in the field of nuclear fusion have not been published since they were a part of secret reports. Only after removing the mark of secrecy from these results it turned out to be that considerable part of the results on the kinetic plasma theory had been obtained by Bogolyubov before they were obtained and published independently by other authors in open literature.

From 1948 M. Bogolyubov along with his work in Kiev began to head the Department of Mathematical Physics at the Institute of Chemical

Figure 4: M.M. Bogolyubov and his disciples, Academicians O.S. Parasyuk (left) and Yu.O. Mytropol's'kyi (right)

Physics in Moscow, and from 1949 – also the Department of Theoretical Physics of the Steklov Mathematical Institute of the Academy of Sciences of the USSR. In 1956 Mykola Bogolyubov became the Director of the Laboratory of Theoretical Physics of the Joint Institute for Nuclear Research (JINR) in Dubna. In January 1965, at the session of the plenipotentiaries of the governments of the member states of the Institute, Mykola Bogolyubov was elected the Director of JINR, which he has headed for over 20 years. Since 1957 Mykola Bogolyubov also headed the Laboratory of the Theory of Atomic Nuclei and Elementary Particles at the Institute of Physics of the Academy of Sciences of the Ukrainian SSR.

Among other results by Bogolyubov concerning the perturbation methods in the quantum field theory, one should also mention the method of renormalization group – the new general approach in theoretical physics, which has found application in various fields.

Mykola Bogolyubov is a founder of a new field of research in the quantum theory that was later called the axiomatic field theory. In particular, the proposed derivation of the dispersion relations has led to the development of a new mathematical approach to the analytical continuation of the generalized functions of many variables. For these studies in 1966 Mykola Bogolyubov was awarded the Danny Heineman Award. In his welcome address, Professor R. Jost said: "You made an unforgettable impression on me. Most theorists at the time were disrespectful of mathematics, and logical deduction was 'trampled'. Only the romantic influence of genius could have value. And then you appeared, a person who knows both mathematics and physics and who is ready to solve complex problems that require their logical combination. It seems to me that this is a reflection of the national character of your great people".

In 1961 M. Bogolyubov introduced the fundamental concept of quasi-averages and thus, in fact, a new theory of phase transformations was

Figure 5: M. Gell-Mann (left) and M.M. Bogolyubov at the conference in Chicago (1967)

created. The spread of these ideas to the physics of elementary particles was called spontaneous symmetry breaking – another fundamental result of Mykola Bogolyubov, which is important for quantum physics.

During the period of 1964–1966 Bogolyubov published important papers on the symmetry theory and quark models of elementary particles. One of the important results in this field is the introduction of the new quantum number for quarks, now known as color, proposed by him and his disciples.

Bogolyubov's scientific activity revealed the unity of the mathematical structure of theories for different branches of physics. The follower of Bogolyubov academician V. Vladimirov noted: "Combination of mathematics and physics in the works of M.M. Bogolyubov made it possible for him to contribute considerably to the development of theoretical physics and in fact to create the foundations of modern mathematical physics, which continues the traditions of Hilbert, Poincare, Einstein, Dirac".

Mykola Bogolyubov had the talent of a great researcher and outstanding organizer of science. An example confirming his organizational skills is the foundation in 1966 of the Institute for Theoretical Physics that since 1993 is called by his name. It should be noted that the creation of an elite physical institute in Kiev was an extremely difficult task. There were several reasons. These include the existence in the USSR of the Institute for Theoretical Physics of the USSR Academy of Sciences in Chornogolovka (now the Landau Institute for Theoretical Physics of RAS), and the inconsistency to the 'general line', according to which the priority in the development of fundamental research belonged obviously not to Ukraine, and also problems with the formation of highly skilled staff capable to perform competitive research. And it was necessary to have the influence and weight of Mykola Bogolyubov to succeed. It was also important that the

Figure 6: Sitting: A.S. Davydov, G. Lipkin, M.M. Bogolyubov, R. Marshak. Stand: A.N. Tavhelidze, Yu.A. Mitropolsky, O.S. Parasyuk, M.K. Polivanov, V.P. Shelest, A.I. Akhiezer (left to right)

First Secretary of the Communist Party of Ukraine Petro Shelest and President of the Academy of Sciences of Ukrainian SSR Borys Paton gave him great help and assistance in this matter. As a result of their joint efforts, on January 5, 1966, the Council of Ministers of the Ukrainian SSR adopted a decree "On the Establishment of the Institute for Theoretical Physics of the Academy of Sciences of the Ukrainian SSR", and in 1970, during the Rochester Conference, a new building of the institute was opened on the site chosen by Bogolyubov.

Everything related to the foundation of the Institute has been done with the direct participation of Bogolyubov – from the choice of the site for the institute building to staff appointments. He formulated the main fields of scientific activity of the Institute, namely: elementary particle theory, theory of nucleus and nuclear reactions, and statistical physics. Mykola Bogolyubov invited to the institute outstanding scientists, including his talented students. Among the scientists with world names whom he invited were academicians A. Davydov, A. Petrov, A. Sitenko, I. Yukhnovsky; students of Bogolyubov: A. Tavhelidze (later academician of the Russian Academy of Sciences), academicians of the NAS of Ukraine O. Parasyuk, D. Petryna, corresponding member of the Academy of Sciences of Ukraine V. Shelest and others. As a result, for the first seven years of the directorship of Mykola Bogolyubov the Institute has become a powerful center of theoretical physics, well known not only in Ukraine but also far beyond its borders.

Bogolyubov paid much attention to the development of international cooperation, in particular the organization of international conferences such

Figure 7: E.C.G. Sudarshan (USA), R.E. Marshak (USA), M.M. Bogolyubov, and V.P. Shelest at the Institute for Theoretical Physics during the XV International Conference on High Energy Physics (Kyiv, 1970)

as Rochester Conference on high-energy physics and international conferences on plasma theory initiated by him together with A. Sitenko. These conferences proved to be so successful that they were called the "Kiev Conferences on Plasma Theory" and were held under this name in many countries around the world, periodically returning to Kiev in 1976, 1987, and 2006.

As already mentioned, the scientific fields formulated by Mykola Bogolyubov have determined the activities of the Institute for many years. Today its main activities are related to high-energy physics and astrophysics, nucleus theory, quantum field theory and symmetry theory, the theory of nonlinear phenomena in condensed matter and plasmas, as well as the kinetic theory of highly nonequilibrium processes. In fact, this corresponds to somewhat extended trends formulated by Bogolyubov. We can say that much of the research activities of the Institute are related to the application and development of the ideas of Mykola Bogolyubov. In particular, in the field of theoretical high-energy physics, this concerns the dynamic generation of masses, spontaneous symmetry breaking, quantum chromodynamics, and the application of symmetry theory in quantum field theory. The same concerns Bogolyubov's ideas in the kinetic theory. As noted above, Mykola Bogolyubov is one of founders of the theory of many-particle systems. Previously, such theory was used to describe gases and plasma.

But Bogolyubov's methods have also proved to be effective for describing much more complex systems, in particular for the study of dusty plasmas i.e. a mixture of plasmas and solid particles. Creative inheritance of Bogolyubov is also used today to solve the problems of condensed matter

Figure 8: First International Conference on Plasma Theory (Kyiv, Institute for Theoretical Physics of the Academy of Sciences of the UkrSSR, 1971)

physics. These include the description of high-temperature superconductivity, the phenomenon of Bose condensation in various systems, nonlinear phenomena in solids and liquids, transport processes in molecular systems, and the kinetics of electron transport in nanoobjects. Methods of quantum field theory are, in turn, widely used in the study of low-dimensional and so-called Dirac structures as well as new materials.

Along with scientific researche and organizational activities, Bogolyubov carried out impressive pedagogical work. In 1936–1941 and 1944–1949 he taught at Kyiv State University, in 1945–1948 he was Dean of the Faculty of Mechanics and Mathematics, where he founded and headed the Department of Mathematical Physics. From November 1943 he was Professor at the Lomonosov Moscow University. In January 1953, Bogolyubov was elected as the Head of the Department of Theoretical Physics of that University, where in 1966 he also founded the well-known Bogolyubov Department of Quantum Statistics and Field Theory.

It is important to note that Mykola Bogolyubov's life and work from the first years the last days were closely connected with Ukraine. Being ethnic Russian in origin, he was brought up in the atmosphere of deep love to Ukraine, he felt great respect for the land where his childhood and adolescence passed, where he made his first steps in science and gained worldwide name. Desiring to share the fate of the Ukrainian people in everything, he considered himself Ukrainian, as he wrote about himself in all questionnaires and personal papers. The same entry was in his Soviet passport. Mykola Mykolayovych's attitude to Ukraine is comprehensively characterized by Alexey Bogolyubov's words about his elder brother: "Mykola Mykolayovych had two homelands – Russia and Ukraine and two native languages – Russian and Ukrainian. Beginning from the Velyka Crucha

years, he became associated with Ukraine, and Shevchenko's poetry was, in fact, the first poetry he became interested in. The young graduate student of the Department of Mathematical Physics wrote the minutes of the seminars of the department in Ukrainian, and his first works were also written in Ukrainian." And further: "Mykola Mykolayovych, in the difficult times for Ukraine, when the Ukrainian intelligentsia started to be destroyed, when the shameful process of the Ukrainian Liberation Union took place in Kharkiv and Ukrainian books burned, he admitted himself to be a Ukrainian and so considered himself for his whole life. It is an indisputable fact that all the development of his personality and the acquisition of features of scientific creativity took place in Ukraine, and were also closely associated with Ukraine. He used to call Kyiv his favorite city, equating to him only Paris". Although these words are well known and have been cited for many times in articles about the Ukrainian period of Bogolyubov and memories of him, we have to mention them here, because they reveal the origins of Bogolyubov's love for Ukraine. Mykola Mykolayovych's attitude to his native Ukrainian land, to the Ukrainian language, should be a good example for many of our compatriots.

Mykola Bogolyubov passed away on February 13, 1992. He has left invaluable scientific heritage, numerous scientific schools, a large cohort of students and followers, with whom he always shared scientific ideas and interesting research.

M.M. Bogolyubov was a scholar of wide international recognition. He was elected a member of 10 foreign academies of sciences and was awarded the honorary doctorate of 10 foreign universities. Foreign state and scientific awards also testify to the recognition of Bogolyubov's contribution to the world science. In particular, he is a winner of the Prize of the Academy of Sciences of Bologna (1930, Italy), Heinemann Prize of the American Physical Society (1966), Helmholtz Gold Metal (1969), Max Planck medal of the Physical Society of Germany (1973), Franklin Prize (1974, USA), Prize of the Slovak Academy of Sciences (1975), Paul Dirac Prize (1992) and others.

In 1987, the International Center for Theoretical Physics in Trieste founded the Bogolyubov Prize for outstanding achievements in mathematics and solid state physics for scientists from developing countries. The National Academy of Sciences of Ukraine has also established the Bogolyubov Prize for the research in mathematics and physics. The Russian Academy of Sciences founded in 1999 the Bogolyubov Gold Medal, for research in the field of mathematical physics and mathematics. The Bogolyubov Gold Medal was also founded in JINR. At 2018 the Bogolyubov Institute for Theoretical Physics started awarding Bogolyubov Prizes for the best works in theoretical and mathematical physics.

The monuments of Academician Bogolyubov were erected in Nizhny Novgorod and Dubna, and his busts were located in Kiev at the Bogolyubov Institute for Theoretical Physics and at the JINR Laboratory of Theoretical Physics. A memorial plaque honoring Mykola Bogolyubov decorates the

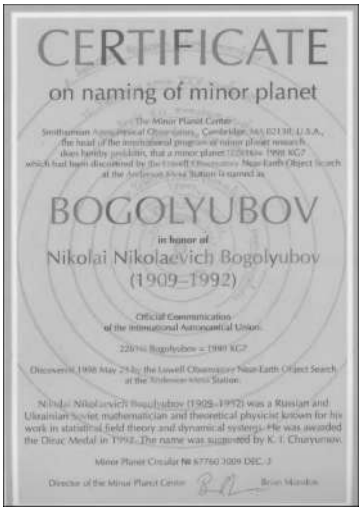


Figure 9: Certificate on naming of a minor planet in the solar system as Bogolyubov

Figure 10: M.M. Bogolyubov (left) and W. Heisenberg (right) after receiving Max Plank medal by Bogolyubov (1973)

Red Building of Taras Shevchenko National University of Kyiv. A memorial sign in honor of Mykola Mykolayovych was erected in the village of Veleka Krucha.

The 100th anniversary of Bogolyubov birth is widely celebrated in Ukraine. The International Bogolyubov Conference "Modern Problems of Theoretical and Mathematical Physics" and the II Ukrainian Mathematical Congress were held in Kyiv, and the anniversary Bogolyubov Conference was also held in Lviv; books and articles about the life and work of the great scientist were published. The anniversary coin and the Bogolyubov Medal of the Ukrainian Mathematical Congress were minted.

On December 3, 2009, at the application of the famous Ukrainian astronomer K.I. Churyumov initiated by the Bogolyubov Institute for Theoretical Physics, the International Astronomical Union adopted the decision to give the minor planet of the Solar System (22616) = 1998 KG7 the name Bogolyubov.

A brilliant scientist continues his life in the works of his students and numerous followers, including those who work at the Bogolyubov Institute for Theoretical Physics, and we are sure that the ideas of Mykola Bogolyubov will inspire many theorists for many years.

Acknowledgement

The author is grateful to the Publishing House "Academperiodyka" for permission to use some of the materials from the book "Founder of Theoretical and Mathematical Physics. To the 110th Anniversary of Academician M.M. Bogolyubov" Academperiodyka, Kyiv, 2019.

From p-Adic to Zeta Strings*

Branko Dragovich†
Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia
Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade, Serbia

June 4, 2020

Abstract

This article is related to construction of zeta strings from p-adic ones. In addition to investigation of p-adic string for a particular prime number p, it is also interesting to study collective effects taking into account all primes p. An idea behind this approach is that a zeta string is a whole thing with infinitely many faces which we see as p-adic strings. The name zeta string has origin in the Riemann zeta function contained in related Lagrangian. The starting point in construction a zeta string is Lagrangian for a p-adic open string. There are two types of approaches to get a Lagrangian for zeta string from Lagrangian for p-adic strings: additive and multiplicative approaches, that are related to two forms of the definition of the Riemann zeta function. As a result of differences in approaches, one obtains several different Lagrangians for zeta strings. We briefly discuss some properties of these Lagrangians, related potentials, equations of motion, mass spectrum and possible connection with ordinary strings. This is a review of published papers with some new views.

1. Introduction

The subject of this paper belongs to String Theory (ST) [1], in particular to p-adic and zeta strings. String theory emerged at the end of 1960th and has been developed as the best candidate for unification of fundamental interactions (gravitational, electromagnetic, strong and weak) and elementary matter constituents in the form of strings. Strings are one-dimensional objects that exist at the very short distances (close to the natural Planck length $\ell_P = \sqrt{\frac{G\hbar}{c^3}} \sim 10^{-35}$m). Although strings are not yet experimentally discovered, string theory has played very important role in an interplay between general physical laws and modern mathematics [1].

* This article is based on a talk given at the first *Conference on Nonlinearity*, held in Belgrade, 11-12 October 2019, and dedicated to the 110th anniversary of Nikolay Nikolayevich Bogolyubov – an outstanding Soviet and Russian scientist.

† e-mail address: dragovich@ipb.ac.rs

p-Adic strings [2, 3] are introduced in string theory in 1987 as a p-adic analog of ordinary strings. In the case of open and closed strings, it was shown that there is a connection between ordinary and p-adic strings in the form of product of their scattering amplitudes which is a constant. By construction of Lagrangian for p-adic strings it occurred that p-adic strings are even simpler for mathematical investigation than ordinary strings. Interesting properties of p-adic strings motivated construction of some other physical models using methods of p-adic analysis, and it has resulted in emergence of *p-Adic Mathematical Physics*, e.g., see reviews [2, 3, 4] and references therein.

Note that in the word "p-adic", p is related to a prime number. Since there are infinitely many prime numbers, it should be infinitely many p-adic strings. Then question arises how to connect p-adic counterparts with ordinary model over real (or complex numbers). Usual way to connect ordinary model with p-adic analogs of the same physical system is by using adelic approach, which is based on adelic analysis. To have insight into adelic aspects of strings, one can consume [2, 3]. An adelic model of the quantum harmonic oscillator is given in [5].

It seems unnatural that there exist infinitely many kinds of p-adic strings – one string for each prime number p. It is more natural to think that there exists one new entity (say zeta string) which has infinitely many faces – one face for each p. Fortunately, in the case of p-adic open strings there is possibility to work in this direction. Namely, there are Lagrangians for p-adic open strings, which have the same form and the difference is practically in number p that serves as a parameter. As it will be shown in the sequel of this paper, one can start with this Lagrangian of p-adic string, extend p to n (natural numbers), and then by "integration" over n to get a new Lagrangian without n. This new Lagrangian is related to a new thing, which we call *zeta string*, since this Lagrangian contains the Riemann zeta function. It occurs that obtaining of this new Largangian is not unique and depends on applied procedure. This review article is devoted to construction of Lagrangians for zeta strings.

The paper is designed as follows. Some basic concepts from p-adic mathematics will be recalled in the next section. A brief review of p-adic open string, emphasizing related Lagrangian, will be given in section 3. Construction of Lagrangians for zeta strings, and some their elaboration, is presented in section 4. Several concluding remarks are given in section 5.

This review is based on a series of papers [6]–[12], some of them published in the journal *Theoretical and Mathematical Physics*, which was founded in 1969 by N. N. Bogolyubov.

2. A Brief Review of p-Adic Mathematical Background

Recall that rational numbers play important role in physics and mathematics. In physics, numerical results of experiments are rational numbers. In

mathematics, rational numbers make a simple infinite number field $\mathbb{Q}$ with respect to summation and multiplication. On $\mathbb{Q}$, in addition to the usual absolute value, there is also well defined p-adic norm (or in other words – p-adic absolute value). For a given prime number p, any nonzero rational number can be presented as $x = \frac{a}{b}p^{\mu}$, where $a, b \neq 0$ are integers not devisable by p, and $\mu \in \mathbb{Z}$. Then, according to definition, p-adic norm of x is $|x|_p = p^{-\mu}$ and $|0|_p = 0$. p-Adic distance between two rational numbers x and y is defined as $d_p(x, y) = |x - y|_p$. p-Adic distance is well known example of ultrametrics, because it satisfies strong triangle inequality, i.e. $d_p(x, y) \leq \max\{d_p(x, z), d_p(y, z)\}$.

Rational numbers, and their completion, with respect to the p-adic norm are p-adic numbers. Any nonzero p-adic number x can be presented in a unique way as

$$x = p^{\nu} \sum_{n=0}^{+\infty} x_n \, p^n, \quad x_0 \neq 0, \quad \nu \in \mathbb{Z}, \quad x_n \in \{0, 1, 2, \ldots, p-1\}, \qquad (1)$$

where x_n are digits. For a given p, all numbers (1) make field $\mathbb{Q}_p$ of p-adic numbers. There are infinitely many different $\mathbb{Q}_p$ – one number field for each prime number p.

There are mainly two kinds of analysis with p-adic arguments: 1) p-adic valued functions and 2) complex (or real) valued functions. On $\mathbb{Q}_p$ two important continuous complex-valued functions are defined [3]:

- 1) multiplicative character $\pi_p(x) = |x|_p^c$, where $x \in \mathbb{Q}_p^* = \mathbb{Q}_p \setminus \{0\}$ and $c \in \mathbb{C}$;
- 2) additive character $\chi_p(x) = e^{2\pi i \{x\}_p}$, where $x \in \mathbb{Q}_p$ and $\{x\}_p$ is fractional part of x.

Real and p-adic numbers have their origin in rational numbers. $\mathbb{Q}$ is dense subset in $\mathbb{R}$ and all $\mathbb{Q}_p$. This fact enables a unification of real and all p-adic numbers as a ring of adeles. An adele (α) is an infinite sequence that takes into account real and all p-adic numbers:

$$\alpha = (\alpha_\infty, \ \alpha_2, \ \alpha_3, \ldots, \alpha_p, \ldots), \qquad (2)$$

where $\alpha_\infty \in \mathbb{R}$ and $\alpha_p \in \mathbb{Q}_p, \ p = 2, 3, 5, 7, \ldots$ with some restrictions.

The following product formulas connect real and p-adic properties of the same rational number:

$$|x| \prod_p |x|_p = 1, \quad \text{if } x \in \mathbb{Q} \setminus \{0\}, \qquad (3)$$

$$e^{-2\pi i x} \prod_p e^{2\pi i \{x\}_p} = 1, \quad \text{if } x \in \mathbb{Q}. \qquad (4)$$

As a comprehensive review on p-adic numbers, adeles and their analysis, e.g., see [2, 3, 13].

3. p-Adic Open Scalar Strings

Recall that ordinary string theory started by construction of the Veneziono amplitude for scattering of two open bosonic strings $A_\infty(a,b)$, which in the crossing symmetric form is

$$A_\infty(a,b) = g_\infty^2 \int_{\mathbb{R}} |x|_\infty^{a-1} \, |1-x|_\infty^{b-1} \, d_\infty x \tag{5}$$

$$= g_\infty^2 \, \frac{\zeta(1-a)}{\zeta(a)} \, \frac{\zeta(1-b)}{\zeta(b)} \, \frac{\zeta(1-c)}{\zeta(c)}, \tag{6}$$

where a, b, c are complex parameters related to momenta of strings and satisfy equality $a+b+c=1$, g_∞ is a coupling constant, $|\cdot|_\infty$ denotes usual absolute value, x is real variable related to the string world-sheet, and ζ is the Riemann zeta function.

Scattering amplitude between two p-adic strings was introduced as p-adic analog of integral (5), i.e.

$$A_p(a,b) = g_p^2 \int_{\mathbb{Q}_p} |x|_p^{a-1} \, |1-x|_p^{b-1} \, d_p x \tag{7}$$

$$= g_p^2 \, \frac{1-p^{a-1}}{1-p^{-a}} \, \frac{1-p^{b-1}}{1-p^{-b}} \, \frac{1-p^{c-1}}{1-p^{-c}}, \tag{8}$$

where now x is world-sheet variable described by p-adic (instead of real) numbers, while parameters a, b, c remain their properties as in the case of ordinary strings. According to this definition of string amplitudes, it follows that p-adic and ordinary strings differ in description of their world-sheet, i.e. by p-adic and real numbers, respectively. Final expressions for amplitudes of ordinary (6) and p-adic (8) strings differ, but they are connected by the Freund-Witten product formula

$$A(a,b) = A_\infty(a,b) \prod_p A_p(a,b) = g_\infty^2 \prod_p g_p^2 = const. \tag{9}$$

In (9) is used the Euler product formula for the definition of the Riemann zeta function

$$\zeta(s) = \prod_p \frac{1}{1-p^{-s}}, \quad s = \sigma + i\tau, \quad \sigma > 1. \tag{10}$$

Importance of product (9) consists in possibility to express complex ordinary string amplitude as product of all inverse p-adic amplitudes which are simpler than the ordinary one, i.e. $A_\infty(a,b) = const. \prod_p A_p^{-1}(a,b)$. It also gives rise to think that not only ordinary strings may exist but also p-adic ones, or a new string (zeta string) that encompasses all p-adic effects.

One of the main achievements in p-adic string theory is finding of an effective field description of p-adic strings without explicit use of p-adic numbers. The corresponding *Lagrangian* is very simple and exact. It describes not only four-point scattering amplitude but also all higher ones at the tree-level.

This Lagrangian for effective scalar field φ, which describes open p-adic string (tachyon), is

$$\mathcal{L}_p(\varphi) = \frac{m_p^D}{g_p^2}\,\frac{p^2}{p-1}\Big[-\frac{1}{2}\,\varphi\, p^{-\frac{\Box}{2m_p^2}}\,\varphi + \frac{1}{p+1}\,\varphi^{p+1}\Big], \tag{11}$$

where p is any prime number, $\Box = -\partial_t^2 + \nabla^2$ is the D-dimensional d'Alembert operator, m_p is string mass and metric has signature $(-\,+\,...\,+)$. Kinetic term in (11) contains nonlocal operator with infinite number of space-time derivatives

$$p^{-\frac{\Box}{2m_p}} = \exp\Big(-\frac{\ln p}{2m_p}\,\Box\Big) = \sum_{k\geq 0}\Big(-\frac{\ln p}{2m_p}\Big)^k\,\frac{1}{k!}\,\Box^k\,. \tag{12}$$

Potential $V_p(\varphi)$ is

$$V_p(\varphi) = -\mathcal{L}_p(\Box = 0) = \frac{m_p^D}{g_p^2}\,\frac{p^2}{p-1}\Big[\frac{1}{2}\,\varphi^2 - \frac{1}{p+1}\,\varphi^{p+1}\Big] \tag{13}$$

and contains nonlinearity of the degree $p+1$.

The equation of motion for the scalar field φ in Lagrangian (11) is

$$p^{-\frac{\Box}{2m_p^2}}\,\varphi = \varphi^p, \tag{14}$$

which has two trivial solutions $\varphi = 0$ and $\varphi = 1$. There are also inhomogeneous solutions resembling solitons, and for any spatial coordinate x^i one has

$$\varphi(x^i) = p^{\frac{1}{2(p-1)}}\,\exp\Big(-\frac{p-1}{2m_p^2\,p\ln p}(x^i)^2\Big). \tag{15}$$

It is worth mentioning that taking limit $p = 1+\varepsilon \to 1$ in Lagrangian (11) one obtains

$$\mathcal{L}_1 = \frac{m^D}{g^2}\Big[\frac{1}{2}\,\varphi\frac{\Box}{m^2}\varphi + \frac{\varphi^2}{2}\,(\log\varphi^2 - 1)\Big], \tag{16}$$

and this new Lagrangian (16) is related to ordinary scalar string, see [14].

Starting from Lagrangian (11), many properties of p-adic string were investigated, e.g. see some references in [6] for more information.

4. Zeta Strings

Recall that starting from Lagrangian (11) and using methods of quantum field theory was obtained p-adic string scattering amplitude

$$A_p(a,b) = g_p^2 \frac{1-p^{a-1}}{1-p^{-a}} \frac{1-p^{b-1}}{1-p^{-b}} \frac{1-p^{c-1}}{1-p^{-c}}, \quad (a+b+c=1) \tag{17}$$

which is the same as that derived from integral expression (7) with p-adic world sheet, see [2] for a review.

Since the scattering amplitude for the whole p-adic sector was obtained in the form

$$A(a,b) = \prod_p A_p(a,b) = g^2 \frac{\zeta(a)}{\zeta(1-a)} \frac{\zeta(b)}{\zeta(1-b)} \frac{\zeta(c)}{\zeta(1-c)}, \tag{18}$$

then the question arises about possibility to construct Lagrangian that might produce amplitude (18). If such Lagrangian exists, then to obtain it, we should start with Lagrangian for p-adic string (11) and then perform suitable summation or multiplication over all primes p in such way to have a new Lagrangian with the Riemann zeta function. Thus, there are two approaches: additive and multiplicative.

4.1. Additive approach

To use additive approach, note that the Riemann zeta function can be introduced in the following ways:

$$\zeta(s) = \sum_{n=1}^{+\infty} \frac{1}{n^s}, \quad s = \sigma + i\tau, \ \ \sigma > 1, \tag{19}$$

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{+\infty} \frac{\mu(n)}{n^s}, \quad s = \sigma + i\tau, \ \ \sigma > 1, \ \ \mu(n) = \text{Möbius function} \tag{20}$$

$$(1-2^{1-s})\zeta(s) = \sum_{n=1}^{+\infty} (-1)^{n-1} \frac{1}{n^s}, \quad s = \sigma + i\tau, \ \ \sigma > 0. \tag{21}$$

It is well known that the above defined Riemann zeta function has analytic continuation to the whole complex s-plane except the point $s = 1$, where it has a simple pole with residue 1.

Note also that there is a sense to replace prime number p by any natural number $n \geq 2$ in Lagrangian (11), i.e. one can introduce

$$\mathcal{L}_n(\varphi) = \frac{m_n^D}{g_n^2} \frac{n^2}{n-1} \Big[-\frac{1}{2} \varphi \, n^{-\frac{\Box}{2m_n^2}} \varphi + \frac{1}{n+1} \varphi^{n+1} \Big], \tag{22}$$

Now we want to introduce a new Lagrangian with a new field ϕ by the following sum:

$$L(\phi) = \sum_{n=1}^{+\infty} C_n \mathcal{L}_n$$
$$= \sum_{n=1}^{+\infty} C_n \frac{m_n^D}{g_n^2} \frac{n^2}{n-1} \Big[-\frac{1}{2} \varphi \, n^{-\frac{\Box}{2m_n^2}} \varphi + \frac{1}{n+1} \varphi^{n+1} \Big], \qquad (23)$$

whose concrete realization depends on the particular choice of the coefficients C_n, masses m_n and coupling constants g_n. To avoid a divergence problem of $1/(n-1)$ when $n = 1$, we take that C_n is proportional to $n-1$. We also assume that m_n and g_n do not depend on n, and denote $m_n = m$ and $g_n = g$. Taking $C_n = \frac{n-1}{n^2} D_n$ we can rewrite (23) as

$$L(\phi) = \frac{m^D}{g^2} \sum_{n=1}^{+\infty} D_n \Big[-\frac{1}{2} \phi \, n^{-\frac{\Box}{2m^2}} \phi + \frac{1}{n+1} \phi^{n+1} \Big]. \qquad (24)$$

One can easily see that term with $n = 1$ is equal to zero and does not contribute to the sum in (24), but its presence is useful to perform procedure required by definition of the Riemann zeta function.

In [6]–[12], we introduced new Lagrangians for the following values of coefficient D_n: $D_n = 1$, $D_n = n+1$, $D_n = \mu(n)$, $D_n = -\mu(n)\ (n+1)$, $D_n = (-1)^{n-1}$, $D_n = (-1)^{n-1}\ (n+1)$.

4.1.1. Case $D_n = 1$, [6, 7].

In this case we have

$$L(\phi) = \frac{m^D}{g^2} \sum_{n=1}^{+\infty} \Big[-\frac{1}{2} \phi \, n^{-\frac{\Box}{2m^2}} \phi + \frac{1}{n+1} \phi^{n+1} \Big]. \qquad (25)$$

Performing summation with application of the Euler formula (19) and taking analytic continuation, we can rewrite (25) in the form

$$L(\phi) = -\frac{m^D}{g^2} \Big[\frac{1}{2} \phi \, \zeta\big(\frac{\Box}{2m^2}\big) \phi + \phi + \frac{1}{2} \log (1-\phi)^2 \Big] \qquad (26)$$

The potential $V(\phi) = -L(\Box = 0)$ is

$$V(\phi) = \frac{m^D}{g^2} \Big[\frac{\zeta(0)}{2} \phi^2 + \phi + \frac{1}{2} \log (1-\phi)^2 \Big], \qquad (27)$$

where $\zeta(0) = -1/2$. Potential (27) has two local maximums, that are unstable points: $V(0) = 0$ and $V(3) \approx 1.443 \frac{m^D}{g^2}$. It has the following singularities: $\lim_{\phi \to 1} V(\phi) = -\infty$, $\lim_{\phi \to \pm\infty} V(\phi) = -\infty$.

The corresponding equation of motion is

$$\zeta\Big(\frac{\Box}{2\,m^2}\Big)\phi = \frac{\phi}{1-\phi} \tag{28}$$

with two trivial solutions: $\phi = 0$ and $\phi = 3$. The solution $\phi = 3$ follows from the Taylor expansion of the Riemann zeta function operator

$$\zeta\Big(\frac{\Box}{2\,m^2}\Big) = \zeta(0) + \sum_{n\geq 1} \frac{\zeta^{(n)}(0)}{n!}\,\Big(\frac{\Box}{2\,m^2}\Big)^n. \tag{29}$$

In the weak-field approximation $|\phi(x)| \ll 1$, equation of motion (28) becomes

$$\zeta\Big(\frac{\Box}{2\,m^2}\Big)\phi = \phi. \tag{30}$$

$\zeta\Big(\frac{\Box}{2m^2}\Big)$ can be regarded as a pseudodifferential operator

$$\zeta\Big(\frac{\Box}{2m^2}\Big)\,\phi(x) = \frac{1}{(2\pi)^D}\int_{\mathbb{R}^D} e^{ixk}\,\zeta\Big(-\frac{k^2}{2m^2}\Big)\,\tilde{\phi}(k)\,dk \tag{31}$$

with singularity at point $k^2 = -2m^2$. Now mass spectrum is $M^2 = \mu m^2$, where $M^2 = -k^2 = k_0^2 - \overrightarrow{k}^2$, is determined by formula

$$\zeta\Big(\frac{M^2}{2\,m^2}\Big) = 1 \tag{32}$$

and gives many tachyon masses ($M^2 < 0$).

Note that one can replace above $D_n = 1$ by $D_n = -1$. Then Lagrangian (26) and potential (27) will change their sign, while equation of motion (28) and mass spectrum formula (32) will remain unchanged. This case may be more interesting than the previous one and will be elaborated elsewhere.

In [7], a more general Lagrangian was considered, i.e.

$$L_h(\phi) = \frac{m^D}{g^2}\Big[-\frac{1}{2}\phi\zeta\Big(\frac{\Box}{2m^2}+h\Big)\phi + \mathcal{AC}\sum_{n=1}^{\infty}\frac{n^{-h}}{n+1}\phi^{n+1}\Big], \tag{33}$$

where h is a real parameter (for $h = 0$ Lagrangian (33) reduces to (26)). The related equation of motion is

$$\zeta\Big(\frac{\Box}{2m^2}+h\Big)\phi = \mathcal{AC}\sum_{n=1}^{\infty} n^{-h}\phi^n. \tag{34}$$

Solution of equation of motion (34) is investigated in [15], where LHS is simplified by an entire function of exponential type.

4.1.2. Case $D_n = n + 1$, [8, 9].
In this case, Lagrangian (24) becomes

$$L(\phi) = \frac{m^D}{g^2}\Big[-\frac{1}{2}\,\phi \sum_{n=1}^{+\infty}\Big(n^{-\frac{\Box}{2m^2}+1} + n^{-\frac{\Box}{2m^2}}\Big)\,\phi + \sum_{n=1}^{+\infty}\phi^{n+1}\Big] \tag{35}$$

and according to (19) and analytic continuation we have

$$L(\phi) = \frac{m^D}{g^2}\Big[-\frac{1}{2}\,\phi\Big\{\zeta\Big(\frac{\Box}{2\,m^2}-1\Big) + \zeta\Big(\frac{\Box}{2\,m^2}\Big)\Big\}\,\phi + \frac{\phi^2}{1-\phi}\Big]. \tag{36}$$

Since $\zeta(-1) = -1/12$ and $\zeta(0) = -1/2$, the corresponding potential is

$$V(\phi) = -L(\Box = 0) = \frac{m^D}{g^2}\,\frac{7\phi - 31}{24\,(1-\phi)}\,\phi^2\,, \tag{37}$$

with properties: $V(0) = V(31/7) = 0\,,\ V(1 \pm 0) = \pm\infty\,,\ V(\pm\infty) = -\infty$. At $\phi = 0$ potential has local maximum.

The equation of motion is

$$\Big[\zeta\Big(\frac{\Box}{2\,m^2}-1\Big) + \zeta\Big(\frac{\Box}{2\,m^2}\Big)\Big]\,\phi = \frac{\phi((\phi-1)^2+1)}{(\phi-1)^2}\,, \tag{38}$$

which has only $\phi = 0$ as a real constant solution.

The weak field approximation of (38) is

$$\Big[\zeta\Big(\frac{\Box}{2\,m^2}-1\Big) + \zeta\Big(\frac{\Box}{2\,m^2}\Big) - 2\Big]\,\phi = 0, \tag{39}$$

which implies condition on the mass spectrum

$$\zeta\Big(\frac{M^2}{2\,m^2}-1\Big) + \zeta\Big(\frac{M^2}{2\,m^2}\Big) = 2\,. \tag{40}$$

From (40) follows that there are no finite solutions $M^2 > 2m^2$ and there are many tachyon solutions $M^2 < 0$.

4.1.3. Case $D_n = \mu(n)$, [10].
Related Lagrangian is

$$L_\mu(\phi) = \frac{m^D}{g^2}\sum_{n=1}^{+\infty}\Big[-\frac{1}{2}\,\phi\,\mu(n)\; n^{-\frac{\Box}{2m^2}}\,\phi + \frac{\mu(n)}{n+1}\,\phi^{n+1}\Big]. \tag{41}$$

where $\mu(n)$ is the Möbius function:

$$\mu(n) = \begin{cases} 0, & n = p^2 m \\ (-1)^k, & n = p_1 p_2 \cdots p_k,\ p_i \neq p_j \\ 1, & n = 1,\ (k = 0) \end{cases} \tag{42}$$

Taking into account zeta function by expression (20) one can rewrite Lagrangian (41) as

$$L_\mu(\phi) = \frac{m^D}{g^2}\Big[-\frac{1}{2}\,\phi\,\frac{1}{\zeta\big(\frac{\Box}{2m^2}\big)}\,\phi + \mathcal{AC}\int_0^\phi \mathcal{M}(\phi)\ d\phi\Big], \tag{43}$$

where $\mathcal{AC}$ denotes analytic continuation and $\mathcal{M}(\phi) = \sum_{n=1}^{+\infty}\mu(n)\phi^n = \phi - \phi^2 - \phi^3 - \phi^5 + \phi^6 - \phi^7 + \phi^{10} - \ldots$

The corresponding potential, equation of motion and formula for mass spectrum are, respectively:

$$V_\mu(\phi) = -\frac{m^D}{g^2}\Big[\phi^2 + \mathcal{AC}\int_0^\phi \mathcal{M}(\phi)\ d\phi\Big], \tag{44}$$

$$\frac{1}{\zeta\big(\frac{\Box}{2m^2}\big)}\ \phi = \mathcal{M}(\phi), \quad \zeta\big(\frac{M^2}{2m^2}\big) = 1. \tag{45}$$

In mass spectrum, there are only tachyons.

4.1.4. Case $D_n = -\mu(n)\ (n+1)$, [11].
Related Lagrangian is

$$L_{-\mu}(\phi) = \frac{m^D}{g^2}\ \sum_{n=1}^{+\infty}\phi\Big[\frac{\mu(n)}{2}\ n^{-\frac{\Box}{2m^2}+1} + \frac{\mu(n)}{2}\ n^{-\frac{\Box}{2m^2}} - \mu(n)\,\phi^{n-1}\Big]\phi, \tag{46}$$

where $\mu(n)$ is the Möbius function. Using procedure as in the previous case, one can rewrite Lagrangian (46) in the form

$$L_{-\mu}(\phi) = \frac{m^D}{g^2}\Big\{\frac{1}{2}\,\phi\,\Big[1/\zeta\big(\frac{\Box}{2m^2} - 1\big) + 1/\zeta\big(\frac{\Box}{2m^2}\big)\Big]\,\phi - \phi^2\mathcal{AC}\ F(\phi)\Big\} \tag{47}$$

where $F(\phi) = \sum_{n=1}^{+\infty}\mu(n)\ \phi^{n-1} = (1 - \phi - \phi^2 - \phi^4 + \phi^5 - \ldots)$. Infinite sum $\sum_{n=1}^{+\infty}\mu(n)\ \phi^{n-1}$ is convergent when $|\phi| < 1$.

Taking into account that $\zeta(-1) = -1/12$ and $\zeta(0) = -1/2$, the potential is

$$V_{-\mu} = \frac{m^D}{g^2}\big[7 + \mathcal{AC}\ F(\phi)\big]\phi^2. \tag{48}$$

The related equation of motion and mass spectrum formula are:

$$\Big[\zeta^{-1}\big(\frac{\Box}{2m^2}-1\big)+\zeta^{-1}\big(\frac{\Box}{2m^2}\big)\Big]\,\phi = 2\phi\mathcal{AC}\ F(\phi)+\phi^2\mathcal{AC}\ F'(\phi), \tag{49}$$

$$\zeta^{-1}\big(\frac{M^2}{2m^2}-1\big)+\zeta^{-1}\big(\frac{M^2}{2m^2}\big)=2. \tag{50}$$

4.1.5. Case $D_n=(-1)^{n-1}$.
A new example is based on (21)

$$\sum_{n=1}^{+\infty}(-1)^{n-1}\frac{1}{n^s}=(1-2^{1-s})\,\zeta(s),\quad s=\sigma+i\tau\,,\quad \sigma>0 \tag{51}$$

which has analytic continuation to the entire complex s plane with the corresponding analytic expansion [16]

$$(1-2^{1-s})\,\zeta(s)=\sum_{n=0}^{\infty}\frac{1}{2^{n+1}}\ \sum_{k=0}^{n}(-1)^k\binom{n}{k}(k+1)^{-s}\,. \tag{52}$$

At point $s=1$, one obtains

$$\lim_{s\to 1}(1-2^{1-s})\ \zeta(s)\ =\ \sum_{n=1}^{+\infty}(-1)^{n-1}\frac{1}{n}\ =\ \log 2. \tag{53}$$

Applying formula (51) to Lagrangian (24) and using analytic continuation, we have

$$L=\frac{m^D}{g^2}\ \Big[-\frac{1}{2}\phi\Big(1-2^{1-\frac{\Box}{2m^2}}\Big)\zeta\Big(\frac{\Box}{2m^2}\Big)\phi\ +\phi-\frac{1}{2}\ \log(1+\phi)^2\Big]. \tag{54}$$

The potential is

$$V(\phi)=-L(\Box=0)=\frac{m^D}{g^2}\Big[\frac{1}{4}\,\phi^2-\phi+\frac{1}{2}\log(1+\phi)^2\Big], \tag{55}$$

which has one local maximum $V(0)=0$ and one local minimum $V(1)\approx -0.057\frac{m^D}{g^2}$. It is singular at $\phi=-1$, i.e. $V(-1)=-\infty$, and $V(\pm\infty)=+\infty$. The equation of motion is

$$\Big(1-2^{1-\frac{\Box}{2m^2}}\Big)\zeta\Big(\frac{\Box}{2m^2}\Big)\phi=\frac{\phi}{1+\phi}, \tag{56}$$

which has two constant solutions: $\phi=0$ and $\phi=1$. Formula for the mass spectrum is

$$\Big(1-2^{1-\frac{M^2}{2m^2}}\Big)\zeta\Big(\frac{M^2}{2m^2}\Big)=1. \tag{57}$$

4.1.6. Case $D_n = (-1)^{n-1}(n+1)$, [12, 11].
Applying $D_n = (-1)^{n-1}(n+1)$ and formula (21) to Lagrangian (24), and using analytic continuation we obtain

$$L = -\frac{m^D}{g^2}\Big\{\frac{1}{2}\,\phi\Big[\Big(1-2^{2-\frac{\Box}{2m^2}}\Big)\,\zeta\Big(\frac{\Box}{2\,m^2}-1\Big) + \Big(1-2^{1-\frac{\Box}{2m^2}}\Big)\,\zeta\Big(\frac{\Box}{2\,m^2}\Big)\Big]\,\phi - \frac{\phi^2}{1+\phi}\Big\}. \tag{58}$$

The related potential, equation of motion and mass spectrum formula are, respectively:

$$V(\phi) = \frac{m^D}{g^2}\Big(\frac{3}{8}-\frac{1}{1+\phi}\Big)\phi^2\,, \tag{59}$$

$$\Big[\Big(1-2^{2-\frac{\Box}{2m^2}}\Big)\,\zeta\Big(\frac{\Box}{2\,m^2}-1\Big)+\Big(1-2^{1-\frac{\Box}{2m^2}}\Big)\,\zeta\Big(\frac{\Box}{2\,m^2}\Big)\Big]\phi = \frac{2\phi+\phi^2}{(1+\phi)^2}, \tag{60}$$

$$\Big(1-2^{2-\frac{M^2}{2m^2}}\Big)\,\zeta\Big(\frac{M^2}{2\,m^2}-1\Big)+\Big(1-2^{1-\frac{M^2}{2m^2}}\Big)\,\zeta\Big(\frac{M^2}{2\,m^2}\Big) = 2. \tag{61}$$

4.2. Multiplicative approach

Let us note that Lagrangian for p-adic strings (11) can be rewritten as follows (see [10]):

$$\mathcal{L}_p(\varphi) = \frac{m_p^D}{g_p^2}\,\frac{p^2}{p-1}\Big[-\frac{1}{2}\,\varphi\,p^{-\frac{\Box}{2m_p^2}}\,\varphi+\frac{1}{p+1}\,\varphi^{p+1}\Big] \tag{62}$$

$$= \frac{m^D}{g^2}\,\frac{p^2}{p^2-1}\Big\{\frac{1}{2}\,\varphi\,\Big[\Big(1-p^{-\frac{\Box}{2m^2}+1}\Big)+\Big(1-p^{-\frac{\Box}{2m^2}}\Big)\Big]\,\varphi - \varphi^2\Big(1-\varphi^{p-1}\Big)\Big\}, \tag{63}$$

where we take mass $m_p = m$ and coupling constant $g_p = g$. Now we introduce a new Lagrangian $\mathcal{L}(\phi)$ by the following steps:

$$\mathcal{L}_p(\varphi) \to \prod_p \mathcal{L}_p(\varphi) \to \mathcal{AC}\prod_p \mathcal{L}_p(\varphi) = \mathcal{L}(\phi), \tag{64}$$

where $\mathcal{AC}$ means analytic continuation, and introduced new scalar field is denoted by ϕ. The explicit form of $\mathcal{L}(\phi)$ is

$$\mathcal{L} = \frac{m^D}{g^2}\,\zeta(2)\,\Big\{\frac{1}{2}\,\phi\Big[\frac{1}{\zeta\Big(\frac{\Box}{2m^2}-1\Big)}+\frac{1}{\zeta\Big(\frac{\Box}{2m^2}\Big)}\Big]\,\phi - \phi^2\mathcal{AC}\;G(\phi)\Big\}, \tag{65}$$

where $G(\phi) = \prod_p (1 - \phi^{p-1})$. Infinite product $\prod_p (1 - \phi^{p-1})$ is convergent if $|\phi| < 1$. One can easily see that $G(0) = 1$ and $G(1) = G(-1) = 0$.

The corresponding equation of motion is

$$\Big[1/\zeta\Big(\frac{\Box}{2m^2} - 1\Big) + 1/\zeta\Big(\frac{\Box}{2m^2}\Big)\Big]\,\phi = 2\phi\mathcal{AC}\;G(\phi) + \phi^2\mathcal{AC}\;G'(\phi) \tag{66}$$

and has $\phi = 0$ as a trivial solution. In the weak-field approximation (i.e. $|\phi(x)| \ll 1$), equation of motion (66) becomes

$$\Big[1/\zeta\Big(\frac{\Box}{2m^2} - 1\Big) + 1/\zeta\Big(\frac{\Box}{2m^2}\Big)\Big]\,\phi = 2\phi. \tag{67}$$

The potential $V(\phi) = -\mathcal{L}(\Box = 0)$ is

$$V(\phi) = \frac{m^D}{g^2}\;\zeta(2)\;[7 + \mathcal{AC}\;G(\phi)]\;\phi^2, \tag{68}$$

where $\zeta(-1) = -1/12$ and $\zeta(0) = -1/2$ are taken into account. This potential has local minimum $V(0) = 0$ and values $V(\pm 1) = 7\,m^D$. To explore behavior of $V(\phi)$ for all $\phi \in \mathbb{R}$ one has first to investigate properties of the function $G(\phi)$.

Mass spectrum of M^2 is determined by solutions of equation

$$\zeta^{-1}\big(\frac{M^2}{2m^2} - 1\big) + \zeta^{-1}\big(\frac{M^2}{2m^2}\big) = 2. \tag{69}$$

There are many tachyon solutions.

Remark. The difference between Lagrangians $\mathcal{L}$ (65) (multiplicative approach) and $L_{-\mu}$ (47) (additive approach) is practically in functions $G(\phi)$ and $F(\phi)$. Since

$$G(\phi) = \prod_p \big(1 - \phi^{p-1}\big) = 1 - \phi - \phi^2 + \phi^3 - \phi^4 + ... \tag{70}$$

and

$$F(\phi) = \sum_{n=1}^{\infty} \mu(n)\phi^{n-1} = 1 - \phi - \phi^2 - \phi^4 + \phi^5 - \dots \tag{71}$$

it follows that these functions approximatively have the same behavior for $|\phi| \ll 1$. Hence, in the weak-field approximation these Lagrangians describe the same scalar field.

5. Concluding Remarks

In this review article, seven possible Lagrangians for zeta strings are presented. These Lagrangians contain non-polynomial nonlinearity and space-time nonlocality. This nonlocality is encoded in the Riemann zeta function with d'Alembert operator $\Box$ as its argument. The corresponding potentials, equations of motion and the mass spectrum formulas are also presented. All of these seven models of zeta strings contain the Riemann zeta function and have some intriguing properties that deserve further investigation.

Zeta-string model presented in Section 4.1.5 is a new one and further investigation will be presented elsewhere. In [6], Lagrangian for coupled open and closed zeta string is constructed.

Instead of the Riemann zeta function one can try construction of Lagrangians with some similar functions. In [15], embedding of the Dirichlet L-function was proposed.

A sum of Lagrangian $\mathcal{L}_1$ (16) with any of the above zeta-string Lagrangians should give some information about connection between ordinary and zeta strings.

Note that an interesting approach towards a field theory and cosmology based on the Riemann zeta function and its generalizations is proposed in [17], see also [4]. In [5], the vacuum state of the adelic harmonic oscillator is connected with the Riemann zeta function in the form of its functional relation.

References

[1] M. B. Green, J. H. Schwarz and E. Witten, *Superstring Theory* (Cambridge University Press, 1987).

[2] L. Brekke and P.G.O. Freund, "p-Adic numbers in physics", *Phys. Rep.* **233** (1993) 1.

[3] V.S. Vladimirov, I.V. Volovich and E.I. Zelenov, *p-Adic Analysis and Mathematical Physics* (World Scientific, Singapore, 1994).

[4] B. Dragovich, A.Yu. Khrennikov, S.V. Kozyrev, I.V. Volovich and E.I. Zelenov, "pAdic mathematical physics: the first 30 years", *p-Adic Numbers Ultrametric Anal. Appl.* **9** (2017) 87–121; arXiv:1705.04758 [math-ph].

[5] B. Dragovich, "Adelic harmonic oscillator", *Int.J.Mod.Phys.* **A10** (1995) 2349-2365; arXiv:hep-th/0404160.

[6] B. Dragovich, "Zeta Strings", arXiv:hep-th/0703008 (2007).

[7] B. Dragovich, "Zeta-Nonlocal Scalar Fields", *Theor. Math. Phys.* **157** (2008) 1671–1677; arXiv:0804.4114v1 [hep-th].

[8] B. Dragovich, "Lagrangians with Riemann Zeta Function", *Romanian J. Physics* **53** (2008) 1105-1110; arXiv:0809.1601v1[hep-th].

[9] B. Dragovich, "Some Lagrangians with Zeta Function Nonlocality"; arXiv:0805.0403v1[hep-th], (2008).

[10] B. Dragovich, "Towards Effective Lagrangians for Adelic Strings", *Fortschr. Phys.* **57** (2009) 546–551; arXiv:0902.0295v1 [hep-th].

[11] B. Dragovich, "Nonlocal Dynamics of p-Adic Strings", *Theor. Math. Phys.* **164** (2010) 1151–1155; arXiv:1011.0912 [hep-th].

[12] B. Dragovich, "The p-Adic Sector of Adelic Strings", *Theor. Math. Phys.* **163** (2010) 768–773; arXiv:0911.3625 [hep-th].

[13] I.M. Gelfand, M.I. Graev and I.I. Pyatetskii-Shapiro, *Representation Theory and Automorphic Functions* (Saunders, Philadelphia, 1969).

[14] A.A. Gerasimov and S.L. Shatashvilli, "On exact tachyon potential in open string field theory", *JHEP* (2000) 034; arXiv:hep-th/0009103.

[15] A. Chávez, H. Prado and E.G. Reyes, "The Borel transform and linear nonlocal equations: applications to zeta-nonlocal field models"; arXiv:1907.02617 [math-ph], (2019).

[16] J. Sondow, "Analytic continuation of Riemanns zeta function and values at negative integers via Eulers transformation of series", *Proc. Amer. Math. Soc.* **120**, 421–424 (1994).

[17] I.Ya. Arefeva and I.V. Volovich, "Quantization of the Riemann zeta-function and cosmology", *Int. J. Geom. Meth. Mod. Phys.* **4** (2007) 881; arXiv:hep-th/0701284.

T-duality in Superstring Theory*

Branislav Sazdović†
Institute of Physics Belgrade, University of Belgrade, Belgrade, Serbia

Abstract

In this article we are going to explain why we need string theory. Then we introduce motivation for so called M-theory and explain the role of S-duality and T-duality in better understanding M-theory. In the second part we will discuss T-duality with more details. We will establish standard approach to T-duality, so called Buscher approach. Then we will show that double space approach, which is permutation of some coordinates in the double space, is equivalent to the Buscher's one. This clearly explains why T-duality in double space is nonphysical transformation. We will also briefly discuss relation between T-duality and non-geometric background.

1. Introduction

The string theory is the attractive idea which attempts to solve some long term problems of fundamental physics. It claims to describe all fundamental forces in nature, and in some most optimistic predictions it is a candidate for the theory of everything [1]-[5] .

In this article we will try to explain what are the unresolved problems in modern physics which encouraged us to introduce new theory. Also we will consider what our expectations, that string theory can solve these problems, are based on.

There exists five consistent versions of superstring theory. It was conjectured that they were different limiting cases of a single theory, known as M-theory. It is candidate for describing all fundamental forces and all forms of matter. Today we still do not understand properly meaning and structure of M-theory. Until this is achieved, a prominent scientist E. Witten has suggested that the M should stand for "Magic", "Mystery", or "Membrane".

Up to now, we know formulation of all five consistent superstring theories, but we do not know formulation of M-theory. One of approaches in understanding M-theory is unification of all T-dual theories in double space.

*Work supported in part by the Serbian Ministry of Education, Science and Technological Development, under contract No. 171031.

† e-mail address: sazdovic@ipb.ac.rs

2. Why string theory?

To understand the motivation to introduce string theory, it is useful to remind ourselves to main achievements and main problems of modern physics. It is well known that today there are two fundamental theories. The first one is quantum field theory which describes electromagnetic, weak, and strong interactions of elementary particles on quantum level. The second is general relativity which describes gravity at classic level. Let us stress that both theories have strong experimental confirmation.

The unified theory should include both of the above theories. Some people call it "Quantum gravity". There are several candidates for "Quantum gravity". The most famous are Kaluza-Klein theory, super-symmetry and supergravity, non-commutative space-time, loop quantum gravity and superstring theory. The superstring theory in some way involves all other approaches. In addition to unify gravity with quantum theory, string theory unites all known interactions.

2.1. Problem with unification of gravity with quantum theory

The main problem of quantum field theory is appearance of ultra-violet divergences at small space-time distances. So, if we want to have theory free from ultra-violet divergences we must search for theory with radical change at small distances. The string theory is just that type of theories. It is based on assumption that elementary objects are not point (particles) but they are one-dimensional (strings).

Note one immediate consequence. A one-dimensional object may have two topological forms. Open strings describe electro-weak and strong interactions while closed strings describes gravity.

It is expected that the strings become relevant at Planck scales $L_P = \sqrt{\frac{\hbar G}{c^3}} \approx 1.6 \times 10^{-35} m$. Note that the characteristic length of the string is much smaller than the scales where today's physics takes place.

2.2. Why string theory is a candidate for the theory of everything?

Why some people believe that string theory is good candidate for the theory of everything? Because it has three important features.

1. The closed string contains the theory of general relativity;
2. the open string contains non-abelian gauge theory and
3. in string theory there is no ultra-violet divergences on small distances.

There is important consistency condition in string theory. This is the requirement that quantum theory keep the symmetries of classical one. In particular, this means that it remains invariant under two-dimensional conformal transformations.

Requirement for absence of conformal anomaly leads to conditions which can be interpreted as target space equations of motion. One of them has

formally the same form as Einstein's equation

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \mathcal{O}(\alpha' R_c^{-2}) = 8 G_N T_{\mu\nu} \,, \tag{1}$$

where $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ is the Ricci curvature tensor, R is the scalar curvature, G_N is Newton's gravitational constant and $T_{\mu\nu}$ is energy-momentum tensor. Here we put the speed of light to be equal to unity, $(c = 1)$. The last term on the left hand side is string theory correction. It is proportional to the Regge slope parameter α'.

Unlike Einstein's equation, which is valid at four dimension, the equation for strings is defined in the critical number of dimensions (26 for bosonic and 10 for superstring). So, we supported item 1. because requirement of consistency in string theory produced generalized Einstein's equation.

There are some degrees of freedom that are localized only on the end points of the open string. They are described by vector fields. There exist new consistency conditions (similar to that in the case of closed strings) for the vector fields. The first nontrivial term of the new nonlinear equation is just Yang-Mills equation, which supports item 2.

In order to calculate the Feynman diagrams it is necessary to integrate over all possible trajectories. In the case of the particles, when the points of interaction become infinitely close, coresponding propagator goes to infinity. There is no similar analogy in string theory, because there is no point interactions. This supports item 3.

2.3. Consistent superstring theories

It is known that there are five consistent versions of superstring theories [4]. These are: type I, type IIA, type IIB, and two heterotic string theories. Since string theory pretends to be unique theory of everything the question is why there is not just one consistent theory? The other way we can ask: which of these five theories is real? It turns out that all five theories are real.

There is interesting suggestion for the solution of the problem. In fact all five superstring theories are special limiting cases of an eleven-dimensional theory called M-theory. All these theories are connected in nontrivial ways. In fact different string theories are related by T-duality and S-duality. The T-duality means that strings propagating on completely different spacetime geometries may be physically equivalent. The S-duality means that system of strongly interacting strings can be viewed as a system of weakly interacting strings. In absence of an understanding structure of M-theory, Witten has suggested that the M should stand for Magic, Mystery, or Membrane.

The important question appears: how to find fundamental formulation of M-theory? The possible answer is to construct the theory that contains initial and all dual theories.

3. T-duality

The strings, as extended objects, can see the geometry of space-time different from particles. Duality connects theories that have different formulations but the same physical content.

T-duality is a fundamental characteristic of string theory [6, 4, 5]. As a consequence of T-duality there is no physical difference between string theory compactified on a circle of radius R and circle of radius $1/R$. This conclusion can be generalized to tori of various dimensions. The T-duality is feature that has no analogy in the world of particles.

Compactification on the radius R has two consequences. First, the translation operator for $2\pi R$ does not affect the states if $e^{-i2\pi Rp} = 1$. So, the momentum is quantized $p = \frac{n}{R}\,, (n \in Z)$ and consequently some conditions are lost. Both strings and particles can see this effect.

On the other hand, some new topological states appear. These are the number of windings m, defined as $x^\mu(\sigma + 2\pi) - x^\mu(\sigma) = 2pRm$, $(m \in Z)$. Only an extended object, such as strings, can see this effect.

In the present article we will investigate only T-duality, which is the subject of the research of the group from Institute of Physics in Belgrade, [8]- [15].

3.1. Action

The bosonic string sigma model, describes the bosonic string moving in a curved background. The dynamics is described by the action [4, 5]

$$S[x] = \kappa \int_\Sigma d^2\xi \sqrt{-g}\Big[\frac{1}{2} g^{\alpha\beta} G_{\mu\nu}[x] + \frac{\epsilon^{\alpha\beta}}{\sqrt{-g}} B_{\mu\nu}[x]\Big] \partial_\alpha x^\mu \partial_\beta x^\nu \,. \tag{2}$$

The integration goes over two-dimensional world-sheet Σ parametrized by ξ^α ($\xi^0 = \tau$, $\xi^1 = \sigma$). There are two background fields: a metric tensor $G_{\mu\nu}$, and a Kalb-Ramond field $B_{\mu\nu}$. Here $x^\mu(\xi)$, $\mu = 0, 1, ..., D-1$ are the coordinates of the D-dimensional space-time and $g_{\alpha\beta}$ is the intrinsic world-sheet metric.

Varying the action (2) over x^μ, one obtains the equations of motion

$$\nabla_\pm \partial_\mp x^\mu + \Gamma^\mu_{\mp\nu\rho} \partial_\pm x^\nu \partial_\mp x^\rho = 0 \,. \tag{3}$$

Here we introduced the covariant derivative

$$\nabla_\pm = \partial_\pm + n\omega_\pm \,, \tag{4}$$

and the generalized connection

$$\Gamma^\mu_{\pm\nu\rho} = \Gamma^\mu_{\nu\rho} \pm B^\mu{}_{\nu\rho}, \tag{5}$$

in terms of the Christoffel symbol $\Gamma^\mu_{\nu\rho} = \frac{1}{2}(G^{-1})^{\mu\sigma}(\partial_\nu G_{\rho\sigma} + \partial_\rho G_{\sigma\nu} - \partial_\sigma G_{\nu\rho})$ and the field strength of Kalb-Ramond field $B_{\mu\nu}$, $B^\mu{}_{\nu\rho} = (G^{-1})^{\mu\sigma} B_{\sigma\nu\rho} = (G^{-1})^{\mu\sigma}(\partial_\sigma B_{\nu\rho} + \partial_\nu B_{\rho\sigma} + \partial_\rho B_{\sigma\nu})$.

The principle of stationary action $\delta S = 0$, beside equations of motion gives the boundary conditions

$$\gamma^{(0)}_{\mu}(x)\delta x^{\mu}/_{\sigma=\pi} \quad - \quad \gamma^{(0)}_{\mu}(x)\delta x^{\mu}/_{\sigma=0} = 0\,, \tag{6}$$

where we introduced useful expression as

$$\gamma^{(0)}_{\mu}(x) \equiv \frac{\delta S}{\delta x'^{\mu}} = \kappa\Big(2B_{\mu\nu}\dot{x}^{\nu} - G_{\mu\nu}x'^{\nu}\Big)\,. \tag{7}$$

3.2. Buscher T-duality procedure for constant background

T-duality [5], interchanges the string momenta and winding numbers, leaving the spectrum unchanged. On the level of string sigma model it was given by Buscher procedure [6]. It can be applied along isometry directions, such that background fields do not depend on coordinates x^{μ}.

The Buscher procedure contains a few steps. First we should gauge global symmetries $\delta x^{\mu} = \lambda^{\mu}$, substituting derivatives $\partial_{\alpha}x^{\mu}$ with covariant ones $\partial_{\alpha}x^{\mu} \to D_{\alpha}x^{\mu} = \partial_{\alpha}x^{\mu} + v^{\mu}_{\alpha}$, where v^{μ}_{α} are gauge field. In such away we can obtain gauge invariant Lagrangian.

Then we introduce field strength for gauge fields, $F^{\mu}_{\alpha\beta} = \partial_{\alpha}v^{\mu}_{\beta} - \partial_{\beta}v^{\mu}_{\alpha}$. In order to obtain physically equivalent theory (the T-dual theory) we will require that gauge fields do not have any degrees of freedom. We can achieve it with demand that field strength is zero, $F^{\mu}_{01} \equiv F^{\mu} = 0$. Therefore, we can introduce invariant action

$$S_{inv}(x,y,v) = \kappa\int_{\Sigma} d^2\xi\left[(\frac{\eta^{\alpha\beta}}{2}G_{\mu\nu} + \varepsilon^{\alpha\beta}B_{\mu\nu})D_{\alpha}x^{\mu}D_{\beta}x^{\nu} + \frac{1}{2}y_{\mu}F^{\mu}\right]\,, \tag{8}$$

where y_{μ} is Lagrange multiplier.

Next, we can use gauge fixing $x^{\mu} = 0$, which produces gauge fixed action

$$S_{fix}(y,v) = \kappa\int_{\Sigma} d^2\xi\left[v^{\mu}_{\alpha}\left(\frac{\eta^{\alpha\beta}}{2}G_{\mu\nu} + \varepsilon^{\alpha\beta}B_{\mu\nu}\right)v^{\nu}_{\beta} + \frac{1}{2}y_{\mu}F^{\mu}\right]\,. \tag{9}$$

Now, we can check that the last action is equivalent to the initial one. Equation of motion with respect to Lagrange multiplier y_{μ} produces $\partial_{\alpha}v^{\mu}_{\beta} - \partial_{\beta}v^{\mu}_{\alpha} = 0$, with the solution $v^{\mu}_{\alpha} = \partial_{\alpha}x^{\mu}$. Substituting it in the previous relation we can conclude that gauge fixed action turns to the initial one, $S_{fix} \to S(x)$.

Elimination of gauge fields v^{μ}_{α} on their equations of motion produces T-dual action

$${}^{\star}S(y) = \kappa\int d^2\xi\ \partial_{+}y_{\mu}\,{}^{\star}\Pi^{\mu\nu}_{+}\,\partial_{-}y_{\nu} = \frac{\kappa^2}{2}\int d^2\xi\ \partial_{+}y_{\mu}\theta^{\mu\nu}_{-}\partial_{-}y_{\nu}\,, \tag{10}$$

where

$$\Pi_\pm \equiv B_{\mu\nu} \pm \frac{1}{2} G_{\mu\nu}\,, \qquad \theta^{\mu\nu}_\pm \equiv \theta^{\mu\nu} \mp \frac{1}{\kappa}(G_E^{-1})^{\mu\nu}\,, \tag{11}$$

and

$$G^E_{\mu\nu} \equiv G_{\mu\nu} - 4(BG^{-1}B)_{\mu\nu}, \qquad \theta^{\mu\nu} \equiv -\frac{2}{\kappa}(G_E^{-1}BG^{-1})^{\mu\nu}\,. \tag{12}$$

The dual Action ${}^\star S(y)$ has the same form as initial one

$${}^\star S(y) = \kappa \int_\Sigma d^2\xi\, \partial_\alpha y_\mu \left(\frac{\eta^{\alpha\beta}}{2} {}^\star G^{\mu\nu} + \varepsilon^{\alpha\beta} {}^\star B^{\mu\nu} \right) \partial_\beta y_\nu\,, \tag{13}$$

but with T-dual background fields

$${}^\star G^{\mu\nu} = (G_E^{-1})^{\mu\nu}, \quad {}^\star B^{\mu\nu} = \frac{\kappa}{2}\theta^{\mu\nu}\,. \tag{14}$$

3.3. T-duality transformation of variables for constant background

If we equate solutions for gauge field v^μ_α (in two cases when we used equations of motion with respect to the Lagrange multiplier y_μ and to the gauge fields v^μ_α) we obtain T-dual transformations of the coordinates

$$v^\mu_\pm \cong \partial_\pm x^\mu \cong -\kappa \Theta^{\mu\nu}_\pm \partial_\pm y_\nu\,. \tag{15}$$

The inverse T-duality transformation takes a form

$$\partial_\pm y_\mu \cong -2\Pi_{\mp\mu\nu}\partial_\pm x^\nu\,. \tag{16}$$

In the canonical approach we can rewrite above expressions as

$$\kappa\, x'^\mu \cong {}^\star\pi^\mu\,, \quad \pi_\mu \cong \kappa y'_\mu\,, \quad -\kappa\,\dot{x}^\mu \cong {}^\star\gamma^\mu_{(0)}(y)\,, \quad \gamma^{(0)}_\mu(x) \cong -\kappa\,\dot{y}_\mu\,. \tag{17}$$

4. Unification of all T-dual theories in doubled space

For better understanding M-theory one should unify all T-dual theories. Up to now there was partial success: unification of all theories with fixed number d of T-dualization ($d = 1, 2, \cdots D$). This approach has attracted a lot of attention, but the problem is that for each d there is a different formulations of unified theory. We still need only one theory, which contains all T-dual theories.

Following [7], [8] and [9] we will start with T-dual transformation laws along all coordinates

$$\partial_\pm x^\mu = -\kappa \Theta^{\mu\nu}_\pm \partial_\pm y_\nu\,, \qquad \partial_\pm y_\mu = -2\Pi_{\mp\mu\nu}\partial_\pm x^\nu\,. \tag{18}$$

Separate parts with ${\varepsilon_\pm}^\pm = \pm 1$ and ${\eta_\pm}^\pm = 1$ we obtain

$$\pm\partial_\pm y = G_E\partial_\pm x - 2(BG^{-1})\partial_\pm y\pm\,, \quad \partial_\pm x = 2(G^{-1}B)\partial_\pm x + G^{-1}\partial_\pm y\,. \tag{19}$$

In doubled space with coordinates

$$Z^M = \begin{pmatrix} x^\mu \\ y_\mu \end{pmatrix}, \tag{20}$$

we can rewrite equations (19) as

$$\partial_\pm Z^M \cong \pm\,\Omega^{MN}\mathcal{H}_{NK}\,\partial_\pm Z^K\,, \tag{21}$$

where $\mathcal{H}_{MN}$ is generalized metric

$$\mathcal{H}_{MN} = \begin{pmatrix} G^E_{\mu\nu} & -2B_{\mu\rho}(G^{-1})^{\rho\nu} \\ 2(G^{-1})^{\mu\rho}B_{\rho\nu} & (G^{-1})^{\mu\nu} \end{pmatrix}, \tag{22}$$

and

$$\Omega^{MN} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{23}$$

4.1. T-duality along all coordinates can be realized by replacement x^μ with y_μ

In the double space we are able to represent the backgrounds of all T-dual theories and T-duality transformations of coordinates in unified manner. This can be realized by exchanging places of some coordinates x^a, along which we want to perform T-duality and the corresponding dual coordinates y_a. In such approach it is evident that T-duality relates physically equivalent theories.

Let us first consider the case when we perform T-duality along all coordinates x^μ. Then we have

$${}^\star Z^M = \begin{pmatrix} y_\mu \\ x^\mu \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x^\mu \\ y_\mu \end{pmatrix} = \mathcal{T}Z\,. \tag{24}$$

Requiring that T-dual transformation for double coordinates ${}^\star Z^M$ has the same form as initial one

$$\partial_\pm {}^\star Z^M \cong \pm\,\Omega^{MN}{}^\star\mathcal{H}_{NK}\,\partial_\pm {}^\star Z^K\,, \tag{25}$$

we find expression for T-dual generalized metric, ${}^\star\mathcal{H} = \mathcal{T}\mathcal{H}\mathcal{T}$.

4.2. T-dualization along arbitrary number of coordinates

Let us now consider the case when we perform T-duality along some subset of coordinates x^a. Then we can realize replacement x^a with y_a $(a = 1, 2, \cdots d)$ as

$$Z_a^M = \mathcal{T}^{aM}{}_N Z^N \quad \begin{pmatrix} y_a \\ x^i \\ x^a \\ y_i \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1_a & 0 \\ 0 & 1_i & 0 & 0 \\ 1_a & 0 & 0 & 0 \\ 0 & 0 & 0 & 1_i \end{pmatrix} \begin{pmatrix} x^a \\ x^i \\ y_a \\ y_i \end{pmatrix} . \qquad (26)$$

It produces T-dual background fields in complete agreement with Buscher approach.

This interpretation of T-duality (as permutation of the coordinates in double space) works for: bosonic string with flat background [8], bosonic string with weakly curved background [10, 9], Type IIA and Type IIB superstrings [11], open bosonic string with Neumann A_a^N and Dirichlet A_i^D gauge fields [12, 13] and for fermionic T-duality [14, 15].

4.3. Example: Three torus

As an example, let us consider three torus with nontrivial components of the background fields

$$G_{\mu\nu} = \delta_{\mu\nu} , \qquad (\mu, \nu = 1, 2, 3) \qquad B_{12} = -\frac{1}{2} H x^3 . \qquad (27)$$

All T-duality transformations between these theories are shown in the diagram

$$\begin{array}{ccccc} & & {}_1S(y_1, x^2, x^3) & \xrightarrow{\mathrm{T}^2} & {}_{12}S(y_1, y_2, x^3) \\ & \overset{\mathrm{T}^1}{\nearrow} & & \nearrow^{\mathrm{T}^1}\searrow^{\mathrm{T}^3} & & \overset{\mathrm{T}^3}{\searrow} & \\ S(x^1, x^2, x^3) & \xrightarrow{\mathrm{T}^2} & {}_2S(x^1, y_2, x^3) & & {}_{13}S(y_1, x^2, y_3, V^3) & \xrightarrow{\mathrm{T}^2} & {}_{123}S(y_1, y_2, y_3, V^3) \\ & \underset{\mathrm{T}^3}{\searrow} & & \nearrow^{\mathrm{T}^1}\searrow^{\mathrm{T}^3} & & \underset{\mathrm{T}^1}{\nearrow} & \\ & & {}_3S(x^1, x^2, y_3, V^3) & \xrightarrow{\mathrm{T}^2} & {}_{23}S(x^1, y_2, y_3, V^3) & . \end{array}$$

In the literature the theories $S, {}_1S, {}_{12}S$ and ${}_{123}S$ are known as theories with H, f, Q and R fluxes respectively.

All theories when we performed T-dualization along x^3 (note that Kalb-Ramond fiels $B_{\mu\nu}$ depend on x^3) are non-geometric, with non trivial R fluxes. These are the theories ${}_3S, {}_{13}S, {}_{23}S$ and ${}_{123}S$ and they depend on non local expression V^3.

5. T-duality and non-geometric background

The Buscher's original procedure works for constant background and i all cases when we perform T-dualization along some coordinates (let us say x^a),

such that background fields do not depend on x^a. In both cases Abelian T-duality has been applied along the coordinates with global shift symmetry.

The most interesting case is when we try to perform T-dualization along non-isometry directions, along some set of coordinates x^a such that background fields do depend on them, [10]. Then we should apply the generalized Buscher's procedure, developed in Refs. [10, 16]. In that case, the expression for T-dual background fields (14) formally remain the same but the argument of T-dual background fields is not simply the T-dual variable y_a. It becomes the line integral of world-sheet gauge fields v_+^a and v_-^a, or exactly $V^a[v_+, v_-] \equiv \int_P d\xi^\alpha v_\alpha^a = \int_P (d\xi^+ v_+^a + d\xi^- v_-^a)$. On the equations of motion the line integral does not depend on the path P. On the solution for gauge fields it turns to the expression $V^a = -\kappa\,\theta^{ab} y_b + G_E^{-1ab}\,\tilde{y}_b$, where $\tilde{y}_a$ is double of T-dual variable y_a which satisfy $\dot{\tilde{y}}_a = y'_a$ and $\tilde{y}'_a = \dot{y}_a$. Let us stress that in such cases T-dual theories becomes locally non-geometric because the argument of background fields is the line integral.

Let us briefly discuss the Buscher procedure for the Lagrangian in canonical form $S = \int d^2\xi[\pi_\mu \dot{x}^\mu - \mathcal{H}(x, x', \pi)]$. Here, $\mathcal{H}$ is Hamiltonian density and x-dependence comes from the arguments of background fields. In generalized Buscher's procedure the corresponding auxiliary action takes the form $S_{aux} = \int d^2\xi[\pi_\mu v_0^\mu - \mathcal{H}(V, v_1, \pi) - \kappa(v_0^\mu y'_\mu - v_1^\mu \dot{y}_\mu)]$. Formally, we substitute $\dot{x}^\mu \to v_0^\mu$, $x^\mu \to V^\mu$ and add the last term with Lagrange multiplier y_μ. It easy to check that on the equation of motion with respect to Lagrange multiplier y_μ, the auxiliary action S_{aux} turns to the initial action S.

If we work only with isometry directions then the background fields and consequently the Hamiltonian density $\mathcal{H}$ do not depend on x^μ. It means that after substitution $x^\mu \to V^\mu$ the Hamiltonian density does not depend on V^μ. Varying S_{aux} with respect to v_0^μ we obtain relation $\pi_\mu \cong \kappa y'_\mu$. This is second relation in our equation (17). When we perform T-dualization along non-isometry directions then the argument of T-dual background fields V^μ is a non-local expression because it is a line integral of both v_0^μ and v_1^μ. Variation with respect to v_0^μ produces a new term and we obtain

$$\kappa y'_\mu = \pi_\mu - \frac{\kappa}{3} B_{\mu\nu\rho} x'^\nu x^\rho \,, \tag{28}$$

where $B_{\mu\nu\rho}$ is field strength for Kalb-Ramond field $B_{\mu\nu}$.

5.1. Nonassociativity of R-flux background

The relation (28) is origin of closed string non-commutativity, see Ref.[17]. In fact coordinates of T-dual space y_μ depend on both initial momenta π_μ and initial coordinates x^μ. Since the initial space is geometric one, the standard Poisson algebra is satisfied

$$\{x^\mu(\sigma), \pi_\nu(\bar{\sigma})\} = \delta^\mu{}_\nu \delta(\sigma - \bar{\sigma})\,. \tag{29}$$

It means that coordinates of T-dual space, y_μ and y_ν have non-trivial commutation relations

$$\{y_\mu(\sigma), y_\nu(\bar{\sigma})\} \;\cong\; -\frac{1}{\kappa} B_{\mu\nu\rho}\big[x^\rho(\sigma) - x^\rho(\bar{\sigma})\big]\theta(\sigma-\bar{\sigma})\,, \tag{30}$$

where the function $\theta(\sigma)$ is defined as

$$\theta(\sigma) \equiv \begin{cases} 0 & \text{if } \sigma = 0 \\ 1/2 & \text{if } 0 < \sigma < 2\pi, \quad \sigma \in [0, 2\pi]. \\ 1 & \text{if } \sigma = 2\pi \end{cases} \tag{31}$$

For $\sigma = \bar{\sigma}$ we obtain that all Poisson brackets vanish, and consequently, coordinates commute. But, taking $\sigma = \bar{\sigma} + 2\pi$ we obtain the closed string non-commutativity relation

$$\{y_\mu(\sigma + 2\pi), y_\nu(\sigma)\} \cong -\frac{2\pi}{\kappa} B_{\mu\nu\rho} N^\rho\,, \tag{32}$$

where $N^\mu = \frac{1}{2\pi}\left[x^\mu(\sigma + 2\pi) - x^\mu(\sigma)\right]$ is winding number of the original coordinates.

We expect full T-dualization along all coordinates corresponds to the so called R-flux. To check it out we are going to calculate characteristic features of R-flux: non-associativity of the coordinates and breaking of Jacobi identity.

Because expression (28) is bilinear in initial coordinates x^μ, the commutator of two T-dual coordinates (30) is linear in x^μ. It is origin of non-associativity of T-dual coordinates, [17]. Reducing our procedure to three dimensions from (28) and (30) we obtain

$$\begin{aligned} &\{\{y_\mu(\sigma_1), y_\nu(\sigma_2)\}, y_\rho(\sigma_3)\} - \{y_\mu(\sigma_1), \{y_\nu(\sigma_2), y_\rho(\sigma_3)\}\} \\ &= \frac{1}{2\kappa^2} B_{\mu\nu\rho}\left[2\theta(\sigma_3-\sigma_2)\theta(\sigma_2-\sigma_1) + \theta(\sigma_1-\sigma_3)\theta(\sigma_3-\sigma_2)\right. \\ &\left.+\theta(\sigma_3-\sigma_1)\theta(\sigma_1-\sigma_2)\right]\,. \end{aligned}$$

The last relation confirms breaking of Jacobi identity

$$\begin{aligned} &\{y_\mu(\sigma_1), y_\nu(\sigma_2), y_\rho(\sigma_3)\} \equiv \{\{y_\mu(\sigma_1), y_\nu(\sigma_2)\}, y_\rho(\sigma_3)\} \\ &+ \{\{y_\nu(\sigma_2), y_\rho(\sigma_3)\}, y_\mu(\sigma_1)\} + \{\{y_\rho(\sigma_3), y_\mu(\sigma_1)\}, y_\nu(\sigma_2)\} \\ &= \frac{1}{\kappa^2} B_{\mu\nu\rho}\left[\theta(\sigma_1-\sigma_2)\theta(\sigma_2-\sigma_3) + \theta(\sigma_3-\sigma_1)\theta(\sigma_1-\sigma_2)\right. \\ &\left.+\;\theta(\sigma_2-\sigma_3)\theta(\sigma_3-\sigma_1)\right]\,. \end{aligned}$$

Consequently, the background of the T-dual theory depends on nonlocal variable V^μ which incorporates main features of the non-geometric spaces. It is clear that T-dualization along non-isometry directions corresponds to non-geometric T-dual theories.

6. Conclusion

At the beginning of this article we discussed the problems of contemporary physics and offer some possible solutions. We emphasized the role of string theory and M-theory. Later we discussed T-duality, as one of the main link in understanding M-theory.

We constructed the theory in double space which contains all T-dual theories. We expect that it can be a good candidate on a path to better understanding M-theory. In the standard formulation, T-duality transforms the initial theory to the equivalent one, T-dual theory. The double space formulation contains both initial and T-dual theories and T-duality becomes the global symmetry transformation.

Let us stress that the $2D$ dimensional space, beside initial D dimensional space-time coordinates x^μ contains the corresponding T-dual coordinates y_μ. This extended space with the coordinates $Z^M = (x^\mu, y_\mu)$ we call double space. Its introduction makes it possible to offer simple formulation for T-duality. It can be realized by exchanges the places of some subset of the coordinates x^a and the corresponding dual coordinates y_a along which we perform T-dualization. This permutation produces exactly the same T-dual background fields and T-duality transformations as in the standard approach of Refs.[8, 9]. So, double space approach is in fact permutation of some coordinates in the double space and consequently explains why T-duality is nonphysical.

We also discus relation between T-duality and non-geometric background. The non-geometric background is usually connected with R-flux, which main characteristics are non-associativity and breaking of Jacobi identity. In our approach, the initial theory is geometric while all the other (T-dual) theories are non-geometric and additionally non-local. The T-dual coordinate transformation laws connecting these theories.

For further understanding M-theory it is important to include S-duality.

References

[1] Michael B. Green, John H. Schwarz and Edward Witten, *Superstring Theory* (Cambridge University Press, Cambridge, 1987.)

[2] J. Polchinski, *String theory*, (Cambridge University Press, Cambridge, 1998.)

[3] C.V. Johnson, *D-branes*, (Cambridge University Press, Cambridge, 2003).

[4] K. Backer, M. Backer and J. Schwarz, *String Theory and M-theory*, (Cambridge University Press, Cambridge, 2007).

[5] B. Zwiebach, *A First Course in String Theory* (Cambridge University Press, Cambridge, 2009).

[6] T. Buscher, *Phys. Lett.* **B 194** (1987) 51; **201** (1988) 466.

[7] M. Duff, *Nucl. Phys.* **B 335** (1990) 610.

[8] B. Sazdović, *Chinese Physics* **C 41** (2017) 053101.
[9] B. Sazdović, *JHEP* **08** (2015) 055.
[10] Lj. Davidović and B. Sazdović, *EPJ* **C 74** (2014) 2683.
[11] B. Nikolić and B. Sazdović, *EPJ* **C77** (2017) 197.
[12] B. Sazdović, *EPJC* **77** (2017) 634.
[13] B. Sazdović, *Chinese Physics* **C 42** (2018) 083106.
[14] B. Nikolić and B. Sazdović, *JHEP* **06** (2012) 101.
[15] B. Nikolić and B. Sazdović, Fermionic T-duality in fermionic double space N̂ucl. Phys. **B 917** (2017) 105.
[16] Lj. Davidović and B. Sazdović, *JHEP* **11** (2015) 119
[17] Lj. Davidović, B. Nikolić and B. Sazdović, *EPJ* **C 74** (2014) 2734.

On the nature of optical rogue waves*

Milivoj Belić†
Texas A&M University at Qatar, 23874 Doha, Qatar

Stanko Nikolić‡
Institute of Physics Belgrade, University of Belgrade,
Pregrevica 118, 11080 Zemun, Serbia

Najdan Aleksić§
Institute of Physics Belgrade, University of Belgrade,
Pregrevica 118, 11080 Zemun, Serbia

Omar Ashour¶
Department of Physics, University of California,
Berkeley CA 94720, USA

Siu A. Chin‖
Department of Physics and Astronomy, Texas A&M University,
College Station TX 77843, USA

Abstract

Rogue waves are giant nonlinear waves that suddenly appear and disappear in oceans and optics. We discuss the facts and fictions related to their strange nature, dynamic generation, ingrained instability, and potential applications. We present rogue wave solutions to the standard cubic nonlinear Schrödinger equation that models many propagation phenomena in nonlinear optics. We propose the method of mode pruning for suppressing the modulation instability of rogue waves. We demonstrate how to produce stable Talbot carpets – recurrent images of light and plasma waves by rogue waves, for possible use in nanolithography. Finally, we display how statistical analysis based on inadequate numerics can lead to misleading conclusions on the nature of rogue waves.

* This work has been supported by Qatar National Research Fund
† e-mail address: milivoj.belic@qatar.tamu.edu
‡ e-mail address: stankon@ipb.ac.rs
§ e-mail address: najdan@gmail.com
¶ e-mail address: ashour@berkeley.edu
‖ e-mail address: chin@physics.tamu.edu

1. Introduction

Nonlinear Schrödinger equations (NLSEs) of different orders continue to elicit intense attention of numerous research groups around the world, for their utility in various branches of mathematics and physics [1, 2, 3, 4, 5, 6, 7, 8]. Here, we are focused on the simplest cubic one-dimensional NLSE arising in many fields of nonlinear and fiber optics. In particular, we are interested in the unstable solutions of the model when the modulation instability sets in.

Modulation instability (MI) is the basic nonlinear optical process in which a weak periodic perturbation of the fundamental pump wave produces an exponential growth of a finite number of spectral sidebands locked to and growing at the expense of the pump [4, 5, 9]. Although commonly known as the Benjamin-Fair instability of Stokes waves that appeared in the nineteen-sixties, the MI of the cubic NLSE debuted already in 1947, in the Bogoliubov's work on the uniform Bose gas [10]. It is widely believed that MI is the root cause of the appearance of rogue waves (RWs) in nonlinear optics. The problem is, how to systematically incorporate the process of MI into the dynamics of generation and observation of RWs in the NLSEs of different types.

A convenient handle in this process is provided by the existence of a family of exact solutions to the basic cubic NLSE in the form of Akhmediev breathers (ABs), Kuznetsov-Ma (KM) solitons, and the Peregrine soliton, which may be regarded as the elementary solutions on a finite background from which higher-order RW solutions can be formed. In this sense, especially relevant seem to be ABs [11, 12], which can be generalized to the doubly periodic solutions (as well as to the extended NLSE models). They allow for an easy systematic buildup of higher order breathers that can be regarded as prototype RWs [13, 14].

The complication is that these basic and higher-order solutions represent homoclinic orbits of unstable Stokes waves in the dynamics of cubic NLSE [15, 16, 17, 18]. The generic long-time dynamics of modulated Stokes waves, for example ABs with two or more unstable modes, is chaotic. Once the system, for a range of relevant parameters and initial and boundary conditions, enters homoclinic chaos, the predictive power of the model diminishes. The question has even been raised whether the chaos seen belongs to the model itself or is induced by the numerical procedure applied [15, 16].

For these reasons a school of thought has emerged which holds that it is not important to follow exact dynamics of individual members of the family of exact solutions, but to look at the statistics of RWs in the chaotic regime [4, 6, 13, 19, 20]. Optical RWs are rare extreme events in the fluctuation of optical fields, therefore their statistical features, such as long-tailed probability distributions, should be considered as their defining features. Thus, one should proceed with the numerical solution of different NLSEs with appropriate initial conditions seeded with noise of various types, and after many runs compare the associated statistics of the resulting field distribu-

tions with the available experimental data.

In this paper, we offer a different point of view. It is our belief that for specific applications one still must perform carefully designed numerical simulations of individual well-defined RW solutions, even when it leads to following their dynamics deep in the chaotic region. The general idea is to discern order from chaos. The specific goal is to investigate the possibility of producing Talbot carpets out of ABs of different orders, with an eye on possible applications in nanolithography. Such an investigation requires launching an exact breather and following its repeated self-imaging recurrences for as long as possible. This goal is in direct collision with the basic tenets of homoclinic chaos; homoclinic solutions are statistically unlikely to be physically observable. A way must be found to mitigate the unavoidable impact of MI and the attending chaos. To this end, we introduce the procedure of mode pruning.

Thus, we examine how AB and RW solutions are obtained for the NLSE and used to accomplish the goal stated. In this endeavor of immense importance are the analytical solutions to the NLSE that are periodic both along the spatial and temporal axes, and can be viewed as Talbot self-images, introduced in [12]. This study was extended in [21, 22], where the nonlinear Talbot carpets of rogue waves were reported for the first time. These solutions are associated with the Talbot effect, first described in the 19th century [23], about the same time the solitary waves were discovered. An interesting feature of the nonlinear Talbot effect is that it only displays the primary and secondary images.

The Talbot effect is a near-field diffraction effect, observed when light beams diffract at some periodic structure (such as gratings) and produce recurrent self-images at equidistant planes. In-between the planes, fractional and even fractal images can be observed, leading to intricate light patterns that are called the Talbot carpets. Later, the self-imaging phenomena have been reported in many areas of physics, such as atomic [24, 25] and quantum [26] optics, waveguide arrays [27], Bose-Einstein condensates [28, 29], photonic lattices [30], and x-ray imaging [31]. Talbot self-images can even be regarded as an example of Fermi-Pasta-Ulam recurrence [9]. Nonlinear Talbot effect from nonlinear photonic crystals was experimentally demonstrated in [32]. An overview of the recent advances of Talbot effect in modern science is presented in [33].

The underlying themes of this paper can be described as follows. We present the dynamical generation of breathers and rogue waves in Talbot carpet-like arrangements, for the cubic NLSE. We calculate the first and higher-order breathers using Darboux transformation (DT) and extract initial conditions in a wide box that is a multiple of the main breather's period [34]. We invent two pruning procedures for Fourier modes, to suppress modulation instability that ruins the double-periodic pattern of high intensity peaks.

But, the central theme of this paper is concerned with the conflicting opinions formed about the nature of RWs: Are they linear or nonlinear;

Random or deterministic; Numerical or physical? A short and in our opinion correct answer to these nagging questions is as follows.

Rogue waves are essentially nonlinear, because their cause is the modulation or Benjamin-Fair instability. They are deterministic, because modulation instability leads to homoclinic chaos, which by its nature is deterministic. They are physical, because they are observed in many experiments and media. Now, there exist reservations to these facts, depending on how one generates and analyzes RWs. This especially holds if the work is statistical and numerical in nature. We'll come to these reservations later in the paper.

2. Rogue wave solutions to the NLSE

To recap, we discuss the nature of optical rogue waves in the cubic NLSE, in view of conflicting opinions expressed in the literature. In particular, as already mentioned, we address three pairs of opposing suppositions on their nature: Linear vs. nonlinear [4, 6]; random vs. deterministic [19, 54]; and numerical vs. physical [15, 16]. A short answer to these suppositions is that rogue waves in optics are essentially nonlinear, deterministic, and physical. They are nonlinear because the major cause of rogue waves is the modulation or Benjamin-Feir instability, which by its nature is the basic nonlinear optical process. Rogue waves are deterministic because modulation instability leads to deterministic chaos; random phenomena are probabilistic and may look chaotic but are not deterministic. Rogue waves are physical because they appear in many experiments and media, with similar statistics. Our opinion is supported by extensive numerical simulations of the weakly nonlinear Schrdinger equation in different regimes that touch upon the aspects of all three conflicting suppositions.

Disturbingly, in numerical simulations optical rogue waves may appear fictitiously, as numerical artefacts. Different numerical algorithms for exactly the same inputs may provide different evolution pictures and distressingly different statistics. An example of different statistics of the probability density distributions, obtained by two different numerical algorithms applied to the same input, is provided in Fig. 6 below. There, the standard beam propagation method, predominantly used in the literature, predicts the appearance of thousands of rogue waves, whereas the higher-order more precise symplectic algorithm predicts significantly fewer. Hence, owing to a vague definition of rogue waves and exponential amplification of numerical errors, there are situations in which optical rogue waves may appear as linear, random, and numerical.

2.1. The influence and suppression of modulation instability

Before proceeding to these more subtle points, we start by discussing the more standard behavior of RWs in the standard NLSE. We study double-

periodic solutions of the cubic nonlinear Schrödinger equation

$$i\psi_x + \frac{1}{2}\psi_{tt} + |\psi|^2\psi = 0, \tag{1}$$

where the transverse variable is denoted by t and the longitudinal variable by x. The wave function $\psi \equiv \psi(t, x)$ represents the slowly-varying envelope that could be optical, plasmonic or other in nature.

Various solutions of the basic NLSE and its extensions, such as solitons [48, 49], breathers [49, 50], and RWs [51], have been discussed in the literature. It is well known that breathers and solitons of arbitrary order can be obtained analytically using the DT technique. These solutions are single-periodic: breathers along t, and solitons along x direction. Single-periodic NLSE solutions, such as ABs, can be utilized to dynamically construct nonlinear Talbot carpets, which are also intimately connected with the double-periodic solutions of the NLSE.

Let us recall the first-order AB solution,

$$\psi(t,x) = \left[1+\frac{2(1-2a)\cosh(\lambda(t-t_0))+i\lambda\sinh(\lambda(t-t_0))}{\sqrt{2a}\cos(\omega x)-\cosh(\lambda(t-t_0))}\right]\exp i(t-t_0), \tag{2}$$

in which both the growth factor λ and the spatial frequency ω are given in terms of the single parameter $0 < a < 0.5$. When $a = 0.5$ AB turns into the Peregrine RW and when $a > 0.5$ it becomes the KM soliton. The period of an AB (first- or higher-order) is also determined by a [9]:

$$L = \frac{\pi}{\sqrt{1-2a}}. \tag{3}$$

The initial condition for dynamical generation is derived from an exact AB solution at a certain value of the evolution variable $x = x_0$, using Darboux transformation [52]. Here, it is essential to adjust the size of the transverse box (t_0, t_1) to an integer multiple M of the fundamental breather's period and apply periodic boundary conditions,

$$\Delta T = t_1 - t_0 = ML. \tag{4}$$

Numerical solutions of NLSEs in this paper are obtained using the second-order split-step fast Fourier transform (FFT) method, except when the statistics of RWs is discussed. When the box size is exactly equal to the breather's fundamental period L, the Fourier harmonics form the basic set of spatial frequencies

$$S_1 = \{\omega_j = j\Omega; |\ 0 \le j < N\}, \tag{5}$$

where $\Omega = 2\pi/L$ is the mode spacing and N the total number of modes. The mode growing out of this basic set will be the stable fundamental breather

mode. However, if the box is larger ($M > 1$), the mode spacing gets smaller $\Omega_M = \Omega/M$, and Fourier modes form a new set S_M with a larger number of modes (NM). All modes from S_M that are not elements of S_1 exponentially grow from infinitesimal amplitudes, owing to the modulation instability. Now, the modes from S_1 are also under MI and grow exponentially, but they interfere constructively and only form the fundamental AB mode.

The key point in generating nonlinear Talbot carpets is to suppress the undesirable unstable Fourier modes. This can be achieved in different ways. In the simplest, after each numerical iteration one simply eliminates the unstable subharmonics, leaving only the ones responsible for the formation of the fundamental AB (labeled as 0, $\pm M$, $\pm 2M$, $\pm 3M$, etc). It effectively eliminates MI. The procedure is illustrated in Fig. 1. Although quite drastic, it apparently works. The other ways include suppressing the unstable modes selectively and to a degree.

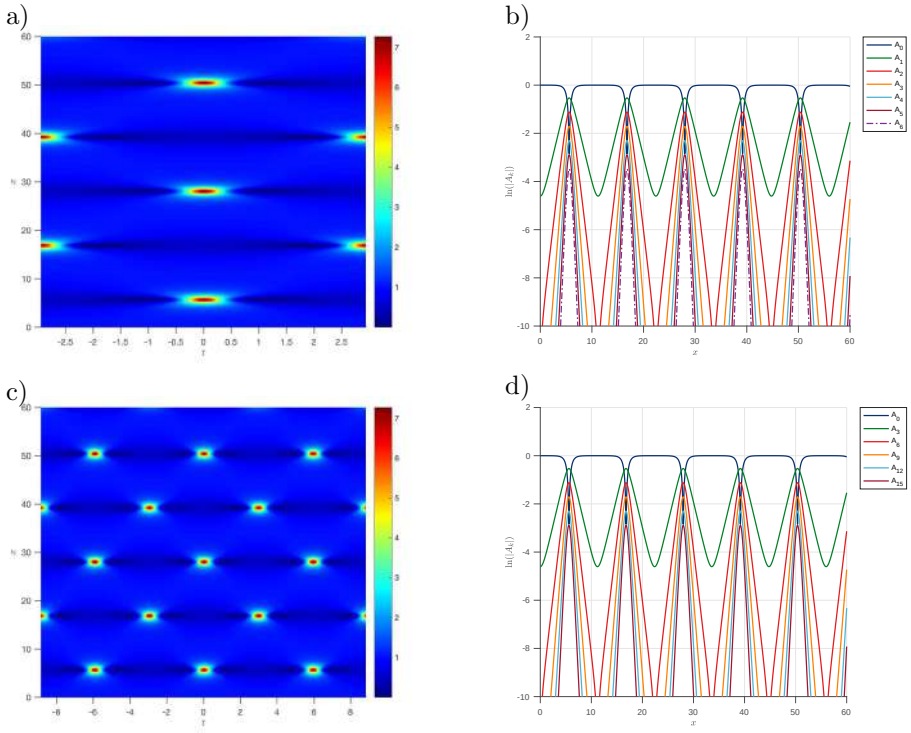


Figure 1: Double-periodic numerical solutions, made of the first-order NLSE breathers, using the pruning procedure in FFT. The breather parameter is $a = 0.36$. (a) One breather in the box, no pruning. (b) Its spectrum. (c) Three breathers (3 periods) in the box, with the pruning. (d) The corresponding spectrum.

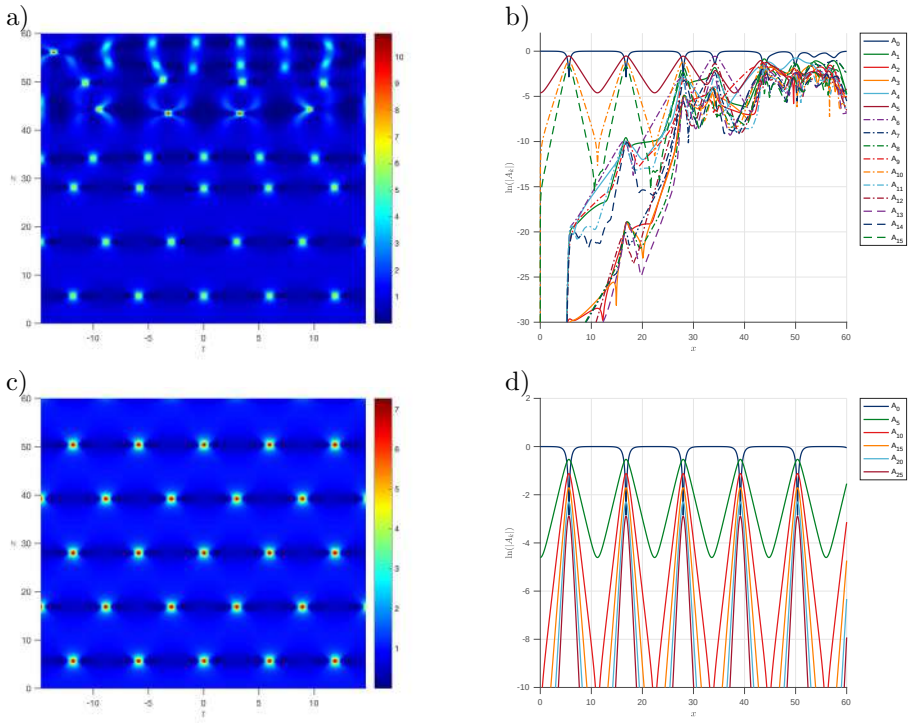


Figure 2: Dynamical generation of Talbot carpet from the first-order breathers using a DT initial condition, with $a = 0.36$ and $M = 5$. (a) A failed attempt, due to MI. (b) The corresponding Fourier spectrum, displaying the loss of Talbot periodicity due to the exponential growth of non-quintuplet modes. (c) Successful generation of the Talbot carpet using the pruning procedure. (d) Fourier spectrum after the pruning is applied.

In Fig. 1a we show numerical evolution of the first-order Akhmediev breather ($a = 0.36$) when the box size is equal to the breather's period ($M = 1$). One can see that the intensity peak at $t = 0$ is repeated along x-axis at the Talbot periods, forming a stable mode. This peak is consecutively shifted for half-a-period along the t-axis, forming the secondary Talbot image at half the Talbot period. The corresponding Fourier spectrum is shown in Fig. 1b. Next, we calculate the same breather over three periods ($M = 3$). We apply the simple pruning algorithm to Fourier modes, setting all unstable mode amplitudes to zero except the triplet modes, indexed as 0, ± 3, ± 6, ± 9, and so on. The result is an extended Talbot carpet with alternate shifting of intensity maxima along x- and t-axes, as presented in Fig. 1c. The spectrum of the triplet mode amplitudes is shown in Fig. 1d.

In Fig. 2 we display how the simple pruning technique actually works. We again choose the first-order breather with $a = 0.36$ and set the numerical

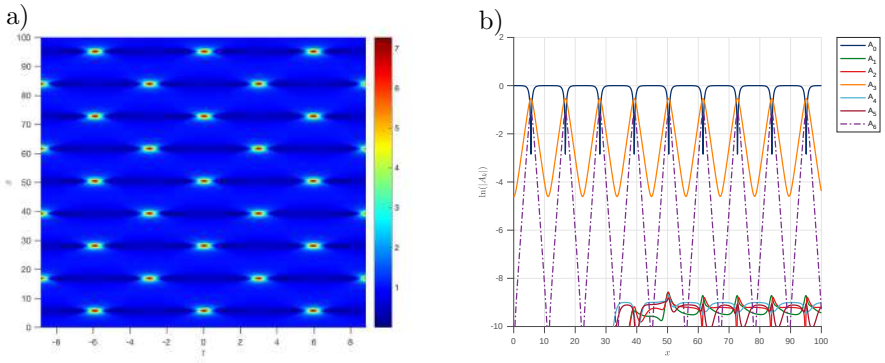


Figure 3: Same as Fig. 1 but with Gaussian pruning. (a) The carpet. (b) Its spectrum. The modes at the bottom are the suppressed unstable modes.

box to be exactly five times the breather's period: $M = 5$. In this case, the AB will be formed by the modes A_1, A_5, A_{10}, ..., A_{NM}. If the pruning algorithm is not applied, the chaotic behavior ruins the carpet after just one full Talbot cycle, as shown in Fig. 2a. This is the MI in action: the unstable modes grow exponentially, and prevent the homoclinic orbit (the initial AB mode) from returning to itself after more than one cycle. This is clearly observed in the buildup of Fourier spectrum of all modes (Fig. 2b), which destroys the spatial Talbot periodicity. Note that after the full cycle, another displaced AB appears but not at half-cycle, interacting with the full mode. As a result of this interaction, or beating of the two modes, two second-order ABs (that can be regarded as the second-order RWs) are formed around $x = 45$, which constitutes the normal channel for the production of RWs through MI. However, when the pruning procedure is applied, all non-quintuplet modes are killed after each iteration. Effectively, the procedure prevents the orbit to wander in the homoclinic tangle, forcing it to stick to itself and return back to the starting point. The result is the perfect nonlinear Talbot carpet (Fig. 2c), with the perfect Fourier spectrum (Fig. 2d).

In Fig. 3 we illustrate the Gaussian pruning algorithm, in which the unstable modes are not eliminated completely but suppressed by a Gaussian factor. Thus, the unstable modes are multiplied by a Gaussian factor that depends on their strength: when weak, they are allowed to grow; but the more they grow the more they are suppressed. Effectively, they can grow only up to a certain level, determined by the Gaussian distribution. In Fig. 3b, the unstable modes, visible at the bottom of the figure, cannot grow above the level of approximately 10^{-9}.

In Fig. 4 we present the nonlinear Talbot carpet consisting of the second-order breathers with consequently higher peak intensity. It can be

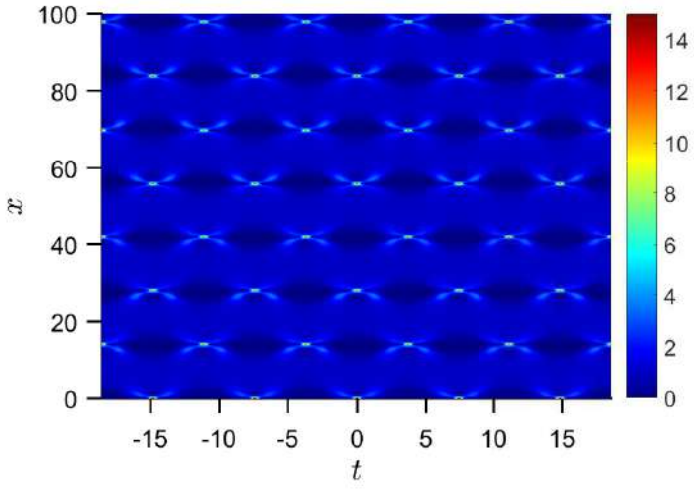


Figure 4: Double-periodic numerical solution of NLSE, made of the second-order breathers, having $a = 0.41$. The box contains 5 breather's periods. The solution is obtained using the pruning procedure.

regarded as a carpet composed of rogue waves. The box size is five times the fundamental breather period, having $a = 0.41$. This breather contains two unstable modes. Initial conditions were derived from DT along a line passing through the breather's maximum. As in the previous figures, the simple pruning algorithm was used, which left only the quintuplet Fourier modes intact.

2.2. The influence of numerics on the statistics of rogue waves

In this subsection we address the questions pertaining to the statistics of RWs. But, before going to the details of statistics, let us briefly stress how we got to this point. To recall, the generating mechanism of optical rogue waves (RWs) is the Benjamin-Feir or modulation instability (MI). It is the basic nonlinear optical process in which a weak perturbation of the background pump wave produces an exponential growth of spectral sidebands that constructively interfere to build RWs. We have produced RWs in numerical simulations of the cubic nonlinear Schrdinger equation (NLSE) with noisy (or other) inputs on the flat background. Optical RWs represent homoclinic orbits of unstable sideband modes that, due to MI, generate homoclinic chaos. The question is then whether the chaos seen belongs to the model itself or is induced by the numerical procedure applied. Namely, different numerical algorithms represent different dynamical systems and may generate different homoclinic chaos.

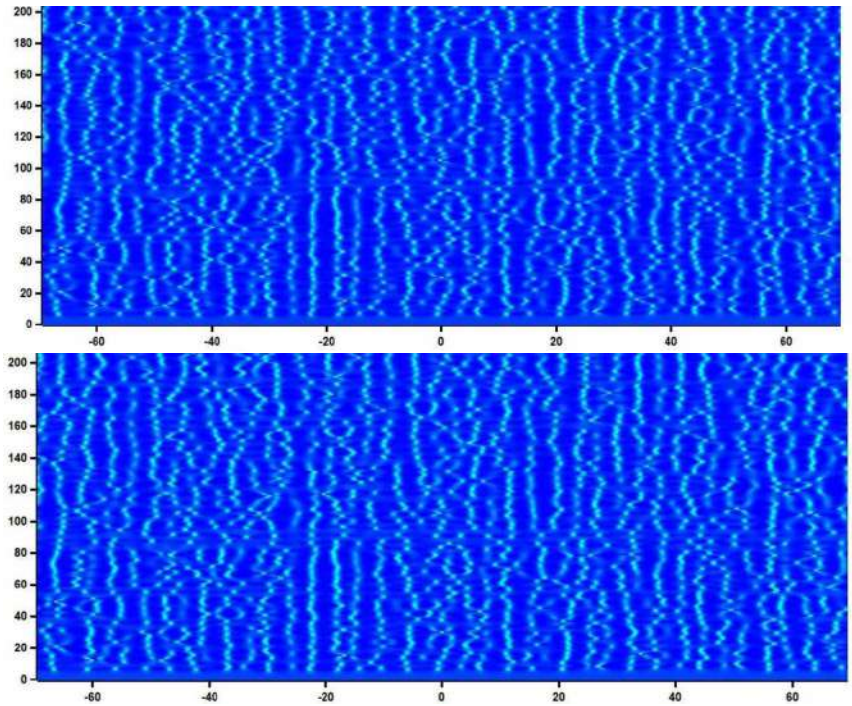


Figure 5: Evolution of the NLSE wave intensity seeded by white noise (5% amplitude) around the background intensity of 1. The wave evolution proceeds from 0 to 6500 steps, but only the first 200 are shown; the transverse box is from -70 to 70. There are 6.5 x 10^6 x 2048 grid points and more than 10^6 peaks above a threshold of 2. Top: Beam propagation method. Bottom: Symplectic algorithm.

We address the question by simulating the same equation, with exactly the same parameters, inputs and boundary conditions, but using different numerical algorithms. We utilize the standard 2nd-order beam propagation method and the 4th-order symplectic algorithm. We demonstrate that, distressingly, different algorithms provide different evolutions and different statistics of the RW peaks. Even the same algorithms may provide different evolutions by simply changing the numerical step size. We find that RWs appear in different algorithms at different places and thus may represent fictitious structures or numerical artefacts.

An example of two different evolution pictures is provided in Fig. 5. The evolution of intensity by the two methods appears consistent to about 100 steps (each step represents 1000 numerical iterations) but after that the distributions become different. The evolution proceeds to 6500 steps, after which the statistics of the probability density of intensity is formed.

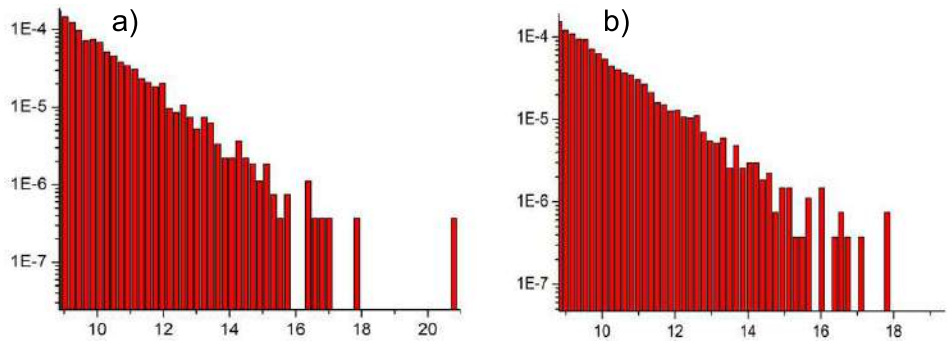


Figure 6: Statistics of high intensity peaks, obtained from Fig. 5. The statistics are formed by using two different numerical algorithms: the beam propagation method (left) and the 4th order symplectic method (right). The intensity scale starts from the intensity of the Peregrine soliton, I_{PS}=9. There are about 2640 peaks on the left and 2350 on the right.

It follows the familiar exponential decay of MI-driven systems, which may not be the case for the density of high intensity peaks, forming RWs (Fig. 6). There are millions of peaks and thousands of RWs here. The statistics are still similar, but the number of peaks, the maximum of intensity, and the slope of distributions, among other things, are different. Hence, in the chaos produced by MI, optical RWs and their statistics may appear as numerical artefacts and cannot be counted as definite defining features of the RW phenomena.

3. Conclusion

In conclusion, we have discussed the facts and fictions related to the strange nature, dynamic generation, ingrained instability, and potential applications of RWs. We have proposed the method of mode pruning for suppressing the modulation instability of rogue waves. We have demonstrated how to produce stable Talbot carpets - recurrent images of light and plasma waves - by rogue waves, for possible use in nanolithography.

We have also discussed the nature of optical rogue waves, in view of conflicting opinions expressed in the literature. In particular, we have addressed the three pairs of opposing suppositions on their nature: Linear vs. nonlinear; random vs. deterministic; and numerical vs. physical. In summary, a correct answer to the three suppositions is that the rogue waves in optics are essentially nonlinear, deterministic, and physical. They are nonlinear because the major cause of rogue waves is the modulation or Benjamin-Feir instability, which by its nature is the basic nonlinear optical process. Rogue waves are deterministic because modulation instability leads to deterministic chaos; random phenomena are probabilistic and may

look chaotic but are not deterministic. Rogue waves are physical because they appear in many experiments and media, with similar statistics.

Nevertheless, in numerical simulations optical rogue waves may appear fictitiously, as numerical artefacts. Different numerical algorithms for exactly the same inputs may provide different evolution pictures and distressingly different statistics. Hence, owing to a vague definition of rogue waves and exponential amplification of numerical errors, there are situations in which optical rogue waves may appear as linear, random, and numerical.

References

[1] Kivshar, Y.S., Agrawal, G.P.: Optical Solitons. Academic Press, San Diego (2003)

[2] Agrawal, G.P.: Applications of nonlinear fiber optics. Academic Press, San Diego (2001)

[3] Fibich, G.: The Nonlinear Schrödinger Equation. Springer, Berlin (2015)

[4] Dudley, J.M., Dias, F., Erkintalo, M., Genty, G.: Instabilities, breathers and rogue waves in optics. Nature Phot. **8**, 755 (2014)

[5] Dudley, J.M., Taylor, J.M.: Supercontinuum Generation in Optical Fibers, Cambridge University Press, Cambridge (2010)

[6] Solli, D.R., Ropers, C., Koonath, P., Jalali, B.: Optical rogue waves. Nature **450**, 1054 (2007)

[7] Ankiewicz A., Kedziora, D.J., Chowdury, A., Bandelow, U., Akhmediev, N.: Infinite hierarchy of nonlinear Schrödinger equations and their solutions. Phys. Rev. E **93**, 012206 (2016)

[8] Kedziora, D.J., Ankiewicz, A., Chowdury, A., Akhmediev, N.: Integrable equations of the infinite nonlinear Schrödinger equation hierarchy with time variable coefficients. Chaos **25**, 103114 (2015)

[9] Chin, S.A., Ashour, O.A., Belić, M.R.: Anatomy of the Akhmediev breather: Cascading instability, first formation time, and Fermi-Pasta-Ulam recurrence. Phys. Rev. E **92**, 063202 (2015)

[10] Bogolubov, N.: On the theory of superfluidity. J. Phys. (USSR) **11**, 23 (1947).

[11] Akhmediev, N.N., Korneev, V.I.: Modulation instability and periodic solutions of the nonlinear Schrödinger equation. Theor. Math. Phys. **69**, 1089 (1986)

[12] Akhmediev, N., Eleonskii, V., Kulagin, N.: Exact first-order solutions of the nonlinear Schrödinger equation. Theor. Math. Phys. **72**, 809 (1987)

[13] Erkintalo, M., Hammani, K., Kibler, B., Finot, C., Akhmediev N., Dudley, J.M., Genty, G.: Higher-order modulation instability in nonlinear fiber optics. Phys. Rev. Lett. **107**, 253901 (2011)

[14] Chin, S.A., Ashour, O.A., Nikolić, N.N., Belić, M.R.: Maximal intensity higher-order Akhmediev breathers of the nonlinear Schrödinger equation and their systematic generation. Phys. Lett. A **380**, 3625 (2016)

[15] Herbst, B.M., Ablowitz, M.J.: Numerically induced chaos in the nonlinear Schrödinger equation. Phys. Rev. Lett. **62**, 2065 (1989)

[16] Ablowitz, M.J., Herbst, B.M.: On homoclinic structure and numerically induced chaos for the nonlinear Schrödinger equation. SIAM J. Appl. Math. **50**, 339 (1990)

[17] Calini, A., Schober, C.M.: Homoclinic chaos increases likelihood of rogue wave formation. Phys. Lett. A **298**, 335 (2002)

[18] Calini, A., Schober, C.M.: Dynamical criteria for rogue waves in nonlinear Schrödinger models. Nonlinearity **25**, R99 (2012)

[19] Toenger, S., Godin, T., Billet, C., Dias, F., Erkintalo, M., Genty, G., Dudley, J.M.: Emergent rogue wave structures and statistics in spontaneous modulation instability. Sci. Reps. **5**, 10380 (2015)

[20] Akhmediev N., et al.: Roadmap on optical rogue waves and extreme events. J. Opt. **18**, 063001 (2016)

[21] Zhang, Y.Q., Belić, M.R., Zheng, H., Chen, H., Li, C., Song, J., Zhang, Y.P.: Nonlinear Talbot effect of rogue waves. Phys. Rev. E **89**, 032902 (2014)

[22] Zhang, Y., Belić, M.R., Petrović, M.S., Zheng, H., Chen, H., Li, C., Lu, K., Zhang, Y.: Two-dimensional linear and nonlinear Talbot effect from rogue waves. Phys. Rev. E **91**, 032916 (2015)

[23] Talbot, H.F.: Facts relating to optical science. Philos. Mag. **9**, 401 (1836)

[24] Wen, J., Du, S., Chen, H., Xiao, M.: Electromagnetically induced Talbot effect. Appl. Phys. Lett. **98**, 081108 (2011)

[25] Zhang, Y.Q., Yao, X., Yuan, C.Z., Li, P.Y., Yuan, J.M., Feng, W.K., Jia, S.Q., Zhang, Y.P.: Controllable Multiwave Mixing Talbot Effect. IEEE Photon. J. **4**, 2057 (2012)

[26] Song, X.-B., Wang, H.-B., Xiong, J., Wang, K., Zhang, X., Luo, K.-H., Wu L.-A.: Experimental Observation of Quantum Talbot Effects. Phys. Rev. Lett. **107**, 033902 (2011)

[27] Iwanow, R., May-Arrioja, D.A., Christodoulides, D.N., Stegeman, G.I., Min, Y., Sohler, W.: Discrete Talbot Effect in Waveguide Arrays. Phys. Rev. Lett. **95**, 053902 (2005)

[28] Deng, L., Hagley, E.W., Denschlag, J., Simsarian, J.E., Edwards, M., Clack C.W., Helmerson, K., Rolston S.L., Phillips, W.D.: Temporal, Matter-Wave-Dispersion Talbot Effect. Phys. Rev. Lett. **83**, 5407 (1999)

[29] Ryu, C., Andersen, M.F., Vaziri, A., d'Arcy, M.B., Grossman, J.M., Helmerson K., Phillips, W.D.: High-Order Quantum Resonances Observed in a Periodically Kicked Bose-Einstein Condensate. Phys. Rev. Lett. **96**, 160403 (2006)

[30] Ramezani, H., Christodoulides, D.N., Kovanis, V., Vitebskiy, I., Kottos, T.: PT-Symmetric Talbot Effects. Phys. Rev. Lett. **109**, 033902 (2012)

[31] Pfeiffer, F., Bech, M., Bunk, O., Kraft, P., Eikenberry, E.F., Brönnimann, C., Grünzweig, C., David, C.: Hard-X-ray dark-field imaging using a grating interferometer. Nat. Mater. **7**, 134 (2008)

[32] Zhang, Y., Wen, J., Zhu, S.N., Xiao, M.: Nonlinear Talbot effect. Phys. Rev. Lett. **104**, 183901 (2010)

[33] Wen, J., Zhang, Y., Xiao, M.: The Talbot effect: recent advances in classical optics, nonlinear optics, and quantum optics. Adv. Opt. Photonics **5**, 83 (2013)

[34] Ashour O.A.: Maximal intensity higher-order breathers of the nonlinear Schrödinger equation on different backgrounds. Undergraduate Research Scholars Thesis, Texas A&M University, USA (2017)

[35] Chowdury, A., Kedziora, D.J., Ankiewicz, A., Akhmediev, N.: Breather-to-soliton conversions described by the quintic equation of the nonlinear Schrödinger hierarchy, Phys. Rev. E **91**, 032928 (2015)

[36] Trippenbach, M., Band, Y.B.: Effects of self-steepening and self-frequency shifting on short-pulse splitting in dispersive nonlinear media. Phys. Rev. A **57**, 4791 (1998)

[37] Wang, D-S., Chen F., Wen X-Y.: Darboux transformation of the general Hirota equation: multisoliton solutions, breather solutions and rogue wave solutions. Advances in Difference Equations, 2016:67 (2016)

[38] Guo, R., Hao, H.Q.: Breathers and multi-solitons solutions for the higher-order generalized nonlinear Schrödinger equation. Commun. Nonlinear. Sci. Numer. Simul. **18**, 2426-2435 (2013)

[39] Mani Rajan, M.S., Mahalingam, A.: Nonautonomous solitons in modified inhomogeneous Hirota equation: soliton control and soliton interaction. Nonlinear Dyn. **79**, 2469-2484 (2015)

[40] Anderson, D., Lisak, M.: Nonlinear asymmetric self-phase modulation and self-steepening of pulses in long optical waveguides. Phys. Rev. A **27**, 1393 (1983)

[41] Backus S., Durfee III C.G., Mourou G., Kapteyn H.C., Murnane M.M.: 0.2-TW laser system at 1 kHz. Optics Letters 22, **1256** (1997)

[42] Mirzazadeh, M., Eslami, M., Zerrad, E., Mahmood, M.F., Biswas, A., Belić, M.: Optical solitons in nonlinear directional couplers by sine-cosine function method and Bernoulli's equation approach. Nonlinear Dyn. **81**, 1933-1349 (2015)

[43] Biswas, A., Khalique, C.M: Stationary solutions for nonlinear dispersive Schroödinger equation. Nonlinear Dyn. **63**, 623-626 (2011)

[44] Hirota, R.: Exact envelope-soliton solutions of a nonlinear wave equation. J. Math. Phys. **14**, 805-809 (1973)

[45] Tao, Y., He, J.: Multisolitons, breathers, and rogue waves for the Hirota equation generated by the Darboux transformation. Phys. Rev. E **85**, 026601 (2012)

[46] Guo, R., Zhao, X-J.: Discrete Hirota equation: discrete Darboux transformation and new discrete soliton solutions. Nonlinear Dyn. **84**, 1901-1907 (2016)

[47] Nikolić, S.N., Aleksić, N.B., Ashour, O.A., Belić, M.R., Chin, S.A.: Systematic generation of higher-order solitons and breathers of the Hirota equation on different backgrounds. Nonlinear Dyn. **89**, 1637-1649 (2017)

[48] Chowdury, A., Kedziora, D.J., Ankiewicz, A., Akhmediev, N.: Soliton solutions of an integrable nonlinear Schrödinger equation with quintic terms. Phys. Rev. E **90**, 032922 (2014)

[49] Lan, Z., Gao, B.: Solitons, breather and bound waves for a generalized higher-order nonlinear Schrödinger equation in an optical fiber or a planar waveguide. Eur. Phys. J. Plus **132**, 512 (2017)

[50] Chowdury, A., Kedziora, D.J., Ankiewicz, A., Akhmediev, N.: Breather solutions of the integrable nonlinear Schrödinger equation and their interactions. Phys. Rev. E **91**, 022919 (2015)

[51] Yang, Y., Yan, Z., Malomed, B.A.: Rogue waves, rational solitons, and modulational instability in an integrable fifth-order nonlinear Schrödinger equation. Chaos **25**, 103112 (2015)

[52] Kedziora, D.J., Ankiewicz, A., Akhmediev, N.: Circular rogue wave clusters. Phys. Rev. E **84**, 056611 (2011)

[53] Nikolić, S.N., N.B., Aleksić, Ashour, O.A., Belić, M.R., Chin, S.A.: Breathers, solitons and rogue waves of the quintic nonlinear Schrödinger equation on various backgrounds. Nonlinear Dyn. doi.org/10.1007/s11071-018-4726-8 (2019)

[54] F.T. Arecchi, U. Bortolozzo, A. Montina, S. Residori, "Granularity and inhomogeneity are the joint generators of optical rogue waves," Phys. Rev. Lett. 106, 153901 (2011)

Transverse Transport of Magnetized Particles in Random Electric Fields*

V. I. Zasenko†, A. G. Zagorodny, O. M. Cherniak
Bogolyubov Institute for Theoretical Physics, Kyiv, Ukraine

Abstract

The study of anomalous particle transport across a magnetic field is an important task of plasma physics. The cause of anomalous transport may be electrostatic turbulence. In the limit of small correlation times of random fields transport is diffusive. In the opposite limit of large correlation times transport occurs as advection. We proposed the statistical approach to description of transport of magnetized particles undergoing random isotropic electric fields in a wide range of correlation times including both limits without use of free parameters. Direct numerical simulations were carried out to verify solutions of the transport equations.
PACS codes: 05.40.-a, 52.25.F

1. Introduction

Plasma in space and laboratory is as a rule in a non-equilibrium state because of gradients of its density and temperature, as well due to fluxes of particles and radiation usually passing through it. For this reason waves with intensity much higher than a thermal level are excited, and nonlinear wave interactions result in plasma turbulent state. Particle collisions with random waves bring to anomalous transport. In plasma with high level of turbulence anomalous transport dominates over classical diffusion caused by pair particle collisions. Study of turbulent transport is important for understanding behavior of plasma systems, in particular estimation of the plasma confinement time in fusion devices.

A great deal of research is focused on a problem of turbulent transport. Here we consider that aspect of the general problem that relates to dependence of transport coefficients on a correlation time of a random electric fields. There are two limits of small and large correlation times characterized by different modes of particle transport across magnetic field. A small correlation time, in compare with time evolution of macroscopic characteristics, corresponds to instantaneous collisions, and behavior of particles

* This work has been supported by the Program of Fundamental Research of the Department of Physics and Astronomy of the National Academy of Sciences of Ukraine, Project 0117U000240.
† e-mail address: zasenko@bitp.kiev.ua

is similar to the classical Brownian motion. In this limit the Corrsin approximation [1] is valid for statistical description of particle ensemble. The approximation is based on the assumption that a spatial distribution of particles in a course of their scattering by random fields remains of the Gaussian form.

The opposite limit of infinitely large correlation time is realized for frozen (not dependent on time) random electrostatic fields. Here particle trapping in closed two-dimensional orbits can occur, and then their spreading becomes rather convective than diffusive. In addition, simulations show that particle distribution is non-Gaussian. To estimate asymptotic diffusion coefficients in this limit an analogy with percolation was proposed [2, 3]. Particle transport in two-dimensional Gaussian incompressible stochastic velocity fields in a wide range of correlation times and with accounts for particle trapping was later studied using the decorrelation trajectory method [4, 5, 6].

The description of evolution of particle ensembles is convenient to carry out in terms of a distribution functions. In [7] it was shown that particle distribution function in a random external field is governed by a nonlinear integro-differential equation. Currently, there are no universal methods for constructing solutions of equations of this type. Here we shall consider an approximation based on a new closure of statistical equations [8, 12]. The solutions will give particle transport coefficients in a wide range of correlation times, including both limits. Comparison of our approach with the decorrelation trajectory method mentioned above and based on a different closure is given in [13].

The structure of the paper is as follows. In Sec. 2 the equations of particle microscopic motion in fields with random phases are given in a drift approximation. They are used for numerical calculation of particle trajectories. Averaging over realizations gives us evolution of statistical characteristics of particle ensemble. In Sec. 3 we introduce a new closure and formulate statistical equations for dispersion of particles in random frozen field with infinitely large correlation time. These statistical equations are derived from the same equations of microscopic motion that were used for simulation. The solution of statistical equations is compared with the result of numerical simulation. In Sec. 4 our approach is applied to random fields with finite correlation times. Conclusions are given in Sec. 5.

2. Frozen field with infinite correlation time

First we look at a frozen random electric field. It is characterized by an infinitely long correlation time and is therefore of particular interest. When constructing a statistical description, it is usually assumed that there are two time scales, namely fast time of collisions and slow time of evolution of a distribution function. The difference in time scales makes it possible to construct a kinetic equation for a distribution function by averaging over fast time. In contrast, for statistical description of a particle ensemble in a

frozen field it is necessary to have a different approach.

Let us consider drift motion of particle guiding centers in random electrostatic field $\mathbf{E} = -\nabla\Phi$ given by a potential $\Phi(\mathbf{r})$ across a constant uniform magnetic field $\mathbf{B}$. Extension of our approach for finite Larmor radius may be found in [9, 10, 11]. Two components of drift velocity $v_{x,y}$ in a plane perpendicular to the magnetic field are governed by the equations

$$v_x = -\frac{\partial}{\partial y}\varphi(\mathbf{r}), \qquad v_y = \frac{\partial}{\partial x}\varphi(\mathbf{r}), \tag{1}$$

where $\mathbf{r} = (x, y)$ in Cartesian and (r, θ) in polar coordinates. $\varphi = c\Phi/B$ is a statistically isotropic stream function. Each realization of φ is supposed to be a superposition of $N_k \times N_\phi$ harmonics

$$\varphi(r, \theta) = \sum_{i=1}^{N_k} \sum_{j=1}^{N_\phi} \varphi_i \cos(k_i r \, \cos(\phi_j - \theta) + \alpha_i + \beta_j) \tag{2}$$

with random phases α_i and β_j. Summation is carried out over wave numbers k_i and polar angles ϕ_j which determine a set of two-dimensional wave vectors. Partial intensities are distributed over wave numbers according to Gaussian with a width Δk

$${\varphi_i}^2 = \varphi_0^2 \, g \, \exp\left(-\frac{k_i^2}{\Delta k^2}\right),$$

where φ_0^2 is a field intensity and g is a normalizing factor. Equations of motion (1) were solved numerically for a number of realizations. The statistical characteristics of particle ensemble such as a particle dispersion, a running diffusion coefficient and a Lagrangian velocity correlation function, which will be introduced in the next section, were calculated by averaging over realizations.

Note that particles move along the streamlines corresponding to a fixed level of potential $\Phi = const$. Fig. 1 shows the potential landscape. Almost all particles move in closed orbits. The size of the orbits is smaller near peaks and pits and larger in valleys. For particles uniformly distributed in space their distribution over potential levels, with account for Eq. (2) and according to the central limit theorem, is Gaussian. Distribution of orbits of different sizes will be similar. These properties of orbits are illustrated in Figs. 2, 3. Here and further on the axes are dimensionless units. The accuracy of the statistical description can be improved by splitting of the whole ensemble of particles into subensembles corresponding to different potential levels.

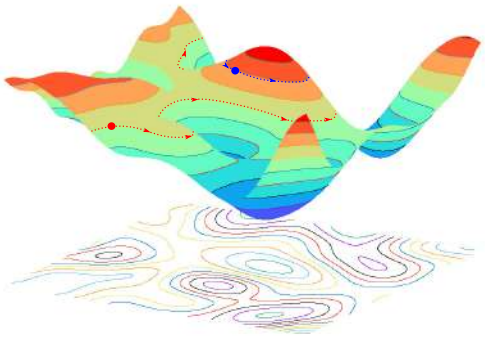


Figure 1: Potential landscape.

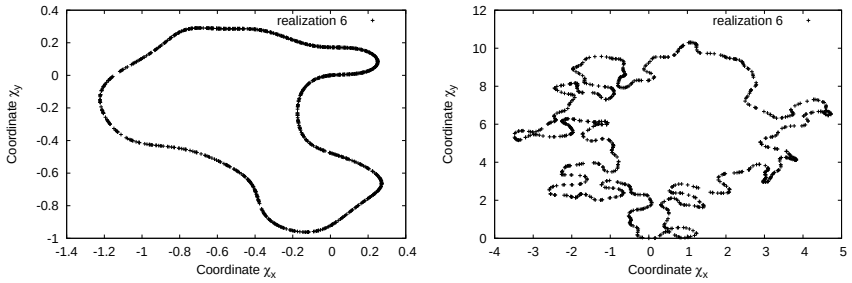


Figure 2: Particle orbits at two potential levels differ in size. Particle population on these levels is shown in Fig. 3.

3. Statistical equations

The equation for distribution function of particles $f(\mathbf{r}, t)$ in given velocity field $\mathbf{v}$ reads

$$\frac{\partial}{\partial t} f + \mathbf{v} \frac{\partial}{\partial \mathbf{r}} f = 0.$$

This equation is linear, but whether velocities are random it is *statistically nonlinear*. To show this we split a distribution function on an averaged over field realizations component $F = \langle f \rangle$ and a fluctuation δf, so that $f = F + \delta f$. Kinetic equation for the averaged distribution function is as follows

$$\frac{\partial}{\partial t} F + \langle \mathbf{v} \frac{\partial}{\partial \mathbf{r}} \delta f \rangle = 0,$$

where the second term is a collision integral given through an unknown fluctuation δf. To obtain the collision integral in terms of averaged dis-

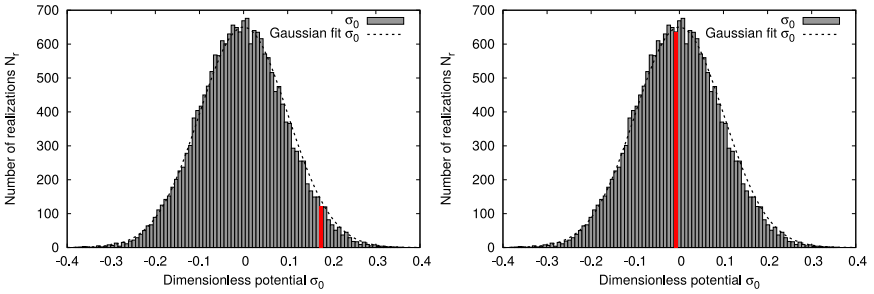


Figure 3: Particle distribution over potential level. The two levels corresponding to the orbits in the previous figure are indicated.

tribution function we follow the paper [7] and express $F(\mathbf{r},t)$ through its initial value $F(\mathbf{r},0)$ and the averaged transition probability $W(\mathbf{r},t;\mathbf{r}',0)$

$$F(\mathbf{r},t) = \int d\mathbf{r}' W(\mathbf{r},t;\mathbf{r}',0) F(\mathbf{r}',0).$$

The transition probability W (also the Green function, or the distribution function with the δ-like initial condition) is governed by the kinetic equation with collision integral given in terms of W and the Eulerian velocity correlation function $\langle v_i(\mathbf{r}) v_j(\mathbf{a})\rangle$

$$\frac{\partial}{\partial t} W(\mathbf{r},t;\mathbf{r}',t') = \frac{\partial}{\partial r_i} \int\limits_{t'}^{t} d\tau \int d\mathbf{a}\, W(\mathbf{r},t;\mathbf{a},\tau) \langle v_i(\mathbf{r}) v_j(\mathbf{a})\rangle \frac{\partial}{\partial r'_j} W(\mathbf{a},\tau;\mathbf{r}',t'), \tag{3}$$

with the initial condition

$$W(\mathbf{r},t';\mathbf{r}',t') = \delta(\mathbf{r}-\mathbf{r}').$$

Eq.(3) describes nonlocal in time and space transport. Moreover, this equation is nonlinear since the nonlocal diffusion coefficient depends on the transition probability. Methods to find exact solutions of such types equations are unknown.

But assuming that the particle distribution function is the Gaussian, the equation (3) is simplified. The assumption was proposed by Corrsin [1], and it is justified in the limit of small correlation times of random fields. In this limit Eq. (3) is reduced to

$$\frac{\partial}{\partial t} W(\mathbf{r},t;\mathbf{0},0) = D(t) \Delta W(\mathbf{r},t;\mathbf{0},0), \qquad W(\mathbf{r},0;\mathbf{0},0) = \delta(\mathbf{r}-\mathbf{0}). \tag{4}$$

The solution of Eq. (4) is the Gaussian

$$W(\mathbf{r},t;\mathbf{0},0) = 1/(\pi\langle r^2(t)\rangle)\ \exp\left(-r^2/\langle r^2(t)\rangle\right), \tag{5}$$

where the mean square particle displacement $\langle r^2(t)\rangle$ is given through the unknown running diffusion coefficient $D(t)$

$$\langle r^2(t)\rangle = 2\int_0^t D(t')dt'. \tag{6}$$

And the diffusion coefficient is obtained by integration of the Lagrangian correlation function of random velocities $V_L(t)$ which is unknown as well

$$D(t) = \int_0^t V_L(t')dt'. \tag{7}$$

The closure of set of equations (5-7) is given by the relation between the Eulerian V_E and Lagrangian $V_L(t)$ correlation functions:

$$V_L(t) = \int d\mathbf{a}\, W(\mathbf{a},t;\mathbf{0},0)\ V_E(a^2). \tag{8}$$

The Eulerian correlation function V_E for a prescribed random field (2) is calculated by averaging over random phases. As far as the field is statistically isotropic it depends on squared distance between two points

$$V_E(r^2) = \ \langle v_x(\mathbf{r})\, v_x(\mathbf{0})\rangle + \langle v_y(\mathbf{r})\, v_y(\mathbf{0})\rangle. \tag{9}$$

The solution of Eqs. (5) - (9) will be illustrated later. The system of equations presented here is obtained in the Corrsin approximation, which is often used in the theory of anomalous transport. However, it should be noted here that the assumption of a small correlation time does not correspond to a frozen field whose correlation time tends to infinity.

For the frozen field, we propose another closure of statistical equations. Direct simulation shows, the distribution function is not Gaussian. So to simplify the problem, we will look not for a particle distribution function, but only for its second moment which describes a mean square particle displacement (dispersion). The first moment (average particle displacement) in isotropic fields is zero.

Equations (6) - (7), first formulated by Taylor [14] for particle diffusion under the influence of random fields, do not depend on a shape of a distribution function. Therefore, they will remain unchanged in the new system of equations. But the closure for arbitrary correlation times will be different. Instead of Eq. (8) we assume the following relation between the Eulerian and Lagrangian correlation functions

$$V_L(t) = V_E(\langle r^2(t)\rangle). \tag{10}$$

Then from Eqs. (6), (7), (9), (10) we obtain the equation for evolution of the particle dispersion

$$\frac{d^2\langle r^2(t)\rangle}{dt^2} = V_E(\langle r^2(t)\rangle), \tag{11}$$

where the Eulerian velocity correlation function corresponding to the random field (2) is of the form

$$V_E(r^2) = v_0^2 \exp(-\alpha)\{(1-2\alpha)I_0(\alpha) + 2\alpha I_1(\alpha)\}. \tag{12}$$

Here I_0 and I_1 are the modified Bessel functions, v_0^2 is an intensity of random veocity field related to φ_0^2, and $\alpha = (1/8)\Delta k^2 r^2$.

The results can be improved by more detailed description of particle orbits. We take into account that particle motion governed by Eq. (1) occurs along equipotential lines $\phi(\mathbf{r}) = const$. Simulations show that particles travel different distances from a starting point at different levels of potential (Fig. 2). A group of particles on a fixed level of potential constitute a subensemble. The amplitude of the partial Euler correlation function in each subensemble is determined by the potential level. The contribution of each subensembles to statistical characteristics of the whole ensemble of particles is proportional to their weights. According to the central limit theorem distribution of particle over subensembles is Gaussian. In addition, a direct numerical test shows that for finite number of particles in numerical simulation the Gaussian distribution over initial values of potential is a good approximation (Fig. 3). Eqs. (11), (12) were solved numerically, and after averaging over subensambles we obtain the Lagrangian velocity correlation function (10), the running diffusion coefficient (7), and the mean square displacement (6). In Fig. 4 the Lagrangian velocity correlation function $V_L(t)$ obtained as a solution of Eqs. (10)-(12) is compared with the Corrsin approximation, Eqs. (5-9), and the result of numerical simulation. It is important the solution based on the new closure, in contrast to the Corrsin approximation, recovers negative values of the correlation function observed in simulation. These negative values reflect particle trapping in closed orbits shown in Fig. 2.

Double time integration of the Lagrangian velocity correlation function gives the particle dispersion shown as log-log plot in Fig. 5. The linear dependence of dispersion on time divides the area of the figure into two parts corresponding to sub- and superdiffusion. The dispersion of particle displacements obtained as a solution based on our closure grows slower than linear law, i.e. manifests the subdiffusive regime caused by particle trapping. This solution is similar to that obtained in a direct numerical simulation. In contrast, the Corrsin approximation gives a superdiffusion. The comparison of our approach with the decorrelation trajectory method is given in the paper [13].

A frozen field serves as a good test for statistical theories as far as the common assumption about small correlation time is inappropriate, and it

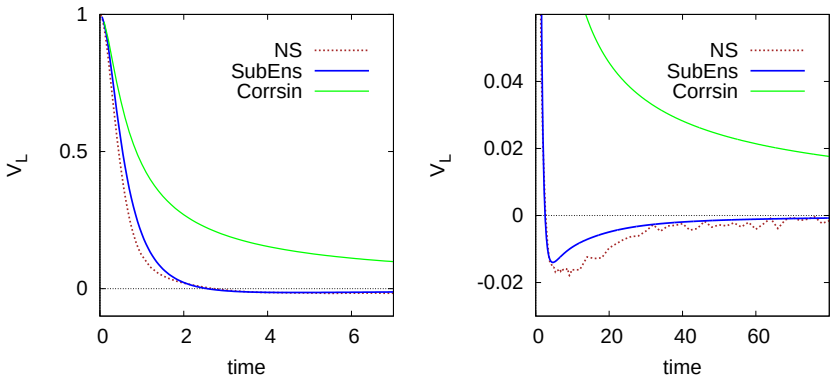


Figure 4: The Lagrangian velocity correlation function (normalized to 1) in a large (left) and small (right) scales. (NS) - numerical simulation, (SubEns) - the solution with our closure and averaging over subensembles, and (Corrsin) - the solution with assumption about the Gaussian particle distribution.

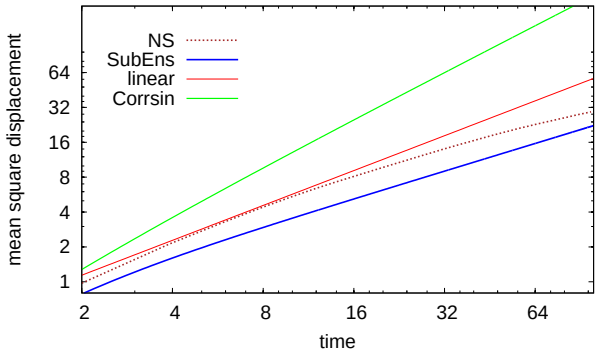


Figure 5: Particle dispersion in a log-log scale. (NS) - numerical simulation, (SubEns) - our closure, the linear dependence of dispersion on time, and the Corrsin approximation.

is not possible to separate two time scales corresponding to collisions and evolution of distribution function. In addition, the particle trapping effect plays a significant role in a frozen field.

The running diffusion coefficient (7) considered in the drift approximation slowly tends to zero for large times. Zero asymptotic value for frozen field is explained by particle trapping in closed orbits. Extension of our method for a finite Larmor radius, when particle orbits are no longer closed, is given in papers [9, 10, 11].

4. Fields with exponential decay of correlations

After testing Eqs.(6),(7),(9),(10) we apply our method to piece-wise frozen fields with a finite correlation time. If the phase of the field (2) is jumping, regularly or randomly, then the correlation function of such field will decay exponentially. Waves with jumping phases were studied experimentally and theoretically. They can penetrate in overdense plasma and initiate a new type of discharge [15, 16]. New features of particle heating by such waves were discussed in [16, 17].

Let us consider a field which is differ from (2) by a randomly jumping phase $\gamma(t)$

$$\varphi(r,\theta,t) = \sum_{i=1}^{N_k} \sum_{j=1}^{N_\phi} \varphi_i \cos\left(k_i r\, \cos(\phi_j - \theta) + \alpha_i + \beta_j + \gamma(t)\right), \qquad (13)$$

where $\gamma(t)$ jumps by an arbitrary value with a frequency ν and the probability p. Then the Eulerian correlation function $V_E(r^2)$ given by Eq. (12) is transformed as follows

$$V_E(r^2) \;\rightarrow\; V_E(r^2) \exp\left(-t/\tau\right). \qquad (14)$$

Depending on the parameters of the phase jumps ν and p time of correlation decay $\tau = -1/\left[\nu \ln\left(1-p\right)\right]$ may vary from zero to infinity.

Eqs. (11), (12), (14) were solved numerically for various correlation times τ. Particle transport initially is subdiffusive in an interval of few τ, and then particle dispersion grows linearly in time. This means that the diffusion coefficient attains a constant asymptotic value. In Fig. 6 the dependence of the asymptotic diffusion coefficient on the correlation time is shown in compare with results of direct numerical simulation. At small correlation times we observe a known diffusion which can be described by the Corrsin approximation as well. At larger correlation times particle transport is slowed down. As a correlation time grows, the effect of particle trapping is becoming increasingly pronounced, and transport looks more like advection. Obtained solutions describe the transition of transport coefficient between two limits of small and large correlation times without use of fitting parameters, and are in agreement with the results of simulations.

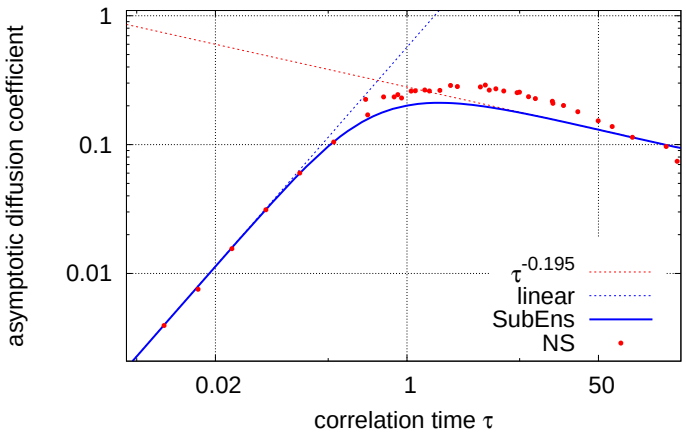


Figure 6: Asymptotic diffusion coefficient $D(t \to \infty)$ vs correlation time τ.

5. Conclusions

An anomalous transport of charged particles across magnetic field caused by random electrostatic field was studied in a drift approximation. In the limit of small correlation times of random fields particle transport is similar to Brownian diffusion. In the opposite limit of long correlation times particles are trapped in closed orbits, and transport occurs as advection. The new closure of statistical equations was considered, and solutions of these equations give asymptotic transport coefficients in a wide range of field correlation times. A transition from diffusion to advection mode was shown without use of free parameters. Obtained results were supported by direct numerical simulations.

References

[1] S. Corrsin, *Atmospheric Diffusion and Air Pollution, edited by H.E. Landsberg and J. van Mieghem*, New York: Academic Press, 1959 p. 441.

[2] A. V. Gruzinov, M. B. Isichenko and Ya. L. Kalda, *Zh. Eksp. Teor. Fiz.* **97** (1990) 476; *Sov. Phys. JETP* **70** (1990) 263.

[3] M. B. Isichenko, *Rev. Mod. Phys.* **64** (1992) 961.

[4] M. Vlad, F. Spineanu, J. H. Misguich and R. Balescu, *Phys. Rev. E* **58** (1998) 7359.

[5] R. Balescu, *Aspects of Anomalous Transport in Plasmas*, Bristol: IOP, 2005.

[6] M. Vlad and F. Spineanu, *New J. Phys.* **19** (2017) 025014.

[7] S. A. Orszag and R. H. Kraichnan, *Phys. Fluids* **10** (1967) 1720.

[8] V. I. Zasenko, A. G. Zagorodny and O. M. Chernyak, *Ukrainian Journal of Physics* **56** (2011) 1007.

[9] O. M. Cherniak, V. I. Zasenko and A. G. Zagorodny, *Problems of atomic science and technology* **106** (2016) 96.

[10] O. M. Cherniak and V. I. Zasenko, *Ukrainian Journal of Physics* **62** (2017) 495.

[11] O. M. Cherniak and V. I. Zasenko, *IOP Conf. Series: Journal of Physics: Conf. Series* **1197** (2019) 012003.

[12] V. I. Zasenko, O. M. Chernyak and A. G. Zagorodny, *Problems of atomic science and technology* **106** (2016) 77.

[13] O. M. Cherniak, *Ukrainian Journal of Physics* **60** (2015) 1196.

[14] G. I. Taylor, Proceedings of the London Mathematical Society **s2-20** (1922) 196.

[15] V. I. Karas, Ya .B. Fainberg *et al.*, *Plasma Physics Reports* (2005) **31** (2010) 748.

[16] A. F. Alisov, A. M. Artamoshkin *et al.*, *Plasma Physics Reports* **36** (2010) 736.

[17] V. I. Zasenko, A. G. Zagorodny, and O. M. Cherniak, *Problems of atomic science and technology* **107** (2017) 60.

Portfolio Optimization – A Review*

Branko Urošević†
Faculty of Economics, University of Belgrade, Serbia

Nikola Vasiljević‡
Department of Banking and Finance, University of Zurich, Switzerland

Abstract

In this paper we review some of the salient results in optimization of financial portfolios. We commence by considering Markowitz portfolio optimization and its generalizations. This is followed by portfolio optimization based on maximization of the expected utility in discrete and continuous time. JEL Code: G11.

1. Introduction

A financial portfolio is a collection of investment assets. In this paper we consider some of the core techniques aimed at finding the best *ex-ante* trade-off between expected return and portfolio risk.

In the first part of the paper (Section 2.) we consider mean–variance optimization problem. We begin with the simple case when portfolio consists only of risky assets and short sales are not prohibited. An investor minimizes portfolio variance, the simplest and most frequently used measure of portfolio risk. At the same time, she aims to achieve a certain level of the expected return. For each target portfolio return one can, then, find a closed-form solution for optimal portfolio weights. Additionally, we consider global minimum variance (GMV) portfolio and argue that truly optimal are only those portfolios which have the expected return not lower than this benchmark portfolio. We then extend the analysis along several dimensions.

In the second part of the paper (Section 3. and 4.) we study portfolio optimization based on expected utility maximization. In Section 3. we consider the case of trading in a discrete-time framework. We commence by introducing the notion of the expected utility and discuss how can one quantify absolute and relative risk aversion. We explain the connection

* This work has been supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Project No 179005.

† e-mail address: urosevic@ekof.bg.ac.rs

‡ e-mail address: nikola.vasiljevic@bf.uzh.ch

between these measures of risk aversion and the optimal investment policy. Next, we consider optimal asset allocation for an investor characterized by a logarithmic utility function and derive—in closed form—the optimal investment policy both for a single period and in case of multiple investment periods. We demonstrate that, when parameters of the model do not change over time, optimal fraction of money invested in risky assets remains constant. This result is consistent with the general theorem stated earlier in that section. In Section 4. we consider portfolio optimization problem in continuous time. As this topic is somewhat more technical we first explain some of the key notions from stochastic calculus used in the later development. Subsequently, we derive optimal portfolios for an investor with logarithmic and more general power utility function. Section 5. concludes.

2. Markowitz portfolio optimization problem

Consider a portfolio of risky assets (e.g., stocks). We can define return on asset i as a relative change of its price p^i in consecutive points in time. Thus, return on asset i at time t is given by the expression:

$$r_t^i = \frac{p_{t+1}^i - p_t^i}{p_t^i}. \tag{1}$$

For simplicity, we assume that prices are adjusted for dividends. Such prices are reported habitually by financial information providers. Therefore, the formula above takes into the account both dividends and capital gains, i.e., it reports total returns.

Suppose, further, that we have partitioned funds at our disposal into fractions corresponding to investments in each asset $i \in \{1, 2, \ldots, N\}$, where $N \in \mathbb{N}$. We denote these fractions by w_i. Then, N-dimensional (column) vector of the weights defining our portfolio is $\mathbf{w} := (w_1, w_2, \cdots, w_N)^\intercal$. While the vector of weights may depend on time, in this section we consider a static investment strategy where investment decision is made at time t and the realization occurs at time $t + 1$. Therefore, for simplicity, in this section we drop the time index for portfolio weights. For an investment portfolio $\mathbf{w}$, its return at time t is given by the weighted average of asset returns:

$$r_p = \sum_{i=1}^{N} w_i r_t^i = \mathbf{w}^\intercal \cdot \mathbf{r}. \tag{2}$$

Here, we have denoted by $\mathbf{r} := (r_1, r_2, \cdots, r_N)^\intercal$ the (column) vector of asset returns and by $\cdot$ the scalar product between the two vectors. It is important to note that, in contrast to $\mathbf{w}$ which is determined by us, the vector of asset returns is determined by market forces. We model market outcome as drawn by chance. How successful are we going to be as investors depends, therefore, on our decisions as well as on luck.

In the early 1950s, Harry Markowitz, then a PhD student in finance at the University of Chicago, has proposed the following strategy to construct an optimal financial portfolio (see [1]). Suppose we have a target portfolio return μ that we expect to receive at the end of our investment horizon (which is one period as described above). This return represents our investment goal. In that case, the expected return on a portfolio needs to satisfy the condition

$$\mathbb{E}\left(r_p\right) = \mathbb{E}\left(\sum_i^N w_i r_t^i\right) = \sum_i^N w_i \mathbb{E}\left(r_t^i\right) = \mathbf{w}^\intercal \cdot \mathbf{m} \geq \mu. \tag{3}$$

Here, $\mathbf{m} := (m_1, m_2, \cdots, m_N)^\intercal$ is a (column) vector of expected returns on individual assets. Apart from this constraint, feasible portfolios need to satisfy also the following:

$$\sum_i^N w_i = \mathbf{w}^\intercal \cdot \mathbf{1} = 1, \tag{4}$$

where $\mathbf{1} := (1, 1, \cdots, 1)^\intercal$ is the N-dimensional (column) vector of ones. Since we do not want to leave any money on the table (we consider just a one-period investment), we need to invest it somehow in the N assets. Markowitz has proposed that out of all portfolios satisfying conditions (3)–(4) we select the one which has the smallest risk.

In the simplest case, portfolio risk is defined as standard deviation of portfolio returns (see [1], [2]). Instead of standard deviation, it is more practicable to minimize its square, i.e., variance. From the properties of variance of a sum of random variables it follows that:

$$\mathbb{V}\left(r_p\right) = \mathbb{V}\left(\sum_i^N w_i r_t^i\right) = \mathbf{w}^\intercal \cdot \mathbf{\Sigma} \cdot \mathbf{w}. \tag{5}$$

Here, matrix $\mathbf{\Sigma}$ is a covariance matrix of asset returns, which is a $(N \times N)$-dimensional, positive-definite, and symmetric matrix. Positive-definiteness ensures that variance of any portfolio is non-negative. Thus, in this simple approach the problem of finding optimal portfolio boils down to solving a quadratic optimization problem for weights $\mathbf{w}$:

$$\begin{aligned} &\min_{\mathbf{w}} \mathbb{V}(r_p), \\ \text{s.t. } &\begin{cases} \mathbf{w}^\intercal \cdot \mathbf{m} \geq \mu, \\ \mathbf{w}^\intercal \cdot \mathbf{1} = 1. \end{cases} \end{aligned} \tag{6}$$

2.1. Solution to the standard Markowitz problem

Problem (6) is called Mean–Variance Optimization (MVO) problem or Markowitz Optimization problem. Positive-definiteness of covariance matrix $\mathbf{\Sigma}$ and linearity of constraints (3)–(4) ensure that a unique global minimum exists. It is clear that the optimal portfolio weights, i.e., the solution to MVO problem, are functions of the parameter μ, which represents the target return. Moreover, since higher expected return leads to higher optimal risk, the problem can be restated as follows:

$$\begin{aligned} &\min_{\mathbf{w}} \mathbb{V}(r_p), \\ &\text{s.t.} \begin{cases} \mathbf{w}^\intercal \cdot \mathbf{m} = \mu, \\ \mathbf{w}^\intercal \cdot \mathbf{1} = 1. \end{cases} \end{aligned} \tag{7}$$

Problem (7) can be solved using the simple Lagrange method. For convenience we introduce a multiplicative factor in front of the objective function:

$$\mathcal{L} = \frac{1}{2}\mathbf{w}^\intercal \cdot \mathbf{\Sigma} \cdot \mathbf{w} - \lambda_1(\mathbf{w}^\intercal \cdot \mathbf{m} - \mu) - \lambda_2(\mathbf{w}^\intercal \cdot \mathbf{1} - 1). \tag{8}$$

The first order conditions with respect to $\mathbf{w}$, λ_1, and λ_2 are given by the system of linear equations:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} &= \mathbf{\Sigma} \cdot \mathbf{w} - \lambda_1 \mathbf{m} - \lambda_2 \mathbf{1} = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda_1} &= \mathbf{w}^\intercal \cdot \mathbf{m} - \mu = 0, \\ \frac{\partial \mathcal{L}}{\partial \lambda_2} &= \mathbf{w}^\intercal \cdot \mathbf{1} - 1 = 0. \end{aligned} \tag{9}$$

Since matrix $\mathbf{\Sigma}$ is positive-definite, it can be inverted. From the first equation we obtain that the optimal vector $\hat{\mathbf{w}}$ satisfies

$$\hat{\mathbf{w}} = \lambda_1 \mathbf{\Sigma}^{-1} \cdot \mathbf{m} + \lambda_2 \mathbf{\Sigma}^{-1} \cdot \mathbf{1}. \tag{10}$$

Plugging this expression back into the two constrains and introducing the notation:

$$\begin{aligned} a &:= \mathbf{m}^\intercal \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{m}, \\ b &:= \mathbf{m}^\intercal \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{1}, \\ c &:= \mathbf{1}^\intercal \cdot \mathbf{\Sigma}^{-1} \cdot \mathbf{1}, \\ d &:= ac - b^2, \end{aligned}$$

the Lagrange multipliers read:

$$\begin{aligned} \lambda_1 &= \frac{c\,\mu - b}{d}, \\ \lambda_2 &= \frac{a - b\,\mu}{d}. \end{aligned}$$

Replacing these expressions into equation (10) for $\hat{\mathbf{w}}$ we finally obtain that:

$$\begin{aligned} \hat{\mathbf{w}} &= \mathbf{g} + \mu\mathbf{h}, \\ \mathbf{g} &= \frac{a}{d}\boldsymbol{\Sigma}^{-1}\cdot\mathbf{1} - \frac{b}{d}\boldsymbol{\Sigma}^{-1}\cdot\mathbf{m}, \\ \mathbf{h} &= \frac{c}{d}\boldsymbol{\Sigma}^{-1}\cdot\mathbf{m} - \frac{b}{d}\boldsymbol{\Sigma}^{-1}\cdot\mathbf{1}. \end{aligned} \tag{11}$$

Given a target expected return μ, equation (11) demonstrates the way how to achieve it (*ex-ante*) with minimal risk. Moreover, $\mathbf{g}$ and $\mathbf{g}+\mathbf{h}$ both have an economic meaning. Namely, vector $\mathbf{g}$ corresponds to optimal portfolio weights when $\mu = 0$, while $\mathbf{g}+\mathbf{h}$ corresponds to portfolio with $\mu = 1$. All other optimal portfolios are just linear combinations (or, in other words, portfolios) of these two portfolios. Thus, knowing just two optimal portfolios we can reconstruct the entire efficient frontier, i.e., the set of all possible optimal portfolios.

2.2. Global minimum variance portfolio

However, this is not the end of the story. Namely, it turns out that not all portfolios implied by equation (11) are worth investing in, as illustrated in Figure 1. To find the full set of *efficient* portfolios we need to compute first the Global Minimum Variance Portfolio (GMVP). This is achieved by solving the optimization problem without constraint (3) related to target return:

$$\begin{aligned} &\min_{\mathbf{w}} \mathbb{V}(r_p), \\ &\text{s.t. } \mathbf{w}^{\intercal}\cdot\mathbf{1} = 1. \end{aligned} \tag{12}$$

Following the solution procedure presented in Section 2.1., we utilize the Lagrange method. The Lagrange function now reads:

$$\mathcal{L} = \frac{1}{2}\mathbf{w}^{\intercal}\cdot\boldsymbol{\Sigma}\cdot\mathbf{w} - \lambda(\mathbf{w}^{\intercal}\cdot\mathbf{1} - 1). \tag{13}$$

The first order conditions are given by the system of linear equations:

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\mathbf{w}} &= \boldsymbol{\Sigma}\cdot\mathbf{w} - \lambda\mathbf{1} = 0, \\ \frac{\partial\mathcal{L}}{\partial\lambda} &= \mathbf{w}^{\intercal}\cdot\mathbf{1} - 1 = 0. \end{aligned} \tag{14}$$

The solution is given by:

$$\mathbf{w}_{\mathbf{g}} = \frac{1}{c}\,\boldsymbol{\Sigma}^{-1}\cdot\mathbf{1} = \frac{\boldsymbol{\Sigma}^{-1}\cdot\mathbf{1}}{\mathbf{1}^{\intercal}\cdot\boldsymbol{\Sigma}^{-1}\cdot\mathbf{1}}. \tag{15}$$

Finally, the corresponding expected return is given by $\mu_g := \mathbf{w_g}^{\intercal}\cdot\mathbf{m}$. With this information, we are fully equipped to state an important results in the modern portfolio theory.

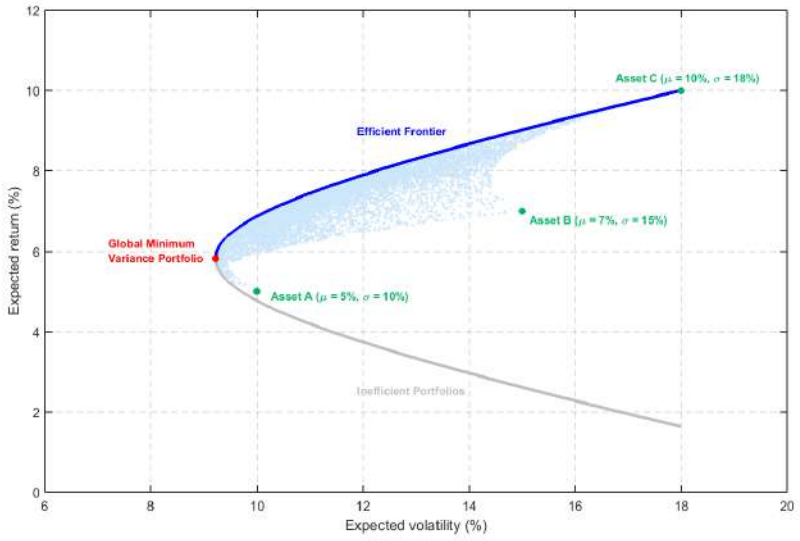


Figure 1: **Unconstrained mean–variance optimization with risk assets.** An example of an unconstrained Markowitz portfolio optimization for a hypothetical case of three risky assets (labeled A, B, and C) with expected return and volatility characteristics (μ, σ): $(5\%, 10\%)$, $(7\%, 15\%)$, and $(10\%, 18\%)$, respectively. The following correlation structure is assumed: $\rho_{AB} = 0.3$, $\rho_{BC} = 0.6$, and $\rho_{AC} = 0.2$. In the μ–σ plane the risky assets are represented by green circles. The global minimum variance portfolio (μ_g, σ_g) is the leftmost point on the mean–variance parabola (i.e., the red circle). The dispersed cloud of light-blue dots represents a set of 10,000 simulated feasible asset mixes. The envelope parabola is the solution to the mean–variance optimization problem for these three risky assets. The upper (solid blue) line represent the efficient frontier. The gray solid line on the lower side of the graph are the inefficient portfolios. For a given level of expected volatility, the asset mix located on the upper (efficient) frontier has a superior expected return compared to the corresponding asset mix on the lower (inefficient) frontier.

Theorem 1 *Portfolios in* (11) *are efficient if and only if* $\mu \geq \mu_g$.

This implies, in particular, that if global minimum variance portfolio has the expected return greater than zero (which, in the past was a normal state of affairs) portfolio $\mathbf{g}$ in (11) would be inefficient since its expected return is zero.

Markowitz optimization is one of the most frequently used asset allocation techniques. Yet, there are some important practical considerations to bear in mind. It turns out that in practice it is much more difficult to precisely forecast expected returns than covariance matrix. In addition, standard Markowitz optimization solutions are not robust with respect to changes in expected return estimation (see [4]). In this respect, using the global minimum variance portfolio can help. Namely, to calculate asset weights for this portfolio does not require estimation of the means. In fact,

this is the only efficient portfolio for which this is the case. As a consequence, global minimum variance portfolio plays an increasingly important role in investment practice.

2.3. Some generalizations of the Markowitz problem

2.3.1. Uncertain investment horizon

Thus far we have considered a problem where investments are made for a fixed and known time horizon (this was implicit in the formulation). Suppose we are mean–variance optimizers but are not sure when we should liquidate our portfolio. Should this exit-time uncertainty worry us, i.e., should we be making different investment decisions with and without such uncertainty? This problem was analyzed in [3].

The authors consider two cases. The first is the so-called independent exit time. There, an investor can exit at a random time but the time of exit does not depend on the portfolio performance. For example, you might sell off your portfolio because you need money to make the down payment on a house or to pay a medical bill. In that case, one can show that under rather general circumstances the optimal portfolio construction can be done by ignoring the exit time uncertainty. In particular, portfolios which are optimal under fixed exit time are optimal under independent stochastic exit time and *vice versa*.

A very different situation occurs when portfolio exit is dictated by portfolio performance. For example, suppose an investor has a simple stop–loss strategy where she liquidates her portfolio after one year if its value at that time falls by more than, e.g., 15 percent of the original level. Otherwise the investor keeps it for another year and then sells her positions. The authors demonstrate that in that case optimal portfolio heavily depends on the structure of the exit time, i.e., on the exit-time policy. Importantly, portfolios efficient under fixed exit time could be inefficient in case of a stop–loss strategy and *vice versa*.

2.3.2. Portfolio optimization constraints

Another class of generalizations of the Markowitz problem consists of imposing additional constraints (see [4]). A rather common situation in practice is the so-called prohibition of short sales. It refers to the requirement that all portfolio weights are non-negative, i.e., that $\mathbf{w} \geq 0$. In some countries such as Serbia, this constraint is imposed on all investors by regulators. Investors are not allowed to borrow an asset from their broker and sell it—thus acquiring a negative amount of a particular asset—only to repurchase it back later to return it to the broker. This is meant by "selling a security short". Even in countries where short selling is not generally prohibited (e.g., in the US), some types of institutions are prohibited in engaging in it (e.g., investment and pension funds).

Given that the total weights need to sum up to one, no-short-selling

constraint effectively requires that each portfolio weight satisfies inequality

$$0 \leq \mathbf{w} \leq 1. \tag{16}$$

Sometimes regulators or principals can impose further restrictions on the feasible set of portfolio weights. For example, there could be a lower and/or an upper limit of, e.g., $\underline{\alpha}_i$ and $\bar{\alpha}_i$ percent, respectively, for investment in particular asset i. In that case, the individual asset constraints are $\underline{\alpha}_i \leq w_i \leq \bar{\alpha}_i$, where $\underline{\alpha}_i, \bar{\alpha}_i \in [0, 1]$. We illustrate the effects of such constrained on an efficient frontier through an example in Figure 2. Similarly to the case of unconstrained optimizations, we denote these fractions by w_i for $i \in \{1, 2, \ldots, N\}$, where $N \in \mathbb{N}$ is the number of assets in the admissible investment universe. Alternatively, there could be restrictions on asset classes or investment styles and strategies (say, no more than 10 percent investment in stocks belonging to the financial sector). While all these constraints make closed-form solution infeasible, it is still a typical quadratic optimization problem which is numerically easy to perform by contemporary solvers. A somewhat more challenging issue is to impose the so-called round lot constraints. In practice, investors do not purchase a fractional number of shares (this would be, generally speaking, implied by the optimization problems we have considered thus far). Instead, they purchase shares in round lots, i.e., in groups of 100 shares. An interesting practical problem is how to perform portfolio optimization where the output of the optimization is a certain number of round lots for each stock in which we want to invest (this can be combined by previously mentioned constraints). The solution involves nonlinear integer programming (see [4]).

2.3.3. Alternative risk measures

Thus far we have considered as portfolio optimization objective to minimize variance or, equivalently, standard deviation of portfolio returns. One problem with an objective function based on portfolio variance is that it is symmetric, i.e., upside and downside returns (relative to the expected return) are equally contributing to market risk. On the other hand, investors typically see no problem with making more money than expected and consider as risk only the downside risk, i.e., making less money than expected. There are many measures of downside risk we can consider. Markowitz [2] himself mentions, for example, semi-variance. This measure is calculated similarly to the portfolio variance, however it takes into the account only returns below the mean.

Other types of downside risk measures that can be used in portfolio optimization are so-called quantile measures such as value at risk (VaR) and conditional value at risk (CVaR). For a given confidence level $1 - \alpha$ where α is a small fraction (e.g., 5 percent), a one-day VaR is defined as the loss that would be exceeded on any given day with probability of not more than α. To illustrate this point, supposed we have calculated that the 5-percent VaR over one-day investment horizon is 7 percent. This means

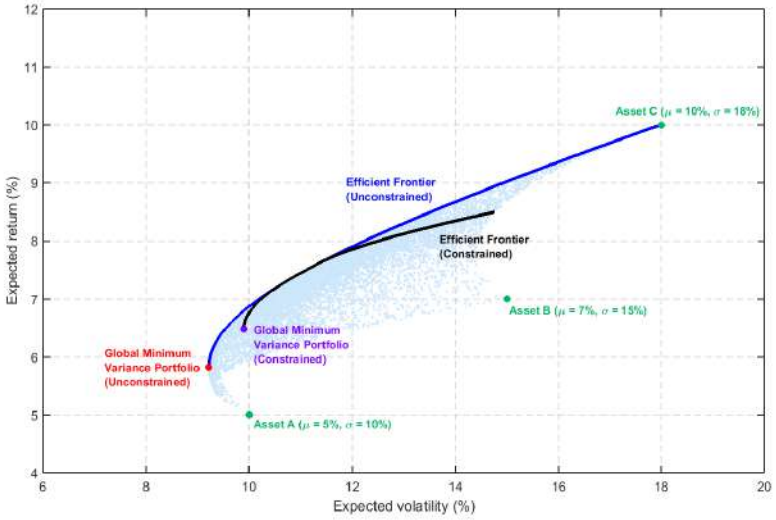


Figure 2: **Constrained mean–variance optimization with risky assets.** An example of a constrained Markowitz portfolio optimization for a hypothetical case of three risky assets introduced in Figure 1. Unconstrained portfolio optimization results presented above are included as a benchmark. For consistency, we follow the same notation and illustration conventions. The constraints applied in this example are: (a) no short selling is allowed, and (b) none of the asset classes is allowed to have portfolio weight greater than 50%. For the constrained optimization, the global minimum variance portfolio (efficient frontier) is represented by a purple circle (solid black line). The key takeaway is that portfolio constraints have a significant impact on the efficient frontier, by shrinking and bending it towards less attractive risk–return ratios. This is expected, due to a constrained (i.e., more limited) investment opportunity set.

that, for our portfolio, realizations leading to losses of more than 7 percent would occur with probability of 5 percent. In other words, in a sample of 100 consecutive trading days we can expect to see around 5 times losses higher than 7 percent. CVaR answers a natural follow-up question: How big of a loss we can expect to have in a given trading day provided that the loss is higher than VaR.

From the asset management point of view a natural generalization of the Markowitz problem is to maximize the expected portfolio return and, at the same time, try to minimize a quantile risk measure such as VaR or CVaR. In [5], [6], [7] the authors develop a way to map the problem of mean-CVaR optimization into an equivalent linear programming problem. As a by-product, their approach provides near-optimal mean-VaR portfolios when VaR is measured using the so-called historical simulation technique, i.e., estimating the corresponding quantile of the empirical portfolio return distribution. There are more accurate VaR estimation methods than the historical VaR. One very frequently used model of VaR estimation is a

univariate VaR with conditional Student's t-distribution and volatility estimation based on generalized autoregressive conditional heteroskedasticity model, e.g., GARCH(1,1). Having higher VaR estimation accuracy comes at a price if we want to find mean-VaR optimal portfolios. This problem cannot be solved using linear programming or standard convex optimization methods. Instead, [8] employ a evolutionary optimization algorithm to address that problem.

2.3.4. Regulatory perspective

Another important generalization stems from financial regulation requirements. Namely, regulated financial institutions such as banks and insurance companies need to keep certain level of capital called regulatory capital. The level of this safety buffer depends on the portfolio of securities that the institution holds. This invites a question: How to optimal trade off expected return on investment portfolio and the corresponding cost of regulatory capital. In case of Basel 2.5, which provides a framework for market risk management in the banking sector, the pioneering work on this topic was done in [9]. Again, an evolutionary optimization is employed in order to construct mean-regulatory capital efficient frontier.

2.4. Efficient frontier in the presence of a risk-free asset

So far we have considered only portfolio optimization using risky assets. Another type of generalization is to stay within the confines of the standard Markowitz setup but add another type of asset into consideration. Namely, suppose that in addition to risky assets (stocks) there is, also, a riskless asset. It has expected return equal to r_f and variance equal to zero. A portfolio consisting of hundred percent invested in riskless asset is an efficient portfolio. This is now the new portfolio with global minimum risk. Consider efficient portfolio consisting only of risky assets when a riskless asset is available. It it obtained by maximizing the so-called Sharpe ratio:

$$\begin{aligned} &\max_{\mathbf{w}} \ \frac{\mu_p - r_f}{\sigma_p}, \\ &\text{s.t. } \mathbf{w}^\intercal \cdot \mathbf{1} = 1. \end{aligned} \tag{17}$$

Sharpe ratio tells us how much premium investors demand with respect to the riskless rate, for each unit of additional risk they are willing to take. The optimization problem (17) simply states that, in the presence of a riskless asset, risky portfolio is optimal if its Sharpe ratio is maximal. Let us denote the standard deviation of this portfolio solving (17) with σ_s and its expected return with μ_s. Then, in a risk–return space, in the presence of a riskless asset two efficient portfolios are $(0, r_r)$ and (σ_s, μ_s). The first one consists only of the riskless asset, while the second one comprises only risky assets. All other efficient portfolios can in this case be obtained as a linear combinations of these two efficient portfolios.

Consider a portfolio which has a fraction of money α invested into the risky portfolio r_s and $1-\alpha$ invested into the riskless asset r_f. Portfolio return, expected return and variance would in this case, respectively, be equal to:

$$\begin{aligned} r_p &= r_f + \alpha(r_s - r_f), \\ \mu_p &= r_f + \alpha(\mu_s - r_f), \\ \sigma_p &= \alpha\,\sigma_s. \end{aligned} \tag{18}$$

From this follows that the fraction of money invested in risky asset is simply the ratio between portfolio volatility and volatility of the maximum Sharpe ratio portfolio, i.e., $\alpha = \sigma_p/\sigma_s$. If we plug this result back into the second equation above, we obtain that

$$\mu_p = r_f + \frac{\mu_s - r_f}{\sigma_s}\,\sigma_p. \tag{19}$$

Thus, the risk premium on an optimal portfolio is equal to the maximum Sharpe ratio times the risk of the portfolio. If we assume that all investors have identical preferences and evaluate risks and returns in the same way (a tall order, admittedly) and if we, furthermore, assume that all investors invest in the entire universe of risky assets, then portfolio s is the market portfolio. It is a portfolio of risky assets weighted by their relative market capitalization. In that case we denote $r_s = r_m$ and the above equation is called the Capital Market Line (CML),

$$\mu_p = r_f + \frac{\mu_m - r_f}{\sigma_m}\,\sigma_p. \tag{20}$$

In summary, if Markowitz optimization is a reasonable, rational thing for all investors to do and if all of them have the same information and think alike, one cannot do better than invest in a portfolio in which a fraction of the money is invested into the riskless asset and the rest into the market portfolio. In that case, all investments lie either on the CML or below it but cannot be above it by the very construction. The reason for this lies in the fact that the CML, by definition, provides the best possible trade-off between risk and return. Under these conditions passive investment—following a market index instead of attempting to beat the market—is the most sensible strategy. An example is provided in Figure 3.

3. Expected utility optimization in discrete time

Consider two investments in the presence of uncertainty. Today is time t. Investment A promises to pay 5 euros in both states of the world at time $t+1$, while investment B promises to pay 9 euros if state 1 materializes and 1 euros if state 2 is realized at time $t+1$. Which of the two investments

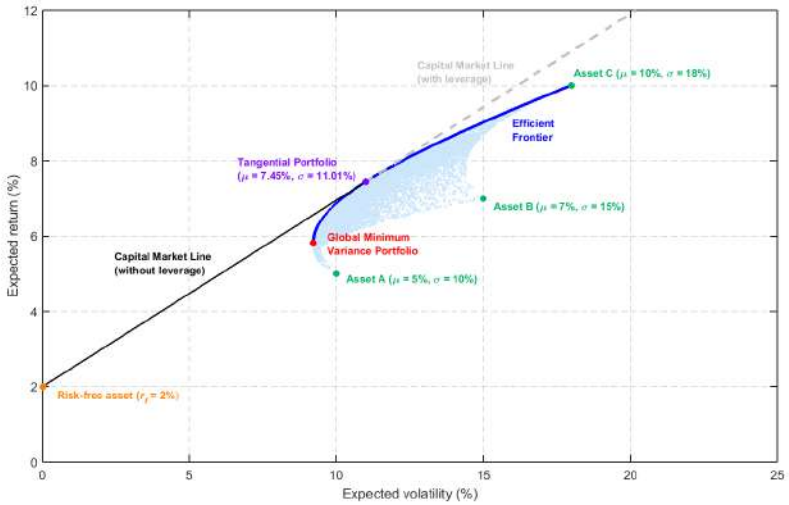


Figure 3: **Mean–variance optimization with risk-free asset and risky assets.** An example of a Markowitz portfolio optimization for a hypothetical case of three risky assets introduced in Figure 1 and Figure 2. However, a risk-free asset ($r_f = 2\%$) is also included in this case. We follow the same notation and illustration conventions as earlier. The only constraint imposed is that the asset weights add up to one. The risk-free asset (tangential portfolio) is represented by the orange (purple) circle. The capital market line connects the risk-free asset and the tangential portfolio (solid black line). Along this line segment, the Sharpe ratio is maximized. If we allow for leverage, an investor could optimally invest also along the capital market line extension beyon the tangential portfolio (dashed gray ray).

should we rationality choose? Suppose, for simplicity, that both states can be realized with equal probability.

If we are risk neutral (i.e., if we do not demand risk premium for taking risk), both investments are worth the same to us. Namely, in that case we would evaluate them using the expected value of each investment:

$$\mathbb{E}\,(c_A) = \frac{1}{2} \times 5 + \frac{1}{2} \times 5 = 5,$$
$$\mathbb{E}\,(c_B) = \frac{1}{2} \times 9 + \frac{1}{2} \times 1 = 5.$$

On the other hand, while both investments have the same expected value, the second one is risky while the first one has no risk at all. Thus, a risk-averse investor would prefer the first investment to the second one. Daniel Bernoulli proposed using the expected utility as the solution to valuing risky investments (he actually studied gambling). This theory was later formalized by von Neumann and Morgenstern [10].

According to the expected utility theory, when comparing investments under uncertainty instead of the expected value we should use the expected

utility, where utility is represented by an increasing and concave function. Suppose that utility function is $u(x) = \sqrt{x}$. In that case:

$$\begin{aligned}\mathbb{E}\left(u(c_A)\right) &= \frac{1}{2} \times \sqrt{5} + \frac{1}{2} \times \sqrt{5} = \sqrt{5},\\ \mathbb{E}\left(u(c_B)\right) &= \frac{1}{2} \times \sqrt{9} + \frac{1}{2} \times 1 = 2.\end{aligned}$$

Obviously, $\mathbb{E}\left(u(c_A)\right) > \mathbb{E}\left(u(c_B)\right)$. Suppose that w is an investor's wealth. It is convenient to introduce the notion of absolute and relative risk aversion as follows:

$$\begin{aligned}A(w) &:= -\frac{u''(w)}{u'(w)},\\ R(w) &:= wA(w).\end{aligned} \tag{21}$$

Both of these functions are related to the degree of concavity of the utility function. A more concave utility function corresponds to a more risk averse investor. In the case of an investor with linear utility function, i.e., $u(w) = a + bw$, both the absolute and relative risk aversion coefficients are equal to zero. Such an investor is risk neutral. On the other hand, when the utility function is given by the natural logarithm, i.e., $u(w) = \log(w)$, it is easy to see that the absolute risk aversion decreases in wealth, while the relative risk aversion is a constant:

$$\begin{aligned}A(w) &\equiv -\frac{u''(w)}{u'(w)} = \frac{1}{w},\\ R(w) &\equiv -w\frac{u''(w)}{u'(w)} = 1.\end{aligned} \tag{22}$$

The situation is qualitatively the same in case of a power utility function $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$, where $0 < \gamma < 1$:

$$\begin{aligned}A(w) &= \frac{\gamma}{w},\\ R(w) &= \gamma.\end{aligned} \tag{23}$$

Thus, logarithmic and power utility belong to the class of constant relative risk aversion (CRRA) utility functions. In order to better understand the meaning of absolute and relative risk aversion, let us state here the following important result:

Theorem 2 *Suppose the investment opportunity set consists of a riskless asset and a risky portfolio. An investor for whom $A(w)$ is a decreasing function would invest more money (in a hard currency, e.g., euros or dollars) into the risky portfolio vis-a-vis the riskless asset the wealthier she*

gets. The opposite is true for an investor for whom $A(w)$ is an increasing function of wealth. Finally, if $A(w)$ does not depend on wealth, investor would keep the same amount of money invested into the risky portfolio no matter how wealthy she gets.

What does this mean? Suppose $A(w)$ is a decreasing function of wealth such as log utility. In that case, when we are dirt-poor we optimally invest smaller amount of money into the risky assets than would be the case if we were wealthier. This intuitively makes sense. On the other hand, if $A(w)$ is constant, then we would optimally invest the same amount of money into stocks whether we have ten or million euros at our disposal. This is not reasonable, of course. Even worse is the situation when $A(w)$ increases in wealth. In case of $R(w)$ the following statement is true:

Theorem 3 *Suppose that the investment opportunity set consists of a riskless asset and a risky portfolio. An investor for whom $R(w)$ is a decreasing function of wealth would invest a higher fraction of her money into the risky portfolio vis-a-vis the riskless asset the wealthier she gets. The opposite would be true for an investor for whom $R(w)$ is an increasing function of wealth. Finally, if $R(w)$ does not depend on wealth, investor would keep the same fraction of money invested into the risky portfolio no matter how wealthy she gets.*

These theorems are called Fundamental Theorems of Portfolio Management. Note that this implies that a log or power utility investor would optimally keep the same fraction of money invested into the risky portfolio no matter how wealthy she is.

It turns out that the Markowitz optimization problem can be rationalized (i.e., presented as a result of an expected utility optimization) for any concave utility function provided that asset returns are normal or for any asset return distribution with finite first and second moments provided that the utility function is quadratic, i.e., it has the form:

$$u(w) = w - \frac{a}{2}w^2,$$

where $w < 1/a$ (see [11]). Unfortunately, these conditions do not bode well for the rationalization of the Markowitz approach. Namely, asset returns are well-known to pose tails fatter than normal (kurtosis in excess of 10 is not uncommon). On the other hand, quadratic utility specified above implies that both $A(w)$ and $R(w)$ are increasing in wealth, which is not realistic. Nevertheless, Markowitz approach and its various generalizations are dominant in asset management practice.

In the rest of the article we consider portfolio optimization via expected utility maximization.

3.1. Single-period expected utility optimization

Suppose there are two assets at our disposal: a risky asset (which can be a stock or a portfolio of stocks) and a riskless asset. Price of the risky asset at time t is denoted as s_t. For simplicity, single-period gross return on the risky asset has two possible outcomes:

$$\frac{s_{t+1}}{s_t} = \begin{cases} u & \text{with probability } p, \\ d & \text{with probability } 1-p. \end{cases} \tag{24}$$

The gross return on riskless asset is $R = 1+r_f$. Importantly, the parameters of the system satisfy the following inequalities $d < R < u$. This condition ensures that risky asset (let us call it a stock) performs better than the riskless asset in the good state and worse than the riskless asset in the bad state of the world. In the absence of this condition market would not support the existence of one of these securities. Namely, if $R < d < u$ it would never pay to invest in the riskless asset. The opposite is true in case when $d < u < R$.

This simple model is called the binomial model of asset prices. Let the investor's wealth at time t be x_t. Suppose, further, that the investor has log utility. The question is: given her initial wealth and attitude towards risk, what is the optimal fraction of money she would invest into the risky asset? Let us denote that fraction as w_t. Given wealth at time t and the investment strategy w_t, portfolio wealth at time $t+1$ is given by:

$$x_{t+1} = \begin{cases} x_t(R + w_t(u-R)) & \text{with probability } p, \\ x_t(R + w_t(d-R)) & \text{with probability } 1-p. \end{cases} \tag{25}$$

This expression is intuitive. Namely, portfolio wealth would grow by the factor of R provided that all our money is invested into the riskless asset. The difference (positive or negative, depending on whether the upper or the lower state is realized at time $t+1$) is determined by the fraction of wealth invested into the risky asset times how much extra return we make by making such investment. Let excess returns on stock be $\tilde{u} = u - R$, $\tilde{d} = d - R$ and probability of down move $q = 1 - p$. Expected utility at time t, then, reads:

$$\begin{aligned} \mathbb{E}_t\left(u(x_{t+1})\right) &:= \mathbb{E}_t\left(\log(x_{t+1})\right) \\ &= \log(x_t) + p\log(R + w_t\tilde{u}) + q\log(R + w_t\tilde{d}). \end{aligned} \tag{26}$$

Next we find the first order condition with respect to w_t and set it equal to zero:

$$\frac{p\,\tilde{u}}{R + w_t\tilde{u}} + \frac{q\,\tilde{d}}{R + w_t\tilde{d}} = 0.$$

Note that the initial wealth cancels out. Thus, the optimal fraction invested into the risky asset should not depend on accumulated wealth at time t.

This is consistent with the result stated above since $R(w) = 1$ is a constant in case of the log utility.

The first order equation is linear in w_t. Its solution reads:

$$\begin{aligned} \hat{w}_t &= -\frac{R}{\tilde{u}\,\tilde{d}}\,(p\tilde{u} + q\tilde{d}), \\ \mathbb{E}_t\left(u(\hat{x}_{t+1})\right) &= \log(x_t) + p\log(R + \hat{w}_t\tilde{u}) + q\log(R + \hat{w}_t\tilde{d}) \\ &=: \log(x_t) + k. \end{aligned} \tag{27}$$

Note that the term in bracket in the expression for the optimal weight is expected excess return on the stock with respect to the riskless asset. Given our definitions, $\tilde{d} < 0$ and $\tilde{u} > 0$. Thus, the sign of the optimal fraction invested into the risky asset is positive provided that a positive risk premium on that asset is paid and vice versa. This is intuitive, of course. We see, also, that the optimal expected utility function of the portfolio wealth at time $t + 1$, with expectation calculated at time t, is equal to the utility function of wealth at time t plus a constant term. This insight is important when we want to solve a multi-period problem.

3.2. Dynamic portfolio optimization in discrete time

Now we consider dynamic portfolio optimization problem which generalizes the example from the previous subsection. Again, we have a stock and a riskless asset. While now there are T trading periods, all the rest stays the same. Namely, risky asset in each period evolves according to (24), investor has log utility, and per-period gross return on riskless asset is R.

We solve the problem backwards in time, using Dynamic Programming technique. Suppose we are at time $t = T - 1$ and want to maximize the expected utility of the final wealth. We assume that the wealth at that time is x_{T-1} and that it is known. Thus, we maximize:

$$\begin{aligned} \mathbb{E}_{T-1}\left(u(x_T)\right) &:= \mathbb{E}_{T-1}\left(\log(x_T)\right) \\ &= \log(x_{T-1}) + p\log(R + w_{T-1}\tilde{u}) + q\log(R + w_{T-1}\tilde{d}). \end{aligned} \tag{28}$$

with respect to w_{T-1}. Clearly, this problem does not differ from (26) except that we need to identify $x_t \to x_{T-1}$ and $w_t \to w_{T-1}$. Using the same procedure as before, we obtain that:

$$\begin{aligned} \hat{w}_{T-1} &= -\frac{R}{\tilde{u}\,\tilde{d}}\,(p\tilde{u} + q\tilde{d}), \\ \mathbb{E}_{T-1}\left(u(\hat{x}_T)\right) &= \log(x_{T-1}) + p\log(R + \hat{w}_{T-1}\tilde{u}) + q\log(R + \hat{w}_{T-1}\tilde{d}) \\ &=: \log(x_{T-1}) + k =: V(x_{T-1}). \end{aligned} \tag{29}$$

In the last step we have introduced the value function, i.e., optimal value of the expected utility calculated at time $t = T-1$. It depends on the state

variable which is, in this case, portfolio wealth at time $t = T - 1$. Next we can go one step backwards in time, to $t = T - 2$. In that case we need to maximize:

$$\begin{aligned}\mathbb{E}_{T-2}\left(V(x_T)\right) &:= \mathbb{E}_{T-2}\left(\log(x_{T-1})\right) + k \\ &= \log(x_{T-2}) + p\log(R + w_{T-2}\tilde{u}) + q\log(R + w_{T-2}\tilde{d}) + k.\end{aligned} \tag{30}$$

Solving for the optimal fraction of wealth and plugging that expression into the objective function we obtain that:

$$\begin{aligned}\hat{w}_{T-2} &= -\frac{R}{\tilde{u}\,\tilde{d}}\left(p\tilde{u} + q\tilde{d}\right) = \hat{w}_{T-1} \equiv \hat{w}, \\ \mathbb{E}_{T-2}\left(u(\hat{x}_{T-1})\right) &= \log(x_{T-2}) + p\log(R + \hat{w}\tilde{u}) + q\log(R + \hat{w}\tilde{d}) + k \\ &= \log(x_{T-2}) + 2\,k =: V(x_{T-2}).\end{aligned} \tag{31}$$

It is easy to check, continuing backwards in time, that the optimal holdings and the value function at time t are given by:

$$\begin{aligned}\hat{w}_t &= -\frac{R}{\tilde{u}\,\tilde{d}}\left(p\tilde{u} + q\tilde{d}\right) =: \hat{w} \\ V(x_t) &= \log(x_t) + k\,(T - t), \\ \text{where } k &= p\log(R + \hat{w}\tilde{u}) + q\log(R + \hat{w}\tilde{d}).\end{aligned} \tag{32}$$

In summary, when an investor has a log utility function, in a stationary economy (i.e., when parameters of the model do not change over time), her optimal fraction invested into the risky asset would stay constant. The value function at time t is equal to the utility function at that time plus a linear function of time. When $t = T$ the value function becomes is equal to the utility function.

4. Portfolio optimization in continuous time

In this section we briefly consider portfolio optimization problem when both risky and riskless asset are traded continuously in time.

4.1. Stochastic processes in continuous time

Consider an economy in which there exist a riskless and a risky asset which can be traded continuously in time. For simplicity, we assume that there are no transaction costs. Riskless asset has continuously compounded interest rate r. Thus, return on the riskless asset in the period $(t, t + dt)$ is given by the following ordinary differential equation:

$$\begin{aligned}\frac{db(t)}{b(t)} &= r\,dt, \\ b(0) &= b_0.\end{aligned} \tag{33}$$

Since there is no risk, this is just an ordinary differential equation. Its solution is $b(t) = b_0 e^{rt}$. In other words, the riskless account grows exponentially in time due to continuously paid interest on interest. The rate of growth is the interest rate r.

Consider now return on a stock in the same infinitesimal interval of time. We model it using Geometric Brownian Motion (GBM) process, as follows:

$$\begin{aligned} &\frac{ds(t)}{s(t)} = \mu\, dt + \sigma dz(t), \\ &s(0) = s_0. \end{aligned} \tag{34}$$

The first part on the right hand side is a deterministic drift term. The second part is stochastic. Namely, $z(t)$ is assumed to be a Wiener process, which satisfies the following properties:

(a) $z(t)$ is a continuous function of time t,

(b) Increments $\Delta z(t) = z(t + \Delta t) - z(t)$ are independent for different t and $\Delta t > 0$,

(c) $z(t)$ are normally distributed with mean 0 and variance t,

(d) $z(0) = 0$.

While sample paths for the Wiener process are continuous, $z(t)$ is not differentiable. This can be seen, e.g., by considering the following limit when $\Delta t \to 0$:

$$\mathbb{V}\left(\frac{z(t+\Delta t) - z(t)}{\Delta t}\right) = \frac{\Delta t}{\Delta t^2} = \frac{1}{\Delta t} \to \infty.$$

Since this is true for any t, $z(t)$ is a nowhere differentiable function. In that sense, differential equation (34) cannot be properly defined using the standard calculus. To tackle this process (and others like it) we need stochastic calculus instead. One version of it is the Itô calculus. Formally, stochastic processes of this type are defined in integral form *only*:

$$s(t) = s_0 + \int_0^t \mu s(u)\, du + \int_0^t \sigma s(u)\, dz(u).$$

Here, the first integral is the standard Riemann integral (the integrand is a continuous function of time) while the second integral is the stochastic or Itô integral (see [12]). Instead of using the integral form, it is both customary and functionally equivalent to use the differential form (34) but to redefine what we mean by a differential. Namely, in calculating differentials—often called Itô differentials—of a function of both t and z, we need to calculate

the second order differential and not the first order one as is customary in regular calculus. In addition, the following mnemonic rules are imposed:

$$\begin{aligned} (dz)^2 &= dt, \\ dz\,dt &= 0, \\ (dt)^2 &= 0. \end{aligned} \tag{35}$$

Thus, we drop from the second order differential terms proportional to $dz\,dt$ and $(dt)^2$ but keep the terms proportional to $(dz)^2$ and replace that construct with $(dz)^2 \to dt$. Note that $\mathbb{E}\left[(dz)^2\right] = \mathbb{V}(dz) = dt$. While this does not constitute the proof of the above rule, it demonstrates its plausibility. The above stochastic differentiation rules are often referred to as Itô's Lemma. Of course, as stressed earlier, Itô's Lemma is properly formulated in an integral form.

Let us now calculate Itô differential of the following expression:

$$s(t) = s_0\, e^{\left(\mu - \frac{\sigma^2}{2}\right)t}\, e^{\sigma z(t)}. \tag{36}$$

We find first order partial derivatives of $s(t)$ with respect to t and z as well as the second order partial derivative with respect to z:

$$\begin{aligned} \frac{\partial s}{\partial t} &= \left(\mu - \frac{\sigma^2}{2}\right) s(t), \\ \frac{\partial s}{\partial z} &= \sigma s(t), \\ \frac{\partial^2 s}{\partial z^2} &= \sigma^2 s(t). \end{aligned} \tag{37}$$

Therefore, the second order differential of $s(t)$ is given by:

$$ds = \frac{\partial s}{\partial t}\,dt + \frac{\partial s}{\partial z}\,dz + \frac{1}{2}\frac{\partial^2 s}{\partial z^2}\,(dz)^2 + \cdots.$$

The omitted terms denoted with dots are those second-order terms proportional to $dz\,dt$ and $(dt)^2$ which we have agreed to drop. Using the first rule in (35) we obtain the following:

$$\begin{aligned} ds &= \frac{\partial s}{\partial t}\,dt + \frac{\partial s}{\partial z}\,dz + \frac{1}{2}\frac{\partial^2 s}{\partial z^2} dt \\ &= \left(\mu - \frac{\sigma^2}{2}\right) s(t)\,dt + \sigma s(t)\,dz(t) + \frac{1}{2}\sigma^2 s(t)\,dt \\ &= \mu s(t)\,dt + \sigma s(t)\,dz(t). \end{aligned} \tag{38}$$

Therefore, we have demonstrated that the expression (36) is the solution to the stochastic differential equation (34).

One can generalize this result by introducing an Itô process $\{x(t),\, t \geq 0\}$ in the form:

$$x(t) = x_0 + \int_0^t \mu(u, x(u))\, du + \int_0^t \sigma(u, x(u))\, dz(u). \tag{39}$$

Here, functions $\mu(u, x(u))$ and $\sigma(u, x(u))$ represent drift and diffusion, respectively. Generally, they depend on the previous realizations of the process $\{x(t),\, t \geq 0\}$. That means that these values are considered known when $dz(u)$ is drawn. Such processes are called adapted processes with respect to the filtration generated by the Wiener process $\{z(t),\, t \geq 0\}$. GBM is a special case of the Itô process with drift and diffusion terms which are linear in x. Wiener process itself is a special case of the Itô process with vanishing drift and diffusion equal to 1. In the absence of a drift, for any adapted process $\sigma(u, x(u))$, $x(\cdot)$ satisfies:

$$\begin{aligned} x(t) &= x_0 + \int_0^t \sigma(u, x(u))\, dz(u), \\ \mathbb{E}_0\left(x(t)\right) &= x_0 + \mathbb{E}_0\left(\int_0^t \sigma(u, x(u))\, dz(u)\right) = x_0, \\ \mathbb{V}_0\left(x(t)\right) &= \int_0^t \sigma^2(u, x(u))\, du. \end{aligned} \tag{40}$$

In the second line we use the fact that σ is adapted process and the expectation of future Wiener increments is zero. As a result, the expectation at time zero of a driftless Itô process is equal to the initial value of the process. More generally, it is easy to see that driftless Itô process is a martingale. That means that the best prediction of its value at time $t + \tau$ is its value at time t (for any t and $\tau > 0$):

$$\begin{aligned} \mathbb{E}_t(x(t+\tau)) &= x_0 + \mathbb{E}_t\left(\int_0^t \sigma(u, x(u))\, dz(u)\right) \\ &\quad + \mathbb{E}_t\left(\int_t^{t+\tau} \sigma(u, x(u))\, dz(u)\right) \\ &= x_0 + \int_0^t \sigma(u, x(u))\, dz(u) + 0 = x(t). \end{aligned} \tag{41}$$

When the expectation in the first line is calculated, we know all of the realizations of the Itô process up to time t. Thus, the expectation is equal to the realization of the process itself. Expectation in the second line is zero for the same reasons as before. Finally, consider the third equation in (40). This is so-called isometric property of the Itô integral. It stems from the fact that Wiener increments have zero mean and variance equal to dt and that they are independent (see [12]).

In order to complete this section let us briefly consider differential form of the general Itô process defined as

$$dx = \mu(t, x(t))\,dt + \sigma(t, x(t))\,dz.$$

Itô's Lemma would now amount to the following mnemonic rules:

$$\begin{aligned} (dx)^2 &= \sigma(t, x(t))^2\,dt, \\ dx\,dt &= 0, \\ (dt)^2 &= 0. \end{aligned} \tag{42}$$

In the next subsection we explain how one can determine an optimal portfolio strategy in continuous time. Our approach follows the pioneering work of Robert Merton (see articles in [13]).

4.2. Hamilton–Jacobi–Bellman (HJB) equation

Consider a portfolio consisting of a riskless asset $b(t)$ and a risky asset $s(t)$ whose dynamics is specified by (33) and (34), respectively. We are interested in the evolution of the portfolio wealth process assuming that wealth generated during the investment process is fully reinvested. Let $w(t)$ be the fraction of money invested into the risky asset at time t. In that case, return on portfolio wealth in time interval $(t, t + dt)$ is given by the following expression:

$$\begin{aligned} \frac{dx(t)}{x} &= (1 - w(t))\frac{db(t)}{b} + w(t)\frac{ds(t)}{s} \\ &= (1 - w(t))r\,dt + w(t)(\mu\,dt + \sigma\,dz(t)) \\ &= r dt + w(t)(\mu - r)\,dt + w(t)\sigma\,dz(t), \\ x(0) &= x_0. \end{aligned}$$

Using Itô's calculus it is easy to check that:

$$\begin{aligned} dx =&\ rx(t)\,dt + w(t)(\mu - r)x(t)\,dt + w(t)\sigma x(t)\,dz(t), \\ (dx)^2 =&\ w(t)^2\sigma^2 x(t)^2\,dt, \\ dx\,dt =&\ 0. \end{aligned} \tag{43}$$

The wealth process (43) is a GBM with variable coefficients. This is the result of the assumption that s follows a GBM and that b is deterministic. In integral form, Itô process $\{x(t),\, t \geq 0\}$ can be represented as

$$x(t) = x_0 + \int_0^t (r + w(u)(\mu - r))x(u)\,du + \int_0^t w(u)\sigma x(u)\,dz(u). \tag{44}$$

The wealth process (44) depends on the trading strategy w. Suppose that we have selected an arbitrary (not necessarily optimal) trading strategy

$\{w(t),\, t \geq 0\}$ and denote the corresponding wealth process as $\{x^w(t),\, t \geq 0\}$. Terminal investment time is $t = T$. At that point we consume all of the accumulated wealth. For a given arbitrary trading strategy w utility of terminal wealth is equal to $u(x^w(T))$.

Suppose that wealth at time t is x. Expected utility at time t of the terminal wealth for a given trading strategy $\{w(t),\, t \geq 0\}$ is given by the expression:

$$J(t, x; w) = \mathbb{E}_{t,x}(u(x^w(T))). \tag{45}$$

We assume here that $w(t)$ has a feedback form $w(t, x^w(t))$ for some deterministic function $w(t, x)$ (see [12] for a discussion of this assumption). With this condition in place $J(t, x)$ is the expected value of a function of the diffusion process $x^w(t)$. This enables us to apply Itô's Lemma to function $J(\cdot, \cdot)$. Consider Itô differential (here subscripts denote partial derivatives with respect to the corresponding variables)

$$\begin{aligned} dJ &= J_t dt + J_x dx + \frac{1}{2} J_{xx} (dx)^2 \\ &= J_t dt + J_x (r + w(t)(\mu - r)) x(t) dt + J_x w(t) \sigma x(t) dz(t) \\ &+ \frac{1}{2} J_{xx} x^2(t) w(t)^2 \sigma^2 dt. \end{aligned}$$

which can be, more conveniently, rewritten as

$$\begin{aligned} dJ = &\left(J_t + (r + w(t)(\mu - r))\, J_x + \frac{1}{2} w(t)^2 x(t)^2 \sigma^2 J_{xx} \right) dt \\ &+ w(t) \sigma x(t) J_x \, dz(t). \end{aligned} \tag{46}$$

Let us integrate both sides of (46) from t to T. On the left hand side we obtain:

$$\int_t^T dJ = J(T, x^w(T)) - J(t, x^w(t)) = u(x^w(T)) - J(t, x).$$

Thus, the integral form of (46) reads:

$$\begin{aligned} u(x^w(T)) - J(t, x) = \int_t^T &\Big(J_u + (r + w(u)(\mu - r)) x^w(u) J_x \\ &+ \frac{1}{2} w(u)^2 \sigma^2 x^w(u)^2 J_{xx} \Big) du + \int_t^T w(u) \sigma x^w(u) J_x dz(u). \end{aligned}$$

Taking the expectation of both sides of this expression at time t, from (45) follows that the left hand side yields zero. On the right hand side, Itô integral term also yields zero. As a result, the following identity has to hold for any t:

$$\int_t^T \left(J_u + (r + w(u)(\mu - r)) x^w(u) J_x + \frac{1}{2} w(u)^2 \sigma^2 x^w(u)^2 J_{xx} \right) du = 0.$$

Taking first derivative of both sides with respect to t it follows that

$$J_t + (r + w(t)(\mu - r))x^w(t)J_x + \frac{1}{2}w(t)^2\sigma^2 x^w(t)^2 J_{xx} = 0. \tag{47}$$

This is an identity that holds for an arbitrary admissible trading strategy $\{w(t),\, t \geq 0\}$. Now we formulate the rule for selecting the optimal trading strategy. Define, first, the value function:

$$V(t,x) := \sup_w J(t,x;w).$$

The value function $V(t,x)$ satisfies the so-called Hamilton–Jacobi–Bellman (HJB) equation:

$$\begin{aligned} 0 &= V_t + \sup_w \left((r + w(\mu - r))xV_x + \frac{1}{2}w^2\sigma^2 x^2 V_{xx} \right), \\ V(T,x) &= u(x). \end{aligned} \tag{48}$$

In this equation we treat w as a simple variable with respect to which we optimize. The optimal trading strategy is given by the first order condition for (48):

$$\hat{w} = -\frac{\mu - r}{\sigma^2}\frac{V_x}{xV_{xx}}. \tag{49}$$

Here, it is assumed that V is the solution of the following nonlinear partial differential equation which is obtained when in (48) we plug in the expression for optimal w (49) and introduce the notation for the market price of risk $\theta := \frac{\mu - r}{\sigma}$:

$$\begin{aligned} 0 &= V_t + rxV_x - \frac{\theta^2}{2}\frac{V_x^2}{V_{xx}}, \\ V(T,x) &= u(x). \end{aligned} \tag{50}$$

Thus, optimal portfolio trading strategy needs to satisfy (49) provided that we have previously solved a rather complicated nonlinear partial differential equation (50). At terminal point in time the value function coincides with the utility function $u(x)$. This is the terminal condition. Note that optimal trading strategy is equal to the Sharpe ratio of the risky asset times the reciprocal value of the relative risk aversion of the derived utility of wealth. It should come as no surprise that when utility function belongs to the class of constant relative risk aversion functions, the optimal trading strategy is a constant.

4.3. Examples of portfolio optimization in continuous time

4.3.1. Logarithmic utility

Consider first the case when $u(x) = \log(x)$. Recall that in a discrete-time setting, the value function is additive-separable. Thus, our ansatz takes the following form:

$$V(t,x) = \log(x) + b(t).$$

From the terminal condition we know that $V(T,x) = \log(x)$. This implies that the function $b(t)$ has to satisfy the terminal condition $b(T) = 0$. The shape of $b(t)$ is determined by plugging the proposed solution into the HJB equation (50). Towards that end, the relevant partial derivatives of the value function read:

$$\begin{aligned} V_t &= b'(t), \\ V_x &= \frac{1}{x}, \\ V_{xx} &= -\frac{1}{x^2}. \end{aligned}$$

Substituting these expressions into (50) we obtain that:

$$\begin{aligned} 0 &= V_t + rxV_x - \frac{\theta^2}{2}\frac{V_x^2}{V_{xx}} = b'(t) + r + \frac{\theta^2}{2}, \\ b(T) &= 0. \end{aligned} \tag{51}$$

The solution to this equation is $b(t) = (r + \frac{\theta^2}{2})(T-t)$. The corresponding optimal value function then reads $V(t,x) = \log(x) + (r + \frac{\theta^2}{2})(T-t)$. To obtain the optimal trading strategy $\hat{w}$ we need to substitute the optimal value function into the expression (49). We obtain:

$$\hat{w}(t) = \frac{\mu - r}{\sigma^2}.$$

Thus, in case of an investor with logarithmic utility, optimal fraction of risky asset is equal to the Sharpe ratio of the risky asset. Since this quantity is a constant, it follows that the investor keeps a constant fraction of her money invested in the risky asset.

4.3.2. Power utility

Consider now the case when $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, where $0 < \gamma < 1$ and propose, this time, a multiplicative rather than additive shape of the value function (this can again be motivated by considering the discrete case first):

$$V(t,x) = \frac{b(t)}{1-\gamma}x^{1-\gamma}.$$

In that case, the relevant partial derivatives read:

$$\begin{aligned} V_t &= \frac{b'(t)}{1-\gamma} x^{1-\gamma}, \\ V_x &= b(t) x^{-\gamma}, \\ V_{xx} &= -\gamma b(t) x^{-\gamma-1}. \end{aligned}$$

From this result it follows that:

$$\begin{aligned} 0 &= V_t + rxV_x - \frac{\theta^2}{2}\frac{V_x^2}{V_{xx}} \\ &= \frac{b'(t)}{1-\gamma} x^{1-\gamma} + rb(t)x^{1-\gamma} + \frac{\theta^2}{2\gamma}\frac{x^{-2\gamma}}{x^{-\gamma-1}} b(t), \\ &\frac{b'(t)}{1-\gamma} + \left(r + \frac{\theta^2}{2\gamma}\right) b(t) = 0, \\ &b(T) = 1. \end{aligned} \tag{52}$$

This differential equation can be rewritten as $b'(t) + ab(t) = 0$, with the terminal condition stating that $b(T) = 1$. The solution is given by the expression $b(t) = e^{a(T-t)}$ and, therefore, the optimal value function for a power utility investor reads:

$$V(t,x) = \frac{e^{a(T-t)}}{1-\gamma} x^{1-\gamma}.$$

Finally, the optimal fraction of money invested into risky assets is given by (49):

$$\hat{w}(t) = \frac{\mu - r}{\gamma \sigma^2}.$$

Note that, again, the optimal fraction of money invested into the risky asset does not change over time. It is equal to the product of the Sharpe ratio of the risky asset and the reciprocal value of the relative degree of risk aversion. Since log-utility implies a higher degree of relative risk aversion than power utility with $0 < \gamma < 1$, a rational investor would hold a smaller fraction of her portfolio in risky asset the more risk averse she is.

5. Conclusions

Financial portfolio optimization is an active field of investigation by researchers and practitioners alike. In this paper we provide a review of portfolio optimization techniques in discrete and continuous trading time. We discuss some of the salient features of optimal portfolios and derive closed-form solutions for several benchmark problems.

References

[1] H. Markowitz *Journal of Finance* **7** (1952) 77.

[2] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, John Wiley and Sons, 1959.

[3] L. Martellini, B. Urošević *Management Science* **52(6)** (2006) 955.

[4] F. Fabozzi, P. Kolm, D. Pachamanova, S. Focardi, *Robust Portfolio Optimization and Management*, John Wiley and Sons, 2007.

[5] R.T. Rockafellar, S. Uryasev *Journal of Risk* **2(3)** (2000) 21.

[6] R.T. Rockafellar, S. Uryasev *Journal of Banking and Finance* **26** (2002) 1443.

[7] P. Krokhmal, J. Palmquist, S. Uryasev *Journal of Risk* **4(2)** (2002) 11.

[8] V. Ranković, M. Drenovak, B. Urošević, R. Jelic *Computers & Operations Research* **72 (Aug.)** (2016) 83.

[9] M. Drenovak, V. Ranković, M. Ivanović, B. Urošević, R. Jelic European Journal of Operations Research **257 (3)** (2017) 1030.

[10] J. von Neumann, O. Morgenstern *Theory of Games and Economic Behavior* Princeton University Press (1943).

[11] J. Ingersoll *Theory of Financial Decision Making* Rowman & Littlefield Pub Inc (1987).

[12] J. Cvitanić, F. Zapatero *Introduction to the economics and mathematics of financial markets* MIT Press (2004).

[13] R. Merton *Continuous-Time Finance* Basil Blackwell Inc (1990).

Modeling of experimental evidences for self-organization of soliton-tweezers in nanosuspensions*

Vladimir Skarka†
Laboratoire de Photonique d'Angers, EA 4464, Université d'Angers, 49045 Angers, France
Institute of Physics, University of Belgrade, 11000 Belgrade, Serbia
Marina Lekić‡
Institute of Physics, University of Belgrade, 11000 Belgrade, Serbia
Stephan Rokotoarimalala§
Laboratoire de Photonique d'Angers, EA 4464, Université d'Angers, 49045 Angers, France

ABSTRACT

The exhaustive experimental investigations allowed us to understand spontaneous self-organization of soliton-tweezers in pure water naturally containing suspended nanobubbles. The laser beam propagation modifies the medium, but due to feedback mechanisms beam is itself altered in a self-organized way generating soliton. In soft matter like body water, suspended nanoparticles are aspirated inside the beam if their refractive index is larger than the background one. In opposite case they are expelled making again soliton self-focusing, thus, compensating defocusing effects. The result is a spontaneously self-collimated soliton-tweezer with a conserved profile as demonstrated by experiments and numerical simulations of a novel complex intensity equation. Coefficients multiplying this equation terms are determined by experimental measurements allowing the predictability important for noninvasive medical applications in inactivation of viruses, bacteria, and cancer cells.

* This work has been supported by the Ministry of Education and Science of the Republic of Serbia, under Projects III 45016, OI 171038, and OI 171006.
† e-mail address: vskarka@yahoo.com
‡ e-mail address: lekic@ipb.ac.rs
§ e-mail address: stephan.rokotoarimalala@univ-angers.fr

1. Introduction

Biomedia like blood, myosin, kinesin, ribosomes, liposomes, and varieties of living cells suspended in body water can be manipulated, tweezed, and controlled using laser beams and pulses[1-7]. The laser beam propagates through the nonlinear medium and alter it. Simultaneously, the changed medium acts by feedback mechanisms to this laser light modifying it in turn. Therefore, the laser light is self-controlled by light through the interaction with nonlinear medium [7-14]. The laser stability and precision are of crucial importance for nondestructive medical applications, *e.g*. photobiomodulation using these feedback mechanisms [6, 15,16]. In order to improve this indispensable dynamical stability, the best way is to leave nature to act by self-organization of solitons localized in space and in time [17,18]. The laser beam becomes soliton whenever antagonist effects, *i.e*., beam focusing and defocusing are balanced [19-21]. In any medium, including vacuum, laser beam is diffracting. In rigid guides beams are confined in the region with higher optical index of refraction. In contrast, the particularity of the propagation in soft matter is that the easily moving fluids adapt their density distribution in such a way that the beams always propagate through the highest index of refraction. Following 2018 Nobel Prize winner Ashkin, the tweezing laser beam collects around its center micro- and nanoparticles with index of refraction n_p higher than the background medium one (n_b), *i.e*., positively polarized particles [1-4,22,23]. Consequently, the beam is always self-focuses. However, in opposite case, when the refractive index of nanoparticles is lower than the background one ($n_p<n_b$), the negatively polarized nanoparticles are expelled from the beam that focalizes too, propagating again through higher index of refraction. Therefore, such a redistribution of nanoparticles density following Gaussian laser intensity induces always the nonlinear beam self-focusing [24-28]. Taking into account that the same nonlinear nanoparticle density distribution causes also defocusing effects, it becomes plausible that self-focusing effects compensate defocusing ones in a self-organized way [29]. As a consequence, the beam is self-collimated and becomes soliton-tweezer with identical widths following two transverse coordinates x and y corresponding to a unique transverse coordinate $r_\perp = \sqrt{x^2 + y^2}$ [19-21,30]. Using paraxial approximation and slowly varying envelop, A(r), the dynamics of electric field, $E=A(r)\exp(ik_0n_0z)$ can be described by two transverse dimensions (2D) nonlinear Schrödinger equation (NLS)

$$i\,\partial E/\partial z + \gamma\nabla_\perp^2 E + \eta E + (\sigma|I| - \nu|I|^2)E = 0 \qquad (1)$$

where$I=E^2$ is the complex intensity and the transverse Laplasian reads $\nabla_{\perp}^2=\frac{\partial^2 E}{\partial x^2}+\frac{\partial^2 E}{\partial y^2}$. The only imaginary term corresponds to the field propagation. The unavoidable diffraction is given by the second one. The third term is linear. The nonlinear self-focusing is given by the fourth term. The higher order negative nonlinearity is necessary in order to avoid the catastrophic collapse [19-21]. Coefficients in front of terms have to be determined from experiment in order to connect it directly to the theory [31]. For this purpose, we realize the laser beam propagation in pure water.

2. Experiments of soliton-tweezers self-organizationin water suspension of nanoparticles:Equation connecting theory with experiments

Taking into account that the total body water volume estimated from simple anthropometric measurements is about 70%, the safety of medical applications of lasers depends on good knowledge of beam propagation in water [32]. The challenge is to investigate the possibility of spontaneous self-generation of soliton-tweezers in pure water. Following exhaustive literature, the pure water contains naturally gas microbubbles and nanobubbles [33-38]. For instance, collapsing microbubbles may generate free-radicals in body water [39]. It is reported recently that drug delivery plasmonic nanobubbles rapidly detect and destroy drug-resistant tumor [40]. It seems that nanobubbles of about 50 to 100 nm are the most stable [33-38]. Taking into account that the water has the index of refraction $n_b=1.33$ and air nanobubbles $n_p=1$, we investigate the case of negative polarizability of nanoparticles ($n_b>n_p$). In order to study soliton tweezing gas particles in body water, *e.g.*, blood transporting oxygen and carbon dioxide, the prerequisite are theoretical, numerical and experimental synergetic investigations of the pure water.

We investigate experimentally the propagation of laser beam in the 20 cm long optically clear u-cuvette filled by pure water [41]. A powerful near infrared Mira 900 laser in continuous regime with wavelength $\lambda_0=727$ nm is used as the light source. The laser beam is spontaneously self-structured and self-collimated propagating in water suspension of nanobubbles as can be seen in Figure 1a. Indeed, the laser beam has the same width along of the cuvette. The beam width conservation implies the zero wave front curvature, as a main property of Schrödinger solitons [19-21]. Therefore, this beam is a 2D spatial soliton that is self-organized due to the balance between the self-focusing tweezing effects due to redistribution of nanobubbles, from the one side, and self-defocusing effects, from the other side [29]. The experiment to establish optical properties of the investigated nanosuspension consists in measuring with precision the beam output intensity as a function of gradually increasing

input one, charted in Figure 1e as dots with uncertainty bars [31]. The third order polynomial fit of the distribution of these dots furnishes the numerical factors in front of linear and quadratic focusing terms (x and x^2) and cubic defocusing term (x^3) in insert of Figure 1e giving respectively values of linear coefficient η, self-focusing coefficient σ, and self-defocusing coefficient ν of Eq.(1). The NLS describes the evolution of laser complex electric field E. However, in experiments is always measured the increase of laser complex intensity $I=E^2=A^2(r)\exp(i2k_0n_0z)$. Therefore, in order to obtain right description of experiment in paraxial approximation it is necessary to multiply Eq.(1) by complex electric field E

$$i\,\partial I/\partial z + \gamma\nabla_{\perp}^2 I + \eta I + (\sigma|I| - \nu|I|^2)I = 0 \qquad (2)$$

obtaining a novel soliton-tweezer complex intensity equation (STCIE).This synergetic equation directly connects the measured quantities with

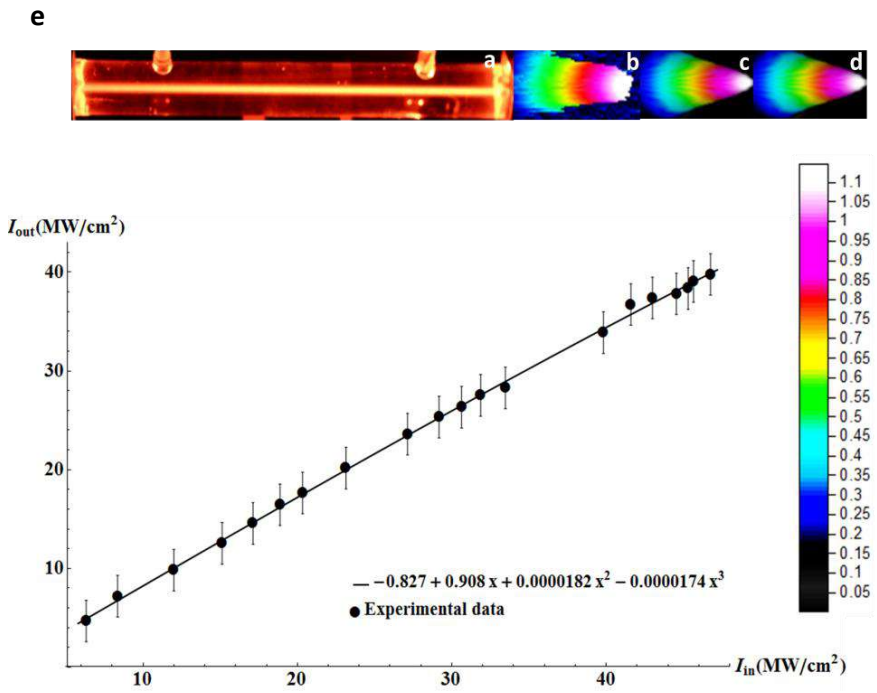


Figure 1a Self-collimated powerful red soliton-tweezer propagating through pure water in 20 cm cuvette. **e** Dots (with uncertainty bars) represent the measured soliton-tweezer output for various input intensities. The third order polynomial fit allows to obtain numerical values for coefficients in Eq.(2). Numerical propagations for z=1000 (**c**) and z=20000 steps (**d**) of this STCIE give identical soliton-tweezer profiles in very good agreement with the experimental one (**b**).

the theory. Indeed, the numerical factors in front of linear (x), quadratic (x^2), and cubic (x^3) terms in insert of Figure 1egive value of coefficients η, σ, and ν of linear (I), quadratic (|I|I), and cubic ($|I|^2I$) intensity terms of Eq. 2. The soliton-tweezers self-trapped propagation is limited by the cuvette length, but can be extended by numerical propagation of STCIE. Indeed, the camera positioned in the output of u-cuvette captured the soliton-tweezer profile (Figure 1b) that looks similar to profiles obtained by z=1000 (Fig. 1c) and z=20000 steps (Fig. 1d) numerical propagation of synergetic STCIE with coefficients from insert of Figure 1e.

Obviously, the beam intensity profile can be captured only in output of 20 cm u-cuvette, in order to visualize it in the middle, a twice shorter u-cuvette is irradiate by the identical laser beam under identical conditions. The laser beam is spontaneously self-collimated in 10 cm u-cuvette too (see Figure 2a). The intensity profiles after 10 cm and 20 cm propagations in pure water naturally containing suspension of nanobubbles are identical (compare Figures 1b and 2b). The third order

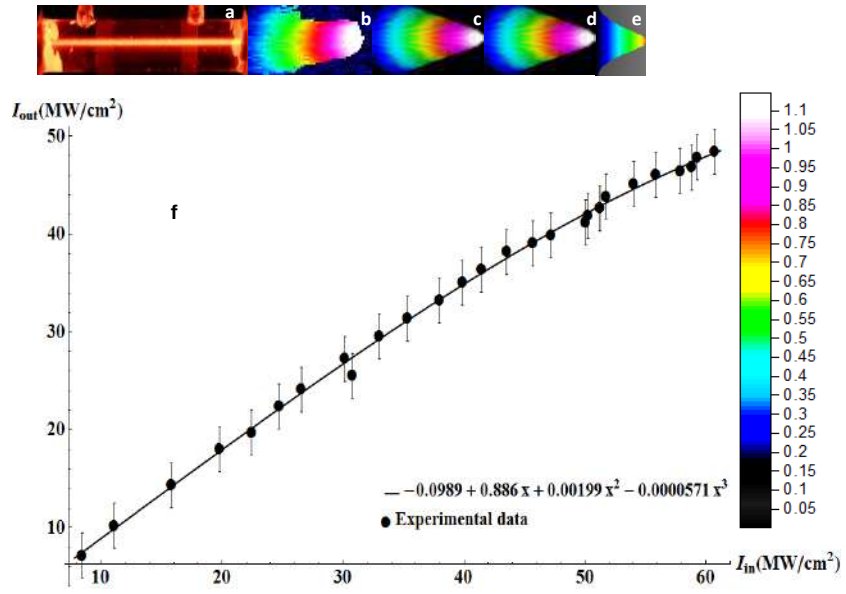


Figure 2a Self-organized propagation of soliton-tweezer in 10 cm cuvette. **b** The output profile resemble to beam profiles after self-trapped numerical propagation of z=600 (**c**) and z=20000 steps (**d**). **e** Diffracting low intensity beam profile. **f** Measured soliton-tweezer output intensities for increasing input ones represented by dots. The full line corresponds to third order polynomial fit furnishing numerical values for coefficients in STCIE. The color scale is common for all cuvettes.

polynomial fit, in insert of Figure 2f, gives different numerical values for coefficients η, σ, and ν, from those in Figure 1e. However, the

numerical propagation of Eq. 2 with new values of coefficients, after z=600 (Figure 2c) and z=20000 steps (Figure 2d) give intensity profiles identical to those of Figures 1c and 1d and very similar to experimental profiles. Therefore, the feasibility of soliton-tweezers in the pure water is confirmed. In contrast, the low intensity beam is diffracting as in Figure 2e.

For security reasons, it is desirable to nondestructively tweeze and manipulate body water nanoparticles using laser of very low intensity, *i.e.*, low power density. Consequently, in order to demonstrate the soliton-tweezer feasibility even for three order of magnitude lower intensities, a weak continuous green laser beam with wavelength λ_0=532 nm is propagated through the pure water in 20 cm cuvette. From the beam lateral view in Figure 3a can be concluded that the width of the beam profile (Figure 3b) is maintained constant through the cuvette. In order to find equivalent of beam profile in the center of 20 cm u-cuvette, the same green beam is now propagated through the 10 cm u-cuvette. It becomes self-collimated due to the soliton-tweezer self-trapping too (see Figure 4a).

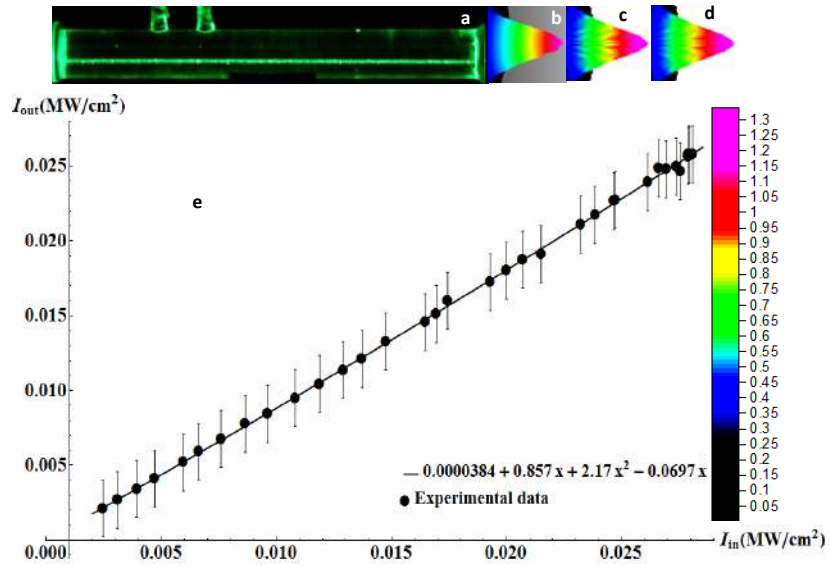


Figure 3a The propagation of a weak green laser beam through the same 20 cm cuvette becomes self-collimated due to the interaction with nanobubbles suspended in water. **b** Its output profile coincide with those obtained numerically after z=800 (**c**) and z=20000 steps(**d**). **e** The curve that fit experimental data corresponds to a third order polynomial with coefficients that are integrated in Eq.(2) allowing the soliton-tweezer numerical propagation.

Therefore, this beam keeps always the same profile as in Figure 4b. This output profile is identical to the output one of 20 cm u-cuvette in Figure 3b. In contrast, in Figure 4e is shown the profile of a very low intensity beam that is diffracting. The variation of output intensities in function of input intensities for 20 cm and 10 cm cuvettes are charted respectively in Figures 3e and 4f. Although fitting of these curves gives different coefficients in front of terms in Eq.(2) its long numerical simulations till z=20000 steps, charted in Figures 3c,d and 4c,d keep the identical form of profile that is the signature of a soliton. These numerical profiles are in very good agreement with experimental profiles in Figures 3b and 4b. Therefore, even for very low intensities, soliton-tweezers self-trapped in water suspension of nanobubbles remain robust allowing photobiomodulation and other noninvasive medical applications in body water colloidal nanosuspensions.

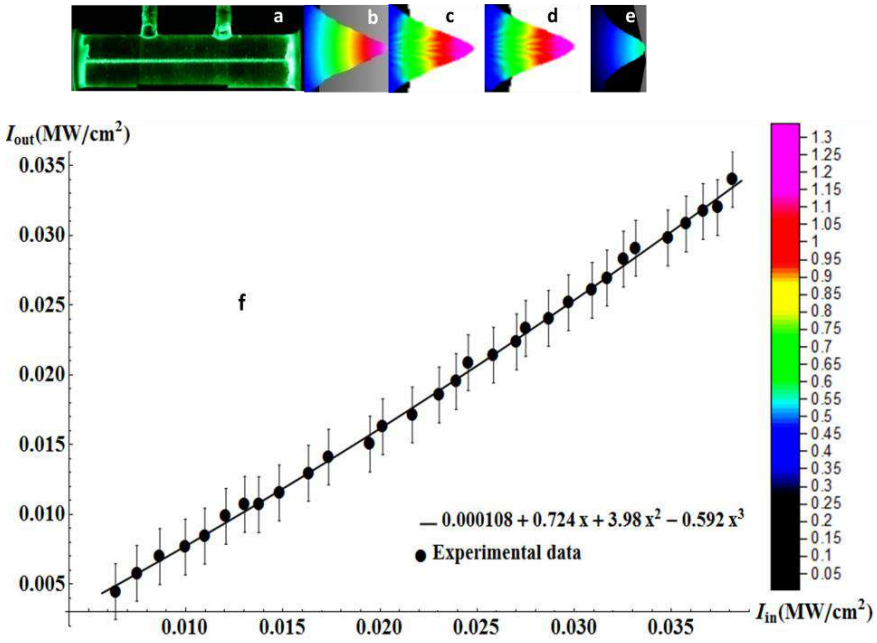


Figure 4a Lateral view of the same green laser beam propagation in 10 cm cuvette shows the self-collimation caused by the self-organized interaction with nanobubbles in water suspension, even though intensities in the self-trapping range are three orders of magnitude lower than those of red laser. **f** Coefficients of the STCIE, experimentally obtained from fitted input-output data, allow long numerical simulations of soliton-tweezer profiles that are conserved without any modification, not only after z=1000 steps as in (**c**), but also after z=20000 steps as in (**d**). **b** These profiles coincide with the experimental one, and those in Figure 6. **e** Out of trapping range the beam is diffracting.

Our synergetic approach establishes a very useful bridge between the theory and experiment via numerical simulations of novel STCIE

allowing a security test of non destructibility prior every medical application.

3. Dissipative soliton-tweezers self-trapped in water suspensions of nanobubbles

The nonlinear Schrödinger equation and the established soliton-tweezer complex intensity equation, both describe conservative systems. However, the energy losses are unavoidable in real dynamical systems and can disintegrate solitons unless gain is also present. Such systems are well described by the complex cubic-quintic Ginsburg-Landau equation (CQGLE) governing complex electric field E[9,14,21,29]

$$\frac{\mathrm{i}\,\partial \mathrm{E}}{\partial \mathrm{z}} + (\eta + \sigma|\mathrm{I}| - \nu|\mathrm{I}|^2 + \gamma\nabla_\perp^2)\mathrm{E} = \mathrm{i}(\delta + \varepsilon|\mathrm{I}| - \mu|\mathrm{I}|^2 + \beta\nabla_\perp^2)\mathrm{E}. \quad (3)$$

The left hand side of this equation corresponds to conservative NSL. The dissipative, thus, imaginary right hand side contains the same terms but with different coefficients to be fixed from the experiment. The parameter δ of the first imaginary term must be negative, in order to ensure the stability of the system [5,9]. It corresponds to linear loss giving with propagation term the linear absorption equation. This term and negative nonlinear quintic term are both compensated by cubic gain, similarly as real terms of self-defocusing and self-focusing. We demonstrated that the self-organized dynamical balance of antagonist effects is in reality more complex, with second order derivatives compensation playing central role [9,42,43]. The positive β diffusive term corresponds to gain. In order to connect this theory with experiments, all terms are multiplied by coefficients. Their values are measured in experiments [31]. Recently, B.N. Aleksić, *et al.* showed that the competing cubic-quintic nonlinearity represents a very good approximation of saturable nonlinearity [44].

For the sake of comparative investigations we propagate through 20 cm u-cuvette again 727 nm Mira laser beam, but this time in femtosecond regime. This beam is spontaneously self-collimated during propagation, having a constant width along u-cuvette, as can be seen in Figure 5a.

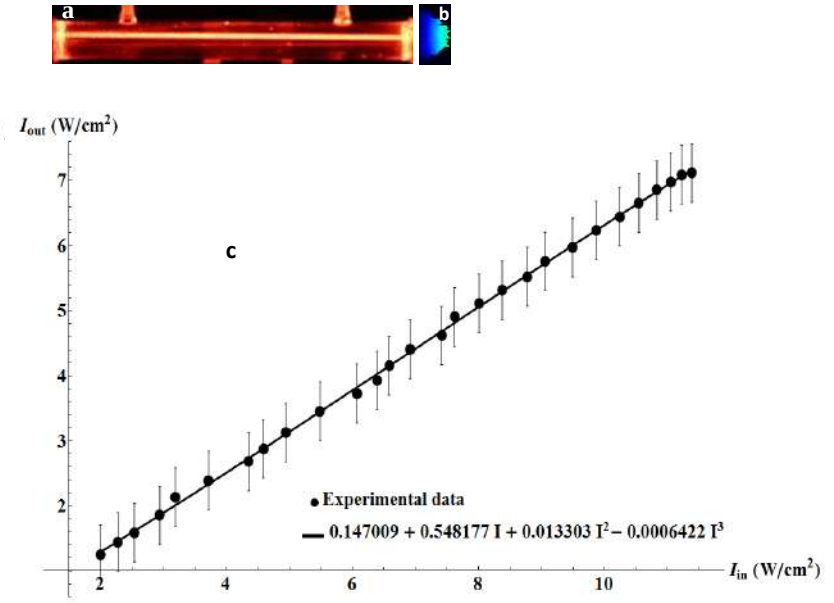


Figure 5 a Self-trapped and self-collimated femtosecond soliton-tweezer propagating through nanosuspension in 20 cm cuvette. **b** Experimentally obtained output profile. **c** Dots represent the measured output for various input intensities. The third order polynomial fit allows to obtain in front of I, I^2,and I^3in insert, numerical values for coefficients η, σ, and ν in left hand side of Eq. (3).

A camera capture from the output of u-cuvette is presented in Figure 5b.In fact, for each increase of input intensity it is taken picture of the output.

Each camera capture is then analyzed using MATLAB in order to extract output intensity from each pixel. The addition of intensities of all pixels in a square framework fitting the light spot gives the value of corresponding global intensity charted as a dot in Figure 5c.Then, the ensemble of dots is fitted by a third order polynomial. The coefficients of this polynomial in insert give experimental values of coefficients in front of terms on the left hand side of Eq. (3) (CQGLE) in the same way as it was done for Eq. (1) (NSL). The imaginary right hand side of CQGLE is related to the balance between gain and loss of power density. The coefficients multiplying corresponding terms can be find from power density diagram in Figure 6 obtained by systematic measure of input and output power after each camera capture. The global power is divided by the surface of the same square framework fitting the light spot, providing the same units of W/cm^2 in both side of CQGLE. These input-output dots are fitted using the same third order polynomial. Indeed, right hand side terms are of the similar order of magnitude as left hand side terms. In the insert of Figure 6 are charted experimental values for the coefficients in

front of CQGLE imaginary terms. Our aim is to establish a direct correspondence between the experiment and the theory. However, theoretical approaches use electrical field, E to describe laser beam, while in experiments is measured the complex intensity, $I=E^2$ [31]. Therefore, the direct correspondence between experiment and theory is established multiplying Eq. (3) by electrical field, E in order to recover the complex intensity

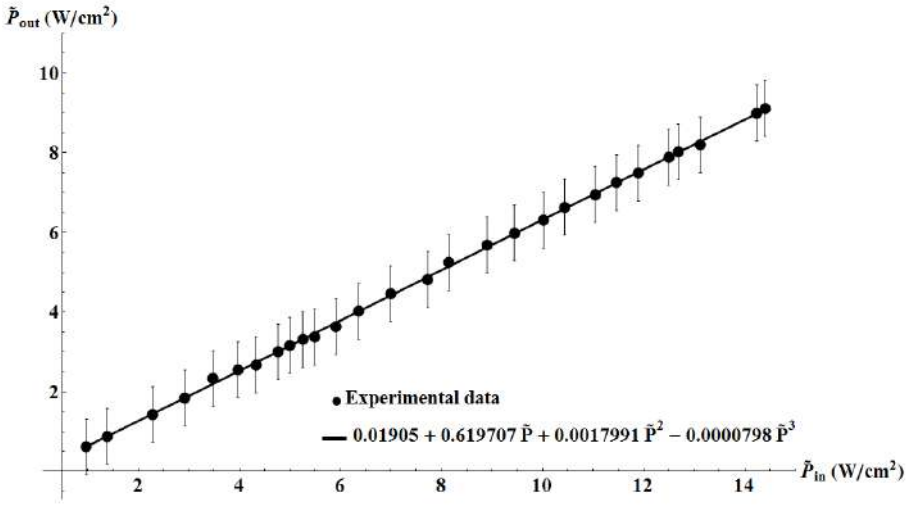


Figure 6 Measures of output versus input power densities of self-trapped dissipative soliton-tweezer, charted by ensemble of dots (with uncertainty bars). Their third order polynomial fit provides us in insert with numerical values in front of $\dot{P}$, $\dot{P}^2$,and $\dot{P}^3$ corresponding to coefficients δ, ε, and μ in the right hand side of Eq. (3).

$$\frac{i\,\partial I}{\partial z} + (\eta + \sigma|I| - \nu|I|^2 + \gamma\nabla_\perp^2)I = i(\delta + \varepsilon|I| - \mu|I|^2 + \beta\nabla_\perp^2)I. \quad (4)$$

This dissipative soliton-tweezer complex intensity equation (DSTCIE) is astonishingly symmetric. The imaginary and real quadratic terms increase intensity that is decreased by cubic terms in a self-organized balance. More generally, the compensation of antagonist effects is what keep the nature in dynamic equilibrium. The advantage of this novel approach is not only to describe correctly experiments but also to predict their results, which is of particular importance for medical applications. Numerical propagation of DSTCIE allows to test the feasibility of dissipative soliton-tweezers.

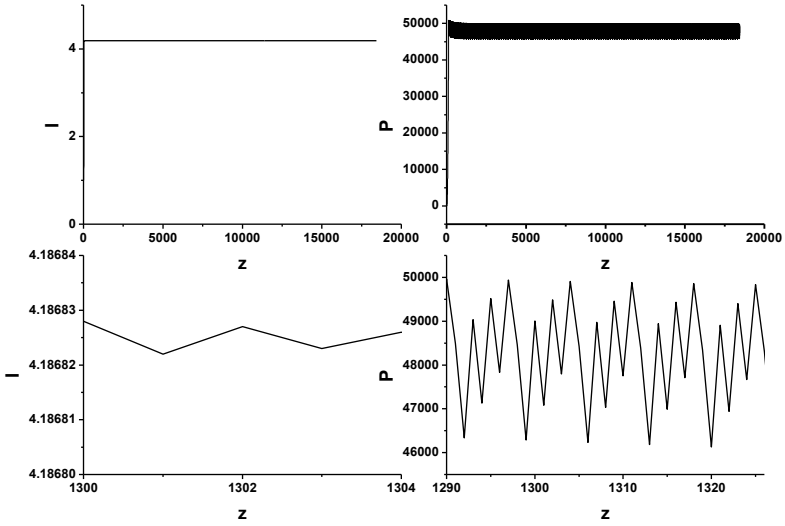


Figure 7 Femtosecond dissipative soliton-tweezer is propagated numerically z=19000 steps using DSTCIE with coefficients from inserts of Figures 5 and 6 showing stability of intensity and power. Only strong zooming revel weak breathing.

A long numerical evolution till z=20000 shows in Figure 7, a very good stability of intensity, I and power, P. A strong zooming revel very weak breathing of intensity and little bit stronger of the power due to its quadratic dependence on beam width in 2D systems[9,14,21,29]. This breathing is more obvious in Figure 8 where the soliton-tweezer 3D profile oscillates between a dome and a "millstone".

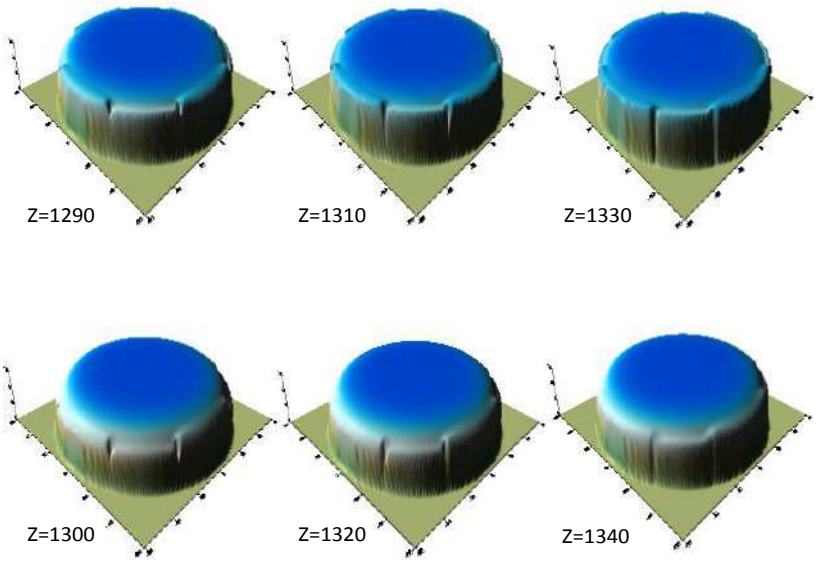


Figure 8 Breathing alternation of soliton-tweezer profiles (in arbitrary units) in form of "millstones" (first row)and domes (second row) corresponding to power maxima and minima in Figure 7.

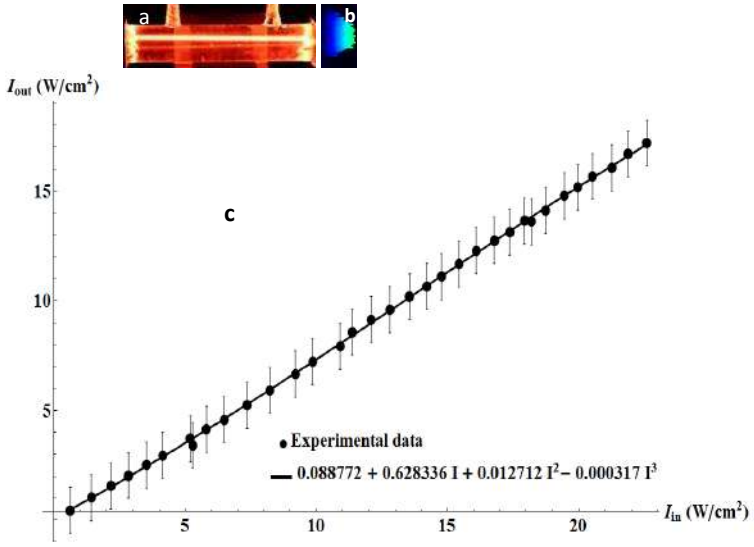


Figure 9a Same self-collimated 727 nm dissipative soliton-tweezer propagating through nanobubbles suspension in 10 cm cuvette. **b** Output profiles after propagation of 10 and 20 cm are identical. **c** Fitting input-output intensity dots yields coefficients in front of linear, quadratic, and cubic intensity to be replaced in dissipative soliton-tweezer complex intensity Eq. (4) that can predict ulterior numerical propagation.

We remarked the same behavior in a previous theoretical publication concerning self-structuring of stable dissipative breathing vortex solitons in a colloidal nanosuspension [29]. It is amazing that such a breathing is confirmed by our approach that connect directly the theory with experiments. The inertia of nanobubbles may be the reason of this breathing.

The self-trapping of spontaneously self-collimated soliton-tweezers is

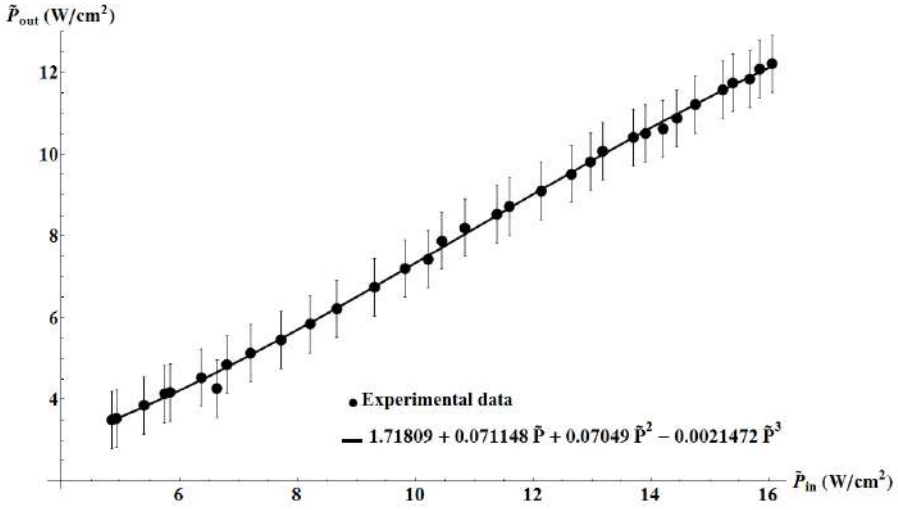


Figure 10 Measured input-output dissipative power density provides values for coefficients of imaginary polynomial terms in Eq. (4).

confirmed using a twice shorter u-cuvette under the identical conditions (see Figure 9a). The camera capture of soliton profile in the output of 10 cm cuvette (Figure 9b) corresponds to the profile of twice longer u-cuvette in Figure 5b. Measured output intensities and powers divided by square surface in Figure 9c and Figure 10 lead to the determination of coefficients in DSTCIE using the third order polynomial fits of corresponding input-output dots. The numerical propagation of DSTCIE gives the same results as in case of 20 cm cuvette (Figures 5 and 6). Therefore, this novel equation correctly describes behavior of dissipative solitons-tweezers in experiments of propagation in water suspension of nanobubbles.

4. Self-generation of vortexsoliton-tweezers in nanosuspensions

Vortex soliton-tweezers are robust steady-state structures with nonzero angular momentum and phase singularity in zero intensity center [14,21,30,45-47]. In conservatives systems v are governed by STCIE

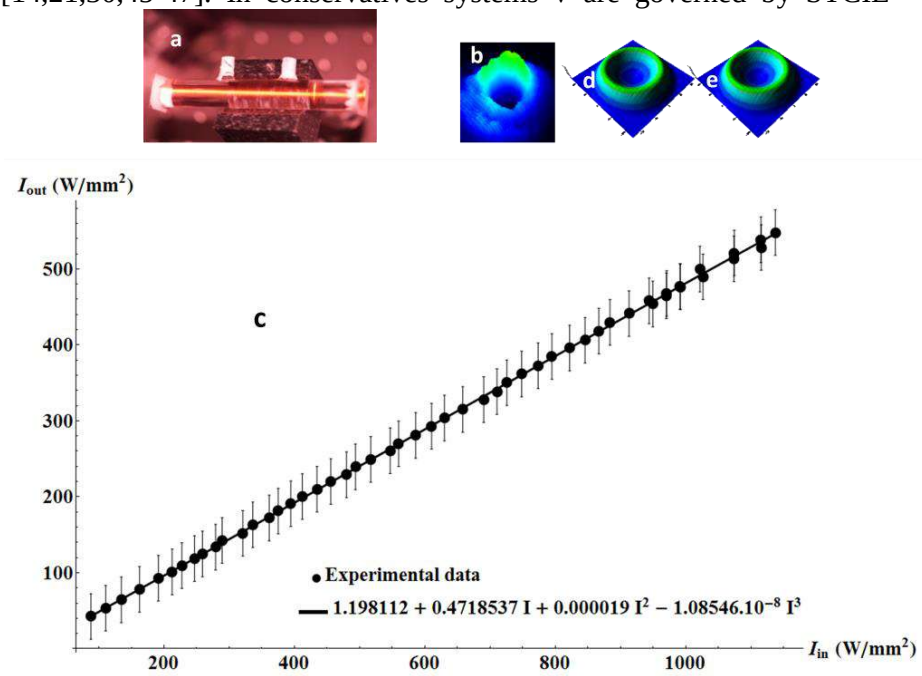


Figure 11a Vortex soliton-tweezer spontaneously self-collimated in 10 cm u-cuvette. **b** Experimentally obtained vortex output profile is similar to profiles obtained by z=1800 (**d**) and z=10000 steps (**e**) numerical simulations of STCIE with coefficients from insert (**c**).

(Eq. 2). The propagation of dissipative vortex soliton-tweezers is described by DSTCIE (Eq. 4). We use a weak Thorlabs HeNe laser beam of 632.8 nm in order to generate vortex in 10 cm u-cuvette containing nanobubbles naturally suspended in pure water. Indeed, a vortex soliton-tweezer is spontaneously self-collimated maintaining along u-cuvette the same width, as shown in Figure 11a. Figure 11b is the camera capture of output vortex profile. Dots with uncertainty bars in Figure 11c are result of systematic and precise measurement of output intensity for each gradually increasing input intensity. The already established procedure of fitting yields the values for STCIE coefficients in insert of Figure 11c. The same values can be used for left hand side DSTCIE coefficients. Those of right hand side are obtained fitting input-output power dots in Figure 12. Hence, the feasibility of vortex soliton-tweezers is established.

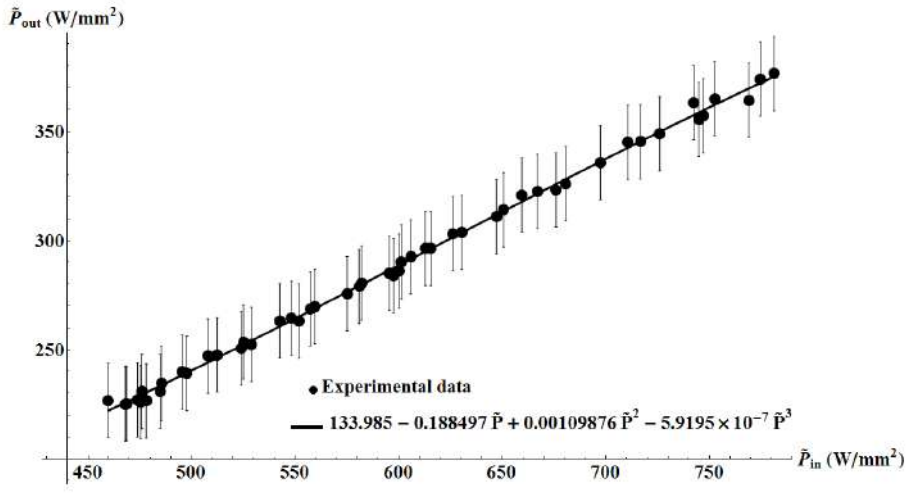


Figure 12 Output versus input power density of self-organized vortex soliton-tweezer fitted in order to obtain, in insert, values for coefficient in imaginary terms in DSTCIE.

5. Novel intensity equation for spatiotemporal tweezing light bullets

Light bullets are spatiotemporal solitons completely localized in space and time [9,19,20,48,49]. In conservative systems they are governed by three transverse dimensions (x, y, and t) NSL

$$\frac{i\,\partial E}{\partial z} + \gamma\Delta E + \eta E + (\sigma|I| - \nu|I|^2)E = 0 \qquad (5)$$

where the Laplasian$\Delta = \partial^2/\partial t^2 + \nabla_\perp^2$ is associated to the time variable t and the transverse coordinate $r_\perp$. The propagation of light bullets in dissipative systems is described using 3D Ginsburg-Landau equation [9, 48,49].

$$\frac{i\,\partial E}{\partial z} + (\eta + \sigma|I| - \nu|I|^2 + \gamma\Delta)E = i(\delta + \varepsilon|I| - \mu|I|^2 + \beta\Delta)E. \qquad (6)$$

In order to establish the direct connection between this spatiotemporal theory involving electric field, E and experiments measuring input and output intensities we multiply Eq. (5) by electric field, E in paraxial approximation

$$\frac{i\,\partial I}{\partial z} + \gamma\Delta I + \eta I + (\sigma|I| - \nu|I|^2)I = 0 \qquad (7)$$

establishing cubic-quintic tweezing light bullet complex intensity equation (TLBCIE). In this novel equation not only diffraction, but also dispersion is compensated by competing cubic-quintic nonlinearity.

In reality, experiments are always localized in three spatial dimensions and in time. They involve always losses and sometimes gain. The adequate theoretical and numerical approaches are the prerequisites for experimental realization of light bullets. Therefore, the complex intensity is the right variable in an equation that pretends to describe the experiment in a realistic

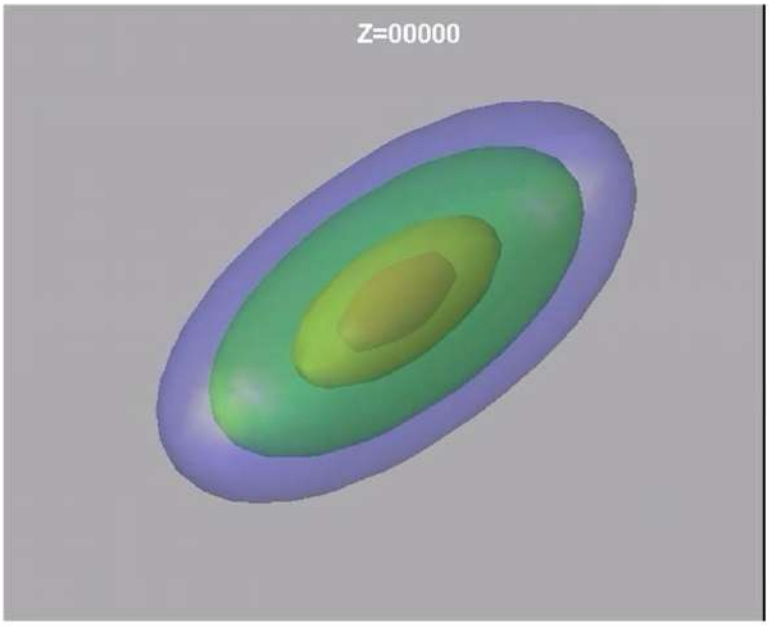


Figure 13 Dissipativelight bullet from non-spherically-symmetric input pulse becomes in the end of propagation spherically-symmetric like in Figure 14.

way. Therefore, Eq. (6) is in paraxial approximation multiplied by electric field

$$\frac{i\,\partial I}{\partial z} + (\eta + \sigma|I| - \nu|I|^2 + \gamma\Delta)I = i(\delta + \varepsilon|I| - \mu|I|^2 + \beta\Delta)I. \qquad (8)$$

In such a way we establish dissipative tweezing light bullet complex intensity equation (DTLBCIE). This equation represent the most general expression of self-organized description of 3D spatiotemporal soliton propagation, taking into account that the competing cubic-quintic nonlinearity represents a very good approximation of saturable nonlinearity [9,44,48,49]. The DTLBCIE is astonishingly symmetric.

Figure 14 Dissipative light bullet with topological singularity in center is effectively a soliton-tweezing shell transporting small energy suitable for secure medical applications.

The quadratic terms amplify the complex intensity through self-focusing and gain. They compensate the linear and cubic losses. The gain diffusive term with positive β balances the diffraction and dispersion. We demonstrated that in reality a cross compensation occurs between the excess of the self-focusing, losses, and gain making the system self-organized [9,48,49]. Numerically simulated light bullet from non-spherically-symmetric input pulse is shown in Figure 13. The light bullet with topological singularity is the light shell (in Figure 14). The powerful femtosecond bullet consisting of ultra-thin light shell transports a very small energy suitable for nondestructive and secure medical applications.

6. Conclusions

The prerequisites for understanding phenomena in nanophotonics are exhaustive experimental investigations coupled with theoretical advances using tools of numerical simulations. Object of our investigations is the water as the essential substance for the life in the earth. On the first glance it is astonishing that water contains naturally air nanobubbles. However, it is to forget that fishes breathe in water. Therefore, the pure water is in fact, nonlinear suspension of relatively stable nanobubbles. The nonlinear behavior of water is revealed using interactions with lasers. Propagating through water laser beams and pulses are spontaneously self-organized into soliton-tweezers that collect positively polarized nanoparticles with index of refraction larger than the water one. In opposite case, negatively polarized nanoparticles, with smaller refractive index, are expelled out, allowing laser light to propagate again through the higher index, thus to be in both cases self-

focused. This self-focusing compensates all defocusing effect, spontaneously self-collimating the soliton-tweezer as our experiments and numerical simulations confirm. To be able to make the numerical simulations, it is necessary to have an adequate equation describing this phenomenon. We solved the paradox that the theoretical descriptions use the electric field, while in experiments is measured the electric intensity, by introducing a novel complex intensity equation for conservative and dissipative systems. Terms of this equation contain coefficients determined from the experimental measurements. Soliton-tweezers profiles obtained by numerical propagation of established equation, coincide with experimentally obtained profiles, confirming the correspondence between experiment and theory. In such a way, the feasibility of soliton-tweezers to control and manipulate nanoparticles in water suspension, is established. The corresponding applications are multiples. They may be medical, biological, industrial... The blood in living organisms can be considered as water suspension of nanoparticles. Soliton-tweezers can be useful in fight against viruses, bacteria, and cancer cells in blood without damaging healthy tissues. The blood passing through our u-cuvette can be irradiate by laser. Depending of their frequency soliton-tweezers or tweezing light bullets can selectively destruct some viruses, bacteria, or cancer cells [50-52]. For instance, if Coronaviruses-19 (COVI-19) have higher resonant frequencies than the frequency of soliton-tweezer they can be inactivated in its centre due to high intensity, heat, and vibrations. If viruses have lower resonant frequencies, they can be expelled out of soliton on the wall of u-cuvette where they stay bind by monoclonal antibodies [53-55]. As a consequence, the blood is purified. It is important to note that the purification is efficient even though we do not know the viruses resonant frequencies. Indeed, for the same frequency of soliton-tweezer, viruses with higher resonant frequency are inactivated in soliton center by high heat, and vibrations. Simultaneously, those viruses with lower resonant frequencies are stick on wall by monoclonal antibodies [53-55]. Viruses are much smaller than other nanoparticles in body water, so that they have much higher resonant frequencies than other. The safe use of low power soliton-tweezers for noninvasive medical applications is in such a way reinforced.

References

[1] A. Ashkin, J. M. Dziedzic and T. Yamane, *Nature***330**(1987) 769.
[2] A. Ashkin, *Science* **210** (1980) 1081.
[3] A. Ashkin *et al*. *Nature* **348** (1990) 346.
[4] R. A. L. Jones, *Soft Condensed Matter*, Oxford University Press, 2002.

[5] N. N. Akhmediev, A. Ankiewicz, *Dissipative Solitons: From Optics to Biology and Medicine*, *Lect. Notes Phys*.751 Springer-Verlag: Berlin, 2008.
[6] L. Santana-Blank *et al*.*Water as a Photoacceptor, Energy Transducer, and Rechargeable Electrolytic Bio-battery in Photobiomodulation*. In *Handbook of Low-Level Laser Therapy*, (2017) 119.
[7] C. Conti, G. Ruocco and S. Trillo, *Phys. Rev.Lett.***95** (2005) 183902.
[8] Y. S. Kivshar and G. P. Agrawal, *Optical Solitons: From Fibers to Photonic Crystals*; Academic Press: San Diego, 2003.
[9] V. Skarka and N. B. Aleksić, *Phys. Rev. Lett*. (2006) 0139031.
[10] Z. Chen, M. Segev and D. N. Christodoulides, *Rep. Prog. Phys.* **75** (2012) 086401.
[11] I. S. Aranson and L. Kramer, *Rev. Mod. Phys*. **74** (2002) 99.
[12] F. T. Arecchi, S. Boccaletti and P. L. Ramazza, *Phys. Rep*. **318** (1999) 1.
[13] D. Mihalache *et al*.*Phys. Rev. Lett.* **97** (2006) 0739041.
[14] V. Skarka *et al*.*Phys. Rev. Lett.***105** (2010) 213901.
[15] M. Sato, M. Ishihara, M. Kikuchi, and J. Mochida, *Lasers Surg. Med*. **43** (2011) 421.
[16] G. R. Schweinsberger *et al*. *Lasers Surg. Med*. **43** (2011) 443.
[17] R. W. Boyd, *Nonlinear Optics*; Academic Press: San Diego, 1992.
[18] G. Nicolis and I. Prigogine, *Self-organization in Nonequilibrium Systems* Wiley: New York 1977.
[19] V. Skarka, V. I. Berezhiani and R. Miklaszewski, *Phys. Rev. E***59** (1999) 1270.
[20] V. Skarka, V. I. Berezhiani and R. Miklaszewski, *Phys. Rev*. *E Rev*. *E***56** (1997) 1080.
[21] V. Skarka *et al. Phys. Rev. A***90**(2014)0238451.
[22] A. Ashkin, *Phys. Rev. Lett.* **24** (1970) 156.
[23] P. W. Smith, A. Ashkinand W. I. Tomlinson, *Opt. Lett*. **6** (1981) 284.
[24] S. Fardad *et al*. *Nano Lett*.**14**(2014) 2498.
[25] R. El-Ganainy *et al.Opt. Express***15**(2007) 10207.
[26] M. Matuszewski, W. Krolikowski and Y. S. Kivshar, *Opt. Express* **16** (2008) 1371.
[27] M. Matuszewski, W. Krolikowski and Y. S. Kivshar, *Phys. Rev*. *A***79** (2009) 023814.
[28] R. El-Ganainy *et al.Phys. Rev*. *A***80** (2009) 053805.
[29] V. Skarka *et al. Opt. Express***25** (2017) 10090.
[30] B. N. Aleksić, N. B. Aleksić, V. Skarka and M. R. Belić, *Physical Review A***91** (2015) 043832.
[31] V. Skarka, M. M. Lekić, A. G. Kovačević, B. Zarkovand N. Ž. Romčević, *Opt. Quant.Electron.***50** (2018) 37.
[32] P. E. Watson, I. D. Watsonand R. D. Batt, *Am. J. Clin. Nutr*.**33** (1980) 27.
[33] S. A. Thorpe, A. R. Stubbs, A. J. Hall and R. J. Turner, *Nature* **296** (1982) 636.
[34] N. Ishida *et al. Langmuir* **16** (2000) 6377.

[35] F.Y. Ushikubo*et al*. *Colloids and Surfaces A*: *Physicochem. Eng. Aspects* **361** (2010) 31.
[36] R. M. Pope and E.S. Fry, *App. Opt*. **36** (1997) 8710.
[37] B. H. Tan, H. An and C.-D. Ohl, *Phys. Rev. Lett*. **120** (2018) 164502.
[38] M. Chaplin, Nanobubbles in *Water Structure and science, Water Site Map,* London South Bank University 2000.
[39] M. Takahashi, K. Chiba and P. Perry, *J. Phys. Chem*. **111**(2007) 1343.
[40] , E. Y. Lukianova-Hleb *et al.Theranostics***2**(2012) 976.
[41] E. L. Falcao-Filho, C. B. de Araujo, G.Boudebs, H. Leblond and V. Skarka, *Phys. Rev.Lett.***110** (2013) 0139011.
[42] V. Skarka, D. V. Timotijević, N. B. Aleksić, *J. Opt. A: Pure Appl. Opt.* **10** (2008) 075102.
[43] D. Mihalache *et al*.*Phys. Rev. A***82** (2010) 023813.
[44] B. N. Aleksić, L. A. Uvarova, N. B. Aleksić and M. R. Belić, *Opt. Quant. Electron.***52** (2020) 175.
[45] V. Skarka, N. B. Aleksić, M. Derbazi and V.I. Berezhiani, *Phys. Rev. B***81** (2010) 035202.
[46] I. Berezhiani, V. Skarka, and N. B. Aleksić, *Phys. Rev. E***64** (2001) 057601.
[47] V. Skarka, V. I. Berezhiani, and N. B. Aleksić, *Phys. Lett. A***291** (2001) 124.
[48] N. B. Aleksić, V. Skarka, D. V. Timotijević, D. Gauthier, *Phys. Rev. A***75** (2007) 061802(R).
[49] V. Skarka, N. B. Aleksić and V. I. Berezhiani, *Phys. Rev. A***81** (2010) 045803.
[50] A. Ashkin *et al. Nature* **348** (1990) 346.
[51] A. R. Singh, A. Košmrlj and R. Bruinsma, *Phys. Rev. Lett*. **124**
(2020) 158101.
[52] Szu-Chi Yang *et al*. *Scientific. Reports* **5** (2016) 18030.
[53] R. G. H. Cotton, and C. Milstein, *Nature* **244** (1973) 42.
[54] Lingshu Wang, *et al*. *Journal of Virology* **92**(2018) e02002.
[55] S. Balamurugan, S. Konlavat, W. Kittikhun and P. Waranyoo, *Asian Pac J Allergy Immunol*.**38** (2020) 10.

The brachistochronic motion of a heavy ball rolling along an imperfect rough surface[*]

Obradović Aleksandar[†]
University of Belgrade, Faculty of Mechanical Engineering,
KraljiceMarije 16, Belgrade 35 11120, Serbia

Šalinić Slaviša[‡]
University of Kragujevac, Faculty of Mechanical and Civil Engineering
in Kraljevo, Dositejeva 19, 36000 Kraljevo, Serbia

ABSTRACT

The problem of brachistochronic motion of a heavy uniform ball rolling without slip along the upper outside surface of an imperfect rough stationary sphere, is solved. The control forces are located in the tangential plane, and their total power equals zero. In the first part of the paper the determination of the brachistochronic motion is solved as the problem of optimal control using Pontryagin's maximum principle. This solution corresponds to the motion of the heavy ball along a perfect rough sphere. The second part provides the case when the constraint between the sphere and the ball is imperfectly rough. Here, the problem of optimal control is formulated in such way that the tangential component of the reaction of constraint is taken for the control, with the restriction resulting from Coulomb's laws of sliding friction. The problem thus formulated belongs to the theory of singular optimal controls, and the solution that satisfies the Maximum principle consists of a singular part and a non-singular part.

[*] This work has been supported by Ministry of Education, Science and Technological Development, Republic of Serbia, Project Nos.451-03-68/2020-14/200105 and 451-03-68/2020-14/200108

[†]e-mail address: aobradovic@mas.bg.ac.rs
[‡]e-mail address: salinic.s@mfkv.kg.ac.rs

1. Introduction

The classical brachistochrone problem of the heavy point in a vertical plane has experienced application to various classes of mechanical systems over the past years. Among the contemporary works, the doctoral dissertation [1] should be noted, which proves the actuality of the problem even today. This recently defended dissertation reveals that the problems of the brachistochronic motions involve the areas not investigated yet. The dissertation provides a very good survey of the results which include several of our papers as well. In creating the task of optimal control, as we do in our papers, some of the reaction of constraints is taken for the control. The same has been done in [2]. The brachistochronic motion of the heavy point under the action of the sliding friction force was also the subject of the research in both [3] and [4], where bilaterally limited normal reaction of the constraint was studied.

In [5] both the classical brachistochrone problem and the unrestrained brachistochrone problem were formulated within the framework of the optimal control theory. Also, the extension to the cases of the brachistochrone problem for a rolling rigid body in a vertical plane, as well as the three-dimensional-minimum-time optimal problem for a disk rolling on the interior surface of a hemisphere was given.

Here, the research focuses on those brachistochrone problems of the nonholonomic mechanical systems where the reaction of the nonholonomic constraint is limited, and the paper is a continuation of the authors' earlier research studies.

In [6] the lateral reaction of the constraint ($|R_\eta| < N_b$) of the Chaplygin sleighis bilaterally limited (Fig. 1), and the brachistochronic motion is realized by subsequent imposition of an ideal holonomic constraint to the mass center C. This type of restriction was considered in a well-known work by Caratheodory [7].

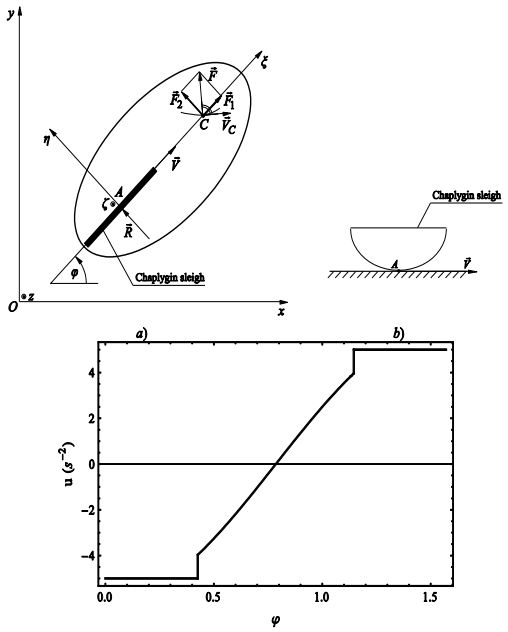


Fig.1 [6] Brachistochronic motion of the Chaplygin sleigh

The brachistochronic motion is realized in such way that on some parts the constraint reaction is on one of its limits $R_\eta = \pm N_b$ and a singular part is in the middle of motion, so that angular acceleration, depending on the angle, is given in Fig. 1c.

In the cases when the disks are rolling along horizontal surfaces Coulomb's laws of sliding friction limit the maximum possible horizontal components of the constraint reactions.

Paper [8] considers the brachistochronic motion of a vertical disk along a horizontal plane. Motion is controlled by three couples and the restrictions arise from the condition that slip will not occur at the contact point of the disk and the surface. Figure 2b) shows change in the horizontal components of the disk constraint reactions in the numerical example, where a non-singular part is in the middle of the interval of motion.

Previous research was extended in [10] to a more complex system of bodies, the simplified model of a vehicle [9], Fig. 3. Wheels slip is prevented based on the restrictions following from Coulomb's laws of friction. For real values of Coulomb's coefficient of friction, in this case too, one obtains non-singular parts of the brachistochronic motion, where a horizontal component of the reaction force is on its limit.

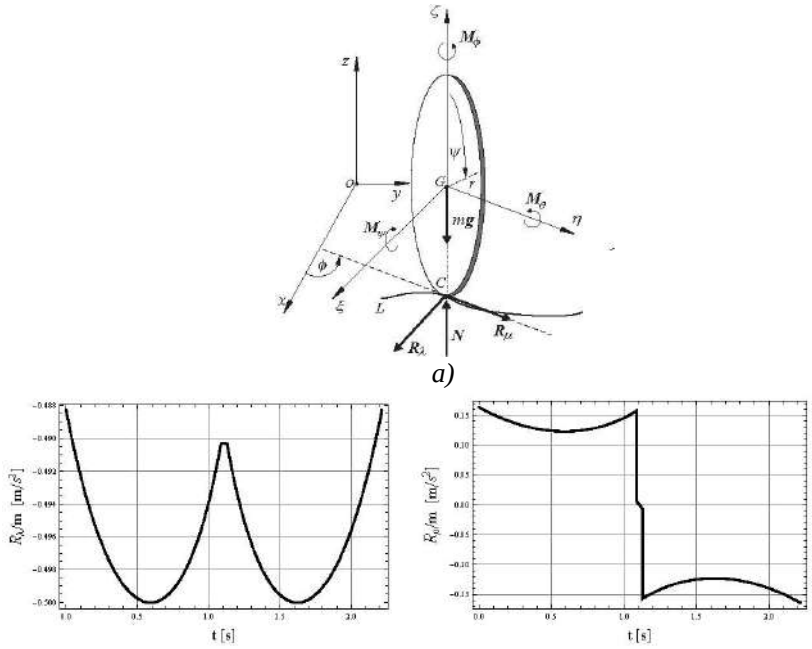
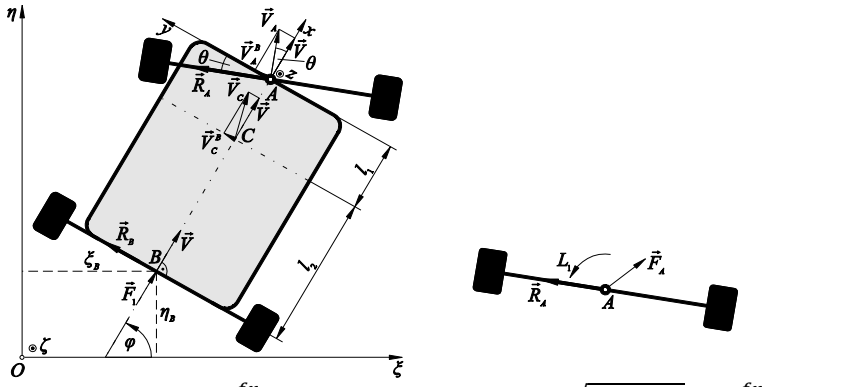


Fig. 2 [8] Brachistochronic motion of a vertical disk



$$\mu_2(t)N_2 = |R_A| < F_2^{fr} = \mu N_2, \quad \mu_1(t)N_1 = F_b = \sqrt{R_B^2 + F_1^2} < F_1^{fr} = \mu N_1$$

Fig. 3 [10] Brachistochronic motion of a simplified vehicle model [9]

The major goal of this paper is to determine the brachistochronic motion of the heavy ball rolling along a real rough sphere. For various real materials that the ball and the sphere can be made of their contact cannot be considered perfectly rough. In real rough contact the quotient of the intensity of horizontal and vertical component of the constraint reaction must be lower than the real value of Coulomb's friction coefficient. It is necessary to apply the procedure similar to that for the rolling disk [8].

2. Problem formulation

Observe the motion of a heavy uniform ball rolling without slip along the upper outside surface of an imperfect rough stationary sphere (Fig. 4)

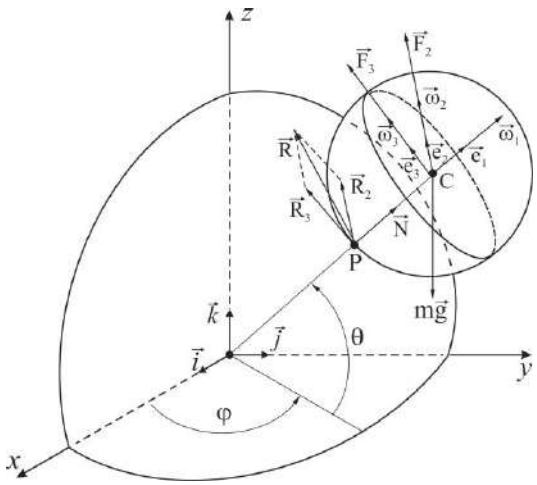


Fig.4 Heavy ball rolling without slip along the upper outside surface of a stationary sphere($\overline{OC} = L,\ \overline{PC} = r$)

The control forces F_2 and F_3 are located in the tangential plane of the other sphere, along which the center of the ball is moving, and their total power equals zero in the brachistochronic motion of the mechanical systems. Thus, during motion the total mechanical energy is maintained. The initial value of the mechanical energy is specified, and the initial and final position of the ball's center is defined by the spherical coordinate system angles φ and θ, the ball orientation (Euler angles) not being considered in this problem. Such mechanical system is nonholonomic, and dynamic equations in this paper are derived using the general theorems of mechanics.
The theorems on the motion of the center of mass and the change of the kinetic moment for the center of mass read:

$$m\vec{a}_C = \vec{F}_R^S, \qquad \dot{\vec{L}}_C = \vec{M}_C^S, \tag{1}$$

where the velocities and accelerations of the mass center in this coordinate system are:

$$\begin{aligned} &v_{C1} = 0, v_{C2} = L\dot{\varphi}\cos\theta\,, v_{C3} = L\dot{\theta} \\ &a_{C1} = -L\left(\dot{\varphi}^2 cos^2\theta + \dot{\theta}^2\right), a_{C2} = L\left(\ddot{\varphi}\cos\theta - 2\dot{\varphi}\,\dot{\theta} sin\,\theta\right), \\ &a_{C3} = L\left(\ddot{\theta} + \dot{\varphi}^2\,sin\,\theta\cos\theta\right), \end{aligned} \tag{2}$$

and dynamic quantities in expressions (1) are given by expressions:

$$\begin{aligned}&\vec{L}_C = J\vec{\omega},\ J = \frac{2}{5}mr^2,\\&\vec{F}_R^S = m\vec{g} + \vec{N} + \vec{F}_2 + \vec{F}_3 + \vec{R}_2 + \vec{R}_3,\\&\vec{M}_C^S = rR_3\vec{e}_2 - rR_2\vec{e}_3.\end{aligned} \tag{3}$$

Nonholonomic constraints are obtained from the condition that there is no slip at the contact point:

$$\vec{v}_C = \vec{\omega} \times (\vec{r}_C - \vec{r}_P) \quad \Rightarrow \quad L\dot{\varphi}\cos\theta = r\omega_3 \ \wedge \ L\dot{\theta} = -r\omega_2, \tag{4}$$

where $\omega_i, i = 1,2,3$ are projections of the ball's angular velocity onto the movable coordinate system at point C (Fig. 4).
Now, dynamic equations can be written in the form:

$$\begin{aligned}&-\frac{mr^2}{L}(\omega_2^2 + \omega_3^2) = N - mg\sin\theta\,,\\&m(r\dot{\omega}_3 + r^2\omega_2\omega_3\tan\theta/L) = R_2 + F_2,\\&m(-r\dot{\omega}_2 + r^2\omega_3^2\tan\theta/L) = R_3 + F_3 - mg\cos\theta,\\&J\dot{\omega}_1 = 0,\\&J(\dot{\omega}_2 - r\omega_3^2\tan\theta/L - r\omega_1\omega_3/L) = rR_3,\\&J(\dot{\omega}_3 + r\omega_2\omega_3\tan\theta/L + r\omega_1\omega_2/L) = -rR_2.\end{aligned} \tag{5}$$

The total power of control forces in generalized brachistochrone problems equals zero [11]:

$$F_2v_{C2} + F_3v_{C3} = 0 \ \Rightarrow \ F_2\omega_3 - F_3\omega_2 = 0, \tag{6}$$

so that from dynamic equations (5) the conservation of the total mechanical energy follows

$$\frac{1}{2}J\omega_1^2 + \frac{1}{2}(J + mr^2)(\omega_2^2 + \omega_3^2) + mgL\sin\theta = E. \tag{7}$$

as well as the maintenance of the projection of the ball's angular velocity onto the radial direction:

$$\omega_1 = C_1. \tag{8}$$

By introducing dimensionless variables:

$$\begin{aligned}&r' = r/r = 1,\ L' = L/r = 3,\ t' = t\sqrt{\frac{g}{r}}\,,\ \omega_i' = \omega_i\sqrt{\frac{r}{g}},\\&N' = N/(mg), F_i' = F_i/(mg)\,,\ R_i' = R_i/(mg)\,,\ E' = E/(mg).\end{aligned} \tag{9}$$

in the text below, the label "prim" will be removed, and all expressions will be in dimensionless variables.

The laws of change in the control forces are of the dimensionless form:

$$F_2 = \frac{2}{15}\omega_1\omega_2 + \frac{7}{15}\omega_2\omega_3\tan\theta + \frac{7}{5}\dot{\omega}_3,$$
$$F_3 = \frac{2}{15}\omega_1\omega_3 + \frac{7}{15}\omega_3^2\tan\theta - \frac{7}{5}\dot{\omega}_2 + \cos\theta\,, \tag{10}$$

and the reactions of constraints are:

$$N = \sin\theta - \frac{1}{3}(\omega_2^2 + \omega_3^2) = \frac{17}{7}\sin\theta - \frac{10}{21}(E - C_1^2/5),$$
$$R_2 = -\frac{2}{15}(\omega_1\omega_2 + \omega_2\omega_3\tan\theta + 3\dot{\omega}_3), \tag{11}$$
$$R_3 = -\frac{2}{15}(\omega_1\omega_3 + \omega_3^2\tan\theta - 3\dot{\omega}_2).$$

Detachment angle ($N(\theta_{perf.rough}) = 0$)for the case of perfect rough sphere and zero initial value of the angular velocity projection ω_1 onto the radial direction:$\sin(\theta_{perf.rough}) = \frac{10E}{51}$.

Condition for non-slip occurrence based on Coulomb's laws of sliding friction:

$$R = \sqrt{R_2^2 + R_3^2} \le \mu N. \tag{12}$$

The next section considers the brachistochrone problem as a problem of optimal control when the sphere is perfectly rough and there are no restrictions (12), and section 3 deals with the case of a real rough sphere, when given restriction must be taken into account.

3. The brachistochrone problem for the case of perfect rough sphere

Let us formulate the problem of optimal control that will be solved using Pontryagin's maximum principle [12].
Let it be known at the initial moment:

$$t_0 = 0,\ \varphi(t_0) = 0,\ \theta(t_0) = \theta_0,\ E(t_0) = E, \tag{13}$$

and let it be known at the final moment:

$$t_f = ?,\ \varphi(t_f) = \varphi_f,\ \theta(t_f) = \theta_f, \tag{14}$$

where:

$$\theta_{perf.rough} < \theta_f \le \theta_0 < \frac{\pi}{2},\ 0 \le \varphi_f < \pi. \tag{15}$$

Differential equations of this problem of optimal control can be obtained from (4), where it has been taken that $L = 3$:

$$\dot{\psi} = \omega_1,\ \dot{\varphi} = \frac{\omega_3}{3\cos\theta},\ \dot{\theta} = \frac{-\omega_2}{3}. \tag{16}$$

Energy integral (7) obtains the form:

$$2\omega_1^2 + 7(\omega_2^2 + \omega_3^2) + 30\sin\theta = 10E. \tag{17}$$

The optimal control problemis as follows: For the mechanical system, for specified differential equations(16) and initial (13) and end conditions (14), determine the motion of the system in minimum time while maintaining the energy integral (17).
In order to apply the maximum principle, cost functional is formed:

$$t_f = \int_{t_0}^{t_f} dt. \tag{18}$$

Pontryagin's function:

$$\begin{aligned} H = &-1 + \lambda_\psi\omega_1 + \lambda_\varphi\omega_3/(3\cos\theta) - \lambda_\theta\omega_2/3 + \\ &+\rho(2\omega_1^2 + 7(\omega_2^2 + \omega_3^2) + 30\sin\theta - 10E), \end{aligned} \tag{19}$$

and costate system:

$$\dot{\lambda}_\psi = 0,\ \dot{\lambda}_\varphi = 0,\ \dot{\lambda}_\theta = -\lambda_\varphi\omega_3\sin\theta/(3\cos^2\theta) - 30\rho\cos\theta. \tag{20}$$

Transversality conditions are of the form:

$$\lambda_\psi(0) = 0,\ \lambda_\psi(t_f) = 0. \tag{21}$$

Maximum principle yields the conditions:

$$\begin{aligned} &\frac{\partial H}{\partial\omega_i} = 0 \ \Rightarrow\ \omega_1 = -\lambda_\psi/(4\rho), \\ &\omega_2 = \lambda_\theta/(42\rho),\ \omega_3 = -\lambda_\varphi/(42\rho\cos\theta). \end{aligned} \tag{22}$$

The final moment is indefinite so that:

$$H = 0 \ \Rightarrow\ \rho = 1/(20(3\sin\theta - E)). \tag{23}$$

There is no rotation around the axis in the radial direction:

$$\left(\dot{\lambda}_\psi(t) = 0 \wedge \lambda_\psi(0) = 0 \wedge \lambda_\psi(t_f) = 0\right) \Rightarrow \lambda_\psi(t) = 0 \Rightarrow \omega_1(t) = 0. \tag{24}$$

On the brachistochronic motion, based on (11) and (20-24), the components of the constraint reaction are:

$$N = (51\sin\theta - 10E)/21\,,\quad R = \sqrt{R_2^2 + R_3^2} = 2\cos\theta/7. \tag{25}$$

When the ball is rolling down, the angle is decreasing, the normal component of the constraint reaction N is decreasing too, whereas tangential component R is increasing, so that in term of the slip, the critical slip is at the end of motion.
Now, the discussion on the possible values of the task parameters θ_f, E, μ can be conducted, where $\theta_0 = 1.5$

1. $N \geq 0 \Rightarrow E \leq 51\sin\theta_f/10$ the point must be below the red colored surface in Fig. 5;
2. $R \leq \mu N \Rightarrow E \leq 3(17\mu\sin\theta_f - 2\cos\theta_f)/(10\mu)$ the point must be below the blue colored surface;
3. $E \geq 3\sin\theta_0$ the point must be above the green colored surface because the initial kinetic energy is non-negative

This means that the representative point (θ_f, μ, E) corresponding to the task parameters must be located simultaneously above the green and below blue surface. The red surface is always above the blue one, which means that if there is no slip at the end point, there will be no detachment either.

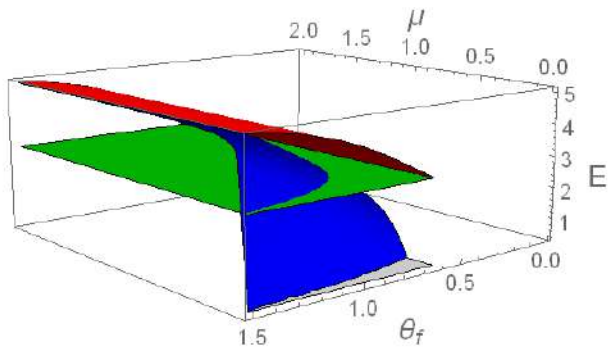


Fig. 5 Discussion on the possible positions of the representative point(θ_f, μ, E)

Also, it is noticeable that identical value R of the constraint reaction in the tangential plane is obtained from (10) and (11) when the motion is non-controlled $F_2 = F_3 = 0$ and when $\omega_1 = 0$, so the above discussion is applicable to non-controlled free rolling of a heavy ball along a real rough sphere.
Numerical solution parameters for the case of perfect rough surface in this example is:

$$\theta_0 = 1.5,\ \ E = 3,\ \ \theta_f = \pi/4\,,\ \varphi_f = \pi/2\,. \tag{26}$$

Two-point boundary value problem of the maximum principle, in this case, has differential equations:

$$\begin{aligned}
&\dot{\varphi} = 10\lambda_\varphi(1-\sin\theta)/(21\cos^2\theta),\\
&\dot{\theta} = 10\lambda_\theta(1-\sin\theta)/21,\\
&\dot{\lambda}_\varphi = 0,\\
&\dot{\lambda}_\theta = -10\lambda_\varphi^2(1-\sin\theta)\sin\theta/(21\cos^3\theta) + \cos\theta/(2(1-\sin\theta)),
\end{aligned} \tag{27}$$

with initial conditions:

$$\begin{aligned}
&t_0 = 0,\ \varphi(t_0) = 0,\ \theta(t_0) = \theta_0,\ \lambda_\varphi(t_0) = ?,\\
&H(t_0) = 0 \Rightarrow \lambda_\theta(t_0) = -\sqrt{21/\left(10(1-\sin\theta_0)\right) - \lambda_\varphi^2(t_0)/\cos^2\theta_0}.
\end{aligned} \tag{28}$$

We're choosing $\lambda_\varphi(t_0)$, t_f and shooting: $\varphi(t_f) = \varphi_f$, $\theta(t_f) = \theta_f$ where $\lambda_\varphi^2(t_0) \le 21\cos^2\theta_0/(10(1-\sin\theta_0)), t_f > 0$ Numerical solution this two-point boundary value problem is: $t_f = 5.882531$, $\lambda_\varphi(0) = 1.08448$ and diagrams of the law of change in polar angles and required coefficient of friction $\mu = R/N = 2\cos\theta/(17\sin\theta - 10)$ are given in Fig. 6.

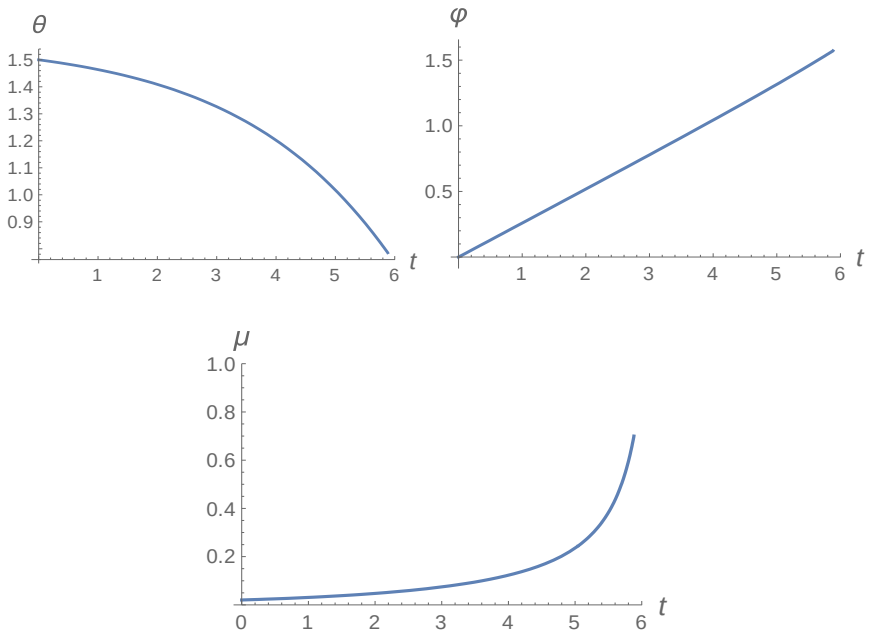


Fig. 6 Brachistochronic rolling of a ball along a perfect rough surface

A new form of differential equations (27):

$$\frac{d\varphi}{d\theta}=\frac{\lambda_\varphi}{\lambda_\theta\cos^2\theta}=\frac{\lambda_\varphi}{-\cos^2\theta\sqrt{21/\left(10(1-\sin\theta)\right)-\lambda_\varphi^2/\cos^2\theta}},$$
$$\frac{dt}{d\theta}=\frac{-3}{\omega_2}=\frac{-21}{10\lambda_\theta(\sin\theta-1)}=\frac{-21}{10(-\sqrt{21/\left(10(1-\sin\theta)\right)-\lambda_\varphi^2/\cos^2\theta})(\sin\theta-1)}, \quad (29)$$

gives us that the solution can be also reached through the squares, given that the angle θ is monotonically decreasing over time:

$$\varphi_f=\int_{\theta_f}^{\theta_0}\frac{\lambda_\varphi(0)d\theta}{\cos^2\theta\sqrt{21/\left(10(1-\sin\theta)\right)-\lambda_\varphi^2(0)/\cos^2\theta}},$$
$$t_f=\int_{\theta_f}^{\theta_0}\frac{21d\theta}{10\sqrt{21/\left(10(1-\sin\theta)\right)-\lambda_\varphi^2(0)/\cos^2\theta}(1-\sin\theta)}. \quad (30)$$

By solving integral equations (30), it is easier to obtain the already obtained numerical solutions for $(t_f,\lambda_\varphi(0))$.
Necessary value of the Coulomb coefficient of friction at the beginning of motion is:

$$\mu^*=R(0)/N(0)=2\cos\theta_0/(17\sin\theta_0-10)=0.0203343\,, \quad (31)$$

whereas necessary value of the Coulomb coefficient of friction at the end of motion is:

$$\mu^{**}=R(t_f)/N(t_f)=2\cos\theta_f/(17\sin\theta_f-10)=0.699823. \quad (32)$$

The assumption that surfaces are perfectly rough in this task of the brachistochronic motion is satisfied only for $\mu\geq\mu^{**}$ (e.g., rubber on rubber and glass on glass). If this is not the case (e.g., wood on wood), slip on the brachistochronic motion would occur earlier. The analysis also holds for non-controlled motion of a heavy ball.

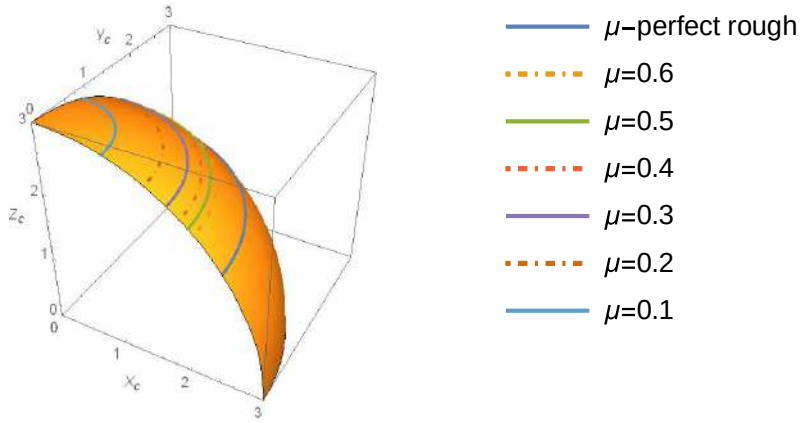


Fig. 7 Slip circles depending on the Coulomb coefficient of friction

Figure 7 shows slip circles that yield the lowest possible end positions during brachistochronic motion, depending on the coefficient of friction. This means that the lower the coefficient of friction, the more restricted the region where the surface is considered perfectly rough. It may even happen that for $\mu < \mu^*$ slip occurs at the beginning of motion. In the case when $\mu^* \leq \mu < \mu^{**}$ the problem of optimal control should include the restriction (12) and the task of optimal control becomes considerably complicated. Such possibility will be analyzed in the section below.

4. The brachistochrone problem for the case of imperfect rough sphere ($\mu^* \leq \mu < \mu^{**}$)

Let us observe the brachistochrone problem with numerical parameters (26) and seek the solutions in a neighborhood of $\mu = \mu^{**}$ where over the entire interval:

$$\omega_2(t) > 0,\ \dot{\omega}_2(t) > 0,\ \omega_3(t) = \sqrt{30\,(1 - sin\,\theta(t))/7 - \omega_2^2(t)} > 0. \tag{33}$$

Differential equations written through the theta angle as independent variables are:

$$\begin{aligned} &\frac{dt}{d\theta} = \frac{-3}{\omega_2}, \\ &\frac{d\varphi}{d\theta} = -\sqrt{30(1 - sin\,\theta)/7 - \omega_2^2}/(\omega_2\, cos\,\theta), \\ &\frac{d\omega_2}{d\theta} = \frac{-3}{\omega_2}\left((30(1 - sin\,\theta)/7 - \omega_2^2)\, tan\,\theta/3 + 5R_3/2\right). \end{aligned} \tag{34}$$

Determination of the set of permissible values of the control $u = R_3$ is based on (5), (6) and (12):

$$\begin{aligned} &R_3\omega_2 - R_2\omega_3 = 2\omega_2 \cos\theta/7 \\ &R_2^2 + R_3^2 \leq \mu^2 N^2 \end{aligned} \Rightarrow$$

$$\begin{aligned} &R_3 \in \left[R_3^{min}, R_3^{max}\right], \\ &R_3^{max} = \frac{2\cos\theta\omega_2^2 + \omega_3\sqrt{49\mu^2N^2\left(\omega_2^2+\omega_3^2\right) - 4\omega_2^2\cos^2\theta}}{7(\omega_2^2+\omega_3^2)}, \\ &R_3^{min} = \frac{2\cos\theta\omega_2^2 - \omega_3\sqrt{49\mu^2N^2\left(\omega_2^2+\omega_3^2\right) - 4\omega_2^2\cos^2\theta}}{7(\omega_2^2+\omega_3^2)}. \end{aligned} \tag{35}$$

In order that the set of permissible controls will not be an empty set:

$$\omega_2^2 \leq \frac{49\mu^2N^2\left(\omega_2^2+\omega_3^2\right)}{4\,cos^2\,\theta} = \frac{30\mu^2(17\,sin\,\theta - 10)^2(1 - sin\,\theta)}{28\,cos^2\,\theta}. \tag{36}$$

For the solutions in a neighborhood of $\mu = \mu^{**}$ it is sufficient to introduce the restriction:

$$\omega_2(t_f) \le \mu(17\sin\theta_f - 10)\sqrt{30(1-\sin\theta_f)/28}/\cos\theta_f. \quad (37)$$

Let us assume that this condition is satisfied in order to avoid a very complex task of optimal control. It will be checked only subsequently after numerical solution is obtained.

Cost functional in this new problem of time minimization is:

$$t_f = \int_{\theta_1}^{\theta_0} \frac{3d\theta}{\omega_2}. \quad (38)$$

whereas Pontryagin's function

$$\begin{aligned} H = {} & \frac{-3}{\omega_2} - \frac{\lambda_\varphi\sqrt{30(1-\sin\theta)/7-\omega_2^2}}{\omega_2\cos\theta} - \\ & -\frac{3\lambda_{\omega 2}}{\omega_2}\left((30\,(1-\sin\theta)/7-\omega_2^2)\tan\theta/3+\frac{5u}{2}\right)+ \\ & +\rho(u^2+\frac{\omega_2^2(u-2\cos\theta/7)^2}{30(1-\sin\theta)/7-\omega_2^2}-\frac{\mu^2(17\sin\theta-10)^2}{49}). \end{aligned} \quad (39)$$

Here, the maximum principle gives the following possibilities for optimal control over some of the intervals

$$\begin{aligned} u_{opt} &= \begin{cases} u_s,\ \lambda_{\omega 2}=0 \\ R_3^{max},\ \lambda_{\omega 2}<0 \\ R_3^{min},\ \lambda_{\omega 2}>0 \end{cases} \\ \rho &= \begin{cases} 0,\ \lambda_{\omega 2}=0 \\ \frac{15\lambda_{\omega 2}(7\omega_2^2-30+30\sin\theta)}{8\omega_2(\omega_2^2\cos\theta+15R_3^{max}(-1+\sin\theta))},\ \lambda_{\omega 2}<0 \\ \frac{15\lambda_{\omega 2}(7\omega_2^2-30+30\sin\theta)}{8\omega_2(\omega_2^2\cos\theta+15R_3^{min}(-1+\sin\theta))}.\ \lambda_{\omega 2}>0 \end{cases} \end{aligned} \quad (40)$$

The costate variable $\lambda_{\omega 2}$ in this task of optimal control has a role of so-called "switching function", so that its sign on a non-singular part determines whether the control will be on the upper or lower limit. On a singular part it equals zero. Calculation of the singular optimal control [13] can be performed in the following manner:

$$\begin{aligned} & \lambda_{\omega 2}=0 \Rightarrow \frac{d\lambda_{\omega 2}}{d\theta} = -\frac{\partial H}{\partial\omega_2} = 0 \Rightarrow\ \lambda_\varphi = -\frac{\sqrt{7}\cos\theta\sqrt{30(1-\sin\theta)-7\omega_2^2}}{10(1-\sin\theta)}, \\ & \frac{d^2\lambda_{\omega 2}}{d\theta^2}=0 \Rightarrow u_s = \frac{2}{7}\cos\theta\,\frac{7\omega_2^2+15\sin\theta-15}{15(\sin\theta-1)} \Rightarrow R = \frac{2}{7}\cos\theta\,. \end{aligned} \quad (41)$$

On a non-singular part:

$$R = \mu N = \mu(17\sin\theta - 10)/7. \quad (42)$$

The assumed structure of the optimal control in a neighborhood of $\mu = \mu^{**}$ is:

$$u_{opt} = \begin{cases} u_s, & \theta_0 \geq \theta > \theta^* \\ R_3^{min}, & \theta^* \geq \theta \leq \theta_f \end{cases}. \tag{43}$$

Numerical solution procedure (two-parameter shooting) consists of the following steps:

1. by choosing $\omega_2(\theta_0)$ and θ^*, numerical integration of the basic system over the interval$[\theta_0, \theta^*]$

$$\begin{aligned} &\frac{d\varphi}{d\theta} = -\sqrt{30(1-\sin\theta)/7 - \omega_2^2}/(\omega_2 \cos\theta), \quad \varphi(\theta_0) = 0 \\ &\frac{d\omega_2}{d\theta} = -\frac{3}{\omega_2}\left((30\,(1-\sin\theta)/7 - \omega_2^2)\tan\theta/3 + 5u_s/2\right). \end{aligned} \tag{44}$$

2. and basic and costate system over the interval $[\theta^*, \theta_f]$

$$\begin{aligned} &\frac{d\varphi}{d\theta} = -\sqrt{30(1-\sin\theta)/7 - \omega_2^2}/(\omega_2 \cos\theta), \\ &\frac{d\omega_2}{d\theta} = -\frac{3}{\omega_2}\left((30\,(1-\sin\theta)/7 - \omega_2^2)\tan\theta/3 + 5R_3^{min}/2\right), \\ &\frac{d\lambda_\varphi}{d\theta} = 0, \qquad \lambda_\varphi(\theta^*) = -\frac{\sqrt{7}\cos\theta^*\sqrt{30(1-\sin\theta^*) - 7\omega_2^2(\theta^*)}}{10(1-\sin\theta^*)}, \\ &\frac{d\lambda_{\omega 2}}{d\theta} = -\frac{\partial H}{\partial \omega_2} = \cdots, \qquad \lambda_{\omega 2}(\theta^*) = 0. \end{aligned} \tag{45}$$

fulfillment of conditions is $\varphi(\theta_f) = \varphi_f$, $\lambda_{\omega 2}(\theta_f) = 0$ ensured.
Numerical example was done for the value of Coulomb's coefficient $\mu = 0.6$ and numerical solutions were obtained $\theta^* = 0.821251$, $\omega_2(\theta_0) = 0.0878961$.

The diagram of the "switching function" is given in Fig. 8, the dashed line ($\mu = 0.6$). The final part of the motion is shown only, the solid line indicating the case when it is possible to a have a singular part over the entire interval ($\mu \geq \mu^{**}$), and the dash-dotted line designating a boundary case ($\mu = \bar{\mu}$).

By gradually decreasing the coefficient of friction and conducting the numerical solution procedure, it can be established that conditions (37) are disrupted for $\bar{\mu} = 0.576383$.

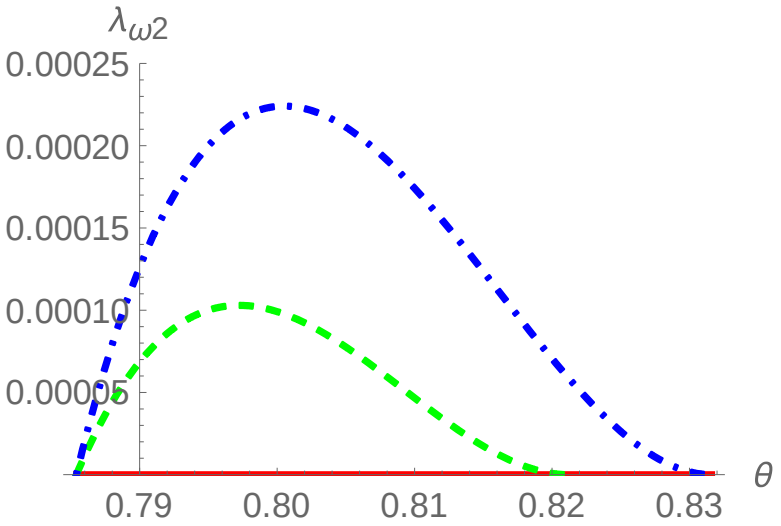


Fig. 8 The switching function $\lambda_{\omega 2}$

Here, it is also necessary to check whether the conditions (40) of the maximum principle are fulfilled after numerical solutions is done

$$\lambda_{\omega 2}(\theta) \geq 0, \quad \theta^* \leq \theta \leq \theta_f, \tag{46}$$

as well as the condition (37).
If the coefficient of friction is lower than $\bar{\mu}$ the structure of optimal control (43) changes, and the numerical solution procedure becomes more complex.

5. Conclusions

The problem of brachistochronic motion of a heavy uniform ball rolling without slip along the upper outside surface of an imperfect rough stationary sphere is solved.
In the first part of the paper the determination of the brachistochronic motion is solved as the problem of optimal control using Pontryagin's maximum principle. Three projections of the ball's angular velocity onto the base vectors of the spherical coordinate system are taken for controls. The two-point boundary value problem, which is reduced to the two-parameter shooting of one coordinate of the conjugate vector and end moment, is solved. It is shown that there is no angular velocity projection onto the radial direction. This solution corresponds to the motion of the heavy ball along a perfect rough sphere, because it is necessary to ensure unrealistically high Coulomb coefficient of sliding friction.
It is shown that mutual detachment of the bodies cannot occur before their mutual slipping at the contact point. A corresponding numerical example is given, with graphical representation of the effects of initial energy values, Coulomb coefficient and ultimate height of the ball center on the solution

structure, in this case. The review highlights regions where it is possible to obtain a singular control across the entire motion or a combination of a singular and non-singular part of the optimal trajectory.

The second part provides the following discussion: if the constraint between the sphere and the ball is imperfectly rough, the formulation of the optimal control problem should include restrictions to the ratio between the tangential and the normal components of the reaction of constraint. Here, the problem of optimal control is formulated in such way that the tangential component of the reaction of constraint is taken for the control, with the restriction resulting from Coulomb laws of sliding friction. The problem thus formulated belongs to the theory of singular optimal controls, and the solution that satisfies the Maximum principle consists of a singular part at the beginning of motion and a non-singular part, during which the ratio between mentioned components has maximum possible value that concrete surfaces can achieve.

Futher research of this problem involves: research on the structure of control and appropriate numerical solutions for $\mu^* < \mu < \bar{\mu}$, generalization of the result to the ball rolling on the stationary rotating surface and generalization of the result to the ball rolling on the moving surface which is rotating about vertical axis at constant angular velocity.

References

[1] A.V. Zarodnyuk, *Optimization of controlled descent and generalized brachistochrone problems (in Russian)*, PhD, Moscow State University, Faculty of Mechanics and Mathematics, 2018.

[2] A. S. Vondrukhov and Yu. F. Golubev, *Optimal trajectories in brachistochrone problem with Coulomb friction, Journal of Computer and Systems Sciences International* **55**(2016) 341–348

[3] A. Sumbatov, *Brachistochrone with Coulomb friction as the solution of an isoperimetrical variational problem, International Journal of Non-Linear Mechanics* **88** (2016) 135-141

[4] S.Šalinić, A. Obradović, Z. Mitrović and S. Rusov, *Brachistochrone with limited reaction of constraint in an arbitrary force field, Nonlinear Dynamics* **69** (2012) 211-222

[5] G. Perantoni and D.J.N. Limebeer, *Time-optimal control of rolling bodies, International Journal of Control* **86** (2013) 2006-2021

[6] S. Šalinić, A. Obradović, Z. Mitrović and S. Rusov, *On the brachistochronic motion of the Chaplygin sleigh, Acta Mechanica* **224** (2013) 2127-2141

[7] C. Caratheodory, *Der Schlitten, ZAMM-Z. Angew. Math. Me.***13** (1933) 71–76

[8] A. Obradović, S. Šalinić , R. Radulović, *Vertical disk rolling on a horizontal plane without slip, Theoretical and Applied Mechanics* **44** (2017), 237-254,

[9] S. K. Soltakhanov, M. P. Yushkov andS. A. Zegzhda,*Mechanicsof Non-Holonomic Systems: A New Class of ControlSystems*, Springer-Verlag, Berlin Heidelberg, 2009.

[10] R. Radulović, A. Obradović, S. Šalinić and Z. Mitrović, *The brachistochronic motion of a wheeled vehicle*, *Nonlinear Dynamics* **87**(2017)191-205

[11] Dj. Djukic, On the Brachistochronic Motion of a Dynamic System, ActaMechanica **32**(1976) 181-186

[12] L. S. Pontryagin, V. G. Boltyansky, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, Interscience Publishers, John Wiley and Sons, New York, 1962.

[13] R. Gabasov, and F. M. Kirillova, Singular Optimal Controls*(in Russian)*, Nauka, Moscow, 1973.

Nonlinear model of fluid diffusion in porous solids

N.M. Ostrovskii*

Euro Gas doo, Senčanski put 75, 24000 Subotica, Serbia

ABSTRACT

Proposed model is based on interpretation of fluid diffusion in porous particles, as quasi-chemical reaction of second order. The diffusion flux is expressed in Onsager form $J_i = \Sigma\ L_i X_{ji}$, with non-linear thermodynamic force X_{ji}. The driving force is considered as a product $X_{ji} = x_j (1–x_i)$ of fullness degree by fluid (x_j) of one type of pores, and an emptiness ($1–x_i$) of neighbor type of pores. The model is designed to describe the drying dynamics of porous polypropylene particles impregnated with a solvent. Three types of "moisture" were assumed inside porous particles, which are corresponded to three sizes of pores. An advantage of such a model is that instead of empirical dependence of diffusivity on moisture (usually used for description of such processes), the measured characteristics of porous structure become parameters of model. Results of experiments and model formulation are presented in this report.

1. Introduction

Diffusion is the physical process by which matter is transported from one part of a system to another as a result of random molecular motions [1]. Diffusion is driven by a gradient in concentration. This was recognized by Fick (1855), adopting to diffusion the mathematical equation of heat conduction derived earlier by Fourier (1822). According to Fick's law, the diffusion flux J_i is proportional to the negative of concentration gradient:

* e-mail address: n-ostr@yandex.ru

$$J_i = -D_i C_s \frac{dy_i}{dx} \qquad \text{(n1)}$$

where y_i – molar fraction of *i*-th component in fluid mixture; J_i – flux per unit area of section; D_i – diffusion coefficient; C_s – total concentration of matter; x – space coordinate normal to the section.

The differential equation of diffusion dynamics is derived from (n1), and for one-dimensional medium, at constant D_i, reducing to [2]:

$$\frac{\partial y_i}{\partial t} = D_i \frac{\partial^2 y_i}{\partial x^2} \qquad \text{(n2)}$$

In multicomponent systems the Stefan-Maxwell model (1867-1871) is more commonly used because it considers the molecular collisions among the different types of gas molecules [3]:

$$\frac{dy_i}{dx} = \sum_{j=1}^{n} \frac{y_i J_j - y_j J_i}{C_s D_{ij}}, \qquad \frac{\partial \overline{y}}{\partial t} = \frac{\partial}{\partial x}\left\{ [B]^{-1} \frac{\partial \overline{y}}{\partial x} \right\} \qquad \text{(n3)}$$

For the description of evolution of dispersed particles population the Fokker-Planck equation (1914-1917) is usually used that is named also as convection-diffusion equation [4]:

$$\frac{\partial f(x)}{\partial t} = -\frac{\partial}{\partial x}\big(F(x)\, f(x)\big) + \frac{1}{2}\frac{\partial^2}{\partial x^2}\big(D(x)\, f(x)\big) \qquad \text{(n4)}$$

where $f(x)$ – probability density function of particles velocity; $F(x)$, $D(x)$ – drift and diffusion coefficients.

The mobility of atoms and ions in solids (metal, alloy, oxide, etc.) is also interpreted as diffusion phenomena. Frenkel (1926) [5] has presented diffusion process in condensed matter as an ensemble of elementary jumps and quasi-chemical interactions of atoms and defects.

Eyring [6] applied his theory of absolute reaction rates to this quasi-chemical representation of diffusion. Such application of mass action law for diffusion leads to various nonlinear versions of Fick's equation. Assuming that all particles can exchange their positions with their closest neighbors, Gorban et al. [7] have derived following nonlinear equation:

$$\frac{\partial y_i}{\partial t} = D_{ij}\left(y_j \Delta y_i - y_i \Delta y_j\right) \qquad \text{(n5)}$$

where Δ is the Laplace operator.

Gorban with coworkers have developed general approach to nonlinear multicomponent diffusion based on the idea of the reaction mechanism borrowed from chemical kinetics [7]. Just their works gave rise to the development of the model that is presented in this paper.

2. Diffusion coefficient

The main parameter in all diffusion models is the diffusion coefficient, or diffusivity (D, cm^2/s). It depends on temperature, pressure (in gases), and on molecular characteristics of diffusion medium. The dependence of diffusion coefficient on these parameters for gases is usually expressed using Chapman-Enskog theory (1939), particular by Hirschfelder formula [8]:

$$D_{ij} = \frac{A\,T^{1.5}}{P\,\Omega_{ij}\,\sigma_{ij}^2}\left(\frac{1}{M_i}+\frac{1}{M_j}\right)^{1/2} \qquad \text{(n6)}$$

where T – temperature; P – pressure; Ω – collision integral; σ – Lennard-Jones' interaction constants; M – molar mass.

The diffusivity in liquids is $\sim 10^4$ times lower than in gases. For Brownian particle in dilute solutions the equation of Stokes-Einstein (1906) is valid:

$$D_i = \frac{k\,T}{6\pi\, r_i\,\mu_i} \qquad \text{(n7a)}$$

where k – Boltzmann constant; r – particle radius; μ – solvent viscosity.

More general equation was proposed by Wilke and Chang (1955) that is widely used for calculation of diffusivity in various liquid mixtures [9]:

$$D_{ij} = A\frac{T}{\mu_j}\left(\varphi M_j\right)^{1/2}\left(V_i\right)^{-3/5} \qquad \text{(n7b)}$$

where φ – characteristic constant of solvent; V – molar volume.

Diffusion in solids is considered as Frenkel jumps of atoms (ions) and vacancies, so that diffusivity D_s characterizes the intensity of jumps [10]:

$$D_S = f\,d^2\,\nu_o\,exp\left(-\frac{E_m}{k_B T}\right) \qquad \text{(n8a)}$$

where f – correlation factor; d – jump distance; ν_o – frequency of atoms vibration; E_m – height of potential barrier (migration energy); k_B – Boltzmann constant.

Similar formula was suggested even early by Dushman and Langmuir (1922) and is also widely used for calculating diffusivity in metals [11]:

$$D_S = \frac{Q d^2}{N_A h} \, exp\left(-\frac{Q}{RT} \right) \qquad \text{(n8b)}$$

where N_A – Avogadro's number; h – Planck's constant; Q – activation energy of self-diffusion; R – ideal gas constant.

Typical intervals of diffusivity are presented below:

Diffusivity	In gases	In liquids	In solids
D_i, cm^2/s	$10^{-1} \div 10^{0}$	$10^{-6} \div 10^{-4}$	$10^{-17} \div 10^{-11}$

3. Diffusivity in porous structure

In many processes such as adsorption, drying, solid fuel combustion, and catalytic reactions, the mass transfer of fluid in porous media is often the rate limiting step. Usually, it is interpreted as diffusion process undergoing to Fick's law, but with special diffusion coefficient.

In a single pore, the Knudsen diffusion (1909) takes place when collisions of gas molecules with the pore walls dominate, rather than intermolecular collisions. Corresponding Knudsen diffusivity is expressed as follows [1]:

$$D_K = \frac{d_p}{3} \left(\frac{8RT}{\pi M} \right)^{1/2} \qquad \text{(n9)}$$

where d_p – pore diameter; T – temperature; M – molar mass.

Despite of heterogeneous porous structure, the total mass transfer in porous particle is often characterized by an effective diffusivity D_e, which is a lumped parameter. There are several dependences of D_e on such parameters as particle porosity ε and tortuosity of pores τ [12, 13]:

$$D_e = D_m \frac{2\varepsilon}{3-\varepsilon}, \quad D_e = D_m \frac{\varepsilon}{\tau}, \quad D_e = D_m \left[1-(1-\varepsilon)^a \right] (1-\gamma)^b \qquad \text{(n10)}$$

where D_m – molecular diffusivity; a, b – empirical constants; γ – fraction of blocked pores.

In bidisperse porous particles, the effective diffusivity depends on D_m in macro- and micropores, according to Wakao-Smith model (1962) [14]:

$$D_e = \varepsilon_M^2 \, D_M + \varepsilon_m^2 \, D_m \frac{1+3\varepsilon_M}{1-\varepsilon_M} \qquad \text{(n11)}$$

where D_m, D_M – diffusivity in micro- and macro-pores; ε_m, ε_M – corresponding porosity.

During modeling of some processes (drying, adsorption, et al.) the effective diffusivity D_e is calculated by fitting the model to experiments. This often requires representing D_e as a function not only of temperature T, but also of concentration C [15, 16]. Three type of such functions, proposed by Crank (1953) and Wagner (1968), are linear, hyperbolic and exponential [2, 16]:

$$D_e = D_o(T)\left(1 \pm bC\right), \quad D_e = D_o(T)\Big/\left(1+bC\right), \quad D_e = D_o(T)\, e^{bC} \qquad \text{(n12)}$$

In case of drying of porous materials, the estimated diffusion coefficient of the system corresponds to an effective moisture diffusivity D_w, which includes the effects of moisture content w, temperature T and porosity ε. Several dependences were proposed by Millington-Quirk (1961), Kiranoudis (1995) and others [15, 17]:

$$D_w = D_o \, w \big/ \varepsilon^{2/3}, \quad D_w = D_o \, e^{-b/w} \cdot e^{-E/RT} \qquad \text{(n13)}$$

The so-called reaction engineering approach to modelling the convective drying of food materials is also reduced to diffusivity as a semi-empirical function of temperature T and moisture w, like following equation of Chen and Putranto [18, 19]:

$$D_w = D_o \, exp\left\{ -\frac{E}{RT}\left(\frac{1+w}{w} \right)^b \right\} . \qquad \text{(n14)}$$

Temperature dependence $D(T)$ is usually exponential, like (n8, n13, n14), and is a natural because it based on Arrhenius equation. But when diffusivity depends on concentration $D(C)$ or $D(w)$, it means that Fick's model of diffusion is not adequate to experiment, and should be revised.

Possible reasons for such occurrences are: (i) polydisperse structure of porous material; (ii) several types of diffusion in particle (molecular, Knudsen, surface, configurative); (iii) phase transition of fluid in pores (condensation – evaporation).

Another approach is based on assumption of heterogeneous porous structure and on models of multi-phase mass transfer. Such models are applied successfully for drying of fruits because are based on their cellular structure [20, 21]. Authors postulate two phases of moisture: (i) free liquid water inside intercellular space; (ii) bound water within the cell structure. An interaction of these types of water within the macroscopic multiphase model is taken as a Mass Action Law (MAL).

Gorban et al. [7] have developed quasi-chemical models of nonlinear multicomponent diffusion and special tools in order to provide the systematic construction of the nonlinear diffusion equations with significant interaction between components. They have proposed also the suitable equations for Fick's diffusion and for Cell-jump formalism, as well as complex balance conditions for MAL application in diffusion.

Based on these approaches and experimental data [22], we have proposed the model of moisture diffusion in polydisperse porous particles. The drying of polypropylene powder was studied, where the porous structure was filled with heptane.

The model of drying has been derived assuming three types of moisture inside porous particles that correspond to three sizes of pores.

4. Experiments

For the study of moisture diffusion the polypropylene porous particles were used, obtained in slurry polymerization process where heptane is used as a solvent. Thus, heptane plays the role of "moisture". The range of particles diameter (d_p) was 50–500 μm, and the range of moisture concentration (w) was 10–20 % wt.

Drying kinetics experiments [22] were carried out by Prof. Mészáros-Szécsényi in University of Novi Sad, using TG–DTA thermal analysis on equipment Q600 SDT from TA Instruments. Measurements were made in nitrogen atmosphere (flow rates: 20 and 100 cm^3/min), with sample masses of 2-4 mg and heating rates of 5, 10 and 20 °C/min using alumina sample pan. One experiment was carried out under stepwise isothermal (SWI) conditions with a heating rate of 10 ^oC/min.

The dependences of solvent evaporation on temperature at different heating rates are presented by the corresponding TG and DTG curves in Figs. 1(a,b), while the enthalpy changes accompanying the heating of the samples is depicted by the DSC curves in Fig. 1(c).

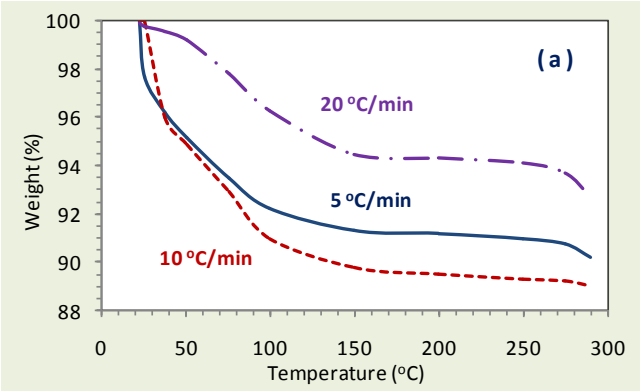


Fig. 1(a). TG curves measured at different heating rates [22].

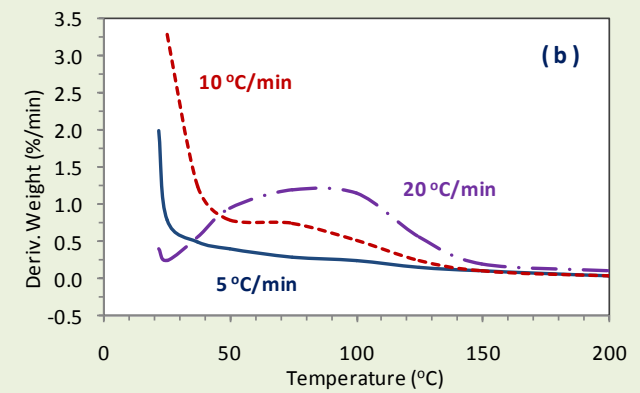


Fig. 1(b). DTG curves measured at different heating rates [22].

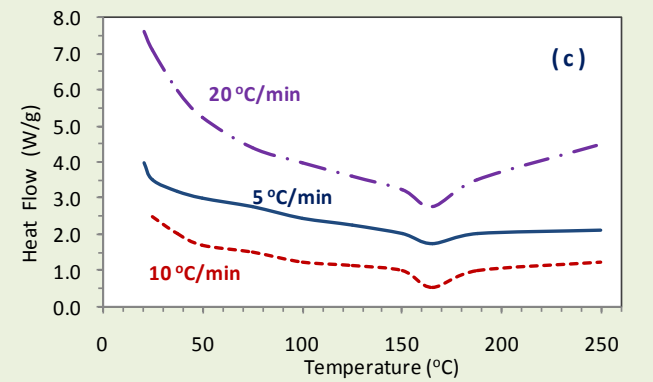
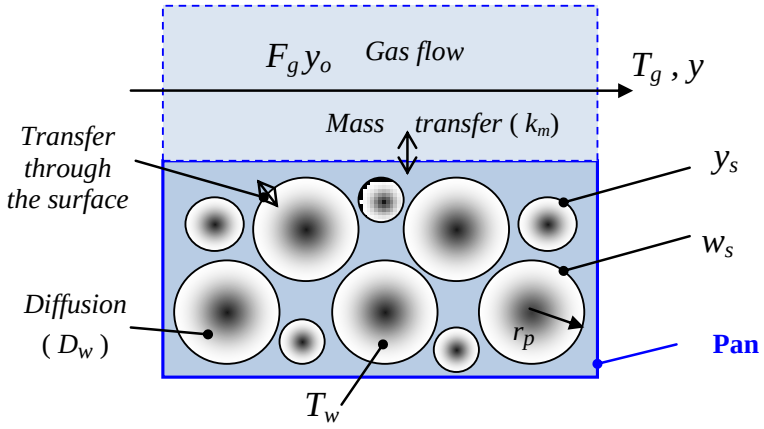


Fig. 1(c). DSC dependences at different rate of heating [22].

5. Formulation of classic model

The model of drying process in thermogravimetric experiments is based on the following scheme:



Differential mass balance of moisture in sample of polymer powder:

[Rate of the mass decreasing] = [Rate of moisture removal with gas]

$$G_{pol}\frac{d\,w}{d\,t} = -F_g\,(y - y_o) \tag{1}$$

and in gas phase flow:

$$V_g\frac{d\,y}{d\,t} + F_g\,(y - y_o) = k_m\,S_s\,(y_s - y) \tag{2}$$

G_{pol} – mass of polimer, mg; w – relative moisture of powder; F_g – gas flow rate, cm^3/min; V_g – gas volume, cm^3; y_o , y – inlet and outlet gas humidity, μg/cm^3; y_s – humidity near particle, μg/cm^3; S_s – cross section of pan, cm^2; k_m – mass transfer coefficient (from the sample to gas phase), cm^3/cm^2 min.

Balance equation for the gas phase near particles surface:

$$V_s\,\varepsilon_s\frac{d\,y_s}{d\,t} + k_m\,S_s\,(y_s - y) = D_w\,V_s\,(1-\varepsilon_s)\,a_p\,G_{pol}\left.\frac{d\,w}{d\,r}\right|_{r=r_p} \tag{3}$$

$$D_w = D_L\,\varepsilon_p / \tau_p\ , \qquad a_p = 3\,(1-\varepsilon_s)/r_p$$

V_s – volume of sample in a pan, cm^3; r_p – radious of particle, cm;
D_w, D_L – moisture diffusivity in particle and in liquid, cm^2/s;
ε_s – voidage of sample; a_p – specific surface area of particles, cm^2/cm^3;
ε_p – particle porosity; τ_p – tortuosity of the particle.

Process in gas phase is much faster than in particles, so we can consider that the drying proceeds in a quasi-steady regime:

$$dy/dt \approx dy_s/dt \approx 0.$$

It means that the change of moisture in gas slaves to the dynamics of moisture inside particles. Therefore, for the modelling only one equation is sufficient:

$$-\frac{dw}{dt} = D_w V_S (1-\varepsilon_S)\, a_p \left.\frac{dw}{dr}\right|_{r=r_p} \tag{4}$$

Because of small particles diameter (50 to 500 μm), the diffusion flux in (4) one can approximate by the expression of linear driving force:

$$dw/dr\,|_{r=r_p} \approx (w-w_r)/r_p \tag{5}$$

w_r – moisture of particles in equilibrium with gas humidity, mg/mg.

Then, an equation for drying rate takes the form:

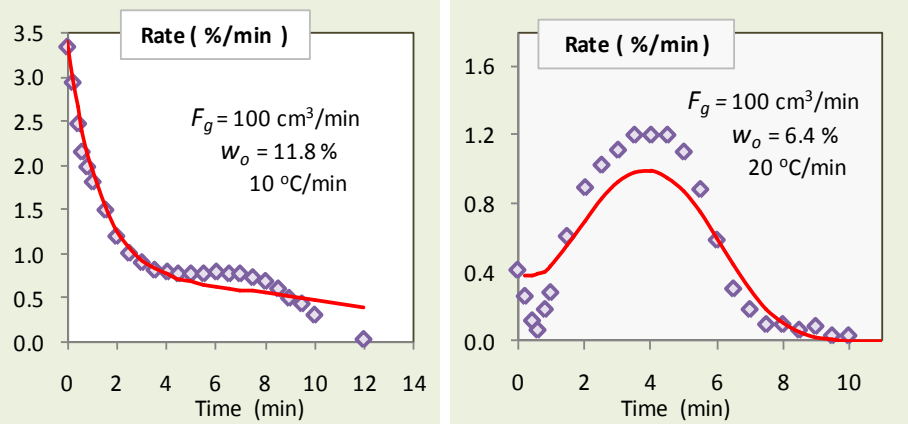
$$-\frac{dw}{dt} \approx D_w(T)\,\frac{a_p}{r_p}\,(w-w_r) \tag{6}$$

Since the diffusivity in liquid D_L depends on temperature, we can write for effective diffusivity:

$$D_w(T) = D_L^o\, exp\left(\frac{E_D}{R}\left[\frac{1}{T_o}-\frac{1}{T}\right]\right)\frac{\varepsilon_p}{\tau_p} \tag{7}$$

Equation (6) with formula (7) represent the model of drying dynamics of porous powder in laboratory pan in thermogravimetric experiments. The temperature increased linearly with time in experiment $T=T_o+bt$, i.e. the rate of heating was constant. Thus, it is convenient to present the drying rate Fig. 1(b) as a function of time Fig. 2.

There are no difference in two experiments at Fig. 2(a,b) except rates of heating, and initial moistures w_o. Therefore, all model parameters (a_p, r_p, D_L, E_D, ε_p, τ_p) should be the same. That is why the quality of description is not acceptable.



Figs. 2(a,b). Rate of drying [22]. Points – experiment; lines – equations (6-7).

There are only two possibilities to improve the description: (i) to include the concentration dependence of diffusivity, which will be empiric; (ii) to change the model of diffusion.

Let us formulate the new non-empiric model on the base of quasi-reaction driving force of moisture diffusion inside particles with heterogeneous porous structure.

6. Nonlinear quasi-reaction model

Polymer particles have various porous structure, depending on polymerization conditions, such as a catalyst type, electron donor, concentration of H_2, etc. In our case the porous structure looks like at micrographs from Scanning Electron Microscope (SEM) Fig. 3.

Most authors consider the polymer particle growing during polymerization (macro-particle) as aggregate of smaller particles (micro-particles, or globules), with a crystallite of catalyst inside [23, 24].

Such a model formally corresponds with bidisperse structure of porous materials. Thus, we can consider that two types of pores take places in particle: macropores, and micropores.

Therefore, we can assume that "three types of liquid" take places in wetted particle distinguished by their localization (Fig. 4), and thus, by the rate of diffusion – fast, slow and configurational.

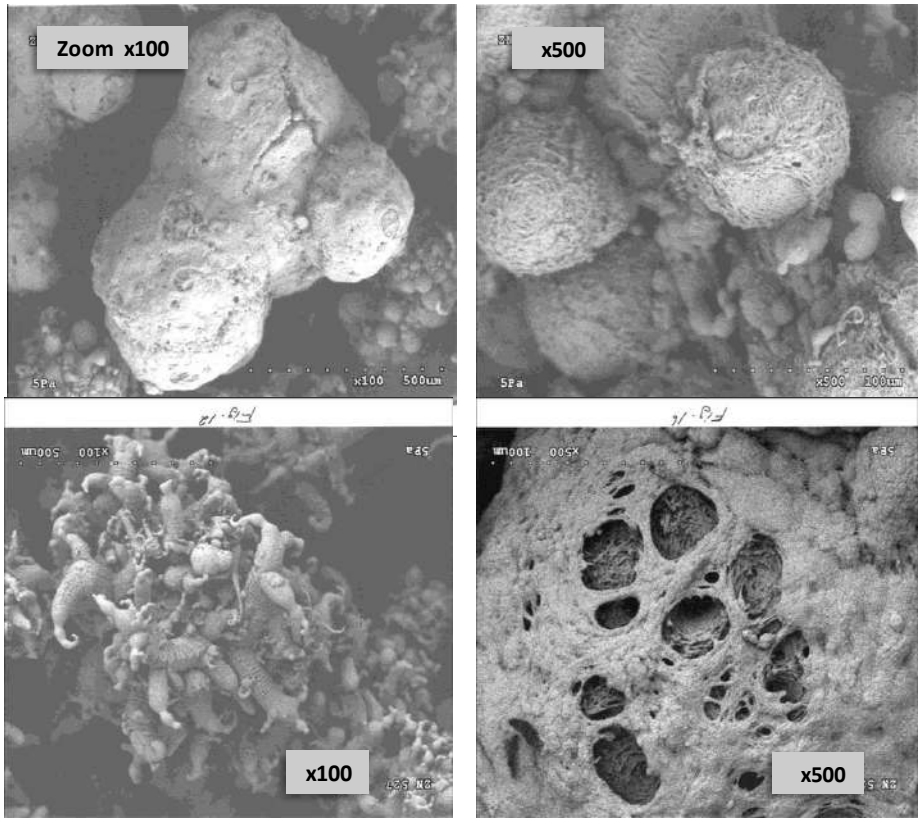


Fig. 3. SEM micrographs of porous structure of polypropylene particles [24].

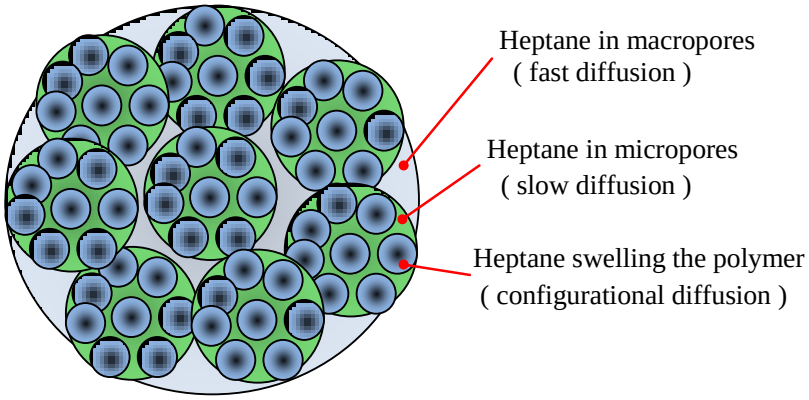


Fig. 4. Schematic model of polypropylene particle wetted by heptane [22].

The moisture of particle (w) one can express via the volume of liquid (V_w), of pores (V_v), of particle (V_p), and of polymer in it (V_{pol}). Because of $V_v / V_p = \varepsilon_p$, and $V_{pol} = V_p\,(1-\varepsilon_p)$, we can write:

$$w = \frac{G_w}{G_{pol}} = \frac{V_w}{V_{pol}}\,\frac{\rho_w}{\rho_{pol}} = \frac{V_w}{V_p\,(1-\varepsilon_p)}\,\frac{\rho_w}{\rho_{pol}} \tag{8}$$

The volume of liquid (V_w) is then a sum of volumes of "three types of liquid" in the corresponding pores ($V_w = \Sigma\, V_i\, x_i$), where (x_i) is a degree of filling the corresponding structures with liquid.

$$\begin{aligned}\frac{V_w}{V_p} &= \frac{1}{V_p}\left(V_{mac}\,x_{mac} + V_{mic}\,x_{mic} + V_{pol}\,x_{pol}\right) = \\ &= \varepsilon_{mac}\,x_{mac} + \varepsilon_{mic}\,x_{mic} + (1-\varepsilon_p)\,x_{pol}\end{aligned} \tag{9}$$

ε_{mac}, ε_{mic} – portion of macro-, and micro-pores;
x_{mac}, x_{mic}, x_{pol} – degree of filling the pores and the polymer with liquid.

It is important that $\varepsilon_{mac} + \varepsilon_{mic} = \varepsilon_p$, and ε_i values remain unchanged with drying, because they characterise the particle porous structure. As for partial moisture (x_i), each changes in the interval from 0 to 1.

With (9), the equation (8) for total moisture becomes as follows:

$$w = \frac{\rho_w}{\rho_{pol}}\,\frac{1}{1-\varepsilon_p}\left(\varepsilon_{mac}\,x_{mac} + \varepsilon_{mic}\,x_{mic} + (1-\varepsilon_p)\,x_{pol}\right) \tag{10}$$

Diffusion fluxes for partial moisture x_i we write using quasi-chemical reaction approach in Onsager form: $J_i = \Sigma\, L_i\, X_{ji}$. The driving force X_{ji} has a nonlinear expression, like in cell-jump diffusion [7] or in Langmuir adsorption kinetics. In our case $X_{ji} = x_j\,(1-x_i)$ where x_j is a degree of saturation by moisture of one type of pores, and $(1-x_i)$ is an emptiness of the nearby pores.

$$\frac{d\,x_{mac}}{d\,t} = -\beta_{mac}\left(x_{mac} - x_r\right) + \beta_{mic}\,x_{mic}\left(1-x_{mac}\right) \tag{11}$$

$$\frac{d\,x_{mic}}{d\,t} = -\beta_{mic}\,x_{mic}\left(1-x_{mac}\right) + \beta_{pol}\,x_{pol}\left(1-x_{mic}\right) \tag{12}$$

$$\frac{d\,x_{pol}}{d\,t} = -\beta_{pol}\,x_{pol}\left(1-x_{mic}\right) \tag{13}$$

Mass-transfer coefficients β_i (min^{-1}) are defined similar to (6) and (7):

$$\beta_i(T) = D_i^o \exp\left(\frac{E_i}{R}\left[\frac{1}{T_o} - \frac{1}{T}\right]\right)\frac{\varepsilon_p}{\tau_p}\frac{a_p}{r_p} \qquad (14)$$

According to the model, the moisture diffusion starts from macropores. Because x_r in (11) is proportional to equilibrium moisture w_r, the diffusion continues until saturation of the gas phase or until emptying of macropores. The diffusion from micropores is possible only toward macropores if they are not filled.

The „liquid“ in a polymer (x_{pol}) is a consequence of polymer swelling in the presence of monomer, oligomers and solvent. It is a very small value but can influence the drying of polymer at high temperature [25].

The model gave a good prediction of the experimental drying rates at different operating conditions and moisture contents (Figs. 5, 6).

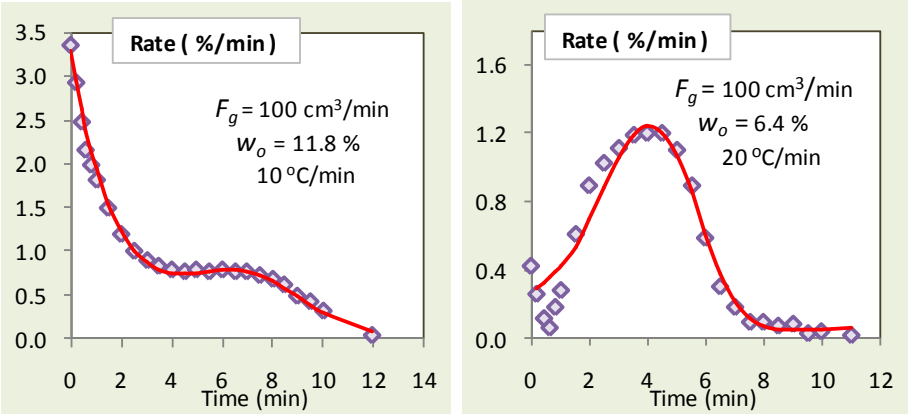


Fig. 5(a,b). Drying dynamics [22]. Points – experiment; lines – model (10-14).

Comparison Fig. 5 and Fig. 2 shows that the new quasi-reaction model provides an excellent description of experiments, in contrast to the classical model. Fig. 6 provides additional confirmation of the model quality.

That is a special experiment with a stepwise temperature rise and the result of its simulation by the proposed model. The red line represents the total drying rate (removal of moisture from the powder sample):

$$\frac{d\,w}{d\,t} = A\sum_{i=1}^{n}\varepsilon_i \frac{d\,x_i}{d\,t}, \quad A = \frac{\rho_w}{\rho_{pol}}\frac{100}{1-\varepsilon_p} \tag{15}$$

At this figure we can see also the dynamic behaviour of each local moisture w_i (in each type of pores):

$$w_i = A\,x_i, \quad \frac{d\,x_i}{d\,t} = \sum_{i,j=1}^{n} \alpha_{ji}\beta_i\,x_i\left(1-x_j\right) \tag{16}$$

$\alpha_{j\,i}$ – stoichiometric matrix of interconnections of pores types.

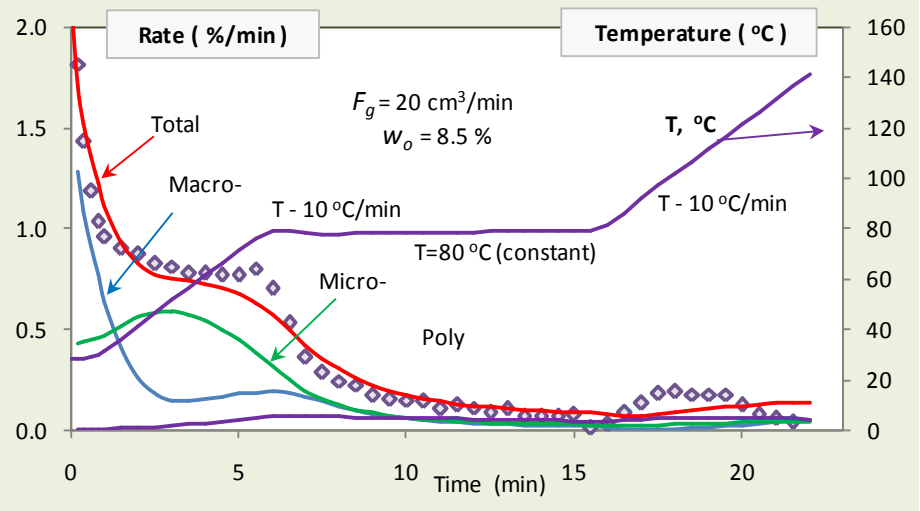


Fig. 6. Drying at SWI [22]. Points – experiment; lines – model (10-14).

The blue line represents the drying rate of macropores, green line - micropores, and violet line - solidification of polymer.

The process begins with diffusion of moisture from macropores into gas phase: $-\beta_{mac}\,(x_{mac} - x_r)$. This "unlocks the door" for the moisture that is in the micropores: $-\beta_{mic}\,x_{mic}\,(1-\,x_{mac})$, because diffusion from micropores cannot begin until the macropores are occupied. Similarly, the diffusion of moisture from the polymer $-\beta_{pol}\,x_{pol}\,(1-\,x_{mic})$ cannot begin until the micropores are occupied.

Let us consider now parameters of the model that are listed in Table. 1. First of all, type of moisture in fact is a type of pores. Characteristics of porous structure are measured parameters. They include the fractions of each pore type (ε_i), pore radii (r_p), total porosity (ε_p), and tortuosity (τ_p). Well known methods, widely used in study of porous materials, are Mercury Intrusion Porosimetry, and Nitrogen Adsorption. All these parameters are measured in polypropylene production, and in its drying in fluidized bed [24, 25].

Only parameters such as diffusion coefficients (D_i^o) and activation energies (E_i) were determined by describing the experimental data given above. Their values are typical for the diffusion of liquid heptane in a bulk and in porous materials. All parameters of model do not depend on moisture concentration, but only on porous structure, which does not change in drying process. In fact, this is the main advantage of quasi-reaction model.

Table 1. Parameters of quasi-reaction model.

Pores and moisture (i)	Pores fraction (ε_i)	D_i^o, cm^2/s at 100 ^oC	Aktiv. energy E_i, kJ/mol
In macropores	0.60	4.0e-05	12
In micropores	0.35	2.0e-06	25
In polymer	0.05	7.0e-08	20-30

Porosity: ε_p = 0.1–0.2 Tortuosity: τ_p = 4–5

7. Final comments

Dependence of the diffusion coefficient on concentration is erroneous in principle if we use the Fick's diffusion law.

It appears when we try to describe the process of mass transfer by equation of Fick, which is strictly correct only for uniform medium.

One of possible way to describe such processes is a quasi-reactions approach, where the driving force is expressed according to action mass law, for example as in adsorption or in *cell*-jump formalism.

In application to moisture diffusion in porous bidisperse particle, the process is considered as a flowing of moisture from macropores to gas phase, and from micropores to releasing part of macropores.

The model demonstrates a good description of experiments, and was used for optimization of industrial dryer with multizone fluidized bed.

References

[1] R.E Cunningham, R.J.J. Williams, *Diffusion in Gases and Porous Media*. Plenum Press, New York & London, 1980.
[2] J. Crank, *The Mathematics of Diffusion*. Clarendon Press, Oxford, 1975.
[3] R. Krishna, J.A. Wesselingh, *Chem. Eng. Sci.* **52** (1997) 861-911.
[4] V.I. Bogachev, N.V. Krylov, M. Rockner, S.V. Shaposhnikov, *Mathem. surveys and monographs*, Amer. Math. Soc., 2015, v. 207.
[5] Y. Frenkel, *Zeitschrift für Physik*, **35** (1926) 652–669.
[6] J.F. Kincaid, H. Eyring, A.E. Stearn, *Chem. Rev.* **28** (1941) 301–365.
[7] A.N. Gorban, H.P. Sargsyan, H.A. Wahab, *Math. Modelling Natural Phenomena*. **6** (2011) 184-262.
[8] J. Hirschfelder, C.F. Curtiss, R.B. Bird, *Molecular Theory of Gases and Liquids*. Wiley, New York, 1954.
[9] C.R. Wilke, Pin Chang, *A.I.Ch.E. Journal*. **1** (1955) 264-270.
[10] A.Yu. Kuksin, D.E. Smirnova, *Phys. Solid State*. **56** (2014) 1214-1223.
[11] A.D. Schwope, L.R. Jackson, *National Advisory Committee for Aeronautics*, NASA, Washington, 1951.
[12] M.A. Silva, P.J.A.M. Kerkhof, W.J. Coumans, *Ind. Eng. Chem. Res.*, **39** (2000) 1443-1452.
[13] M.M. Mezedur, M. Kaviany, W. Moore, *AIChE Journal*. **48** (2002) 15-24.
[14] N. Wakao, J.M. Smith, *Chem. Eng. Sci*. **17** (1962) 825-834.
[15] C.T. Kiranoudis, Z.B. Maroulis, D. Marinos-Kouris, *Int. Journ. Heat & Mass Transfer*. **38** (1995) 463–480.
[16] T. Loulou, B. Adhikari, D. Lecomte, *Chem. Eng. Sci*. **61** (2006) 7185-7198.
[17] H. Chou, L. Wu, L. Zeng, A. Chang, *Water Resour. Res*. **48** (2012) W10539.
[18] X.D. Chen, *Drying Technology*. **25** (2007) 1203–1213.
[19] A. Putranto, X.D. Chen, *AIChE Journal*. **59** (2013) 55-67.
[20] N. Shahari, *Ph.D. Thesis*, University of Nottingham, 2012.
[21] N. Shahari, S. Hibberd, *WSEAS Transactions on Mathematics*. **13** (2014) 840-851.
[22] N.M. Ostrovskii, K. Mészáros-Szécsényi, *Proc. of 13th Int. Conf. "Physical Chemistry"*, Belgrade, 2016, 191-198.
[23] G.D. Bukatov, V.I. Zajkovsky, V.A. Zakharov, et al., *Vysokomol. Soedin.* **24** (1982) 542-548.
[24] N.M. Ostrovskii, D. Stoiljković, *Theoretical Foundation of Chemical Engineering*. **45** (2011) 40-52.
[25] N.M. Ostrovskii, *Chemical Industry*. **68** (2014) 661-671.

Instability region in models of nonlinear reaction systems. The Stoichiometric Network Analysis [*]

Željko Čupić [†]
University of Belgrade, Institute of Chemistry, Technology and Metallurgy, Department of Catalysis and Chemical Engineering, Belgrade, Serbia;

Stevan Maćešić [‡]
University of Belgrade, Faculty of Physical Chemistry, Belgrade, Serbia;
University of Szeged, Department of Physical Chemistry and Materials Science, Szeged, Hungary;

Ljiljana Kolar-Anić [§]
University of Belgrade, Faculty of Physical Chemistry, Belgrade, Serbia;

ABSTRACT

Stability analysis of reaction systems is described by the application of the Stoichiometric Network Analysis to the three-variable-autocatalator. Although simple, this model is complex enough to describe complex forms of nonlinear dynamics phenomena, like mixed-mode oscillations and chaos. Therefore, stability analysis of such model is not a trivial task. Using the Stoichiometric Network Analysis for this purpose makes the process clear and leads to the reliable result.

The method is described briefly in few general steps and all of them are further clarified through the application to the chosen example. First, the reaction rates in steady state are decomposed to contributions of independent pathways, called extreme currents. Then, linearized operator is constructed. Finally, through the analysis of the principal minors of the essential part of this operator, simple stability criterion is identified.

---

[*] This work was partially supported by the Ministry of Education, Science and Technological Development of Republic of Serbia (Grant Numbers OI 172015 and III 45001and Contract numbers: 451-03-68/2020/14/200026 and 451-03-68/2020-14/200146).

[†] e-mail address: zcupic@ihtm.bg.ac.rs
[‡] e-mail address: stevan.macesic@ffh.bg.ac.rs
[§] e-mail address: ljiljana.kolar.anic@ffh.bg.ac.rs

1. Introduction

Reaction systems are wide but specific class of dynamical systems where state variables are usually concentrations of some reactive species. In these systems the contributions of the individual reaction rates to the overall rate of changes are governed by the stoichiometric relations between mentioned species. [1] The number of individual reaction rates (reaction steps) can be more or less large. In biochemical reaction systems it is generally very large. [2] Hence, the stability analysis of such systems is also specific and require use of adequate tools like the Stoichiometric Network Analysis (SNA). [3]

In models of reaction systems, various rate laws [4] may be used, but mass action principle is the most common. It is based on fundamental principle, that the rate of reaction is proportional to the concentrations of the reacting substances. As a result, rates of individual steps are power functions of concentrations as the state variables, and overall rates are obtained as linear combinations of such simple monomial terms. Nevertheless, final expressions are nonlinear as a rule. Moreover, the number of independent variables (the concentrations of independent species) and number of related equations that describe their evolution in time, may be very large.

Numerical simulations based on efficient algorithms for integration of systems of ordinary differential equations are often the best way to analyze dynamical states of the reaction systems. However, the main model parameters – rate constants of all individual reaction steps are generally unknown. Therefore, more general approach is required to evaluate possible dynamic states depending on unknown values of the rate constants. SNA is probably unique tool that may provide such general results on so complex objects as reaction systems are. Large number of reaction steps and reaction species may be limiting for application of the SNA, but several approximations, specific only for SNA, are available to attain some result on instability condition even for very large systems with dozens of reactions and reaction species.

Analytical expressions for the instability condition can be easily calculated as a function of the rate constants, and then tested by comparing them with the results of the numerical simulations obtained for selected parameter values. [5] For this purpose, bifurcation analysis may lead to crucial results. [6]

Simple model known as three variable autocatalator will be used to illustrate the method. [7]

2. The model

Our examinations are illustrated on a reaction model known as three-variable-autocatalator. [7] This model consisting of five reaction species (R, X, Y, Z and P) and six reactions having rate constants k_i where i = 0, 1, ... 5, is given in Table 1. The system may readily be reduced since P is only the product of the reaction which does not influence the rate of any reaction step. We will also assume that reactant R is in large excess so that changes in its concentration may be neglected and dynamical state of the system depends just on its initial value as a control parameter. Hence, only three variables remained (concentrations *x*, *y* and *z* of species X, Y and Z, respectively).

Table 1. The three-variable-autocatalator reaction network model [7]

$R \xrightarrow{k_0} X$	$v_0 = k_0\, r_0$	(M.0)
$X \xrightarrow{k_1} Y$	$v_1 = k_1\, x$	(M.1)
$X+2Y \xrightarrow{k_2} 3Y$	$v_2 = k_2\, x\, y^2$	(M.2)
$Y \xrightarrow{k_3} Z$	$v_3 = k_3\, y$	(M.3)
$Z \xrightarrow{k_4} P$	$v_4 = k_4\, z$	(M.4)
$R+Z \xrightarrow{k_5} X+Z$	$v_5 = k_5\, r_0\, z$	(M.5)

The dynamics of the model can be represented by set of ordinary differential equations:

$$\mathrm{d}x/\mathrm{d}t = v_0 - v_1 - v_2 + v_5 \tag{1}$$

$$\mathrm{d}y/\mathrm{d}t = v_1 + v_2 - v_3 \tag{2}$$

$$\mathrm{d}z/\mathrm{d}t = v_3 - v_4 \tag{3}$$

and if we use rates of reaction steps from Table 1 set of differential equations has the form:

$$\mathrm{d}x/\mathrm{d}t = k_0\, r_0 - k_1\, x - k_2\, x\, y^2 + k_5\, r_0 z \tag{4}$$

$$\mathrm{d}y/\mathrm{d}t = k_1\, x + k_2\, x\, y^2 - k_3\, y \tag{5}$$

$$\mathrm{d}z/\mathrm{d}t = k_3\, y - k_4 z \tag{6}$$

Species Z does not change the concentration in reaction (A.5) since it appears there both, as reactant and product.

From equations (4)-(6), we can calculate steady-state concentrations:

$$x_{\mathrm{SS}} = \frac{k_0 k_3^{\,2} k_4\, r_0\, (k_4 - k_5\, r_0)}{k_1 k_3^{\,2} (k_4 - k_5\, r_0)^2 + k_2 k_4^{\,2} (k_0\, r_0)^2} \tag{7}$$

$$y_{\mathrm{SS}} = \frac{k_4 k_0\, r_0}{k_3\, (k_4 - k_5\, r_0)} \tag{8}$$

$$z_{\mathrm{SS}} = \frac{k_0\, r_0}{k_4 - k_5\, r_0} \tag{9}$$

The stability analysis of three variable system can be performed by several methods. However, the stability analysis of a system with more than four variables can be done only by SNA. Nevertheless, we intend to present the stoichiometric network analysis with its advantages on three variable system for pedagogical purpose.

3. Stoichiometric network analysis of considered model

Stability analysis of complex nonlinear reaction mechanisms is very complicated task. For models that have large number of independent intermediate species, the classical tools for studying the stability conditions are quite ineffective. To avoid this problem it is necessary to use specialized methods. At present, the most powerful one is the stoichiometric network analysis (SNA). [3, 8]

In SNA, the kinetic equations of any stoichiometric model presented by a set of differential equations (such as (4), (5) and (6)) is written in the matrix form:

$$\dot{\boldsymbol{c}} = \mathbf{S} \times \boldsymbol{v} \tag{10}$$

where $\dot{c}$ is the time derivative of the $m \times 1$ concentration vector c comprising the change in concentrations of m independent intermediate species, known as internal ones in SNA. **S** is the stoichiometric matrix and $\boldsymbol{v}$ the so-called reaction or flux vector with reaction rates as components. The stoichiometric matrix **S** is an $m \times n$ matrix where n is the number of reactions in the reaction network (in the model considered $m = 3$ and $n = 6$.). The S_{ik} element of the stoichiometric matrix corresponds to the stoichiometric coefficient of reactive species i (i = 1,2, ..., m) in reaction (R_k) corresponding to column k and row i. The reaction vector v is $1 \times n$ vector whose elements describe the reaction rates.

Using the matrix representation given in equation (10), the model given in Table 1 corresponds to the following system of differential equations:

$$\dot{c} = \begin{matrix} \mathrm{M.0} & \mathrm{M.1} & \mathrm{M.2} & \mathrm{M.3} & \mathrm{M.4} & \mathrm{M.5} \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{bmatrix} \times \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{bmatrix} \tag{11}$$

Row above the matrix **S** in equation (11) indicates correspondence between matrix columns and reactions in the model, and it is not part of the matrix.

Now, we want to obtain information about the interplay between the concentrations of independent intermediate species and the dynamics of the network as a whole. As a first step, we look for conditions where the network is in a quasi-steady-state. The rates at a steady state v_{ss} are solutions of the relation:

$$0 = \mathbf{S} \times \boldsymbol{v}_{ss} \tag{12}$$

Equation (12) represents system of homogenous equations, and we need to find all positive solutions. Method for finding all positive solutions of equation (12) depends on the size of examined model. For simpler

models equation (12) can be solved manually. However, if the number of reactions is large, solving equation (12) becomes much more complex, and only suitable way is to use computer programs based on algorithms developed for this purpose. There exist several algorithms. [9-13]

The solutions of (12), known as extreme currents [3, 8], are reaction pathways in steady state. They offer important information about the consistency of the model, and correlations between individual reactions like mutual exclusion or coupling. [14] The extreme currents E_i are usually presented as the columns (in any order) of a new extreme current matrix **E**. In the case considered, it is:

$$\mathbf{E} = \begin{array}{c} \begin{matrix} E_1 & E_2 & E_3 & E_4 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \begin{matrix} \text{M.0} \\ \text{M.1} \\ \text{M.2} \\ \text{M.3} \\ \text{M.4} \\ \text{M.5} \end{matrix} \tag{13}$$

As before, row above and column beside the matrix **E** in equation (13) indicate correspondence between matrix rows and reactions and between matrix columns and extreme currents, and they are not part of the matrix.

In any specific steady state, each extreme current contributes to reaction rates with its own, distinct, extent. The contributions of the extreme currents, denoted as the current rates j_i, are given as the components of the corresponding vector ***j***. Elements of vector ***j*** are nonnegative numbers. The basic equation of the SNA gives relation between rates at a steady state $v_{ss,k}$ and current rates j_k:

$$v_{ss} = \mathbf{E}\boldsymbol{j} \tag{14}$$

Using equation (14), the particular $v_{ss,k}$ can be written in the form:

$$\begin{aligned} v_{ss,0} &= k_0 r_0 = j_1 + j_2 \\ v_{ss,1} &= k_1 x_{SS} = j_1 + j_3 \end{aligned} \tag{15}$$

$$v_{ss,2} = k_2 x_{SS} y_{SS}^2 = j_2 + j_4$$
$$v_{ss,3} = k_3 y_{SS} = j_1 + j_2 + j_3 + j_4$$
$$v_{ss,4} = k_4 z_{SS} = j_1 + j_2 + j_3 + j_4$$
$$v_{ss,5} = k_5 r_0 z_{SS} = j_3 + j_4$$

The next step in our analysis is to examine stability of a steady state. In other words, we should like to find instability condition. The stability analysis of the particular steady state is usually performed on the linearized form of the general equation of motion of a system around the steady state. Namely, when the system is in a steady state little perturbation of the concentrations of intermediate species can be given as linear deviation from the steady state concentrations. [15]

$$c = c_{ss} + \Delta c \tag{16}$$

We can expand time derivative of concentration vector c in Taylor series near a steady state c_{ss} and keep leading terms only. Hence,

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \frac{\mathrm{d}(c_{ss} + \Delta c)}{\mathrm{d}t} = \frac{\mathrm{d}\Delta c}{\mathrm{d}t} = \mathbf{M}\,\Delta c \ . \tag{17}$$

The leading term **M** is Jacobian matrix which in SNA has the form

$$\mathbf{M}(\boldsymbol{h}, \boldsymbol{v}_{ss}) = \mathbf{S}\,\mathrm{diag}(\boldsymbol{v}_{ss})\,\mathbf{K}^{\mathbf{T}}\mathrm{diag}(\boldsymbol{h}) \tag{18}$$

where diag(***h***) is a diagonal matrix whose elements h_i are the reciprocal of the steady state concentrations of the species *i*, for *i* = 1, 2, 3, and **K** is the matrix of the order of reactions with its transpose $\mathbf{K}^{\mathbf{T}}$. Hence, the stability is defined by the sign of the real part of the eigenvalues of the Jacobian matrix. A general derivation of Jacobian matrix **M** is given in references [8] and [15]. For the model under consideration, matrix **K** is:

$$\mathbf{K} = \begin{array}{c} \begin{matrix} M.0 & M.1 & M.3 & M.4 & M.5 & M.6 \end{matrix} \\ \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array} \begin{matrix} \\ X \\ Y \\ Z \end{matrix} \tag{19}$$

As usual, row above and column beside the matrix **K** in equation (16) indicate correspondence between matrix rows and reaction species and between matrix columns and reaction steps, and they are not part of the matrix.

According to equation (14), the equation (18) can be transformed to

$$\mathbf{M}(\boldsymbol{h}, \boldsymbol{j}_{ss}) = \mathbf{S}\,\mathrm{diag}(\mathbf{E}\boldsymbol{j})\,\mathbf{K}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{h}) \tag{20}$$

The matrix **M** written as a function of the SNA parameters j_i and h_i has particular advantages for the stability analysis since these parameters are non-negative and each element of **M** is linear function of ***j*** and ***h*** parameters, which is an essential feature of the SNA. The steady-state stability is determined by the sign of eigenvalues of **M**, which are the roots λ of the characteristic polynomial: [16]

$$\left|\lambda\mathbf{I} - \mathbf{M}\right| = \sum_{i=0}^{n} \alpha_i \lambda^{n-i} \tag{21}$$

where, for considered model, n = 3. If real parts of all eigenvalues are negative a steady state is stable. If one or more eigenvalues has positive real parts the steady state is unstable. The number of eigenvalues with positive real parts can be determined by Routh - Hurwitz criterion. According to this criterion the number of eigenvalues with positive real parts is equal to the number of the sign changes in the Routh array [16]

$$\mathrm{R} = \left(1, \Delta_1, \frac{\Delta_2}{\Delta_1}, \frac{\Delta_3}{\Delta_2}, \ldots, \frac{\Delta_n}{\Delta_{n-1}}\right) \tag{22}$$

where Δ_i, i = 1,…,n, is i-th Hurwitz determinant, defined as the determinant of the leading principal minor of the Hurwitz matrix **H**, where leading principal minor of dimension i is matrix constructed from the first i rows and columns of matrix **H**.

$$\mathbf{H} = \begin{bmatrix} \alpha_1 & \alpha_3 & \alpha_5 & \alpha_7 & \dots & \alpha_{2n-1} \\ 1 & \alpha_2 & \alpha_4 & \alpha_6 & \dots & \alpha_{2n-2} \\ 0 & \alpha_1 & \alpha_3 & \alpha_5 & \dots & \alpha_{2n-3} \\ 0 & 1 & \alpha_2 & \alpha_4 & \dots & \alpha_{2n-4} \\ 0 & 0 & \alpha_1 & \alpha_3 & \dots & \alpha_{2n-5} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \alpha_n \end{bmatrix} \tag{23}$$

Obviously, $\alpha_i = 0$ for $i > n$. Steady state is stable if all Hurwitz determinants are positive. If there is only one sign change in the Routh array (22), this indicates that only one eigenvalue has positive real part. Such instability occurs when the largest Hurwitz determinant changes its sign keeping all others positive, and this point presents saddle-node bifurcation. From Hurwitz matrix (23) we can see that the largest Hurwitz determinant Δ_n can be written as

$$\Delta_n = \alpha_n \times \Delta_{n-1} \tag{24}$$

Since the sign of the largest Hurwitz determinant is in this case determined by the sign of the largest coefficient of the characteristic polynomial α_n, the saddle-node bifurcation can be identified from [16]

$$\alpha_n = 0 \tag{25}$$

The Hurwitz matrix gives us also condition for appearance of Andronow-Hopf bifurcation which is of great importance, because it is source of oscillations in the system. The Andronow-Hopf bifurcation occurs when: [16, 17]

$$\Delta_{n-1} = 0 \tag{26}$$

For considered model where $n = 3$, the Hurwitz matrix is

$$\mathbf{H} = \begin{bmatrix} \alpha_1 & \alpha_3 & 0 \\ 1 & \alpha_2 & 0 \\ 0 & \alpha_1 & \alpha_3 \end{bmatrix} \tag{27}$$

and condition for appearance of Andronow-Hopf bifurcation given in equation (26) is

$$\Delta_2 = \begin{bmatrix} \alpha_1 & \alpha_3 \\ 1 & \alpha_2 \end{bmatrix} = \alpha_1\alpha_2 - \alpha_3 = 0 \tag{28}$$

Application of an instability condition obtained by this method becomes limited by the number of independent internal species and it is often very difficult to be determined by classical procedure, but this method is convenient for numerical evaluation of stability of steady state. Much simpler method to examine the steady-state stability is the use of the matrix of current rates **V**(***j***), where:

$$\mathbf{V}(\boldsymbol{j}) = \mathbf{S}\,\mathrm{diag}(\mathbf{E}\boldsymbol{j})\,\mathbf{K}^{\mathbf{T}} \tag{29}$$

The steady state is considered to be unstable if there is at least one negative diagonal minor of **V**(***j***). [8] Although it is an approximation, this criterion often gives satisfactory results. The matrix **V**(***j***) for the model considered is

$$\mathbf{V}(\boldsymbol{j}) = \begin{bmatrix} j_1 + j_2 + j_3 + j_4 & 2j_2 + 2j_4 & -(j_3 + j_4) \\ -(j_1 + j_2 + j_3 + j_4) & j_1 - j_2 + j_3 - j_4 & 0 \\ 0 & -(j_1 + j_2 + j_3 + j_4) & j_1 + j_2 + j_3 + j_4 \end{bmatrix} \tag{30}$$

We examined all diagonal minors and detected those with negative terms, since only these minors can be negative. They are negative when polynomial obtained from corresponding determinant is negative. The calculated polynomials have to be compared between one another, the core of instability must be recognized, and essential one ought to be selected. The aim is to find the polynomial with less possible order. Such obtained polynomials are expressed in function of j_i. In the case considered the

selected polynomial is coming from the minor corresponding to the second and third rows-columns:

$$2j_1j_3 - 2j_2j_4 + j_1^2 - j_2^2 + j_3^2 - j_4^2 < 0 \tag{31}$$

It may be transformed to

$$(j_1 + j_3)^2 < (j_2 + j_4)^2 \tag{32}$$

Since all j_i parameters are non-negative condition (32) can be rewritten as:

$$j_1 + j_3 < j_2 + j_4 \tag{33}$$

According to (15) instability condition (33) becomes

$$k_1 x_{ss} < k_2 x_{SS} y_{SS}^2 \tag{34}$$

which is equivalent to

$$y_{SS} > \sqrt{\frac{k_1}{k_2}} \tag{35}$$

Inserting (8) into (34) we obtain

$$\frac{k_4 k_0 r_0}{k_3 (k_4 - k_5 r_0)} > \sqrt{\frac{k_1}{k_2}} \tag{36}$$

which may be transformed to

$$r_0 > \frac{k_1^{1/2} k_3 k_4}{k_4 k_0 k_2^{1/2} + k_1^{1/2} k_3 k_5} \tag{37}$$

Equation (37) gives the critical condition, which need to be fulfilled for periodic temporal evolution of intermediate species that are involved in

dynamics of HPA system model, However, this result is only the approximation based on selection of the most significant minors of the matrix V(***j***), which are responsible for the sign of the corresponding coefficient in the characteristic polynomial (21). In large number of tested models this approximation works very well, even surprisingly well. However, in some cases it is not enough and complete condition for Andronov-Hopf bifurcation (28) must be used. It may be applicable only in the case of small models like the one used here. For the given model it is given by:

$$\begin{aligned}\Delta_2 = \begin{bmatrix} \alpha_1 & \alpha_3 \\ 1 & \alpha_2 \end{bmatrix} = \alpha_1\alpha_2 - \alpha_3 = h_1^2 h_2 \left(j_1 + j_2 + j_3 + j_4\right)^3 + h_1^2 h_3 \left(j_1 + j_2 + j_3 + j_4\right)^3 \\ + h_1 h_2^2 \left(j_1 + j_2 + j_3 + j_4\right)^2 \left(j_1 - j_2 + j_3 - j_4\right) \\ + h_1 h_2 h_3 \left(j_1 + j_2 + j_3 + j_4\right)^2 \left(j_1 - j_2 + j_3 - j_4\right) \\ + h_2^2 h_3 \left(j_1 + j_2 + j_3 + j_4\right)\left(j_1 - j_2 + j_3 - j_4\right)^2 \\ + h_1 h_2 h_3 \left(j_1 + j_2 + j_3 + j_4\right)^3 + h_1 h_3^2 \left(j_1 + j_2 + j_3 + j_4\right)^3 \\ + h_2 h_3^2 \left(j_1 + j_2 + j_3 + j_4\right)^2 \left(j_1 - j_2 + j_3 - j_4\right) \\ + h_1 h_2 h_3 \left(j_1 + j_2 + j_3 + j_4\right)^2 \left(j_3 + j_4\right) = 0\end{aligned} \tag{38}$$

Numerical simulations (Figure 1) were done for three values of the parameter r_0 and fixed values of rate constants.

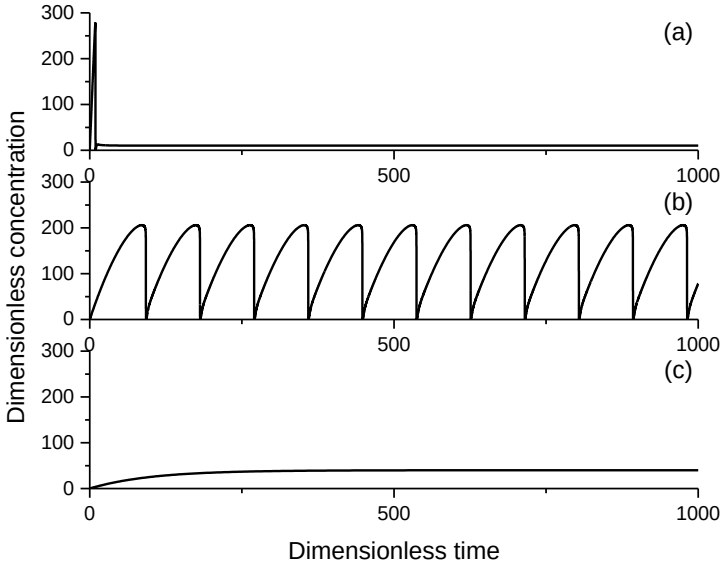


Figure 1. Numerical simulations of the three-variable-autocatalator model given in Table 1. Dimensionless rate constants: k_0=40; k_1=0.01; k_2= 2.439 10^{-3} ; k_3 = 2; k_4 = 0.5; k_5 = 0.5. Reactant dimensionless concentrations: (a) r_0 = 0.8; (b) r_0 = 0.1; (c) r_0 = 0.01.

In Figures 1(a) and 1(c) numerical simulation leads to the stable steady state, while in Figure 1(b) periodic oscillations are obtained. Accordingly, if Δ_2 given in (38) is evaluated with parameter values used in Figures 1(a) and 1(c), positive values are obtained and with parameter values used for Figure 1(b) negative value for Δ_2 is obtained. Hence, two bifurcation points of Andronov-Hopf type may be found between these extreme cases. However, approximate instability condition (37) predicts only one bifurcation value as a function of the r_0. This condition is fulfilled for both sets of parametar values, in Figure 1(a) and 1(b), and we can now say that used approximation was too large.

4. Conclusions

The three-variable-autocatalator was submitted to the Stoichiometric Network Analysis. Four elementary reaction pathways – extreme currents of the model were identified. Linearized operator of the model near the steady state was constructed and matrix of extreme current rates $\mathbf{V}(\boldsymbol{j})$ was extracted from it as the essential part. From the principal minors of the matrix $\mathbf{V}(\boldsymbol{j})$ the instability condition was identified which restricts the ratios between some reaction rates at the steady state, and consequently between the parameter values. Exact instability condition was identified from the Hurwitz matrix and it was validated by numerical simulations.

References

[1] Lj. Kolar-Anić and Ž. Čupić, *Linear and Nonlinear Reaction Systems* in: K. R. Hedrih and Žarko Mijajlović, Eds., *Scientific Review Series: Scientific and Engineering - Special Issue Nonlinear Dynamics* **S2** (2013) 67.

[2] B. O. Palsson, N. D. Price and J. A. Papin, *Trends in Biotechnology* **21** (2003) 195.

[3] B. Clarke, *Stability of complex reaction networks*, in: I. Prigogine and S. Rice, Eds., *Adv. Chem. Phys.* Wiley, 1980, p.p. 1.

[4] E. N. Yeremin, *The foundations of Chemical Kinetics*. Mir Publishers, 1979.

[5] Lj. Kolar-Anić, Ž. Čupić, G. Schmitz and S. Anić, *Chem. Eng. Sci*. **65** (2010) 3718.

[6] S. Maćešić, Ž. Čupić, A. Ivanović-Šašić, S. Anić, M. Radenković, N. Pejić and Lj. Kolar-Anić, Reac Kinet Mech Cat **123** (2018) 31.

[7] B. Peng, S. K. Scott, K. Showalter, *Journal of Physical Chemistry* **94** (1990) 5243.

[8] B. L. Clarke, *Cell Biophys.* **12** (1988) 237.

[9] B. L. Clarke, *J Chem. Phys.* **75** (1981) 4970.

[10] B. L. Clarke, *Qualitative Dynamics and Stability of Chemical Reaction Networks*, in R. B. King, Ed., *Chemical Applications of Topology and Graph Theory*, Elsevier, 1983, p.p. 322.

[11] P. E. Lehner and E. Noma, *Psychometrika* **45** (1980) 135.

[12] R. Schuster and S. Schuster, *Bioinformatics* **9** (1993) 79.

[13] R. Urbanczik and C. Wagner, *Systems biology* **21** (2005) 1203.

[14] J. Gagneur and S. Klamt, *BMC Bioinformatics* **5** (2004) 175

[15] Lj. Kolar-Anić, Ž. Čupić, V. Vukojević and S. Anić, *Dinamika nelinearnih procesa*, Fakultet za fizicku hemiju, 2011.

[16] B. L. Clarke and W. Jiang, *J. Chem. Phys.* **99** (1993) 4464.

[17] W. Liu, *J. Math. Anal. and App.* **182** (1994) 250.

A review of recent studies on nonlinear dynamics of microtubules and DNA*

Slobodan Zdravković†
Vinča Institute of Nuclear Sciences, University of Belgrade, P.O. Box 522, Atomic Physics Laboratory (040), 11001 Beograd, Serbia

ABSTRACT

Nonlinear dynamics of two biomolecules is studied. These are a microtubule and DNA molecule. Two mathematical procedures are explained, yielding to three kinds of solitary waves moving through the systems. These waves are kinks, modulated solitary waves called breathers and bell-type solitons.

1. Introduction

Biomolecules are nonlinear systems due to inevitable presence of weak chemical bonds. Namely, strong interactions yield to small amplitudes and we can assume that intensities of attractive and repulsive forces are equal. This means that these interactions can be modelled by harmonic potential energies, i.e. by functions of the type $f(x) = kx^2$, where $k = \text{const.}$ Its first derivative is a force, obviously a linear function, which brings about a linear differential equation (DE). As for weak interactions, they should be modelled by enharmonic potential energies, which yield to nonlinear DEs.

Two examples are studied in this article. These are microtubule (MT) and DNA. The structure of DNA is known. It consists of two mutually interacting strands. Each strand is a series of covalently interacting nucleotides. A covalent bond is the strongest chemical interaction, which means that the strands are linear structures.

*This work has been supported by funds from Project within the Cooperation Agreement between the JINR, Dubna, Russian Federation and Ministry of Education, Sciences and Technological Development of Republic of Serbia: Theory of Condensed Matter Physics.
† e-mail address: szdjidji@vin.bg.ac.rs

The strands are connected by weak hydrogen interactions and, of course, DNA as a whole is a nonlinear system.

MT is a long hollow cylindrical polymer structure that spreads between a nucleus and cell membrane [1,2]. Its surface is usually formed out of 13 long structures called protofilaments (PFs), as shown in Fig. 1. Each PF represents a series of electric dipoles called dimers, whose mass and length are $m = 1.8 \times 10^{-22}$ kg and $l = 8$nm, respectively [3,4]. The lengths of MTs vary from a few hundred nanometers up to meters in long nerve axons [5]. The longitudinal, tangential and radial components of electric dipole moment are: $p_z = 337\text{Debye}$, $p_\theta = 198\text{Debye}$ and $p_r = -1669\text{Debye}$, respectively [6]. Hence, p_z is in the direction of MT.

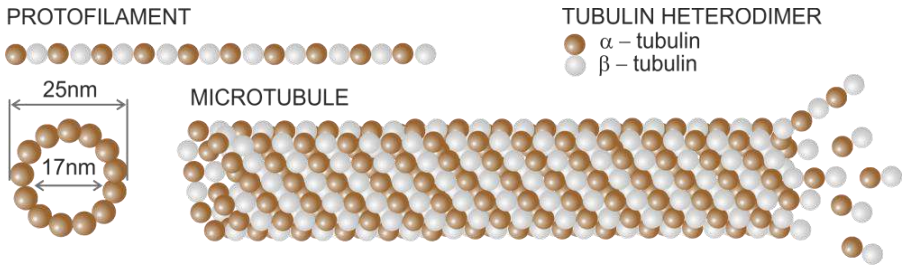


Fig. 1. Microtubule

A head-to-tail binding of dimers, resulting in PFs, appear to be much stronger than those between adjacent PFs [7,8]. This means that a single PF can be seen as a linear system and modelled by the harmonic potential energy. On the other hand, the interaction with the remaining dimers is usually modelled by the function of the type $F(x) = -ax^2 + bx^4 - cx$, where $a > 0$, $b > 0$ and $c \geq 0$ are assumed. For $c = 0$, the function $F(x)$ is symmetric, corresponding to so-called W-potential energy. Therefore, MT as a whole is a nonlinear system.

It was pointed out that the two biological systems are studied in this paper. To model them, two mathematical procedures will be used. They are semi-discrete approximation (SDA) and continuum approximation (CA). The combinations MT-SDA and DNA-CA are explained in Sections 2 and 3, respectively. In Section 4, we deal with DNA-RNA transcription, where SDA is used, while Section 5 is devoted to concluding remarks. It is interesting that the final results, i.e. the solutions of the mentioned DEs, depend on the used mathematical

method rather than on the studied system. The SDA yields to modulated solitary waves called breathers, while the common solution corresponding to the CA is a kink soliton, or kink for short.

A general procedure is equal for both systems and both mathematical procedures. The first step is Hamiltonian. This is nothing but a collection of energies, describing existing interactions. We use generalized coordinates and well-known Hamilton`s equations to obtain dynamical equation of motion. The terms in this equation are forces. Finally, we solve this equation using aforementioned mathematical approximations.

2. Tangential model of microtubules and semi-discrete approximation

There are a couple of models describing nonlinear MT dynamics. Depending on a coordinate which determines a dimer`s displacement they can be either longitudinal or angular. Of course, two component models are also possible. We can monitor a certain evolution of the models through papers [9-17]. Ref. [9] describes the first model, a longitudinal one, where W-potential energy was introduced. Its improved version, which we call *u*-model, was described in Refs. [10-12]. It was shown that Morse potential may be used instead of the W-one [13]. An angular so-called φ-model, that does not comprise the W-potential energy, was introduced in Refs. [14] and [15], while its improved version, including this term, is called a general model [16].

This section is based on a recently introduced two component model that we call a tangential model (TM) [17]. The first that should be clarified is the W-potential energy, modelled by the function $F(x)$, as explained above. This function obviously has two minima, which means that there are two directions of electric field around which the dimer can oscillate. Let the appropriate electric field strengths be $\vec{E}_1$ and $\vec{E}_2$. A resultant internal electric field $\vec{E} = \vec{E}_1 + \vec{E}_2$, coming from all dimers, is in the direction of MT. In principle, the dimer can oscillate around this direction, but any displacement would move it towards the directions of either $\vec{E}_1$ or $\vec{E}_2$. This means that the dimer's position in the direction of $\vec{E}$ is not stable and corresponds to the maximum of the W-potential energy.

Fig. 2 shows oscillation of the dimer around $\vec{E}_1$ only. The orientation of this field is determined by $\theta_{01} \equiv \theta_0$, which is the angle between the direction of the PF and $\vec{E}_1$. A coordinate determining a displacement from the direction of $\vec{E}_1$ is φ, while the dimer`s position with respect to the direction of the PF is θ. It is obvious that

$$\theta = \theta_0 + \varphi \quad . \tag{1}$$

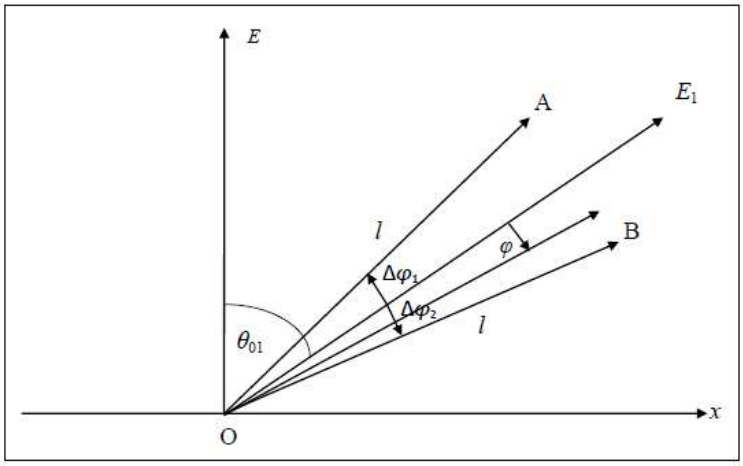


Fig. 2. A schematic representation of the dimer's oscillation

The three components of the electric dipole moment of the single dimer were defined above. They are in the direction of the MT ($\vec{p}_z$), radial direction ($\vec{p}_r$) and in tangential one ($\vec{p}_\theta$). The model assumes that oscillation of the dimer is in the tangential, that is $z-\theta$ plane, which means that $\vec{p}_r \cdot \vec{E}_1 = \vec{p}_r \cdot \vec{E}_2 = 0$ and the relevant moment, used in this paper, is $p = \sqrt{p_z^{\,2} + p_\theta^{\,2}} = 391\text{Debye}$ [17].

The Hamiltonian for a MT can be written as [17]

$$H = \sum_n \left[\frac{I}{2}\dot{\varphi}_n^{\,2} + \frac{k}{2}(\varphi_{n+1} - \varphi_n)^2 - \frac{A}{2}\theta_n^{\,2} + \frac{B}{4}\theta_n^{\,4} - C\theta_n - pE_1 \cos\varphi_n \right], \tag{2}$$

where *n* determines a position of the dimer. The first term is kinetic energy, the dot means the first derivative with respect to time and

I is a moment of inertia of a single dimer. The second term is the potential energy of the interaction between adjacent dimers belonging to the same PF in the nearest neighbour approximation, where k is the inter-dimer stiffness parameter. The next three terms in Eq. (2) represent a non-symmetric W-potential energy, where $A>0$, $B>0$ and $C>0$ are assumed. This potential determines the directions around which the dimer can oscillate. The very last term in Eq. (2) comes from the fact that the dimer is an electric dipole existing in the field of all other dimers. Of course, p is an electric dipole moment, while E_1 is our arbitrary choice. It is assumed that $p>0$ and $E_1>0$.

From Eqs. (1) and (2) and using generalized coordinate $q_n=\varphi_n$ and momentum $p_n=I\dot{\varphi}_n$, as well as Hamilton's equations of motion $\dot{q}_n=\partial H/\partial p_n$, $\dot{p}_n=-\partial H/\partial q_n$, we straightforwardly obtain the following dynamical equation of motion:

$$I\ddot{\varphi}_n=k\left(\varphi_{n+1}+\varphi_{n-1}-2\varphi_n\right)-A_0\varphi_n-C_0{\varphi_n}^2-B_0{\varphi_n}^3+D_0, \tag{3}$$

where $A_0=-A+3B{\theta_0}^2+pE_1$, $B_0=B-pE_1/6$, $C_0=3B\theta_0$ and $D_0=A\theta_0-B{\theta_0}^3+C=0$ [17].

As was mentioned above, we solve Eq. (3) using the SDA [18]. This mathematical procedure was explained including a lot of details in Ref. [19]. Its mathematical basis is a multiple-scale method or a derivative-expansion method [20,21].

According to the SDA, we assume small oscillations, i.e.

$$\varphi_n=\varepsilon\Phi_n, \quad \varepsilon<<1, \tag{4}$$

which changes Eq. (3) into

$$I\ddot{\Phi}_n=k\left(\Phi_{n+1}+\Phi_{n-1}-2\Phi_n\right)-A_0\Phi_n-\varepsilon C_0{\Phi_n}^2-\varepsilon^2B_0{\Phi_n}^3. \tag{5}$$

A key point in the procedure is that we expect the solution to be a modulated wave, i.e. in the form

$$\Phi_n(t)=F(\xi)e^{i\theta_n}+\varepsilon\left[F_0(\xi)+F_2(\xi)e^{i2\theta_n}\right]+\text{cc}+\text{O}(\varepsilon^2), \tag{6}$$

$$\xi=(\varepsilon nl,\varepsilon t), \qquad \theta_n=nql-\omega t, \tag{7}$$

where $F(\xi)$ and $e^{i\theta_n}$ represent an envelope and a carrier components, respectively. The function $e^{i\theta_n}$ obviously includes discreteness, while the envelope will be treated in a continuum limit. As the frequency of the carrier wave is much higher than the frequency of the envelope, we need two time scales, t and εt, for those two functions. Of course, the same

holds for the coordinate scales. In Eqs. (6) and (7), ω is the optical frequency of the linear approximation, $q = 2\pi/\lambda$ is the wave number, cc stands for complex conjugate terms and F_0 is real.

A rather tedious mathematics [17] shows that the functions F_0 and F_2 can be expressed through F, that is $F_0 = \mu|F|^2$ and $F_2 = \delta F^2$, while F is a solution of nonlinear Schrödinger equation (NLSE). The expressions for μ and δ are given in Ref. [17]. All this brings about a final result

$$\varphi_n(t) = 2A'\mathrm{sech}\left(\frac{nl - V_e t}{L}\right)\left\{\cos(\Theta nl - \Omega t) + A'\mathrm{sech}\left(\frac{nl - V_e t}{L}\right)\right.$$
$$\left.\times\left[\frac{\mu}{2} + \delta\cos(2(\Theta nl - \Omega t))\right]\right\}, \qquad (8)$$

where

$$V_e = V_g + U_e, \qquad U_e = \frac{P}{1-\eta}\left[-q + q\sqrt{1 + \frac{2(1-\eta)}{Pq^2}(\omega - qV_g)}\right]. \qquad (9)$$

The parameters P and Q are the dispersion coefficient and coefficient of nonlinearity, respectively. They are given in Ref. [17], as well as the expressions for A', L, Θ and Ω. The second expression in Eq. (9) was obtained based on the idea of a coherent mode (CM) [15,17], assuming that the envelope and carrier wave velocities are equal, that is $V_e = \Omega/\Theta$. Due to the equality of these velocities, the function $\varphi_n(t)$ is the same at any position *n*. This tempting idea has been used for years. However, recent numerical calculations show that this might not always be the case [17]. This interesting problem certainly requires further research. The meaning of the parameter η was explained in Ref. [15]. Its allowed interval is

$$0 \le \eta < 0.5. \qquad (10)$$

To plot the function $\varphi_n(t)$ the values of all parameters should be known or estimated. This rather tedious job was performed in Ref. [17]. The function is shown in Fig. 3 for a certain allowed combination of the parameters. This is obviously a localized modulated wave. One can see that, for the chosen values of the parameters, the angle φ takes the values from about $-6°$ to $8°$, while the wave covers about 112 dimers. The solitonic speed, corresponding to this example, is $V_e = 445\,\mathrm{m/s}$ [17].

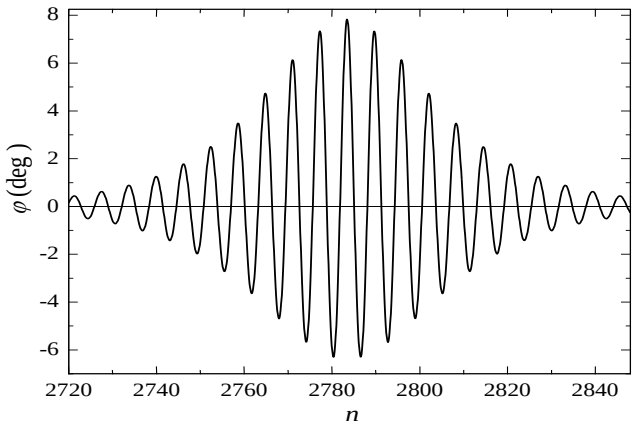


Fig. 3. The function $\varphi_n(t)$ as a function of n for $t = 50\text{ns}$, $k = 12\text{eV}$, $q = \pi/7l$ and $\eta = 0.495$. The CM is assumed

As a conclusion, we should point out that the model explained here is a two component one in a sense that the variable φ describes the oscillation of the dimer around the direction of the electric field, while θ determines the orientation of the field. The used SDA assumes a continuum limit [17]. A question if MT is predominantly discrete or continuum system was studied in Ref. [22].

The model explained here is obviously a mechanical one, but this does not mean that MTs are mechanical systems only. There are experiments which indicate electrodynamic activity of variety of cells in the frequency region from kHz to GHz, expecting MTs to be the source of this activity [4,23]. Therefore, MTs are both mechanical and electrical systems and, regarding their modelling, the best that should be done is to work towards more component models taking both characteristics into consideration. One such attempt is the model introduced in Ref. [24].

MTs can also be modelled as nonlinear RLC transmission lines [25-27]. Electrical activities of MTs are very important in fighting some diseases [28]. It is known that MTs can behave as biomolecular transistors capable of amplifying electrical information [29]. This may affect some crucial neuronal computational capabilities, such as memory and consciousness [29,30].

Finally, it might be interesting to mention kinocilium, a component of vestibular hair cells of the inner ear, comprising 10 pairs of MTs [31].

This is a sensory apparatus that receives the environmental signals and transmit them via collectively excited conformational changes in MTs [31]. Of course, this is possible due to the fact that MTs are capable of specific type of wave propagation, as explained above.

3. Kinks and bell-type solitons in DNA

There are a lot of models describing complex DNA dynamics [32,33]. The first nonlinear one was introduced in 1980, suggesting that nonlinear effects may focus the vibration energy of DNA into localized soliton-like excitations [34].

In this section, we rely on the well-known helicoidal Peyrard-Bishop (HPB) model for DNA dynamics [19,35]. This is an extended version of the PB model, which does not take helicoidal structure into consideration [36]. It might be important to mention that, in some papers, the HPB model is called Peyrard-Bishop-Dauxois (PBD) model. However, there is a similar model [37,38] called the PBD one and, consequently, it is more convenient to name it the HPB model.

As was explained above, the first step is Hamiltonian, from which we obtain the dynamical equation of motion. In Section 2, we demonstrated the SDA for solving it, the method that has been used for years to study DNA dynamics [19,35]. However, the CA was used recently [39] and this is what we explain in this section.

Fig. 4 shows a segment of DNA chain. Interactions along the strands are strong and the longitudinal displacements are neglected. The relevant ones at the position n are u_n and v_n, obviously along the weak hydrogen bonds. Keeping all this in mind, we can write the Hamiltonian as [19,35,39]

$$H=\sum\left\{\frac{m}{2}(\dot{u}_n^{\,2}+\dot{v}_n^{\,2})+\frac{k}{2}[(u_n-u_{n-1})^2+(v_n-v_{n-1})^2]\right.$$
$$\left.+\frac{K}{2}[(u_n-v_{n+h})^2+(u_n-v_{n-h})^2]+D[e^{-a(u_n-v_n)}-1]^2\right\}, \qquad (11)$$

where $m=5.4\cdot10^{-25}\text{kg}$ is the average nucleotide mass, a dot means the first derivative with respect to time, the parameters k and K are coupling constants of the harmonic longitudinal and helicoidal springs, respectively. The first term obviously represents kinetic energy, the

second one is the potential energy of the covalent bond, while the third term describes helicoidal interactions. Namely, due to the helicoidal structure, a nucleotide belonging to one strand at the position n comes close to the $n+h$ nucleotide from the other strand. We assume $h=5$ because the helix has a helical pitch of about 10 base pairs per turn [40]. The last term in Eq. (11) is Morse potential energy, describing the weak interaction, where the parameters D and a are the depth and inverse width of the Morse potential well, respectively.

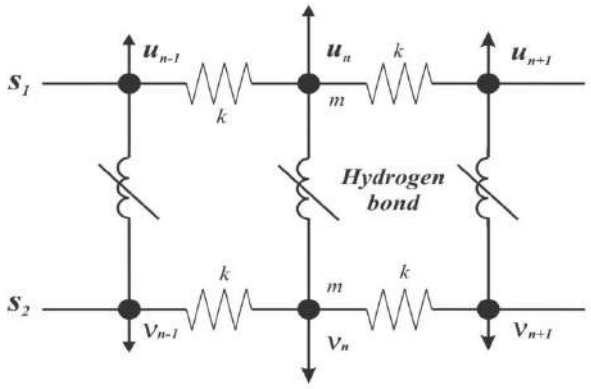


Fig. 4. A short portion of DNA molecule

It is convenient to introduce new coordinates $x_n=(u_n+v_n)/\sqrt{2}$ and $y_n=(u_n-v_n)/\sqrt{2}$, representing the in-phase and out-of-phase transversal displacements, respectively. In other words, $x_n(t)$ describes oscillation of the centre of mass of the nucleotide pair, while $y_n(t)$ represents their stretching. From a point of view of DNA activity (breathing, transcription, replication,…) the pair stretching is crucial, which means that we should see DNA molecule as a collection of nucleotide pairs rather than a collection of single nucleotides.

As was explained above, we use Eq. (11) and the Hamilton`s equations of motion, which brings about the following two completely decoupled dynamical equations of motion [19,35,39]

$$m\ddot{x}_n = k(x_{n+1}+x_{n-1}-2x_n)+K(x_{n+h}+x_{n-h}-2x_n), \tag{12}$$

$$m\ddot{y}_n = k(y_{n+1}+y_{n-1}-2y_n)-K(y_{n+h}+y_{n-h}+2y_n) +2\sqrt{2}aD(e^{-a\sqrt{2}y_n}-1)\,e^{-a\sqrt{2}y_n}. \tag{13}$$

The first dynamical equation is a standard linear discrete equation, whose solution is a linear wave (phonon). So, in what follows, we solve Eq. (13) to which we add a viscosity force $-\gamma \dot{y}_n$ on the right side, where γ is a viscosity coefficient [9,39,41-44]. We use both the CA $y_n(t) \to y(x,t)$ and appropriate series expansions, which yields to the following nonlinear partial DE

$$m\frac{\partial^2 y}{\partial t^2} - l^2\left(k - Kh^2\right)\frac{\partial^2 y}{\partial x^2} + Ay - By^2 + \gamma\frac{\partial y}{\partial t} = 0\,, \tag{14}$$

where $l = 3.4\,\mathring{\mathrm{A}}$ is a distance between the two neighbouring nucleotides in the same strand, $A = 4\left(K + a^2 D\right)$ and $B = 6\sqrt{2}a^3 D$ [39].

It is well known that, for a given wave equation, a travelling wave $y(\xi)$ is a solution which depends upon x and t through a unified variable $\xi \equiv \kappa x - \omega t$, where κ and ω are constants. This brings about the following ordinary DE [39]

$$\alpha\psi'' - \rho\psi' + \psi - \psi^2 = 0\,, \qquad \psi' \equiv \mathrm{d}\psi/\mathrm{d}\xi\,, \tag{15}$$

where

$$y = (A/B)\psi\,, \quad \alpha = \frac{m\omega^2 - l^2\kappa^2\left(k - Kh^2\right)}{A}\,, \quad \rho = \frac{\gamma\omega}{A}\,. \tag{16}$$

There are many procedures for solving Eq. (15). Some of them are: standard procedure [9,45], modified extended tanh-function (METHF) method [46-48], method of factorization [49-51], procedure based on Jacobian elliptic functions [52,53], the simplest equation method (SEM) [54-56], modified SEM [57], exponential function procedure [58,59], (G'/G)-expansion method [60,61], etc. Except the standard procedure and method of factorization, in all mentioned methods the function ψ is expected to be a series of known functions. However, the series expansion in terms of unknown functions is also possible [62,63]. In this section, we explain METHF method, probably the simplest procedure representing series expansions in terms of known functions. According to this procedure, we look for possible solutions of Eq. (15) in the form

$$\psi = a_0 + \sum_{i=1}^{M}\left(a_i\Phi^i + b_i\Phi^{-i}\right), \tag{17}$$

where the function $\Phi = \Phi(\xi)$ is a solution of the well-known Riccati equation [46-48]

$$\Phi' = b + \Phi^2\,. \tag{18}$$

The parameters a_0, a_i, b_i and b are real constants that should be determined, as well as the cut off integer M. One can easily show that, for Eq. (15), $M = 2$ [39]. We are looking for the solutions having physical sense and assume

$$\Phi = -\sqrt{-b}\tanh\left(\sqrt{-b}\,\xi\right), \qquad b_i = 0, \tag{19}$$

which is the solution of Eq. (18) for $b < 0$.

According to Eqs. (17) and (18), we obtain expressions for ψ', ψ'' and ψ^2 and Eq. (15) becomes $A_1\Phi + A_2\Phi^2 + A_3\Phi^3 + A_4\Phi^4 + A_0 = 0$, where A_i, $i = 0,...,4$, are coefficients depending on the parameters b, a_0, a_1, a_2 and α [39]. Of course, this equation is satisfied if all these coefficients are simultaneously equal to zero, which gives a system of five equations. This system brings about the following two solutions [39]

$$a_0^{(1)} = 1/4,\ \alpha^{(1)} = 6\rho^2/25,\ a_0^{(2)} = 3/4,\ \alpha^{(2)} = -6\rho^2/25,\ a_2{}^{(i)} = 6\alpha^{(i)}, \tag{20}$$

while the remaining parameters are common for both of them, i.e.

$$a_1 = -\frac{6\rho}{5}, \qquad b = -\frac{25}{144\rho^2} < 0. \tag{21}$$

Finally, according to Eqs. (17), (19), (20) and (21), we easily obtain the solutions we are looking for

$$\psi_1(\xi) = \frac{1}{4}\left(1 + 2\tanh w + \tanh^2 w\right),\ \ \psi_2(\xi) = \frac{1}{4}\left(3 + 2\tanh w - \tanh^2 w\right), \tag{22}$$

where $w = 5\xi/(12\rho)$. These functions are shown in Fig. 5 for $\rho = 5/12$. Numerically and analytically derived kink profiles nicely fit to each other. The numerical solutions were generated applying the simple Runge-Kutta procedure to Eq. (15), which was firstly transformed into the set of two ordinary first order DEs [56]. One can show that ψ_1 describes a supersonic kink, while ψ_2 corresponds to subsonic one [39]. This is related to the parameter α. Namely, Eq. (16) can be written as

$$\alpha = m\kappa^2\left(V^2 - c^2\right)/A, \qquad c^2 = l^2\left(k - Kh^2\right)/m, \tag{23}$$

where $V = \omega/\kappa$ and c are the solitonic and linear sound velocities, respectively. As A in Eq. (23) is positive, we conclude that the negative α corresponds to the subsonic soliton and vice versa. Some estimations of the velocities V and c can be found in Ref. [39].

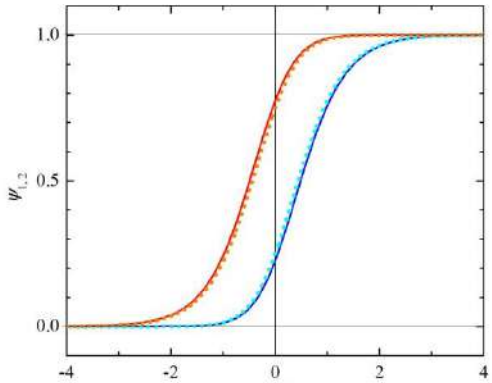


Fig. 5. Solutions $\psi_1(\xi)$ (blue) and $\psi_2(\xi)$ (red) for $\rho = 5/12$. The solid and dotted lines correspond to the numerically and analytically derived kinks, respectively

It might be interesting to study the solutions of Eq. (15) when viscosity is neglected. Applying the same procedure as above for $\rho = 0$ we obtain the following two solutions [39]

$$\psi_{10}(\xi) = \frac{1}{2}\left[-1 + 3\tanh^2\left(\sqrt{\frac{3}{2a_2^{(1)}}}\xi\right)\right], \qquad a_2^{(1)} > 0, \tag{24}$$

$$\psi_{20}(\xi) = \frac{3}{2}\left[1 - \tanh^2\left(\sqrt{-\frac{3}{2a_2^{(2)}}}\xi\right)\right], \qquad a_2^{(2)} < 0. \tag{25}$$

Obviously, these solutions are expressed through the parameters $a_2^{(1)}$ and $a_2^{(2)}$, and they are shown for $a_2^{(1)} = -a_2^{(2)} = 3/2$ in Fig. 6. These are bell-type solitons. The numerical solutions were generated applying the same procedure as for Fig. 5. The functions $\psi_{10}(\xi)$ and $\psi_{20}(\xi)$ represent the supersonic and subsonic solitons, respectively.

What has been shown so far is that the kinks and bell-type solitons may exist in DNA under certain conditions. A key question is which one really, or, at least, very likely exists. In other words, we should deal with stability of the mentioned solutions. We have performed a series of numerical simulations of Eq. (13) with and without viscosity term [39]. Eqs. (22), (24) and (25) represent initial conditions and the system was checked during 10ps [39]. Fig. 7 shows the function $y_1(\xi)$

corresponding, of course, to $\psi_1(\xi)$ in Fig. 5. The function $y_2(\xi)$ is almost indistinguishable from $y_1(\xi)$, as expected from Fig. 5. Basically, this numerical solution matches analytical one but the obtained kink decreases in time.

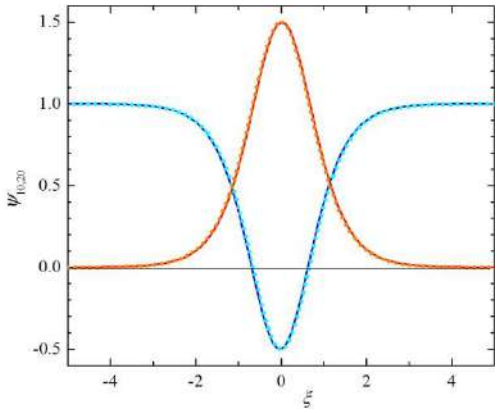


Fig. 6. Solutions $\psi_{10}(\xi)$ (blue) and $\psi_{20}(\xi)$ (red) for $\left|a_2^{(i)}\right| = 3/2$, $i = 1,2$. The solid and dotted lines correspond to the numerically and analytically derived solitons, respectively

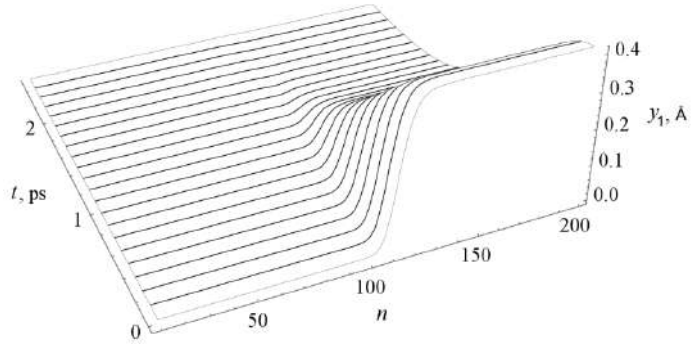


Fig. 7. Solution $y_1(\xi)$ for $\gamma = 2.8\times10^{-12}$ kg/s

When viscosity is neglected, the solutions (24) and (25) completely change their structures in time. An example is shown in Fig. 8. The solution $y_{20}(\xi)$, as well as $y_{10}(\xi)$, is obviously unstable. This should not bother us because these functions describe non-realistic case when viscosity is neglected. Therefore, if we compare Figs. 7 and 8 we can

conclude that the solitons $y_1(\xi)$ and $y_2(\xi)$ are acceptable, while $y_{10}(\xi)$ and $y_{20}(\xi)$ are not. This certainly shows that viscosity is crucial for the wave stability. However, our positive attitude towards Fig. 7 should be discussed. First of all, the soliton decreases in time, which means that it is not stable, at least it is not mathematically stable. However, it exists during a certain period of time and, from biological point of view, a question is if it can perform a required biological task during its lifetime. Let us assume that the kink y_1 lives about 1.5ps, which is suggested by Fig. 7. Its speed was estimated to be $V_2 = 1350\,\text{m/s}$ [39]. During this period of time the kink passes over the distance of about 6 nucleotide pairs. This value can be compared with the experimental value for RNA:DNA hybrid, which is about 8 pairs [64]. These two values match rather well, which means that the kinks $y_1(\xi)$ and $y_2(\xi)$ are biologically acceptable. We are going to return to the RNA:DNA hybrid in the next section.

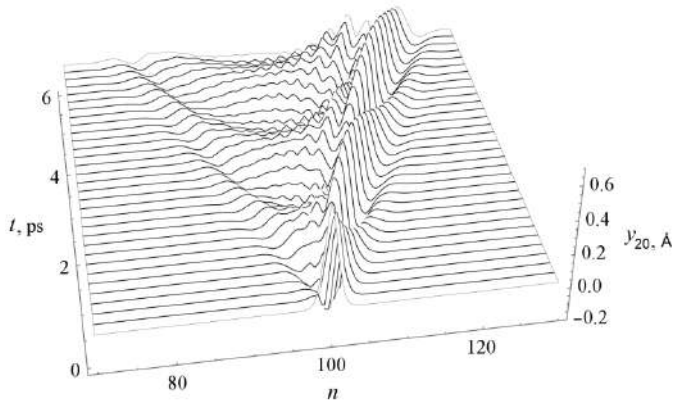


Fig. 8. Solution $y_{20}(\xi)$

4. Demodulated standing solitary wave and DNA-RNA transcription

In the previous two sections, we studied two biological systems and two mathematical methods. The combinations were MT-breather and DNA-kink. In this section, we study DNA using the SDA, i.e. the combination DNA-breather. Our goal is to study DNA-RNA transcription and it turned out that it may be possible within the idea of existence of the breathers in the chain [65]. We rely on the HPB model again and follow Ref. [65].

It is known that the transcription occurs at the segments of DNA chain that are surrounded by RNA polymerase molecules (RNAP), which is shown in Fig. 9 [66,67]. Let us call these segments transcription segments (TSs). One can see that one of the two DNA strands serves as a template for synthesis of a new RNA strand. It is important to know that the transcription is possible because DNA molecule opens locally at these segments, which implies significantly smaller coupling between base pairs. It was shown that the local opening could be seen as DNA breathing mode with extremely high amplitude [68], which, otherwise, can be conceived as a resonance mode [69].

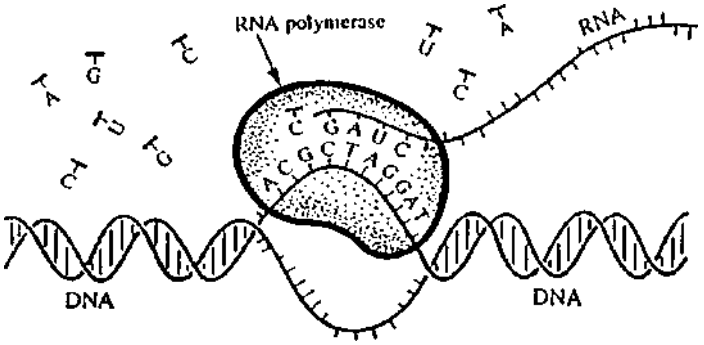


Fig. 9. DNA-RNA transcription (Taken from Ref. [66]).

The main goal of this section is to study DNA breathing at TSs in the context of two ideas. We explain why it would be biologically convenient if the soliton were demodulated at TSs. Our second idea is that the soliton becomes a standing one at the TSs. Hence, we can talk of demodulated standing solitary (DSS) mode. We believe that this mode decreases probability for genetic mistakes and yields to successful transcription.

Therefore, we deal with DNA and Eqs. (11) and (13) hold again. We use the SDA explained in Section 2, which means that we assume small oscillations ($y_n = \varepsilon \Phi_n$, $\varepsilon << 1$) as well as Eqs. (6) and (7). The final result is the function

$$y_n(t) = 2A\,\mathrm{sech}\left(\frac{nl - V_e t}{L}\right)\left\{\cos(\Theta nl - \Omega t) + A\,\mathrm{sech}\left(\frac{nl - V_e t}{L}\right)\left[\frac{\mu}{2} + \delta\cos(2(\Theta nl - \Omega t))\right]\right\}, \tag{26}$$

which is, practically, the same as Eq. (8) except that $\varphi_n(t)$ has been replaced by $y_n(t)$. Of course, the expressions for P, Q, μ and δ are different and given in Ref. [65], while Eqs. (9) and (10) hold again.

Let us get back to transcription and study a certain TS. When DNA gets copied into RNA, RNAP attaches itself to one of the two DNA strands, as shown in Fig. 9. This means that RNAP pulls nucleotides out of solution and form RNA according to DNA order of basis. Therefore, we can talk of DNA and RNA nucleotides.

Let us concentrate on one DNA adenine, for example. Normally, it is bonded with DNA thymine belonging to other strand but also interacts with RNA nucleotides, as can be seen from Fig. 9. The final positioning of RNA nucleotides should be a certain copy of the DNA segment, which means that our DNA adenine should attract a certain RNA uracile and repel the remaining RNA nucleotides [65]. This can be efficiently done only if the DNA adenine is far enough from its DNA partner during transcription, which is really the case due to the local opening.

We argued that the local opening is necessary but not sufficient condition for successful transcription [65]. The stretching of DNA, i.e. the distance between the DNA nucleotides belonging to the same pair, is described by Eq. (26). This obviously means that the respective DNA thymine and adenine are far from each other only during short periods of time and the chosen adenine does not have enough time to attract one RNA uracile. The carrier wave is crucial for soliton movement along DNA chain but is redundant when transcription occurs. Also, it makes sense to believe that only the envelope of Eq. (26) corresponds to local opening. All this suggests that the breather should be demodulated when it reaches a TS. This, practically, means that we should get rid of the cosine functions in Eq. (26), which means that the conditions

$$\Theta = 0, \qquad \Omega = 0 \tag{27}$$

should be satisfied at TSs [65]. A crucial question is how demodulation happens at these segments. A simple explanation is that RNAP changes chemical milieu for DNA nucleotides, i.e. the values of relevant parameters, especially D and a, which yields to the values accommodating Eq. (27). That DNA surrounding, which is, practically, viscosity, can lead to demodulation was shown in Ref. [70]. This section could be understood as a mathematical analysis of this discovery.

One more idea was suggested recently [65]. Both local opening and demodulation increases time during which the DNA and RNA

nucleotides interact. This is probably not enough but there is one more mechanism to increase this time. Namely, this time is bigger if the soliton velocity is smaller. Hence, biologically convenient soliton is the one which is as slow as possible at the TSs and we have proposed the idea that the soliton wave becomes a standing one at these segments. By the standing wave we assume the one for which the envelope velocity is equal to zero, that is

$$V_e = 0. \tag{28}$$

There have been some suggestions how to experimentally determine the soliton speed, width and even its character [71,72]. They are based on micromanipulation experiments on the single DNA molecule [73-81]. Unfortunately, the expressions (27) and (28) have been neither approved nor disapproved so far. What theoreticians can do is to study if the DSS mode is possible [65]. In particular, we investigate if there exists a certain value of q satisfying Eqs. (27) and (28). We introduce new parameters x and p defined as $K = xk$ and $a^2 D = pk$ [65] and use $h = 5$, as explained earlier. Both x and p should be much less than one because k determines the strong covalent interaction. There are a couple of requirements that should be satisfied, such as $\eta < 0.5$, $P > 0$, $Q > 0$, etc. [65]. For each of them we find intervals for ql satisfying it. For example, we plot the function $P(ql)$ for different values of x and determine the accepted intervals for ql. In the end, we compare all these intervals and obtain the final result, which is [65]

$$\left.\begin{array}{rl} 0.577 < ql < 0.578 & \text{and} \\ 0.65 < ql < 0.84 \quad \text{for} & x = 1/30, \\ 0.47 < ql < 0.81 \quad \text{for} & x = 1/50, \\ ql < 0.77 \quad \text{for} & x = 1/80 \end{array}\right\}. \tag{29}$$

One can notice an extremely narrow interval for $x = 1/30$. Such intervals do not exist for $x < 1/39.7$ [65]. Also, the lower limit for ql does not exist for $x < 1/62.5$ [65].

Two examples are shown in Fig. 10 for $D = 0.07\text{eV}$ and $k = 12\,\text{N/m}$ [19]. We see the demodulated waves whose amplitudes are $A_1 = 6.1\,\mathring{\text{A}}$ and $A_2 = 1.6\,\mathring{\text{A}}$. The appropriate wave widths are $\Lambda_1 = 8l$ and

$\Lambda_2 = 7.8l$, respectively [65]. This means that these waves cover about 8 base pairs, which perfectly matches the experimental value for the DNA:RNA hybrid [64]. Notice close result regarding the kinks in the previous section. The big amplitudes are in agreement with the local opening of the chain. The solitons in Fig. 10 have almost equal widths but their amplitudes vary remarkably. This is so because the amplitude depends on the arbitrary and still unknown parameter k. An idea how to experimentally determine k was offered in Ref. [65].

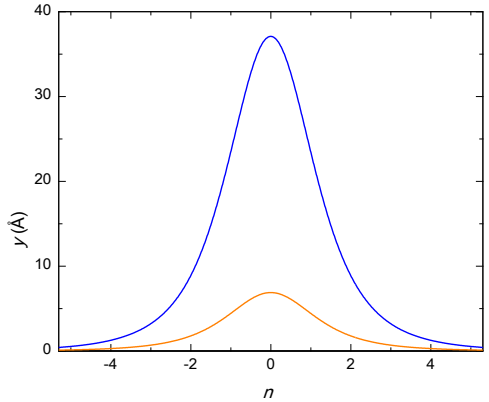


Fig. 10. Demodulated solitary wave for $ql = 0.47\text{rad}$, $x = 1/50$ (blue) and $ql = 0.15\text{rad}$, $x = 1/80$ (orange)

Therefore, we showed that the values for ql, satisfying our postulates explained above, exist. The results are in excellent agreement with the experimental values. A patient reader may have noticed that the big amplitudes are not in agreement with the HPB model, which assumes small amplitudes. This means that the used model predicts the local opening but is not adequate for quantitative analysis. Also, viscosity has been neglected in this section

5. Conclusion

Nonlinear dynamics of biological nanosystems is very interesting and developing branch of science. We here studied two of them and explained two mathematical methods. Nonlinearity has been manifested through the solitary waves.

Internal structures of nucleotides and dimers were neglected. As these are relatively big particles, classical physics was used. However, if their

internal structures are taken into consideration then quantum mechanics becomes relevant. A common example could be ab initio calculations. Also, if we study charge transfer processes in these systems we should use quantum mechanical approach [82,83].

References

[1] P. Dustin, Microtubules, Springer, Berlin, 1984.

[2] S. Zdravković, *J. Serb. Chem. Soc.* **82** (2017) 469.

[3] S. Sahu, S. Ghosh, K. Hirata, D. Fujita and A. Bandyopadhyay, *Appl. Phys. Lett.* **102** (2013) 123701.

[4] D. Havelka, M. Cifra, O. Kučera, J. Pokorný and J. Vrba, *J. Theor. Biol.* **286** (2011) 31.

[5] S. Hameroff and R. Penrose, *Phys. Life Rev.* **11** (2014) 39.

[6] J. E. Schoutens, *J. Biol. Phys.* **31** (2005) 35.

[7] E. Nogales, M. Whittaker, R. A. Milligan and K. H. Downing, *Cell* **96** (1999) 79.

[8] P. Drabik, S. Gusarov and A. Kovalenko, *Biophys. J.* **92** (2007) 394.

[9] M. V. Satarić, J. A. Tuszyński and R. B. Žakula, *Phys. Rev. E* **48** (1993) 589.

[10] S. Zdravković, L. Kavitha, M. V. Satarić, S. Zeković and J. Petrović, *Chaos Solitons Fract.* **45** (2012) 1378.

[11] S. Zdravković, S. Zeković, A. N. Bugay and M. V. Satarić, *Appl. Math. Comput.* **285** (2016) 248.

[12] S. Zdravković and G. Gligorić, *Chaos* **26** (2016) 063101.

[13] S. Zdravković, A. N. Bugay and A. Yu. Parkhomenko, *Nonlinear Dynam.* **90** (2017) 2841.

[14] S. Zdravković, M. V. Satarić, A. Maluckov and A. Balaž, *Appl. Math. Comput.* **237** (2014) 227.

[15] S. Zdravković, A. N. Bugay, G. F. Aru and A. Maluckov, *Chaos* **24** (2014) 023139.

[16] S. Zdravković, M. V. Satarić and V. Sivčević, *Nonlinear Dynam.* **92** (2018) 479.

[17] S. Zdravković, S. Zeković and A. N. Bugay, Tangential model of microtubules and semi-discrete approximation, *in preparation.*

[18] M. Remoissenet, *Phys. Rev. B* **33** (1986) 2386.

[19] S. Zdravković, *J. Nonlin Math. Phys.* **18**, Suppl. 2 (2011) 463.

[20] R. K. Dodd, J. C. Eilbeck, J. D. Gibbon and H. C. Morris, Solitons and Nonlinear Wave Equations, Academic Press, Inc., London, 1982.

[21] T. Kawahara, *J. Phys. Soc. Japan* **35** (1973) 1537.

[22] S. Zdravković, A. Maluckov, M. Đekić, S. Kuzmanović and M. V. Satarić, Appl. Math. Comput. **242** (2014) 353.

[23] M. Cifra, J. Pokorný, D. Havelka and O. Kučera, *BioSystems* **100** (2010) 122.

[24] A. N. Bugay, *Nonlin. Phenomena Complex Sys*. **18** (2015) 236.

[25] M. V. Satarić, D. I. Ilić, N. Ralević and J. A. Tuszynski, *Eur. Biophys. J*. **38** (2009) 637.

[26] P. G. Ghomsi, J. T. T. Berinyoh and F. M. M. Kakmeni, *Chaos* **28** (2018) 023106.

[27] F. II Ndzana and A. Mohamadou, *Chaos* **29** (2019) 013116.

[28] F. T. Ndjomatchoua, C. Tchawoua, F. M. M. Kakmeni, B. P. Le Ru and H. E. Z. Tonnang, *Chaos* **26** (2016) 053111.

[29] A. Priel, A. J. Ramos, J. A. Tuszynski and H. F. Cantiello, *Biophys. J*. **90** (2006) 4639.

[30] M. R. Cantero, C. V. Etchegoyen, P. L. Perez, N. Scarinci and H. F. Cantiello, *Sci. Rep.-UK* **8** (2018) 11899.

[31] M. V. Sataric, D. L. Sekulic, B. M. Sataric and S. Zdravković, *Prog. Biophys. Mol. Bio.* **119** (2015) 162.

[32] L. V. Yakushevich, Nonlinear Physics of DNA, Wiley Series in Nonlinear Science, John Wiley, Chichester, 1998.

[33] G. Gaeta, C. Reiss, M. Peyrard and T. Dauxois, *Riv. Nuovo Cimento* **17** (1994) 1.

[34] S. W. Englander, N. R. Kallenbach, A. J. Heeger, J. A. Krumhansl and S. Litwin, *Proc. Natl. Acad. Sci. (USA)* **777** (1980) 7222.

[35] T. Dauxois, *Phys. Lett. A* **159** (1991) 390.

[36] M. Peyrard and A. R. Bishop, *Phys. Rev. Lett.* **62** (1989) 2755.

[37] T. Dauxois, M. Peyrard and A. R. Bishop, *Phys. Rev. E* **47** (1993) R44.

[38] M. Peyrard, *Nonlinearity* **17** (2004) R1.

[39] S. Zdravković, D. Chevizovich, A. N. Bugay and A. Maluckov, *Chaos* **29** (2019) 053118.

[40] T. R. Strick, M. N. Dessinges, G. Charvin, N. H. Dekker, J. F. Allemand, D. Bensimon and V. Croquette, *Rep. Prog. Phys.* **66** (2003) 1.

[41] S. Zdravković and M. V. Satarić, *Chin. Phys. Lett.* **24** No.5 (2007) 1210.

[42] C. B. Tabi, A. Mohamadou and T. C. Kofané, *Chin. Phys. Lett.* **26** (2009) 068703.

[43] J. B. Okaly, A. Mvogo, R. L. Woulaché and T. C. Kofané, *Commun. Nonlinear Sci. Numer. Simulat*. **55** (2018) 183.

[44] V. Vasumathi and M. Daniel, *Phys. Rev. E* **80** (2009) 061904.

[45] A. Gordon, *Physica B* **146** (1987) 373.

[46] E. Fan, *Phys. Lett. A* **277** (2000) 212.

[47] S. A. El-Wakil and M. A. Abdou, *Chaos Solitons Fract.* **31** (2007) 840.

[48] A. H. A. Ali, *Phys. Lett. A* **363** (2007) 420.

[49] O. Cornejo-Perez and H. C. Rosu, *Prog. Theor. Phys.* **114** (2005) 533.

[50] O. Cornejo-Pérez, J. Negro, L. M. Nieto and H. C. Rosu, *Found. Phys.* **36** (2006) 1587.

[51] W. Alka, A. Goyal and C. N. Kumar, *Phys. Lett. A* **375** (2011) 480.

[52] C. Dai and J. Zhang, *Chaos Solitons Fract.* **27** (2006) 1042.

[53] S. Zeković, S. Zdravković, L. Kavitha and A. Muniyappan, *Chin. Phys. B* **23** (2014) 020504.

[54] N. A. Kudryashov, *Phys. Lett. A* **342** (2005) 99.

[55] N. A. Kudryashov, *Chaos Solitons Fract.* **24** (2005) 1217.

[56] S. Zdravković and G. Gligorić, *Chaos* **26** (2016) 063101.

[57] N. A. Kudryashov and N. B. Loguinova, *Appl. Math. Comput.* **205** (2008) 396.

[58] M. N. Alam, M. G. Hafez, M. A. Akbar and H.-O. Roshid, *J. Sci. Res.* **7** (2015) 1.

[59] M. N. Alam and F. B. M. Belgacem, *Mathematics* **4** (2016) 1.

[60] M. N. Alam, M. G. Hafez, F. B. M. Belgacem and M. A. Akbar, *Nonlinear Stud.* **22** (2015) 613.

[61] M. N. Alam, F. B. M. Belgacem and M. A. Akbar, *J. Appl. Math. Phys.* **3** (2015) 1571.

[62] A. J. M. Jawad, M. D. Petković and A. Biswas, *Appl. Math. Comput.* **217** (2010) 869.

[63] S. Zdravković and S. Zeković, *Chin. J. Phys.* **55** (2017) 2400.

[64] J. Gelles and R. Landick, *Cell* **93** (1998) 13.

[65] S. Zdravković, M. V. Satarić, A. Yu. Parkhomenko and A. N. Bugay, *Chaos* **28** (2018) 113103.

[66] C. R. Calladine, H. R. Drew, B. F. Luisi and A. A. Travers, Understanding DNA-The Molecule and How It Works, Elsevier Academic Press, Third Edition, 2004.

[67] T. Lipniacki, *Phys. Rev. E* **60** (1999) 7253.

[68] S. Zdravković and M. V. Satarić, *Europhys. Lett.* **78** (2007) 38004.

[69] S. Zdravković and M. V. Satarić, *Europhys. Lett.* **80** (2007) 38003.

[70] S. Zdravković, M. V. Satarić and Lj. Hadžievski, *Chaos* **20** (2010) 043141.

[71] S. Zdravković and M. V. Satarić, *Phys. Rev. E* **77** (2008) 031906.

[72] S. Zdravković and M. V. Satarić, *Phys. Lett. A* **373** (2009) 4453.

[73] S. B. Smith, L. Finzi and C. Bustamante, *Science* **258** (1992) 1122.

[74] G. U. Lee, L. A. Chrisey and R. J. Colton, *Science* **266** (1994) 771.

[75] S. B. Smith, Y. Cui and C. Bustamante, *Science* **271** (1996) 795.

[76] U. Bockelmann, B. Essevaz-Roulet and F. Heslot, *Phys. Rev. Lett.* **79** (1997) 4489.

[77] J. F. Allemand, D. Bensimon, R. Lavery and V. Croquette, *Proc. Natl. Acad. Sci. USA* **95** (1998) 14152.

[78] H. Clausen-Schaumann, M. Rief, C. Tolksdorf and H. E. Gaub, *Biophys. J.* **78** (2000) 1997.

[79] T. Lionnet, S. Joubaud, R. Lavery, D. Bensimon and V. Croquette, *Phys. Rev. Lett.* **96** (2006) 178102.

[80] E. A. Galburt, S. W. Grill and C. Bustamante, *Methods* **48** (2009) 323.

[81] F. Mosconi, J. F. Allemand, D. Bensimon and V. Croquette, *Phys. Rev. Lett.* **102** (2009) 078301.

[82] D. Čevizović, S. Galović and Z. Ivić, *Phys. Rev. E* **84** (2011) 011920.

[83] D. Čevizović, Z. Ivić, Ž. Pržulj, J. Tekić and D. Kapor, *Chem. Phys.* **426** (2013) 9.

Computer simulation of virus COVID-19 spreading and deep learning for medical images*

Nenad Filipovic†
Faculty of Engineering, University of Kragujevac, Sestre Janjica 6, 34000 Kragujevac, Serbia,

ABSTRACT

In this chapter, we coupled finite element analysis with deep learning approach in order to analyze CT findings associated with COVID-19 and infection model of virus spreading inside the lung. The spreading model starts from binding to receptors of the host cell where the RNA of the virus is uncoated in the cytoplasm. Then, transcription/translation processes generate a new viral RNA material and proteins. After that, the virus is assembled within vesicles followed by virus release and starting to infect other cells.

We modeled with finite element method the airway coupled to a parenchyma model where virus spreading starting from alveoli to other parts of airway. The results are compared with observation with CT (GGO particularly) for different time (days of patient starting infection with COVID-19). Also, these finite element simulations are used for training and validated of Unet neural network architecture for pneumonia detection. This coupled bioengineering and data mining approach can open a new strategy for prediction of virus spreading in the specific patients.

1.1 Introduction

Up to current knowledge there are three different stages of COVID-19 disease: stage I, which stand for a period of asymptomatic incubation with or without detectable virus; stage II, which is a period with symptoms that are not severe, with the presence of virus and stage III known as stage of severe respiratory symptoms with high viral load (Wang et al 2020).

* This research was funded by Ministry of Education, Science and Technological Development of Serbia, grants OI 174028 and III 41007.
† e-mail address: fica@kg.ac.rs

In Wuhan, among over 1000 patients analyzed, all the age groups were infected evenly, except children and adolescents where the virus appeared only occasionally. The results showed that approximately 15% of the confirmed cases develop severe phase of the virus, with higher incidence in patients over 65 (Guan et al 2020). The question remains open regarding the development of the severe phase of the virus in some patients, while other remain in milder phases. There are two phases of the immune responses induced by SARS-CoV-2 infection. The first phase comes when a specific adaptive immune response is required to eliminate the virus during the incubation and non-severe stages. This will preclude disease progression to severe stages. If the patient is generally of a good health and if he/she has an appropriate genetic background (e.g. HLA) then the patients exhibits specific antiviral immunity. If an immune response is not strong, the virus will induce the destruction of the affected tissues, which specifically refers to organs that have high ACE2 expression. This is second phase of the virus, when it induces innate inflammation in the lungs that is largely mediated by pro-inflammatory macrophages and granulocytes. So, lung inflammation leads to life-threatening respiratory disorders (Xu et al 2020). With the onset of severe lung damage, efforts should be made to reduce inflammation and the symptoms. A good general health did not always show as an advantage for patients who acquired the severe stage of the virus. It was found that HA has the ability to absorb water up to 1000 times its molecular weight. This implies that inhibiting the production of HA may help COVID-19 patients breathe. Medical staff can give patients medical grade hyaluronidase to reduce the accumulation of HA and, consequently, clear the jelly in the lung. CT images of patients showed characteristic white patches called "ground glass", containing fluid in the lungs (Wang et al 2020). Shi et al 2020 proposed approaches to the treatment of COVID-19 believing that the two-phase division is very important: the first immune defense-based protective phase and the second inflammation-driven damaging phase. They suggested that doctors should try to boost immune responses during the first phase, and suppress it in the second. They proposed Vitamin B3 to be used with the start of coughing, as a vitamin having highly lung protective function. When breathing becomes very difficult, hyaluronidase can be used intratracheally and at the same time 4-MU can be given to inhibit HAS2 (Shi et al 2020).

1.2 Method

1.2.1. Image analysis

The SARS-CoV-2 can be specifically detected in respiratory secretions or plasma samples RT-PCR. Sometimes, patients with COVID-19 may have negative initial RTPCR results. It can be due to insufficient cellular material

used for detection or due to improper extraction of nucleic acids from clinical materials. Imaging examination can effectively and conveniently observe the pulmonary parenchymal abnormalities, solve some challenges associated with clinical diagnosis in suspected patients with a false-negative result, especially symptomatic patients with dyspnea and respiratory distress. Radiological examination is of great importance in the early detection and treatment COVID-19 patients. Radiological examinations are relatively easy to perform and can produce fast diagnosis. Because COVID-19 infection is caused by the unprecedented SARS-CoV-2, there are only few related imaging studies. Among the studies that have been conducted, multifocal consolidation or ground-glass opacity (GGO) was observed on chest radiographs and CT (Lei et a 2020, Chen et al 2020, Huang et al 2020).

The virus is developing fast and it has varying symptoms. Consequently, there is a need to systematically analyze the chest CT findings associated with this disease, for the prevention and control measures. According to Fleischner Society recommendations, there are several observations which can be detected in CT examination of COVID-19 patients. There are:

- GGO (appears as hazy increased opacity in the lung, with the preservation of bronchial and vascular margins), consolidation (appears as a homogeneous increase in pulmonary parenchymal attenuation that obscures the margins of vessels and airway walls),
- crazy-paving pattern (appears as thickened interlobular septa and intralobular lines superimposed on a background of GGO),
- halo sign (appears as GGO surrounding a nodule or mass) (Wormanns and Hamer 2015, Hansell et al 2008).

There are also a reticular pattern, nodules, rounded opacities, cavitation, a crazy-paving pattern, an air bronchogram, a halo sign, bronchial wall thickening, a subpleural curvilinear line, and pulmonary fibrosis.

New coronavirus (COVID-19) has some unique features – the tests to prove a patient being positive use polymerase chain reaction (PCR), however, infected patients with pneumonia may have specific patterns that are only moderately characteristic for the human eye on chest X-ray and computed tomography (CT) images (Ng et al 2020). COVID-19's transmission rate depends on the capacity to identify infected patients reliably with a low rate of false negatives. Additionally, a false positive should be reduced as much as possible as it is required to avoid further burden on the healthcare system by putting patients to unnecessary quarantine. All this shows that proper infection control is necessary and timely detection of the disease would help the supportive care, which is necessary for the patients affected by COVID-19.

In January 2020, a paper was published with COVID-19 clinical and paraclinical feature details by a Chinese team of researchers. Abnormalities in chest CT images with most bilateral involvement are present (Huang et al

2020). Intensive care unit (ICU) patients on admission show multiple bilateral lobular and subsegmental areas of consolidation as the typical findings in chest CT images (Huang et al 2020). In comparison, chest CT images of non-ICU patients show bilateral ground-glass opacity and subsegmental areas of consolidation (Huang et al 2020). These patients express bilateral ground-glass opacity with resolved consolidation in later chest CT images (Huang et al 2020). Some indications show that COVID can be better diagnosed using radiological imaging (Fang et al 2020, Ai et al 2020).

The major characteristics of COVID-19 are GGO, followed by GGO with consolidation, rounded opacity, a crazy-paving pattern, and an air bronchogram. Although pathological changes in COVID-19 patients' lungs have not yet been studied, recent studies have reported that it is closely related (with 88% identity) to two bat-derived severe acute respiratory syndrome (SARS)–like coronaviruses, with approximately 79% homology with SARS-CoV and approximately 50% homology with MERS-CoV (Zhou et al 2020, Lu et al 2020).

Pathological changes of SARS patients include injury to pulmonary epithelial cells, hyaline membrane formation on in the inner wall of the alveoli, and a large number of tissue cells and mass thrombi formed by proliferative fibrous tissue that block the small airway and air cavity (Koo et al 2018). These pathological changes may be the main pathological basis of the CT findings, especially the GGO.

1.2.2 Spreading of COVID-19 virus

There are three ways in which virus can spread in tissues: virus transport, virus multiplication in host cells and the virus-induced immune response. It is well known that cytotoxic T cells remove infected cells with a rate determined by the infection level. The mathematical model consists of reaction-diffusion equations which describe the different regimes of infection spreading. It can be directed of a low level infection, a high level infection or a transition between both are determined by the initial virus load and by the intensity of the immune response. Viruses are non-cellular organisms who need cells to replicate their genomes and produce progeny. They will expand locally around the entry site of a newly infected organism, depending on the mode of transmission. The virus could not divide itself automatically (Zhai et al 1997). As an organism consisted by certain special nucleic acid and protein, the denaturation of virus can be regarded as a temperature dependent rate problem, just as any natural substances do. It is very important to understand COVID-19 replication cycle (Fig. 1.1) and its interactions with the immune system. This cycle contains four stages. The first stage is the binding to receptors of the host cell where the virus

RNA is uncoated in the cytoplasm. The second stage is transcription/translation processes which generate new viral RNA material and proteins. The third stage is virus assembly within vesicles followed by virus release. The fourth stage is infecting of other cells.

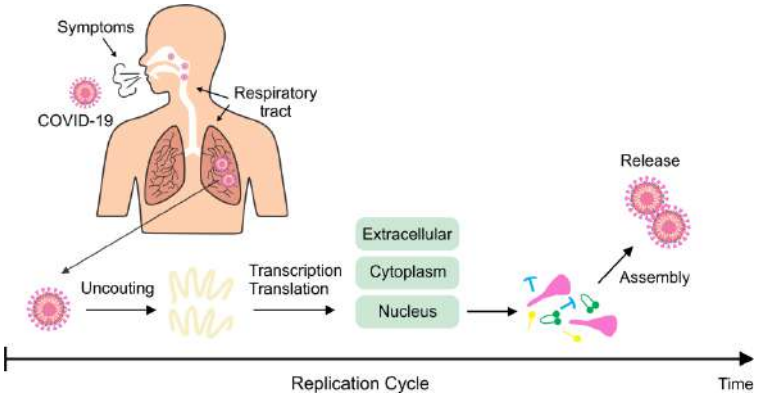


Fig 1.1. Replication Cycle of COVID-19. The process starting from binding to receptors of the host cell where the virus RNA is uncoated in the cytoplasm. Then transcription/translation processes generate a new RNA material of the virus and proteins. Then virus is assembled within vesicles followed by virus release and starting to infect other cells.

There are several approaches of virus spreading where the target cell model has served to represent several diseases such as HIV, Hepatitis virus, Ebola (Nguyen et al 2015, 2017), influenza (Baccam et al 2006, Hernandez-Vargas et al 2014, Pawelek et al 18), among many others. One of the detailed references for viral modelling can be found in Hernandez-Vargas et al 2019. Zou et al. 2020 presented the viral load in nasal and throat swabs of 17 symptomatic patients. COVID-19 replication cycles showed to last longer than flu. The period lasts about 10 days or more after the incubation period (Anderson et al 2020, Zou et al 2020). Here, we contribute to the mathematical study of COVID-19 dynamics at within-host level based on data presented by (Wolfel et al. 2020). We used ordinary differential equations (ODEs), based on the viral kinetics data reported by Woelfel et al. 2020 in infected patients with COVID-19. Viral load was sampled from throat swab cultures and measured in Copies/mL, g Swab, at Log10 scale. In order to make a quantitative recapitulation of experimental data Differential Evolution (DE) algorithm (Storn and Price 1997) was applied.

1.2.3 Exponential Growth and Logarithmic Decay Model

Based on the experimental data (Woelfel et al 2020), the viral dynamic is divided into two parts, exponential growth (Vg) and decay (Vd) modelled by equations (1) and (2), respectively.

$$\frac{dV_g}{dt} = \rho V_g \tag{1.1}$$

$$\frac{dV_d}{dt} = -\rho\mu V_d \tag{1.2}$$

It is assumed that the growth of the virus starts at the onset of symptoms, with initial viral concentration Vg(0). The parameter ρ is the growth rate of the virus. The parameter μ quantifies the decay rate of the virus, while Vd(0) the initial value of the virus in decay phase. The mathematical model used here to represent coronavirus dynamics is based on the target cell-limited model (Hernandez-Vargas et al 2019, Perelson 2002, Ciupe and Heffernan 2017). Coronavirus can replicate in a variety of cell types, including epithelial cells. The coronavirus infection model is as follows:

$$\frac{dU}{dt} = -\beta UV \tag{1.3}$$

$$\frac{dI}{dt} = \beta UV - \delta I \tag{1.4}$$

$$\frac{dV}{dt} = pI - cV \tag{1.5}$$

The states of host cells can be susceptible (U) and infected (I). Viral particles (V) infect susceptible cells with a rate β ((Copies/mL)-1 day-1). When the cells are productively infected, they release the virus at a rate p (Copies/mL day-1 cell-1) and virus particles are cleared with rate c (day-1). Infected cells are cleared at rate δ (day-1) as consequence of cytopathic viral effects and immune responses. Coronaviruses infects mainly in differentiated respiratory epithelial cells (Tyrrell and Myint 1996). Previous mathematical model for influenza (Hernandez-Vargas et al 2014) has considered about 107 initial target cells (U(0)). Initial values for infected cells (I(0)) are taken as zero.

1.2.4 Physiology of human respiratory system

The human respiratory system consists of two lungs irrigated with fresh air through a dyadic structure called the tracheo-bronchial tree. The respiratory muscles are around the lung. These muscles contract during inspiration creating lung expansion, and relax at expiration resulting in lung deflation. Lung

physiology and mathematical modeling of ventilation are given in (Bates 2009), (Vannier 2012) and (Viebel 1963). The fresh air is reach with dioxygen (O2) which goes through the lung. When O2 comes to alveola, it exchanges with vascularization. Further it dissolved into blood to supply body tissue. In the other direction, carbon dioxide leaves blood circulation, passing through the lung in opposite direction and exhaled to the air. Tracheo-bronchial tree in human is a non-symmetric dyadic branching structure. The starting point is trachea, that split into two airways irrigating the right and left lung. It contains approximately 24 generations (Viebel 1963). Airway dimensions started from upper airways with centimetric diameter up to 15 networks of millimeter size airways. In the first 17 generations, the air flow is convective; in lower branches of the tree, Reynolds numbers are low and the flow regime is diffusive. There are about 30000 acini, which are dyadic terminal sub-trees. They end in alveolar sacs where gas exchange takes place. Three hundred millions alveoli representing an exchange surface of about 100 m2 (Bates 2009).

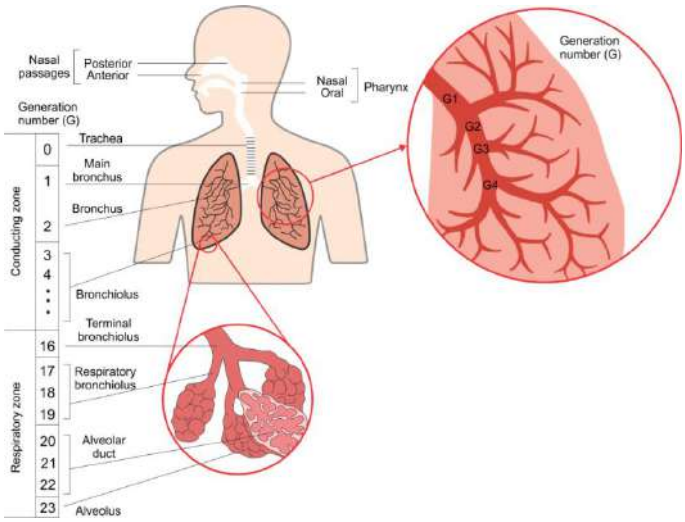


Fig 1.2. Schematic of the human respiratory system, from extra-thoracic components to alveoli, and cast of human airways

1.2.5 Finite element method for airway and lobes

On a macroscopic level, airflow into and out of the lung is driven by pressure differences between the alveoli and the outside environment. Activity of the muscles surrounding the parenchyma induces lung volume variations while breathing. Lungs are divided into units called lobes that are not mechanically

attached and can slide with respect to one another. The human left lung contains two and right lung contains three lobes. Our model is the airway coupled to a parenchyma model (Figure 1.3). Upper airways geometry is segmented from a CT scan. This framework is used to mimic virus spreading from alveoli to other airway geometry. We have been used around 500,000 finite elements to model both airway and lobes (Filipovic at el 2012, 2013, Tsuda et al 2008, 2012). (Figure 1.3).

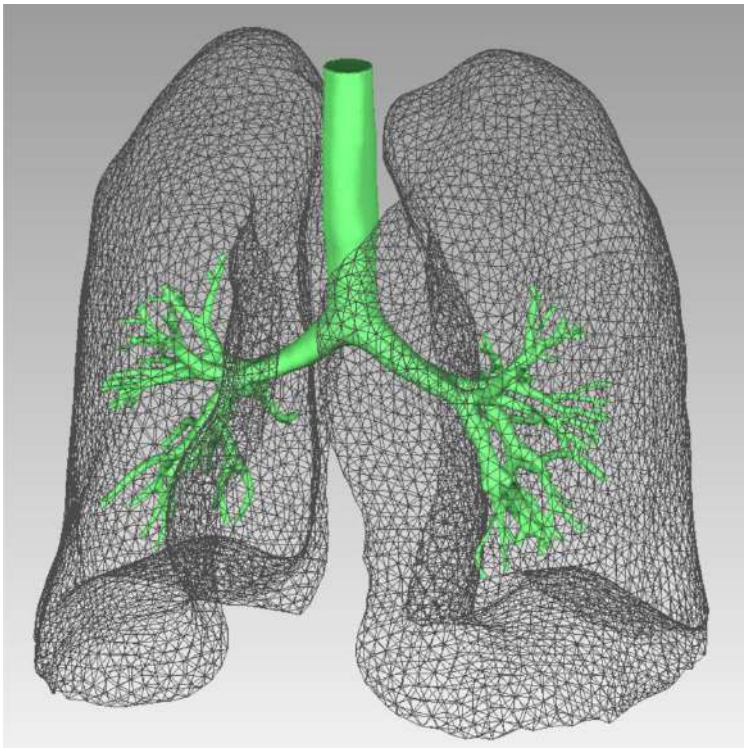


Fig. 1.3. Finite element mesh of upper airways with the lung mesh and all lobes

1.3 Results

1.3.1 Simulation of virus spreading

Our initial simulation results taken into account virus exponential growth (Vg) and decay (Vd) equations (1) and (2) with dynamics infection model defined by equations (3-5) after 5 and 10 days (Figure 1.4). The red color object occupied virus. When virus propagates through airway system through alveoli they

connect to their wall and cause high ACE2 expression. It induces innate inflammation by pro-inflammatory macrophages and granulocytes and makes liquid in the alveoli space and bronchi. It seems that liquid can be seen in CT images as GGO. Our simulation mimic several days of this severe process which can cause the growing of the space which liquid occupies inside the airway system. This process takes a several days. We compare our finite element simulation with spreading COVID-19 with CT finding as GGO particularly (Figure 1.5).

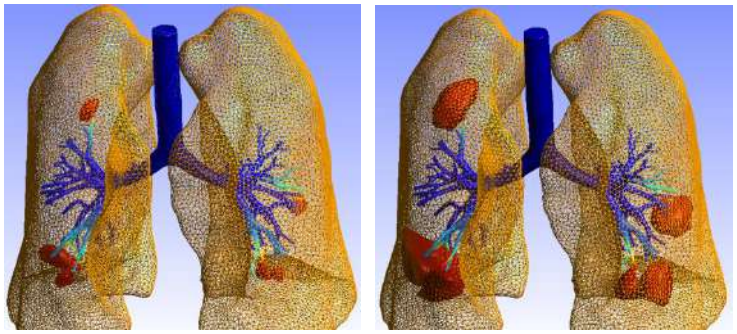


Fig. 1.4. Simulation of spreading of COVID -19 virus after 5 days and 10 days of infection starting. The red object simulates virus spreading inside lobes and airways.

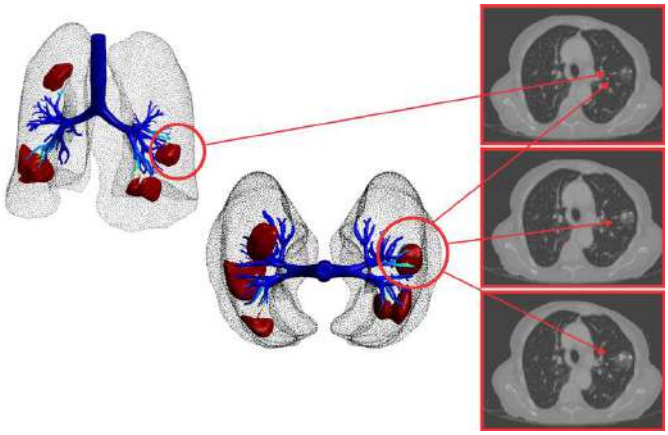


Fig. 1.5. Results of virus spreading and comparison with CT images. Comparison of finite element simulation of virus spreading and CT findings (GGO particulary)

1.3.2 Deep learning with X-rays and CT images

There are several challenges related to the implementation of deep learning for detecting pneumonia from X-ray images. Kaggle database provides 1GB of chest X-Ray images of patient with pneumonia (Kaggle 2019). More into the specifics of the task, there are datasets updated every day related to CT (Kaggle 2020) and X-ray images of patients with COVID 19 and non-COVID-19. Additionally, the White House prepared the COVID-19 Open Research Dataset (CORD-19) with leading research groups. CORD-19 contains over 50000 articles, majority of which is with full text, about COVID-19, SARS-CoV-2, and related coronaviruses. This data base is free for the research community to apply recent advances in natural language processing and other AI techniques in order to produce new insights in support of the ongoing fight against this infectious disease (kaggle.com). GitHub also has its database with chest X-ray or CT images for COVID-19 patients, as well as for other viruses such as MERS, SARS, and ARDS (GitHub 2020). Taking into account all the available datasets, in addition to the dataset from the Clinical Centre Kragujevac, one solid dataset can be created in order to investigate the COVID-19 presence in images. An example of two X ray images infected with COVID-19 virus admitted to Clinical Centre Kragujevac are given in Fig. 1.6.

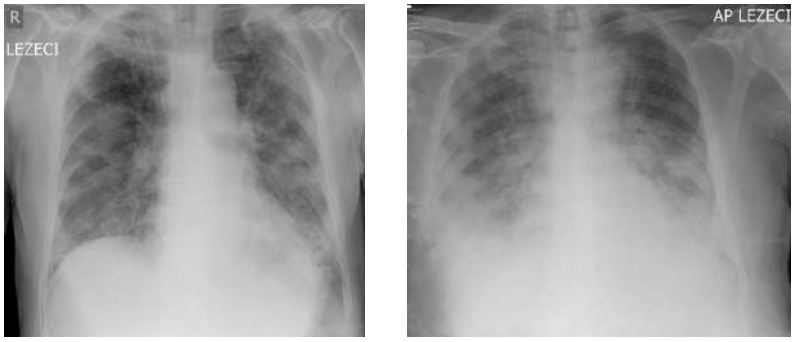


Fig. 1.6. Two examples of X ray images of patients form Clinical Centre Kragujevac infected with COVID-19

In order to overcome some of the drawbacks of the mentioned traditional methods, deep neural networks are introduced and they have started to make an impact on all aspects of computer vision including medical image analysis starting form image to image translation (Shiri et al 2019a), image reconstruction (Shiri et al 2019b), as well as super resolution image generation (Shiri et al 2019c). It has been concluded that convolutional neural networks (CNNs) are able solve various segmentation problems very well in the area of

deep learning-based method including US image segmentation challenges (Moradi et al 2019). This is mainly because of its strong capability in image feature extraction and shift invariance characteristics. Because of the fact that U-net showed promising results in segmentation of various tissues in medical images, we propose the use of U-net in segmentation and classification of COVID-19 cases.

U-net neural network architecture for COVID-19 pneumonia detection that that is proposed is shown in Fig. 7. This architecture has proven to be applicable to various medical image segmentation issues (Moradi et al 2019, Ronneberger 2015). The U-net consists of contraction path and expansion path. Contraction path consists of two 3x3 convolutional layers and 2x2 max pooling. This can help to extract more advanced features and it reduces the size of feature maps. Expansion path consists of consecutive 2x2 up-conv and two 3x3 convolutional layers. This recovers the size of segmentation map, but with loss of the localization information. After each up-conv, we have concatenation of feature maps. This helps to give localization information from contraction path to expansion path. Because fine-grained features may be lost in down-sampling stage, there are cross-over connections used by concatenating feature maps that are equally sized. After scale removal in images, as well as patient data in corners of the image, resulting images were of size 708 x 708 pixels. We propose the network with convolutions filter size of 3×3. The network requires a fixed size input which will be determined later, depending on the resolution and size of the dataset available. The images can be resized depending on the problem definition and demand for computer power. Pixel intensities of the X-ray images used as masks will be rescaled to a range (0, 1).

Neural network will be implemented using Python Tensorflow and Keras (Tensorflow 2019). The training process is presumed to last for about 10 epochs, stochastic gradient descent, learning rate and regularization factor will be varied to determine the most suitable one. ReLU activation function will be used. The data is fed to the network, which then propagates along the described paths (contraction, expansion, and concatenation). The final result is a binary segmented image.

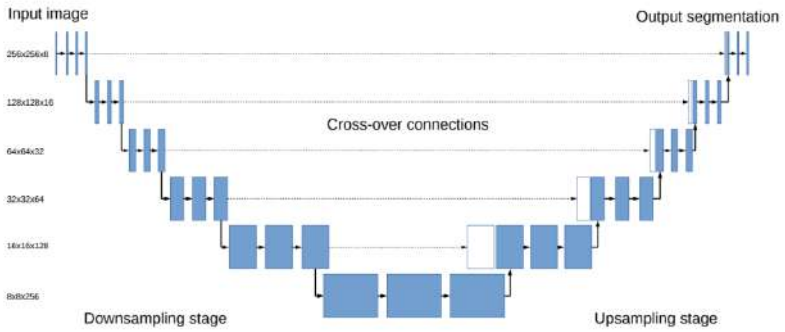


Fig. 1.7. U-net architecture (Smistad et al 2017) - Downward and upward parts show downsampling and upsampling, respectively. The straight cross over connections are represented with dashed arrows

Available dataset will be divided into training, validation and testing datasets in a ratio yet to be determined. By data augmentation, it is possible to increase the size of the training set allowing the network to observe more diversified data points. If necessary, additional images can be created by data augmentation using mirroring, scaling, zooming and flipping.

Additionaly, beside segmentation purpose, the aim is to label the output belonging to one of the categories 1-5 which will mark degree of illnes. Also, healthy subject data will be avalable, meaning that the methodology should be able to distingush bteween the non-covid and covid images.

For the evaluation metrics, segmentation accuracy of the proposed automatic method will be compared to the manual segmentation. General practice is to use dice similarity coefficient D (Lin et al 2003) to calculate the overlapping regions between the automatic segmentation marked as S and the ground truth marked as G:

$$D = \frac{2|S \cap G|}{|S| + |G|} \tag{1.6}$$

Other evaluation metrics can be used, i.e. Hausdorff distance H, which is calculated in millimeters:

$$H = \max\left(\max_{i\in[0,O-1]} d(i,G,S),\ \max_{i\in[0,M-1]} d(i,G,S)\right) \tag{1.7}$$

where $d(i,G,S)$ represents the distance from contour point i in G which is the closest to the contour point in S. O and M represent the number of pixels on the contours G and S, respectively.

Jaccard coefficient (JC) is calculated similarly to DM and is generally used to compare the similarity and diversity of two segmented areas. It is defined as the number of pixels of the intersected area, divided by the number of pixels that represent the union area

$$JC = \frac{|S \cap G|}{|S \cup G|} \tag{1.8}$$

For the evaluation of the label matching, confusion matrix and statistical measures precision, recall, F1 score and accuracy has been used.

1.4 Discussion and conclusion

In this chapter we analysed CT findings associated with COVID-19 which has been detected as GGO, consolidation, crazy-paving pattern and halo sign. Also Xrays images are analysed from local hospital in Clinical Center in Kragujevac, Serbia. U-net neural network architecture for pneumonia detection has been proposed. The infection model of virus has been mathematically defined. Once cells are productively infected, they release virus at some rate and virus particles are cleared with another rate. Infected cells are cleared at rate δ (day-1) as consequence of cytopathic viral effects and immune responses.

We modeled with finite element method the airway coupled to a parenchyma model. The results show virus spreading starting from alveoli to other parts of airway. The idea was to match observation with CT (GGO particularly) for different time (days of patient starting infection with COVID-19). Also these finite element simulations can replace a number of images for training and validated Unet neural network architecture for pneumonia detection. This coupled approach can open a new strategy for prediction of virus spreading in specific patients. Further analyses are necessary to go into this direction.

References

Ai T, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv and L. ... Xia, "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology*, p. 200642, 2020.

Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet. 2020;doi:10.1016/S0140-6736(20)30567-5.

Baccam P, Beauchemin C, Macken Ca, Hayden FG, Perelson AS. Kinetics of influenza A virus infection in humans. Journal of virology. 2006;80(15):7590-9.

Bates JHT, "Lung Mechanics," 2009.

Chen N, Zhou M, DongX et al (2020) Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395(10223):507–513

Ciupe SM, Heffernan JM. In-host modeling. Infectious Disease Modelling. 2017;2(2):188-202.doi:10.1016/j.idm.2017.04.002.

Fang Y, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang and W. Ji, "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, p. 200432, 2020.

Filipovic N, B Gibney, D Nikolic, M Konerding, S Mentzer, ATsuda, Computational analysis of lung deformation after murine pneumonectomy, Computer Methods in Biomechanics and Biomedical Engineering, Vol.17, No.8, pp. 838-844, Doi 10.1080/10255842.2012.719606, 2012

Filipovic N, Gibney Barry C, Kojic Milos R, Nikolic Dalibor, Isailovic Velibor, Ysasi Alexandra, Konerding Moritz A, Mentzer Steven J, Tsuda Akira, Mapping cyclic stretch in the postpneumonectomy murine lung, Journal of Applied Physiology, Vol.115, No.9, pp. 1370-1378, Doi 10.1152/japplphysiol.00635.2013, 2013.

GitHub "COVID X ray-dataset," 2020. (Online). Available: https://github.com/ieee8023/covid-chestxray-dataset. (Accessed 19 4 2020).

Graw F, Perelson AS. Modeling Viral Spread. Annual Review of Virology. 2015;(July):1-18. doi:10.1146/annurev-virology-110615-042249.

Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. The New England journal of medicine. 2020.https://doi.org/10.1056/NEJMoa2002032.

Handel A, Longini IM, Antia R. Neuraminidase inhibitor resistance in influenza: Assessing the danger of its generation and spread. PLoS Computational Biology. 2007;3(12):2456-2464.

Hansell DM, Bankier AA,MacMahon H,McLoud TC, Müller NL, Remy J (2008) Fleischner Society: glossary of terms for thoracic imaging. Radiology 246:697–722

Hernandez-Vargas EA, Middleton RH. Modeling the three stages in HIV infection. Journal of Theoretical Biology. 2013;320:33-40.

Hernandez-Vargas EA, Wilk E, Canini L, Toapanta FR, Binder SC, Uvarovskii A, et al. Effects of aging on influenza virus infection dynamics. Journal of Virology. 2014;88(8):4123-31.

Hernandez-Vargas EA, Wilk E, Canini L, Toapanta FR, Binder SC, Uvarovskii A, et al. Effects of aging on influenza virus infection dynamics. Journal of Virology. 2014;88(8):4123-31.

Hernandez-Vargas EA. Modeling and Control of Infectious Diseases: with MATLAB and R. 1st ed. ELSEVIER Academic Press; 2019.

Hernandez-Vargas EA. Modeling and Control of Infectious Diseases: with MATLAB and R. 1st ed. ELSEVIER Academic Press; 2019.

Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506

Huang C, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu and Z. ... Cheng, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497-506, 2020.

Kaggle "COVID CT," Kaggle, 2020. (Online). Available: https://www.kaggle.com/luisblanche/covidct. (Accessed 19 4 2020).

Kaggle "COVID-19 Open Research Dataset Challenge," 2020. (Online). Available: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge. (Accessed 19 4 2020).

Kaggle, "Chest X ray pneumonia," 2019. (Online). Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia. (Accessed 19 4 2020).

Koo HJ, Lim S, Choe J, Choi SH, Sung H, Do KH (2018) Radiographic and CT features of viral pneumonia. Radiographics 38:719–739

Lei J, Li J, Li X, Qi X (2020) CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia. Radiology. https://doi.org/10.1148/radiol.2020200236:200236

Lin N, W. Yu and J. Duncan, "Combinative multi-scale level set framework for echocardiographic image segmentation," *Med Image Anal*, vol. 7, pp. 529-537, 2003.

Lu R, Zhao X, Li J et al (2020) Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395(10224):565–574

Moradi S, A. Alizadehasl and J. e. a. Dhooge, "MFP-Unet: A Novel Deep Learning Based Approach for Left Ventricle Segmentation in Echocardiography," *Physica Medica*, pp. 58-69, 2019.

Ng MY, E. Y. Lee, J. Yang, F. Yang, X. Li, H. Wang and C. K. M. ... Hui, "Imaging profile of the COVID-19 infection: radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.

Nguyen VK, Binder SC, Boianelli A, Meyer-Hermann M, Hernandez-Vargas EA. Ebola virus infection modeling and identifiability problems. Frontiers in Microbiology. 2015;6:1-11.

Nguyen VK, Hernandez-Vargas EA. Windows of opportunity for Ebola virus infection treatment and vaccination. Scientific reports. 2017;7(1):8975.

Pawelek KA, Dor D, Salmeron C, Handel A. Within-host models of high and low pathogenic influenza virus infections: The role of macrophages. PLoS ONE. 2016;11(2):1-16. doi:10.1371/journal.pone.0150568.

Perelson AS, Ribeiro RM. Modeling the within-host dynamics of HIV infection. BMC Biology. 2013;11(1):96. doi:10.1186/1741-7007-11-96.

Perelson AS. Modelling Viral and Immune System Dynamics. Nature Reviews Immunology. 2002;2(1):28-36.

Pinkevych M, Kent SJ, Tolstrup M, Lewin SR, Cooper DA, S_gaard OS, et al. Modeling of Experimental Data Supports HIV Reactivation from Latency after Treatment Interruption on Average Once Every 5-8 Days. PLoS Pathogens. 2016;12(8):8-11.doi:10.1371/journal.ppat.1005740.

Reluga TC, Dahari H, Perelson AS. Analysis if Hepatitis C Virus Infection Models with Hepatocyte Homeostasis. SIAM journal on applied mathematics. 2009;69(4):999-1023.

Rong L, A ASP. Modeling HIV persistence , the latent reservoir , and viral blips. Journal of Theoretical Biology. 2009;260(2):308{331. doi:10.1016/j.jtbi.2009.06.011.

Ronneberger O, P. Fischer and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.*, 2015, October.

Shi Y, Wang Y, Shao C, Huang J, Gan J, Huang X, Bucci E, Piacentini M, Ippolito G, Melino G, COVID-19 infection: the perspectives on immune responses, Cell Death & Differentiation 2020, https://doi.org/10.1038/s41418-020-0530-3

Shiri I, K. Leung and P. e. a. Geramifar, "PSFNET: ultrafast generation of PSF-modelledlike PET images using deep convolutional neural network," *J Nucl Med*, vol. 60, p. 1369, 2019c.

Shiri I, K. Leung and P. e. a. Ghafarian, "Simultaneous Attenuation Correction and Reconstruction of PET Images Using Deep Convolutional Encoder Decoder Networks from Emission Data," *J Nucl Med,* vol. 60, p. 1370, 2019b.

Shiri I, P. Ghafarian, P. Geramifar, K.-Y. Leung, M. Ghelichoghli and e. a. Oveisi M, "Direct attenuation correction of brain PET images using only emission data via a deep convolutional encoder-decoder (Deep-DAC)," *Eur Radiol* , pp. 1-13, 2019a.

Smistad E, A. Ostvik, B. Haugen and L. Lovstakken, "2D left ventricle segmentation using deep learning," *IEEE Int Ultrason Symp*, pp. 1-4, 2017.

Storn R, Price K. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization. 1997; p. 341-359. doi:10.1023/A:1008202821328.

Tensorflow, (Online). Available: https://www.tensorflow.org/. (Accessed 5 10 2019).

Tsuda A, FS Henry, S Haber, D Haberthür, N Filipovic, D Milasinovic, J Schittny, The simultaneous role of an alveolus as flow mixer and flow

feeder for the deposition of inhaled submicron particles,J Biomech Eng, Vol.134, No.12, Doi 10.1115/1.4007949, 2012.

Tsuda, A., Filipovic, N., Haberthür, D., Dickie, R., Matsui, Y., Stampanoni, M. and Schittny, J.C., Finite element 3D reconstruction of the pulmonary acinus imaged by synchrotron X-ray tomography, *J Appl Physiol* 105: 964-976, 2008

Tyrrell DAJ, Myint SH. Coronaviruses. University of Texas Medical Branch at Galveston; 1996. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21413266.

Vannier C, "Modélisation mathématique du poumon humain," PhD thesis Hal Id: tel-00739462, 2012.

Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of Hospitalized Patients With Novel Coronavirus-Infected Pneumonia in Wuhan, China. Jama.2020.https://doi.org/10.1001/jama.2020.1585.

Weibel AR, *Morphometry of the human lung*.: Springer, 1963.

Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mueller MA, et al. Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. medRxiv. 2020; p. 2020.03.05.20030502 doi:10.1101/2020.03.05.20030502.15/19

Woelfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Mueller MA, et al. Clinical presentation and virological assessment of hospitalized cases of coronavirus disease 2019 in a travel-associated transmission cluster. medRxiv. 2020; p. 2020.03.05.20030502.doi:10.1101/2020.03.05.20030502.

Wormanns D, Hamer OW (2015) Glossary of terms for thoracic imaging–German version of the Fleischner Society recommendations. Rofo 187(8):638–661

Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al.Pathologicalfindings of COVID-19 associated with acuterespiratory distress syndrome. The Lancet Respiratory medicine.2020.https://doi.org/10.1016/S2213-2600(20)30076-X

Zhai ZH (1997) Cell Biology (in Chinese), pp.13–18, Beijing: Higher Education Press

Zhou P, Yang XL, Wang XG et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. https://doi.org/10.1038/s41586-020-2012-7

Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. New England Journal of Medicine. 2020; p.NEJMc2001737. doi:10.1056/NEJMc2001737.

An Essay on the Origins and Development of Nonlinear Dynamics, Chaos and Complex Systems

Miguel A. F. Sanjuán*
Department of Physics
Universidad Rey Juan Carlos
28933 Móstoles, Madrid, Spain

ABSTRACT

We can find dynamics in every field of science, including economics, chemical reactions, physiology or neurodynamics, showing its intrinsic interdisciplinarity. The interactions between the parts of a system and their feedback mechanisms constitute a source of nonlinearity and complexity, which added to the sensitive dependence on initial conditions, hallmark of chaotic behavior, lead to a crucial change of perspective with serious consequences in the understanding of science. Relevant problems such as the three-body problem in celestial mechanics, turbulence in fluid dynamics, irreversibility in statistical physics; or the logistic equation in population dynamics, have been at the origins of this fascinating field of nonlinear dynamics, chaos, and complex systems. A knowledge of its origins, as well as the many schools of mathematics and physics that have contributed to its development, allow us to better understand the discipline and the breadth of its many applications to science.

1. Introduction

Nonlinear dynamics is the discipline that aims to study nonlinear dynamical systems, which are those systems defined by one or more variables evolving with time where the response is not proportional to the stimulus. Chaos is one of the three kinds of motion, in addition to the periodic and quasi-periodic motions. Naturally, there are as many dynamical systems as there are variables that have a temporal evolution, which gives us an idea of the interdisciplinary nature and scope of nonlinear dynamics [1-4].

* e-mail address: miguel.sanjuan@urjc.es

Many of the ideas and concepts of complexity, such as a science of complex systems, pose a real challenge for the integration of various disciplines, among which we must point out nonlinear dynamics and chaos theory, statistical physics, stochastic processes, information theory, network theory, engineering science, life sciences, and computer sciences. This listing is naturally not complete, but it gives an idea of the challenge behind the idea of complexity. This intended goal means more than the idea of crossing disciplinary boundaries, but rather to integrate disciplines in a common background.

Much has been discussed in recent years about a fruitful dialogue between different scientific disciplines, not only to solve old problems, but also as a source of inspiration for new problems. For the study of complexity this is one of the fundamental elements, since its object of study covers problems related to both the so-called hard and soft sciences. Complex systems exist in biology, chemistry, physics, sociology, economics, etc. In any case, the true dialogue between disciplines so necessary for the advancement of knowledge of complex systems in particular, and of science in general, is still lacking.

Different paths that have led to the understanding of chaos as we understand it today. Among them, I like to point out: (1) The logistic map and population dynamics (2) Nonlinear oscillators (3) The three-body problem in celestial mechanics (4) Turbulence in fluid dynamics and (5) Irreversibility in statistical mechanics. All of them will be discussed throughout this article.

2. Nonlinear Dynamics and Deterministic Chaos

As previously discussed, dynamics is the science that studies the variation in time of different variables, that is, its motion. Basically, there are three types of motion: stationary and equilibrium; periodic and quasi-periodic; and finally chaotic motion. Considering the notion of motion in a broad sense, it is easy to understand that we can find dynamical systems in any scientific discipline. That is why it is customary to say that one of the characteristics of nonlinear dynamics is its interdisciplinarity, since with its methods we can approach the study of many different phenomena that evolve over time.

We use the term "nonlinear" to logically contrast it with the term "linear", since the linear approach is the one traditionally used in science due to its mathematical simplicity. The linear approach implies the assumption of properties such as: (1) Proportionality: small causes cause small effects (2) Additivity: the whole is equal to the sum of its parts (3) Replication: the same action under the same conditions produces the same result and (4) clear relationships between cause and effect: it is enough to know a little about the behavior of a system to fully know it.

However, when nature's relationships are not linear, it leads us to very different situations. A proportional relationship between two variables x and y,

where $y = kx$, indicates a linear relationship. Therefore, any relationship between two variables that does not respond to a proportional relationship like the previous one will be nonlinear. It is easy to figure out that most dynamical systems are nonlinear.

When there are relationships of nonlinearity, there can be chaotic behavior that has the following properties: (1) There is no proportionality: small causes can cause large effects (2) Emergence: additivity does not exist, so the whole is greater than the sum of its parts (3) Sensitive dependence on initial conditions: which can make that the same experiment can never be reproduced exactly; and finally (4) Nonlinearity that can generate instabilities, discontinuities and unpredictability, which requires flexibility, adaptability, dynamic change, innovation and reaction capacity.

Possibly one of the deepest ideas about the nature of what is known as chaotic behavior is the idea of sensitive dependence on the initial conditions. That is, trajectories of a chaotic system move away from each other as time progresses when they start from very close initial points. This fact has very drastic consequences on the predictability of a system.

From this viewpoint, it is somehow surprising to read the following sentence from the Chapter XIV of *The Origin of Species* (1859) [5] by Charles Darwin:

"More individuals are born than can possibly survive. A grain in the balance will determine which individuals shall live and which will shall die, which variety or species shall increase in number, and which shall decrease, or finally become extinct",

that in a certain sense shows already the true notion of sensitive dependence on initial conditions.

In this regard, it is also interesting to bring up a famous rhyme traditionally associated with Benjamin Franklin, although antecedents of the same idea date back to the 15th century, and which is known as *"For Want of a Nail "* as shown in Fig. 1.

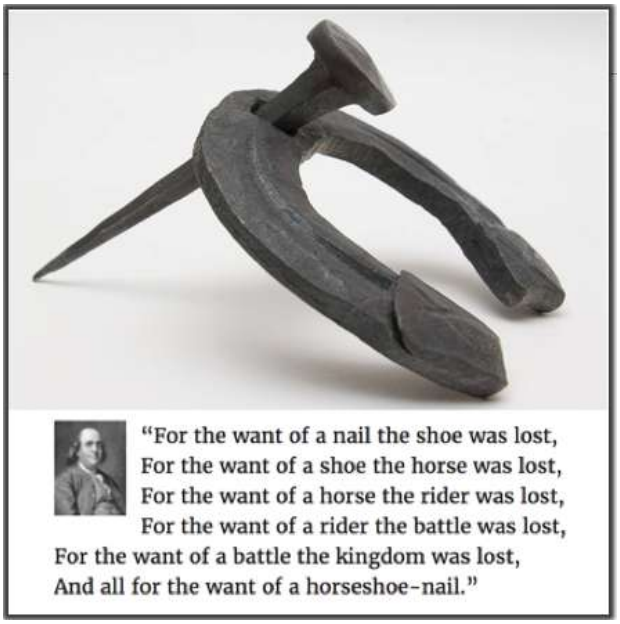


Fig. 1: The rhyme *"For Want of a Nail..."* offers an intuitive and poetic image of the idea of sensitive dependence on initial conditions, which is the hallmark of chaos.

We can define chaotic behavior or chaos as a type of motion that is derived from deterministic temporal dynamics of simple systems that can in fact be described in terms of few variables and whose fundamental characteristics are: (1) Being irregular in time, and given its nonlinear character, of course, cannot be the superposition of periodic motions, being in fact of aperiodic nature and bounded (2) Be unpredictable in the long term and very sensitive on the initial conditions and (3) Be complex, but ordered in the phase space, presenting a geometry of a fractal nature. If we compare the chaotic motion with the regular motion, we can say that the latter is repetitive, periodic, predictable and with a simple geometry, while the former is irregular, unpredictable and with a complicated geometry.

There are different types of chaotic motions. It is fundamentally called permanent chaos when once a dynamical system finds this state it remains in it forever. On the other hand, it is called transient chaos when this chaotic behavior occurs only in a certain period of time and the system subsequently behaves differently. Furthermore, dynamical systems generally distinguish between dissipative and conservative based on whether or not they conserve

energy. Well, for dissipative systems permanent chaos occurs in what is called a chaotic attractor in the phase space. However, in the case of transient chaos, chaotic transients occur in a fractal set. In the conservative case, on the one hand, permanent chaos occurs in bounded regions of the phase space and transient chaos is associated, for example, with the phenomenon of chaotic scattering that occurs in numerous physical phenomena, giving rise to very complex fractal structures. These concepts will be explained in more detail throughout the article.

Dynamical systems are usually classified as discrete and continuous depending on whether time is measured discretely or continuously. A paradigm for discrete dynamical systems is the logistic map, defined as

$$x_{n+1} = rx_n(1 - x_n),$$

which is an iterative equation where the index n indicates an iteration that is linked to the discrete way of measuring time. Figure 2 shows a Feigenbaum bifurcation diagram corresponding to the logistic map, where the final state of the system is displayed as a function of the variation of parameter r.

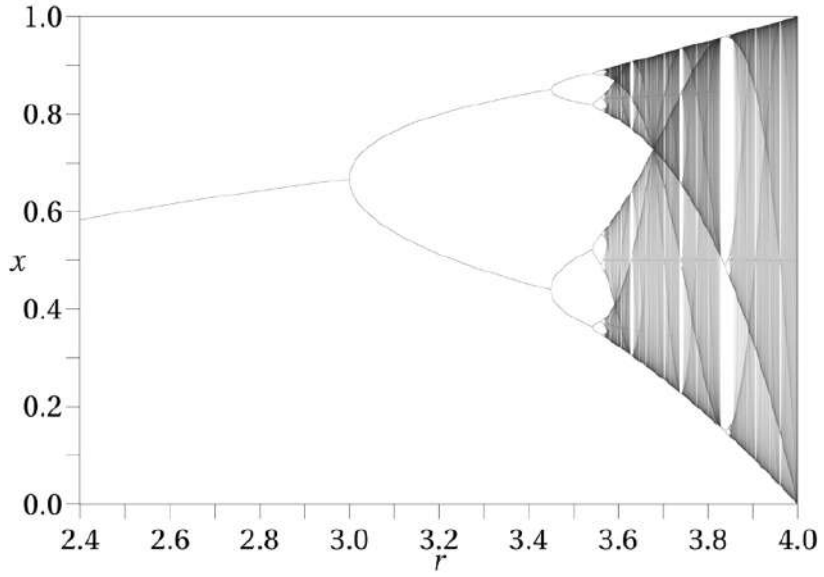


Fig. 2: Feigenbaum bifurcation diagram corresponding to the logistic map. It indicates how the final state of the system varies depending on the value of parameter r.

A paradigm for continuous systems is the simple pendulum (Fig. 3). It consists of a body of mass m that hangs on a cord that is in principle inextensible and of negligible mass, and whose suspension point moves periodically.

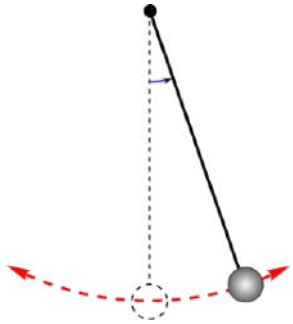


Fig. 3. Oscillatory motion of a pendulum

In this system, time is measured continuously, and therefore it can be modeled, once normalized, using a differential equation such as

$$\ddot{x} + \mu\dot{x} + \sin x = F\cos\omega t \,.$$

This equation contains in addition to the inertia term (the second derivative of the position), the friction of intensity μ which is proportional to the velocity, the nonlinear sinusoidal term and an external periodic forcing of amplitude F and frequency ω. If we consider the friction with the air and assuming that the suspension point remains fixed, then the motion will gradually dampen until it stops in its stable equilibrium position. When the suspension point moves periodically, it has the effect of introducing energy into the system, causing oscillations to be maintained. However, it is also possible to give rise to another type of motion of an irregular and unrepeatable nature on a periodic basis, which is chaotic motion.

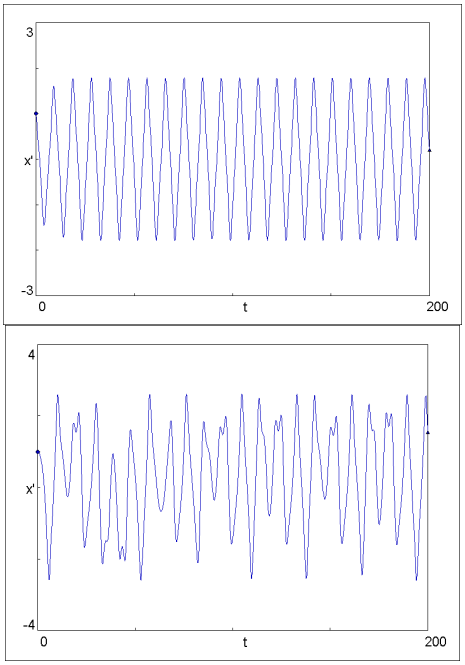


Fig. 4. Evolution of velocity over time for periodic and chaotic motions.

Figure 4 indicates the time evolution of the velocity of a pendulum. In one of them clearly the periodic nature of the oscillations can be observed, that is, after a certain period of time the same motion is repeated. In the other figure, however, an irregular behavior is shown, which turns out to be chaotic, where it can be observed that the same type of motion is not reproduced after any period of time. This is precisely one of the characteristics of chaotic motion, its lack of periodicity.

A simple example of periodic system is the mass-spring system formed by a body that is attached by a spring to a wall (Fig. 5). If the displacement with respect to the equilibrium position is very small, then the spring recovery force is proportional (linear) to the displacement, so that the result of motion is regular, oscillatory, and periodic.

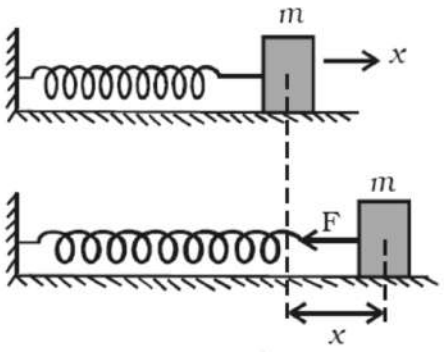


Fig. 5. Oscillatory motion of a system formed by a body of mass *m* attached to a spring.

When the deformation is greater, the spring recovery force is not linear, leading to irregular spring responses. In this situation, the resulting motions can be very irregular, and may be chaotic in nature where there are no regularities or periodicities and where the long-term predictability is lost.

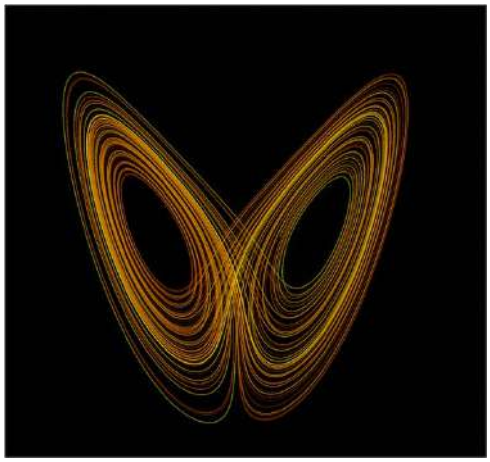


Fig. 6: The figure represents the chaotic attractor of the Lorenz system.

Perhaps one of the more well-known chaotic systems is the Lorenz system, which at the same time is one of the most studied chaotic systems. Figure 6 shows a chaotic attractor of the Lorenz system in phase space. It was introduced by the meteorologist Edward Lorenz to study the thermal convection in a fluid and by means of computer numerical simulations he was able to

observe the property of sensitive dependence on initial conditions, the hallmark of chaotic behavior.

Figure 7 shows the idea of sensitive dependence on initial conditions in the Lorenz chaotic system. The figure shows the temporal evolution in phase space of two orbits (one red and one blue) whose initial conditions are very close. After a certain time, approximately 24 time-units, the corresponding orbits start to drift apart, turning out to be very different at long times.

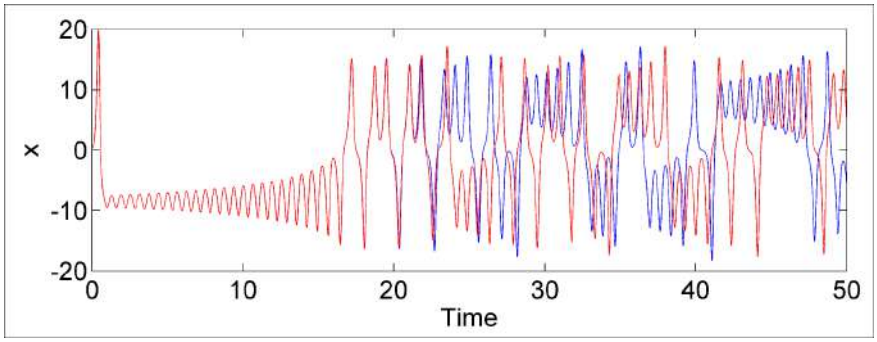


Fig. 7. Time evolution in phase space of two orbits of the Lorenz system initially very close, where the property of sensitive dependence on initial conditions is shown.

A very important tool in nonlinear dynamics is the geometric notion of phase space. The notion of phase space [6] is attributed to the American physicist Josiah Willard Gibbs (1839-1903), who was one of the pioneers of kinetic theory and is also considered one of the founding fathers of statistical mechanics, a term that he also coined. The concept of phase space plays a crucial role in nonlinear dynamics, from whose analysis we can obtain much information about a given dynamical system.

Studying the phase space of a given dynamical system allows complex fractal structures to be obtained whose physical consequences are reflected in uncertainty when determining the subsequent state of the system (Fig. 8).

3. A Historical Overview of Nonlinear Dynamics

Throughout the 19th century, certain limitations appeared around the myth of determinism. On the one hand, it is essential to have a complete knowledge of the initial conditions of the problem. On the other hand, notable difficulties arose in solving the dynamics of a physical system made up of a large number of particles. The latter led to the introduction of concepts related to probability theory in the study of the physical laws of systems made up of

many particles, such as gases, liquids and solids, giving rise to the birth of statistical mechanics. The founding fathers of the discipline include Ludwig Boltzmann (1844-1906), Josiah Willard Gibbs (1839-1903), and James Clerk Maxwell (1831-1879).

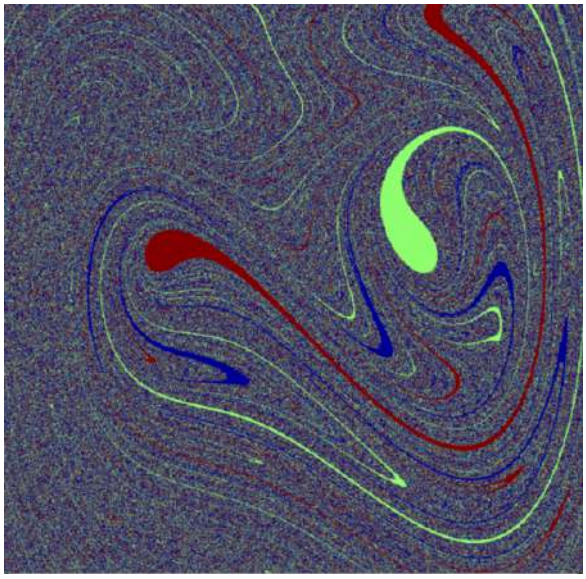


Fig. 8. Fractal structures in the phase space of a chaotic nonlinear oscillator. The variables of the phase space are the position on the *x* axis and the velocity on the *y* axis.

Scottish physicist James Clerk Maxwell (1831-1879), is fundamentally known for having unified the laws of electricity and magnetism. However, his contributions to physics have been among the most prolific in history. Among his great scientific work, it is important to mention that he is considered the father of automatics and statistical mechanics. However, the role he played in the development of modern chaos theory is largely unknown.

Precisely in one of his writings: *Does the progress of physical science tend to give any advantage to the opinion of necessity (or determinism) over that of the contingency of events and the freedom of the will?* from a lecture he gave at Cambridge on February 11, 1873 are the following excerpts showing to what extent Maxwell was familiar with the idea of sensitive dependence on initial conditions, of which we have spoken earlier.

"Much light may be thrown on some of these questions by the consideration of stability and instability. When the state of things is such that an infinitely small

variation of the present state will alter only by an infinitely small quantity the state at some future time, the condition of the system, whether at rest or in motion, is said to be stable; but when an infinitely small variation in the present state may bring about a finite difference in the state of the system in a finite time, the condition of the system is said to be unstable. It is manifest that the existence of unstable conditions renders impossible the prediction of future events, if our knowledge of the present state is only approximate, and not accurate."

Due to the enormous consequences on determinism in physics that quantum mechanics has brought about through Heisenberg uncertainty principle, the idea of indeterminism has been directly related to quantum mechanics. This has led somehow to consider classical mechanics as completely deterministic and predictable, which is not entirely true [7].

It is fascinating to corroborate that the idea of sensitive dependence on initial conditions was considered in detail by the German physicist Max Born (1882-1970), Nobel Prize in Physics in 1954, in an article entitled *Is Classical Mechanics in fact deterministic?* [8]. In it he presented a study of a two-dimensional Lorentz gas initially proposed by the Dutch physicist Hendrik A. Lorentz (1853-1928) in 1905 as a model for the study of electrical conductivity in metals. In this model, a particle moves in a plane that is full of hard spheres and collides with them so that a small change in the initial conditions will significantly alter the trajectory of the particle. This fact led Born to conclude that determinism traditionally related to classical mechanics is not real, since it is not possible to know with infinite precision the initial conditions of a physical experiment.

Furthermore, in the lecture [9] he gave on the occasion of the awarding of the Nobel Prize in 1954 the following words appear:

"Newtonian mechanics is deterministic in the following sense: If the initial state (positions and velocities of all particles) of a system is accurately given, then the state at any other time (earlier or later) can be calculated from the laws of mechanics. All the other branches of classical physics have been built up according to this model. Mechanical determinism gradually became a kind of article of faith: the world as a machine, an automaton. As far as I can see, this idea has no forerunners in ancient and medieval philosophy. The idea is a product of the immense success of Newtonian mechanics, particularly in astronomy. In the 19th century it became a basic philosophical principle for the whole of exact science. I asked myself whether this was really justified. Can absolute predictions really be made for all time on the basis of the classical equations of motion? It can easily be seen, by simple examples, that this is only the case when the possibility of absolutely exact measurement (of position, velocity, or other quantities) is assumed. Let us think of a particle moving

without friction on a straight line between two end-points (walls), at which it experiences completely elastic recoil. It moves with constant speed equal to its initial speed v_0 backwards and forwards, and it can be stated exactly where it will be at a given time provided that v_0 is accurately known. But if a small inaccuracy Δv_0 is allowed, then the inaccuracy of prediction of the position at time t is $t\Delta v_0$ which increases with t. If one waits long enough until time $t_c = l/\Delta v_0$ where l is the distance between the elastic walls, the inaccuracy Δx will have become equal to the whole space l. Thus it is impossible to forecast anything about the position at a time which is later than t_c. Thus determinism lapses completely into indeterminism as soon as the slightest inaccuracy in the data on velocity is permitted."

The American physicist Richard Feynman (1918-1988), who won the Nobel Prize for Physics in 1965 (Fig. 9), makes similar reflections in his well-known book *Lectures in Physics* [10], where he explains that indeterminism does not belong exclusively to quantum mechanics, it is a basic property of many physical systems.

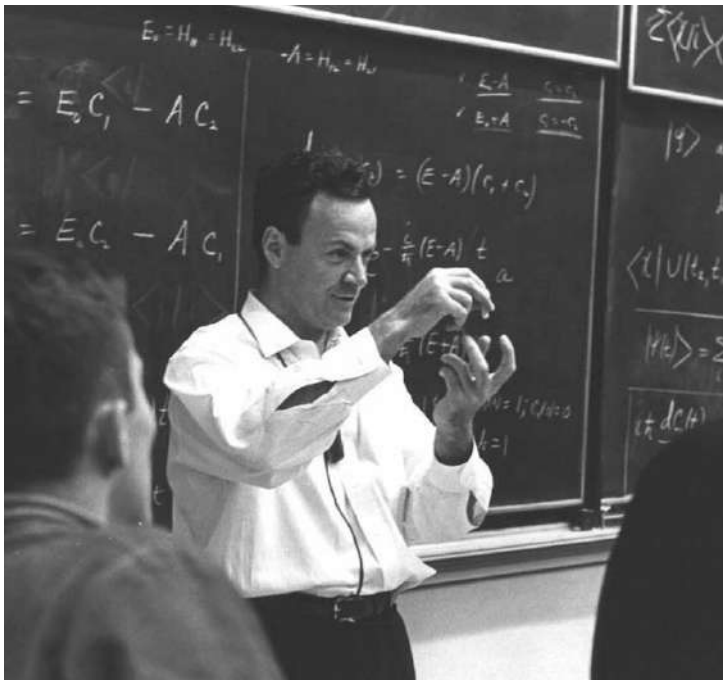


Fig. 9. Richard Feynman (1918-1988), Nobel Prize in Physics in 1965.

In section 38-6, entitled "Philosophical Implications", in the first volume of his *Lectures in Physics*, a masterful description of indeterminism in classical mechanics is made. The fundamental idea is the uncertainty in accurately setting initial conditions to predict the final state of a physical system. Finally affirming: "Because in classical mechanics there was already indeterminism from a practical point of view".

3.1 Poincaré, the Three-Body Problem and the Birth of Chaos

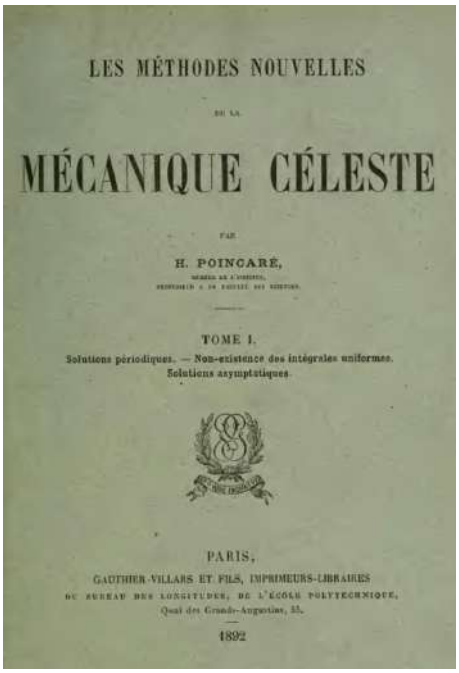
To understand the three-body problem, we go back to the beginnings of modern science with Isaac Newton's works on the gravitational field and the universal law of gravitation. The so-called two-body problem basically consists of analyzing the motion of a system formed by two bodies that attract each other under the action of gravitational forces. Newton solves the problem by reducing the motion of the two bodies to the motion of each of them around the so-called center of mass, which is a point whose mass is the total mass of the system.

Later, an attempt was made to solve the three-body problem, which can be formulated in a simple way: Let 3 bodies of arbitrary masses m_1, m_2 and m_3 be mutually attracted by Newton's law of gravitation. Assuming that they can move freely in a three-dimensional space and with arbitrary initial conditions, determine the evolution of the motion.

Despite the simplicity of its formulation, its resolution has caused real headaches for many scientists. Among them we may highlight Isaac Newton (1642-1727), Alexis Clairaut (1713-1765), Leonhard Euler (1707-1783), Pierre-Simon Laplace (1749-1827), Joseph-Louis Lagrange (1736-1813), Carl Jacobi (1804-1851), George Hill (1838-1914) and Henri Poincaré (1854-1912).

It is precisely the latter who wrote a famous memoir in 1889 on *the three-body problem and the equations of dynamics*, after winning the prize of the contest on the stability of the Solar System that had been summoned by King Oscar II of Sweden and Norway on the occasion of his 60th anniversary. This competition [11] had been proposed by the Swedish mathematician Gösta Mittag-Leffler, who had received it from the German mathematician Karl Weiertrass, who had been his teacher, the idea that the contestants write an original work facing one of four questions. One of Weiertrass's four questions had to do with Celestial Mechanics. The question was born out of a suggestion formulated by the mathematician Peter Gustav Lejeune Dirichlet at the University of Göttingen, who in 1858 had told his student Leopold Kronecker that he had discovered a new method of solving certain differential equations and pointed out that by applying them to the equations of celestial mechanics he could prove with all rigor that the solar system was stable. The committee that evaluated this competition was made up of mathematicians Karl Weiertrass, Frenchman Charles Hermite and Swedish Gösta Mittag-Leffler.

Subsequently, in 1892, Poincaré published his great work *Les Méthodes nouvelles de la Mécanique Céleste* (Fig. 10) in three volumes where numerous new concepts appear that have given rise to the development of the theory of dynamical systems, as mathematicians usually call it or nonlinear dynamics, a term more used by physicists, as well as other mathematical disciplines such as topology. Poincaré is considered to be one of the fathers of chaos theory, as many fundamental ideas of the theory are contained in this book.


LES MÉTHODES NOUVELLES

DE LA

MÉCANIQUE CÉLESTE

PAR

H. POINCARÉ,

TOME I.

Solutions périodiques. — Non-existence des intégrales uniformes.
Solutions asymptotiques.

PARIS,

GAUTHIER-VILLARS ET FILS, IMPRIMEURS-LIBRAIRES

1892

Fig. 10. *Les Méthodes Nouvelles de la Mécanique Céleste* was published by Henri Poincaré in 1892.

The general three-body problem is of enormous difficulty and only in recent years notable advances have been made, without being definitive. However, there is a case that is called restricted, circular and plane, which is the one that has been studied by many of the scientists to whom I have previously referred. Fundamentally, it is considered that the system is not made up of any three masses, but one of them is considered much larger than the others and the third of them is of negligible mass compared to the rest. The analogy certainly comes from considering systems like the Sun, Earth, and Moon, or Earth, Moon, and a satellite, where the approximation of moving in a plane is also

correct. In these circumstances and with an appropriate reference system, the equations of motion can be found without difficulty, from which a potential is derived that gives us an idea of the equilibrium positions in which a third body can be found. These are five equilibrium positions that Lagrange found, which is why they are currently known as the Lagrange points (Fig. 11).

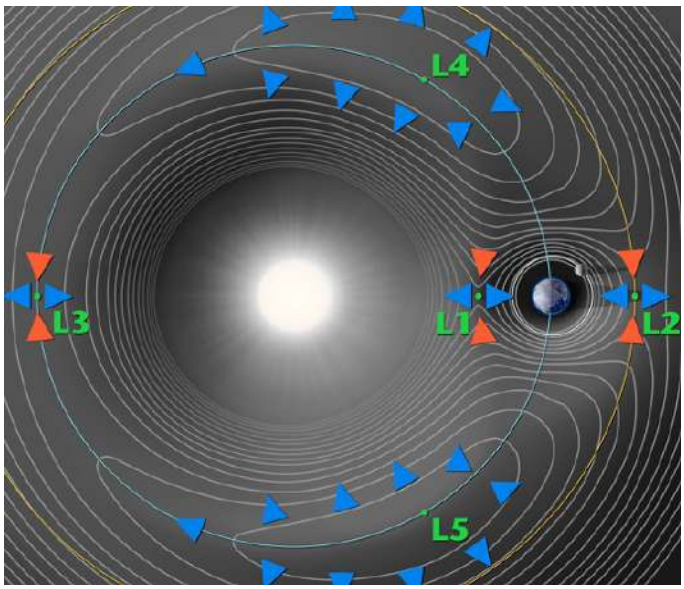


Fig. 11. The figure shows the equipotential curves of the restricted three-body problem, in this case, Sun-Earth-Moon where the five Lagrange points are illustrated.

Knowledge of the Lagrange points is very useful. In fact, at point L1 is the Solar and Heliospheric Observatory (SOHO), which is a space probe to study the Sun. At Lagrange point L2, the Wilkinson Microwave Anisotropy Probe (WMAP) was positioned to study radiation from microwave background of the universe, getting it to stay in place with minimal fuel consumption, always keeping its sensors pointed away from the Earth and the Sun. The James Webb Space Telescope (JWST) is planned to be launched in 2021, which is a developing space observatory that will study the sky in infrared frequency, and that will orbit around the L2 Lagrange point.

As pointed out above, Poincaré did not approach the three-body problem in a general way, but focused on studying what is known as the "restricted three-body problem", which is a particular case in which it is considered that one of the masses is very small compared to the others. In this

study he found what he called doubly asymptotic or homoclinic orbits, which are characterized by having a homoclinic point in the phase space. The presence of one of these points have very serious implications on the dynamic complexity of the system. After studying the problem, Poincaré wrote:

"One will be struck by the complexity of this picture that I do not even dare to sketch. Nothing is more appropriate to give us an idea of the intricateness of the three-body problem and in general all problems of dynamics…"

And it is that when trying to solve this problem he created a method or a geometric approximation by means of which he glimpsed that this problem had a very complex dynamics that is basically what we now call deterministic chaos.

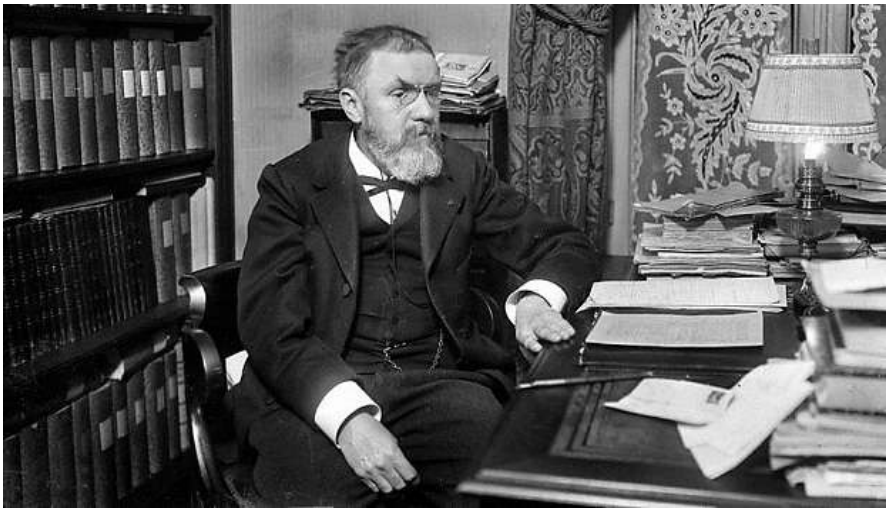


Fig. 12. French mathematician and physicist Henri Poincaré (1854-1912)

The influence of Poincaré (Fig. 12) on the development of Hamiltonian systems is enormous and in this sense it is interesting to mention that his witness was taken by the American mathematician George David Birkhoff (1884-1944), who coined the term dynamical systems, since in turn it had an enormous influence on Edward Lorenz who would rediscover the sensitive dependence on initial conditions in the middle of the 20th century.

Within this stream of thought, and in the American context, it is necessary to mention the mathematician Steven Smale (Fig. 13), deserving of the Fields medal in 1966 for his great contributions to the theory of dynamical systems. It is precisely to him that the concept of Smale horseshoe is due, which

was an important step in understanding the relationship between the existence of a homoclinic point and the notion of deterministic chaos, through the simple idea of symbolic dynamics using the so-called Bernoulli shift map.

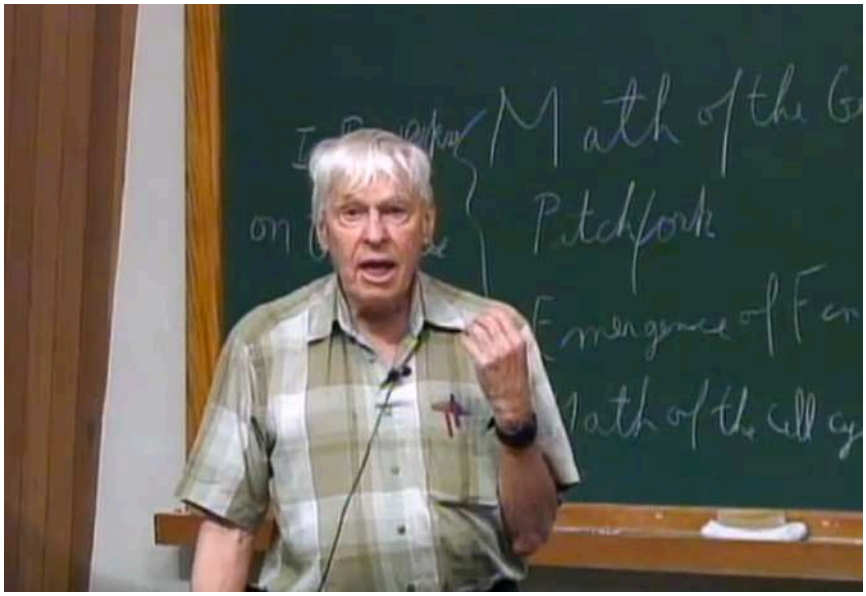


Fig. 13. Steve Smale, Fields Medal 1966.

As for the Russian tradition, we must go back to Alexander M. Lyapunov (1857-1918), who had been a doctoral student of the famous mathematician Pafnuti L. Chebychev (1821-1894), and whose thesis on the stability of motion has exerted an enormous influence on Physics. From Lyapunov we have inherited concepts such as the stability of dynamical systems and also Lyapunov's useful exponents, which help us to characterize when a given dynamical system is chaotic or not.

One of the main schools within the Russian tradition is that of Leonid I. Mandelstam (1879-1944), continued by his disciples Alexander A. Andronov (1901-1952) and Lev S. Pontryagin (1908-1988). Another key school within this same tradition is that of Andrei N. Kolmogorov (1903-1987). All of them developed new methods and made notable contributions to the construction of nonlinear dynamics as we know it today.

In the year 1954, at the International Congress of Mathematics that took place in Amsterdam, Kolmogorov enunciated a theorem for Hamiltonian systems that was subsequently proved by his student Vladimir I. Arnold and by the German Jürgen Moser (1928-1999), who has turned out to be of

considerable importance. This theorem is currently known as the KAM theorem (Kolmogorov-Arnold-Moser) [12] and it has to do with the problem of the stability of invariant tori in the integrable systems of Hamiltonian mechanics under the action of small perturbations.

This work, in fact, naturally links with Poincaré's pioneering works on celestial mechanics, since he had brought out the idea of the complexity of orbits in the three-body problem, and the KAM theorem can be considered as a culmination of these ideas. As we have already seen, the stability of the solar system is a problem of special importance in celestial mechanics and the KAM theorem shows that under certain conditions these orbits remain confined in certain regions.

3.2 Complexity in Fluid Motion

The phenomenon of turbulence in fluid motion is one of the most spectacular cases of chaotic behavior. Although the fundamental equations of fluid motion, the Navier-Stokes equations, have been known since the end of the 19th century, it should be remembered that the form of their solutions in turbulent regime is not yet known.

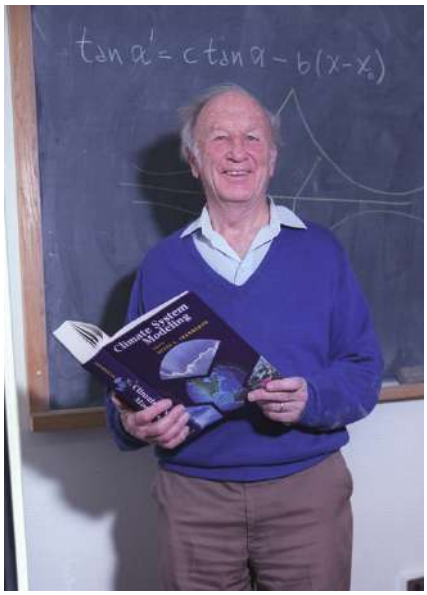


Fig. 14. Edward N. Lorenz (1917–2008)

In 1963 the meteorologist at the Massachusetts Institute of Technology (MIT) Edward N. Lorenz (Fig. 14) developed a model of three ordinary differential equations to describe the motion of a fluid under the action of a thermal gradient. When it came to finding numerical solutions with the help of a computer, he again encountered the phenomenon of sensitive dependence on initial conditions. That is, the system was inherently unpredictable, such that small variations in determining the initial conditions led to drastically different solutions.

At the time, very few gave importance to this fact, perhaps because the results of Lorenz's work were published with a somewhat cryptic title, *"Deterministic Nonperiodic Flow"* [13], in a meteorology journal and went unnoticed by many scientists.

The theory of the Russian physicist Lev D. Landau, and the German Eberhard Hopf that proposed the existence of an infinite set of incommensurable frequencies to explain the turbulence, was surpassed in the 1970s by the theoretical contributions of David Ruelle and Floris Takens, who introduced in 1971 the fundamental concept of strange attractor. It is an attractive geometric object, different from the previously known cases of periodic fixed points, quasi-periodic fixed points or limit cycles, hence the name "strange", and which also has a non-integer or fractional (fractal) dimension.

On the other hand, the development of fractal geometry started by Benoit Mandelbrot [14], who had been a student of the French mathematician Gaston Julia, has played a fundamental role in the understanding and analysis of the complex behavior of nonlinear dynamical systems. In any case, it is important not to forget the role played in many aspects of the development of nonlinear dynamics by the German mathematician Georg Cantor (1845-1918),

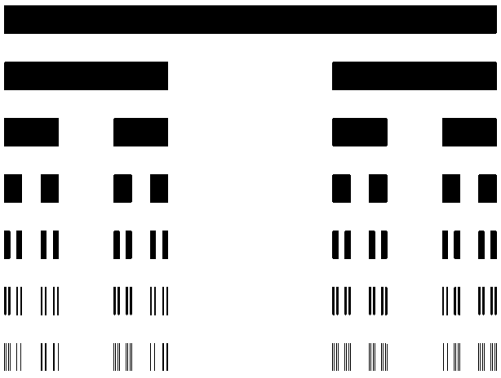


Fig. 15. The middle-third Cantor set is one of the simplest fractal sets

particularly with regard to the middle-third Cantor set (Fig. 15) and its constant appearance in many dynamic problems. This will be discussed in more detail in section 3.6.

3.3 Statistical Mechanics, Origin of Irreversibility and Ergodic Theory

Statistical mechanics is an essential part of theoretical physics whose purpose is to describe the macroscopic properties of a very large system of particles in terms of their averaged properties. It is a discipline that combines the basic laws of dynamics for a particle system along with the laws of statistics, especially those concerning the law of large numbers. The discovery of deterministic chaos has stimulated some physicists to reconsider from a new perspective the foundations of statistical mechanics. This is because deterministic chaos implies that not only systems with a large number of particles, but even deterministic systems with very few degrees of freedom can present behaviors that require statistical tools for their study. Many efforts have been made during this last century to give a correct interpretation of the dynamical origins of irreversibility. Despite all the efforts made to date, there is still no general agreement on what are the essential ingredients needed to support statistical mechanics.

Fig. 16. The founding fathers of Statistical Mechanics: Ludwig Boltzmann (1844-1906), James Clerk Maxwell (1831-1879 and Josiah W. Gibbs (1839-1903)

The problem of irreversibility was one of the major concerns of one of the "founding fathers" of statistical mechanics (Fig. 16), the Viennese physicist Ludwig Boltzmann (1844-1906). The objection raised by Josef Loschmidt (1821-1895) to Boltzmann's program, consisting in deriving the laws of thermodynamics directly from mechanical behavior, revealed the paradoxical of a situation in which, while the laws of mechanics are reversible under temporal

inversion, the thermodynamic behavior of the systems is fundamentally irreversible. There has certainly been great progress in this century in the attempt to clarify the dynamic origin of the kinetic equations, although the problem remains to some extent open. Following Boltzmann, the first attempts to substantiate classical statistical mechanics were based on the supposed validity of the ergodic hypothesis, which, after making considerable theoretical efforts, led to a true impasse.

Following the work of Maxwell and Boltzmann, Gibbs introduced the concept of a "Mixing" associated to a system using the simile of an oil drop in an immiscible fluid, a small region in the phase space that simulates the oil drop, the dynamical evolution would help fill the entire phase space. This idea implies that for a given dynamical system, two sufficiently close points would separate exponentially after a certain period of time. This concept is linked to the notion of sensitive dependence on initial conditions that is at the base of chaotic dynamics in nonlinear dynamics and that leads to define the so-called Lyapunov exponents [15]. The concept of Lyapunov exponent indicates that if a dynamical system has any positive Lyapunov exponent then these initial points or conditions would separate exponentially and this type of systems are called chaotic systems, since the prediction of the evolution of the system in the long term is impossible.

In this sense, scientists like George Birkhoff (1884–1944) stand out, who proposed the ergodic theorem, which was later proved by the German mathematician Eberhard Hopf (1902-1983) using the fact of the ergodicity of the trajectories on surfaces of constant negative curvature that French mathematician Jacques Hadamard had pointed out a few years earlier. However, these results had little impact on the foundation of nonequilibrium statistical mechanics.

The importance of the Lorentz gas, which was previously mentioned when talking about Max Born, is that it shows thermodynamic physical properties, is ergodic and has a positive Lyapunov exponent. The great achievement of Russian American mathematician Yakov Sinai, who received the Abel Prize in 2014 (Fig. 17), was to show the connection between the classical Boltzmann-Gibbs set for an ideal gas and a chaotic Hadamard billiard.

Ideas from chaos theory have been used for the foundation of statistical mechanics, finding deep connections between the dynamical properties of a system, such as its Lyapunov exponents and its transport properties. Knowledge of both the nonequilibrium statistical mechanics and nonlinear dynamics is essential to understand works on nonequilibrium states. Despite numerous efforts and apparent new perspectives to support the nonequilibrium statistical mechanics based on chaos theory, the extraordinary conceptual difficulties of such an undertaking have so far prevented its achievement.

Fig. 17. The American-Russian mathematician Yakov G. Sinai with Crown Prince Haakon of Norway when he received the 2014 Abel Prize.

3.4 The Path towards Chaos through Nonlinear Oscillators

The construction of nonlinear dynamics is like that of a large river to which numerous tributaries contribute. One of these tributaries is the study of nonlinear oscillators. Among the pioneers in this path we can find the English physicist John William Strutt, Lord Rayleigh (1842-1919), motivated by his interest in understanding the physics of musical instruments. For this type of system, a first approximation based on the use of linear oscillators is not effective because the real instruments do not produce a simple tone, as it happens to a linear oscillator, so it is necessary to add friction on one side and terms not linear recovery by another. That is, it is necessary to use an elastic force different from that provided by Hooke's law: *ut tensio sic vis*. By clever use of the basic dynamical elements of the problem, Lord Rayleigh created models that explained the sound emitted by musical instruments. In his famous book *The Theory of Sound* published in 1877, Rayleigh introduced a series of fairly general methods such as the notion of a limit cycle, which is a periodic motion that has the physical system regardless of the initial conditions.

German engineer Georg Duffing (1861-1944) is known primarily for his symmetric nonlinear oscillator model with a cubic nonlinearity: Duffing

oscillator. This model is a paradigmatic model for the study of many phenomena in nonlinear dynamics. The theory was later developed in the late 1940s, just after World War II, by two English mathematicians at Cambridge University: Mary L. Cartwright (1900-1998) and John E. Littlewood (1885-1977) who showed that many of the experiments of experimental physicists and many of the conjectures of theoretical physicists were derived directly from the analysis of differential equations of motion. In fact, these mathematicians had followed the ideas of George Birkhoff.

The school of nonlinear thought in Russia was started by the work of Leonid I. Mandelstam (1879-1944) on nonlinear oscillators, who had trained with the German physicist August Kundt (1839-1894) in Strasbourg, well known for his works on acoustics and the Kundt tube. This line of work was continued by Alexander A. Andronov (1901-1952) (Fig. 18) and by Lev S. Pontryagin (1908-1988), who introduced the notion of structural stability of a system of equations, a concept associated with that of bifurcations of dynamical systems.

Fig. 18. The Russian mathematician Alexander A. Andronov (1901-1952) one of the pioneers in nonlinear science.

The concept of bifurcation of limit cycles that had been suggested by Poincaré in 1892, was tested by Andronov in 1930 and by Hopf in1940, and is called the Andronov-Hopf bifurcation, although it is better known simply as a Hopf bifurcation. This school continued later in the 50s and 60s in Gorky,

current Nizhnii Novgorod, obtaining parallel results to the development of the theory in the West. Many methods of nonlinear physics were developed under the paradigm of nonlinear oscillators and self-oscillations.

Another important school on nonlinear thought in Russia was the Kiev School of Nonlinear Oscillations Research that was initiated by Nikolai M. Krylov (1879–1955) and his student Nikolai N. Bogolyubov (1909–1992) (Fig. 19), at the beginnings of the 1930s. They developed much fundamental work on quasi-periodic solutions for non-autonomous systems and established the discipline of Nonlinear Mechanics as a part of Physics. Most of their work was published in the book *Introduction to Nonlinear Mechanics* (1937) [16] in Russian. An English version was published in 1943 by Princeton University Press after the translation carried out by the Russian mathematician Solomon Lefschetz (1884-1973) who led the Nonlinear Oscillation Project (ONR) in Princeton, and translated work available in Russian for the English-speaking world.

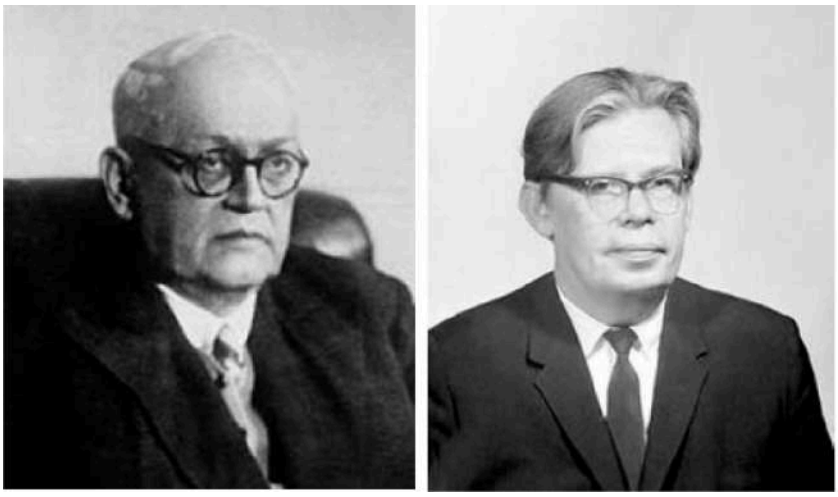


Fig. 19. Nikolai M. Krylov and his student Nikolai N. Bogolyubov who develop the Nonlinear Mechanics School in Kiev in the 1930s.

In Japan, the theory of nonlinear oscillators and their applications to radiophysics were developed at the school of Japanese engineer Chihiro Hayashi (1911-1986) at Kyoto University. Hayashi made notable contributions to the study of nonlinear oscillators and their practical applications in electrical engineering, publishing his famous book *Nonlinear Oscillations in Physical Systems* in 1964 [17].

In 1961 a remarkable event takes place on the part of the Japanese engineer Yoshisuke Ueda, who was a doctoral student of Chihiro Hayashi. Ueda studied the dynamics of various nonlinear oscillators such as the van der Pol oscillator and the Duffing oscillator, and it is precisely in a particular model of the latter that he apparently found solutions for the first time that we now designate as chaotic solutions.

3.5 Population Dynamics and the Logistic Map

The logistic map was popularized by Robert M. May (1936-2020) (Fig. 20) after the publication of his influential paper *"Simple mathematical models with very complicated dynamics"* [18], and constitutes one of the paradigms of the chaotic behavior of nonlinear dynamical systems. Robert May started his scientific career as a physicist, but soon he moved into biology becoming one of the pioneers in theoretical ecology, what led him to become a pioneer in chaos theory. Despite the apparent simplicity of the logistic map, it displays complex dynamics including chaotic behavior. Its formulation derives from the logistic equation, introduced in 1838 as a model of growth in population dynamics by the Belgian mathematician Pierre François Verhulst (1804-1849) in his writing *"Notice sur la loi que la population poursuit dans son accroissement"*. The quadratic map, very similar to the logistic map, had also been extensively studied in other contexts by the French Gaston Julia (1893-1978), by the Hungarian-American John von Neumann (1903-1957), and by the American Norbert Wiener (1894-1964).

Fig. 20. Robert M. May (1936-2020), Baron May of Oxford. A physicist and pioneer of theoretical ecology who led him to contribute to chaos theory.

One of the most influential articles in the field was undoubtedly the previously quoted article by Edward Lorenz *Deterministic nonperiodic flow* [13]. The American mathematician and physicist at the University of Maryland Prof. James A. Yorke (Fig. 21) immediately recognized the implications of such a discovery, as well as its philosophical repercussions, making Lorenz's work known to the scientific community.

Fig. 21. The American mathematician and physicist James Yorke. He coined the term chaos in the modern scientific literature and made considerable achievements in chaos theory.

Later he introduced the term *chaos* in the article entitled *Period Three Implies Chaos* [19] published together with his PhD student Tien-Yien Li in The American Mathematical Monthly magazine in 1975. A few years earlier, in 1963, the Ukrainian mathematician A. N. Sharkovskii had proved a theorem (now known as Sharkovskii's theorem), which was published in Russian in the Ukrainian Mathematics Journal, and where part of the Li and Yorke's result appeared as a corollary. However, one of the fundamental novelties in the article by Li and Yorke is that they wrote that the appearance of a period three orbit implied the appearance of all the others, including the chaotic orbits, while Sharkovskii did not talk about the chaotic orbits.

Subsequently, the American physicist Mitchell Feigenbaum (1944-2019) discovered the existence of universal critical exponents that characterized the transition from periodic to chaotic motion in one-dimensional maps with the

property of period doubling. Simultaneously, the same discovery was made by the French Pierre Coullet and Charles Tresser, who at the time were doctoral students at the University of Nice, and by the German physicists at the University of Marburg, Siegfried Grossmann and Stefan Thomae.

The renormalization group concept had previously been applied in the field of statistical mechanics to study the so-called critical phenomena and phase transitions and its development in these fields had earned the Nobel Prize for the American physicist Kenneth Wilson in 1982. These methods were applied by Feigenbaum and others to develop the mathematical theory of period doubling bifurcations. Until the beginning of the eighties, most of the works were of a theoretical nature or the result of numerical explorations with computers. In any case, the important consequences that these theoretical discoveries had for physics were always considered, as well as the possible importance for understanding the transition to fluid turbulence.

French physicist Albert Libchaber (Fig. 22), currently at Rockefeller University in New York, carried out one of the first experiments where the phenomenon of period-doubling was shown when studying Rayleigh-Bénard convective cells in the late 1970s. American physicist Robert Shaw of the University of California at Santa Cruz performed a simple and particularly relevant experiment with a simple dripping faucet. Another important experimental milestone was carried out by the American physicists Jerry Gollub and Harry Swinney (Fig. 22), who also found the period doubling phenomenon by reproducing the classical Taylor-Couette experiment of fluid motion. Their contributions to the experimental verification of some of some of the ideas derived from chaos theory have stimulated much experimental work in nonlinear dynamics and chaos.

Fig. 22. The physicists Albert Libchaber, Harry Swinney and Jerry Gollub have been pioneers on experimental work on chaos.

3.6 Fractional dimensions, fractals and chaos

There are many complex geometric shapes in nature such as shorelines, river beds, the biological forms and even the complex curves of the financial markets. A common feature in all of them is self-similarity. This is the property

that consists in that when a part of this form is increased, the same type of structure appears. To characterize objects with this universal property, the use of fractional dimensions is necessary, which led to the physicist and mathematician Benoit Mandelbrot (1924-2010) (Fig. 23) to call these objects "fractals". His work in collecting the enormous work that had been done by mathematicians before him such as the French Gaston Julia, the Swedish Helge von Koch, the Polish Wacław Sierpiński, as well as the works on dimensions by the German Felix Hausdorff and the Russian Abram S. Besikovich had a remarkable influence that he gave to the field of fractal geometry.

Fig. 23. Benoit Mandelbrot (1924-2010) and the famous set that bears his name.

The notion of dimension is fundamental when measuring geometric objects. There are several ways to define the concept of dimension, but it is clear that a point has dimension zero, a straight line has dimension one, a plane has dimension two and a cube has dimension three. However, and as strange as it may seem, there are geometric objects whose dimensions are not an integer, turning out to be a fractional value.

This is a simple notion of what is meant by a fractal dimension or a Hausdorff dimension, such that the Cantor set, mentioned above, has a dimension from $\log 2 / \log 3 \approx 0.63$, the Koch curve has a dimension of $\log 3 / \log 4 \approx 1.26$ and the Sierpinski set has a dimension of $\log 3 / \log 2 \approx 1.585$. All of them are self-similar fractal sets, since they are obtained by means of an iterative rule so that the basic structure is repeated at all scales.

The Koch curve (Fig. 24) was devised by the Swedish mathematician Helge von Koch (1870-1924), and is constructed as follows: We start with an interval that we divide into three equal pieces, and in the middle piece we build a triangle and equal sides as it appears in the figure.

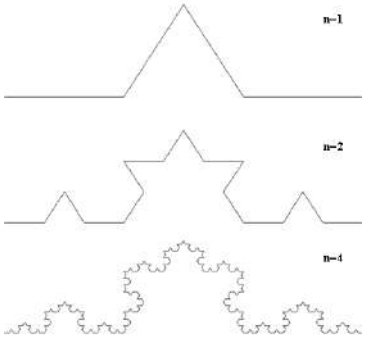


Fig. 24. The Koch curve is a fractal set

Next, we repeat the same strategy in each of the four pieces, giving rise to the figure in the middle, and we repeat the process in successive iterations giving place to a figure that resembles a snowflake. The Sierpinski fractal is due to the Polish mathematician Wacław Sierpiński (1882-1969) and is constructed as follows. We consider a triangle with equal sides, like the one shown in Fig. 25. Next, we remove the white triangle from inside it and in each of the remaining triangles we remove the white triangle and so on, finally giving rise to successive iterations to the Sierpinski triangle, which is a self-similar fractal object.

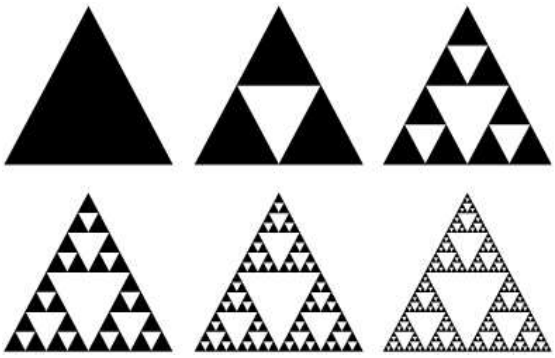


Fig. 25. Sierpinski fractal set

Although in principle fractal geometry and nonlinear dynamics are two disciplines that apparently have nothing to do with it, nevertheless, as previously noted, chaos and fractals are intimately linked. One of the main ideas is due to the fact that associated with the notion of chaos exists that of the chaotic attractor that constitutes a geometric object of a fractal nature that lives in the phase space, so that it is impossible to speak of chaos without speaking of fractals and vice versa.

4. On the Origins of Complexity

At the beginning of the 20th century, fundamental developments took place in two new fields of research in Physics that represent a huge conceptual revolution in the development of science. On the one hand, the theory of relativity that helped us understand the world on cosmic scales and quantum mechanics that involved the knowledge and exploration of the microscopic world at the atomic and subatomic levels. On the other hand, during the second half of the 20th century, we have been able to see how nonlinear dynamics and chaos theory emerged as one of the very fruitful fields of activity in research. Likewise, the discipline of complexity, or the physics of complex systems, has received a huge push, including new lines of research and bringing a new way of doing things.

Fig. 26. Warren McCulloch (1898-1969). One of the pioneers in complexity and mathematical neurosciences.

Talking about the origins of things is never easy and of course the origins of complexity are no exception. In spite that for many it is a relatively new notion, since its use has become widespread in recent years, its origins date back to much earlier times. When exploring certain ideas and activities that have contributed to the development of this set of ideas that complexity encompasses, it is worth mentioning the American neuroscientist Warren McCulloch (Fig. 26), who together with the mathematician Walter Pitts, proposed in 1943 the well-known McCulloch-Pitts neuron model to analyze brain properties. McCulloch also played a prominent role in the organization in the 1940s of the Macy Conferences, with the support of the Macy Foundation, where numerous scientists from various disciplines participated in a highly interdisciplinary environment, among which we can mention the psychiatrist William Ross Ashby, the anthropologist Gregory Bateson; mathematicians John von Neumann, Walter Pitts and Norbert Wiener, biophysicist Max Delbrück, information theorist Claude Shannon and Warren McCulloch himself as moderator.

On the other hand, it is of special interest the figure of the American scientist Warren Weaver (1894-1978) (Fig. 27), who among other things was co-author with Claude E. Shannon of the famous book *The Mathematical*

Theory of Communication published by The University of Illinois Press in 1949.

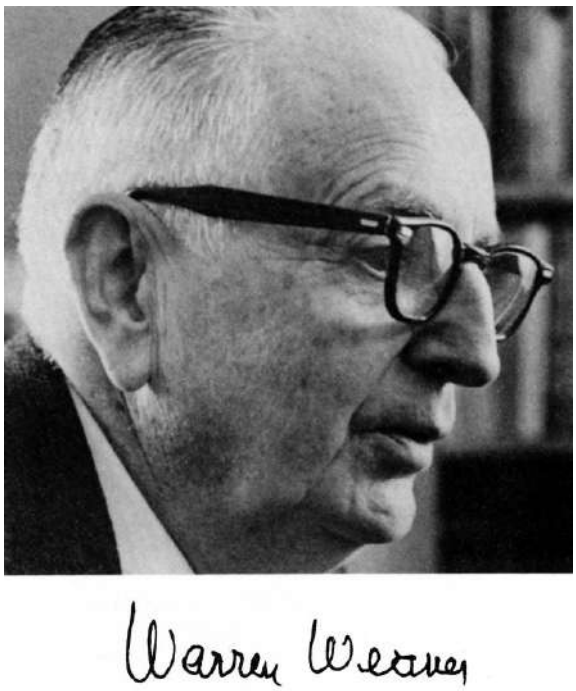


Fig. 27. Warren Weaver (1894-1978) pioneer in the use of computers in scientific research.

In 1948 he published a very interesting article, considered foundational, entitled *Science and Complexity* [20] in the American magazine American Scientist. In fact, he used material that had been published in 1947 and the most important thing is that it is premonitory of many aspects of the complexity that have been discussed in recent years.

4.1 Physics and Emergence

One of the fundamental ideas in complexity is the idea of emergence. In physics there are numerous examples of systems where emerging properties are evident, such as superconductivity and superfluidity. It should also be noted that there is all a fundamental research that seeks to investigate complex phenomena, where instead of resorting to reductionism, which has been the

approach that has governed the evolution of Physics in recent years, the primary engine of this research is the emergence. A fundamental point is that these emerging complex phenomena do not derive from the underlying microscopic laws, although of course they do.

Some of these ideas were masterfully presented by the physicist Philip W. Anderson (1923-2020) (Fig. 28), 1977 Nobel Prize in Physics, in an article published in the journal Science in 1972 and entitled *More is different* [21], where he leaves very clear the idea that:

"At each level of complexity entirely new properties appear, and the understanding of the new behaviors requires research which I think is as fundamental in its nature as any other."

Fig. 28. Philip W. Anderson (1923-2020), 1977 Nobel Prize in Physics.

Philip W. Anderson introduces some aspects of the physics of complex systems in the article entitled Physics: The Opening to Complexity [22], where he points out, among other things:

"But another large fraction are engaged in an entirely different type of fundamental research: research into phenomena that are too complex to be analyzed straightforwardly by simple application of the fundamental laws. These physicists are working at another frontier between the mysterious and the understood: the frontier of complexity. At this frontier, the watchword is not reductionism but emergence. Emergent complex phenomena are by no means in

violation of the microscopic laws, but they do not appear as logically consequent on these laws."

In relation to the physics of emergence, it is also worth mentioning Robert Laughlin, 1998 Nobel Prize in Physics and professor at Stanford University, who proposed to his best students the problem of deducing the laws of superfluidity from first principles, knowingly that it is impossible. Precisely to show them the importance of emergent properties in physics, which is the fundamental argument of his book *A Different Universe: Reinventing physics from the bottom down* [23].

The book is based on an interesting article entitled *The Science of Everything* [24], where among the many questions he points out we can highlight the following two paragraphs:

The central task of theoretical physics in our time is no longer to write down the ultimate equations but rather to catalogue and understand emergent behavior in its many guises, including potentially life itself. We call this physics of the next century the study of complex adaptive matter. For better or worse we are now witnessing a transition from the science of the past, so intimately linked to reductionism, to the study of complex adaptive matter, firmly based in experiment, with its hope for providing a jumping-off point for new discoveries, new concepts, and new wisdom.

"End of Reductionism, for it is actually a call to those of us concerned with the health of physical science to face the truth that in most respects the reductionist ideal has reached its limits as a guiding principle. Rather than a Theory of Everything we appear to face a hierarchy of Theories of Things, each emerging from its parent and evolving into its children as the energy scale is lowered. The end of reductionism is, however, not the end of science, or even the end of theoretical physics."

In fact, when one looks at the world what one observes is of amazing complexity. Although, for the moment, there are no laws of complexity, as there are laws of physics, the authors cited above list a number of simple lessons on complexity that derive from the analysis and observation of numerous complex systems that exist in the universe.

Hungarian physicist Tamas Vicsek from the Department of Biophysics at Eötvös University in Budapest argues in an essay published in Nature [25] that when a concept is not well defined, as is the case with complexity, there is a danger of abusing it. It is true that on many occasions the term can be used indiscriminately as a sign of modernity. However, the fundamental idea derived from this essay is that the laws that describe the behavior of complex systems

are qualitatively different from those that govern the units of which they are composed.

4.2 Complexity and Life Sciences

The enormous development of scientific activity in recent years has caused many disciplines to find fields of application in other sciences. This is what, among many other cases, has happened with the application of disciplines such as physics, mathematics and engineering in the development of some aspects of the life sciences, in which we could include not only biology, but also biomedical sciences and biotechnology. You might think that it is a simple fashion and something that for some reason has been happening for just a few years. However, it is important to note that the influence of these sciences and their contributions to the life sciences are very old.

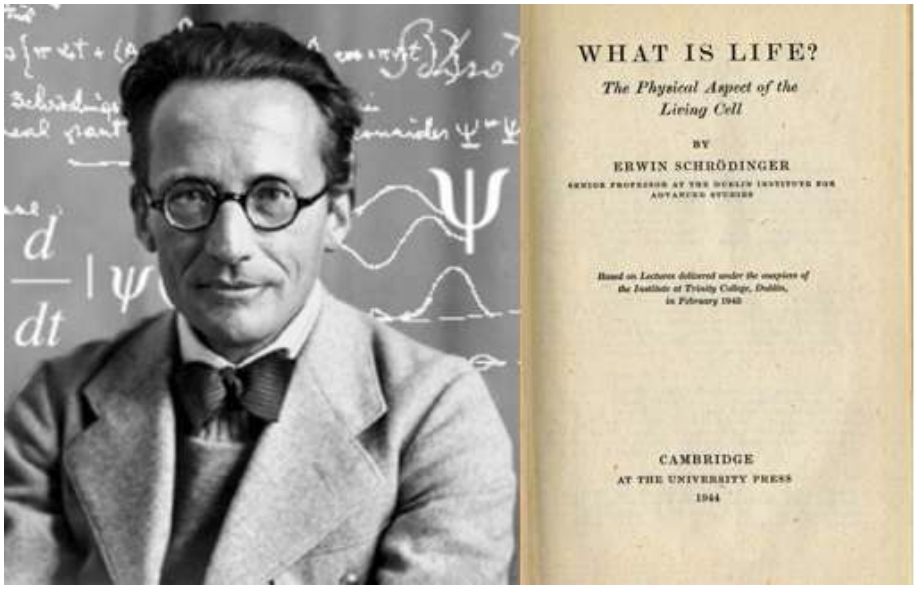


Fig. 29. Erwin Schrödinger, 1933 Nobel Prize in Physics and known above all for his contributions to quantum mechanics, who authored the influential book *What is life?* published in 1944.

There are many eminent physicists, including Nobel Laureates, whose work has been related in some of the aspects related to the issues of complexity. Among them are: Erwin Schrödinger, 1933 Nobel Prize in Physics and known above all for his contributions to quantum mechanics, who authored the

influential book *What is life?* published in 1944 (Fig. 29). Physicist Max Delbrück (1906-1981), Nobel Prize in Medicine in 1969 for his pioneering work in Molecular Biology. Philip W. Anderson, 1977 Nobel Prize in Physics, well known for his work in condensed matter physics, has also played a relevant role in the development of some ideas related to complexity, especially emergence. Physicist Murray Gell-Man, 1969 Nobel Prize in Physics; who coined the term *quark*.

Another fundamental character in this relationship that we are making is the mathematician Norbert Wiener (1894-1964) (Fig. 30), professor at the Massachusetts Institute of Technology (MIT), who was one of the founders of Cybernetics, and knew how to create a highly interdisciplinary environment around him with numerous applications to life sciences. We could continue quoting numerous physicists, such as Nicholas Metropolis, George Gamow, Leo Szilard, Jack Cowan or Geoffrey West.

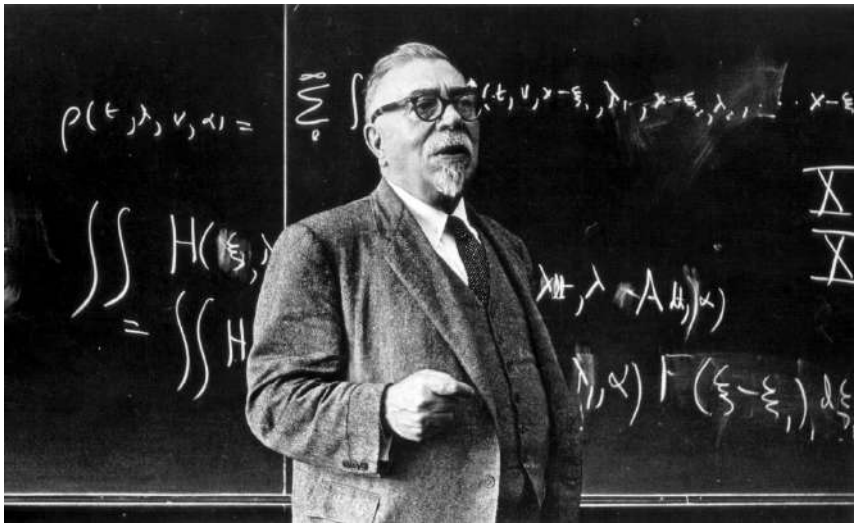


Fig. 30. Norbert Wiener (1894-1964) one of the founders of Cybernetics.

Among the most widely used mathematical models in computational neuroscience, which aim to analyze the brain as a complex system, we can consider the Hodgkin-Huxley model. In 1952 Alan L. Hodgkin and Andrew. F. Huxley (Fig. 31) wrote a series of five articles [26] in which they described the experiments they carried out to determine the laws of ion motion in nerve cells during an action potential.

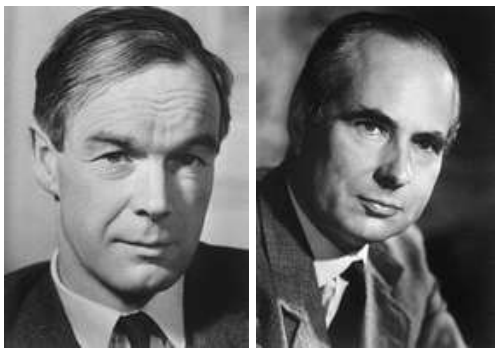


Fig. 31. Alan L. Hodgkin and Andrew F. Huxley received the Nobel Prize in Nobel Prize in Physiology or Medicine in 1963 for his neural model.

They formulated a mathematical model to explain the behavior of nerve cells in a giant squid. Remarkably, this model was formulated long before the existence of electron microscopes and computer simulations and allowed scientists basic knowledge of how nerve cells function without knowing how membranes behaved. They received the Nobel Prize in Physiology or Medicine in 1963, along with Sir John C. Eccles for their discoveries regarding the ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane.

5. Conclusions

One of the key ideas that should be highlighted here is that although the physics of complex systems is currently one of the frontiers of current physical research, the ideas of complexity go back to the beginning of the 20th century and have been developed along various paths until we reach the vision we have as of today, although their evolution and development throughout the 21st century are quite open.

The concept of emergence versus reductionism is another of the fundamental ideas in the physics of complex systems. The ones concerning emergence go back even to the origins of thermodynamics, and they appear in various phenomena studied by that science. Concepts such as chaos and fractals are mentioned in a special way, which have been a catalyst for many of the notions around which complexity moves. Without a doubt, interdisciplinarity is of the utmost importance in this context, since, as it has been pointed out, many ideas associated with complexity help to integrate disciplines, as well as breaking traditional disciplinary barriers.

At all times it has been wanted to put on record that many of the ideas discussed in this article have been beating in the thought and action of many

physicists in the past and present, who have been open to problems about the complexity of life and nature, including some Nobel Prizes.

In recent years, numerous scientists have contributed to the development of chaos theory. In 2003 the Japan Prize, which is awarded each year by the Japanese government through the *The Japan Prize Foundation* was dedicated to Complexity Science and Technology. The award was won by the scientists Benoit Mandelbrot for his contributions to fractals and James A. Yorke for his contributions to the foundation of chaos theory. This award was very special to the community of scientists working in these fields, since for the first time an award of this magnitude was awarded to scientists working on complex science issues.

Following the efforts of numerous scientists, as we have just shown, the entire field of research covering nonlinear dynamics, chaos theory and complexity continues to develop and influence numerous disciplines with new methods and novel ideas, showing great prospects for the future.

References

[1] James Gleick. Chaos: The Creation of a Science. Chaos: Making a New Science. Penguin Books. (2008).
[2] Lenny Smith. Chaos: A Brief Introduction. OUP, Oxford (2007) .
[3] Ian Stewart. Does God Play Dice?: The New Mathematics of Chaos. Penguin Books. (1997).
[4] Manuel de León; Miguel A.F. Sanjuán. Mathematics and Physics of Chaos. Madrid. La Catarata (2010) (In Spanish).
[5] Charles Darwin. The Origin of Species. 1859. http://www.talkorigins.org/faqs/origin.html
[6] D.D. Nolte. The tangled tale of phase space. Physics Today **63** (4) (2010) 33–38.
[7] An interesting and recent discussion of classical indeterminism in the context of quantum mechanics can be seen in: Manabendra Nath Bera et al. Randomness in quantum mechanics: philosophy, physics and technology. Rep. Prog. Phys. **80** (2017) 124001.
[8] Max Born. Is Classical Mechanics in Fact Deterministic? Phys. Blätter **11** (9) (1955) 49.
[9] Max Born. Nobel Lecture. The Statistical Interpretations of Quantum Mechanics. Nobel Lecture, December 11, 1954. https://www.nobelprize.org/uploads/2018/06/born-lecture.pdf
[10] R.P. Feynman, RB Leighton, and M. Sands. The Feynman Lectures on Physics. Vol. I Mainly Mechanics, Radiation and Heat, Addison-Wesley, Reading, Massachusetts. (1963).
[11] M. Mikael Rågstedt. From Order to Chaos: The Prize Competition in Honor of King Oscar II. http://www.mittag-leffler.se/library/prize-competition
[12] H. Scott Dumas. The KAM Story. A Friendly Introduction to the Content, History, and Significance of Classical Kolmogorov – Arnold – Moser Theory. World Scientific. (2014).
[13] E.N. Lorenz (1963). Deterministic Nonperiodic Flow. Journal of Atmospheric Sciences **20** (1963) 130-148.
[14] Benoit B. Mandelbrot. The Fractalist: Memoir of a Scientific Maverick. Vintage. (2014).
[15] Information on Lyapunov exponents can be obtained at: https://es.wikipedia.org/wiki/Lyapunov_Exponent

[16] N. M. Krylov and N. N. Bogolyubov. Introduction to Nonlinear Mechanics. Ukrainian Academy of Sciences Publishers. Kiev. (1937). English translation by S. Lefschetz. Introduction to Nonlinear Mechanics. Princeton University Press (1943).
[17] Chihiro Hayashi. Nonlinear Oscillations in Physical Systems. Princeton University Press (2014).
[18] Robert M. May. Simple Mathematical Models with very Complicated Dynamics. Nature **261** (1976) 459.
[19] Tien-Yien Li; James A. Yorke (1975). Period Three Implies Chaos.The American Mathematical Monthly **82** (1975) 985-992.
[20] Warren Weaver. Science and Complexity. American Scientist **36** (4) (1948) 536-544. https://www.jstor.org/stable/pdf/27826254.pdf
[21] Philip W. Anderson. More is different. Science **177** (1972) 393-396. http://bit.ly/2qmKnt2
[22] Philip W. Anderson. Physics: The Opening to Complexity. Proceedings of the National Academy of Sciences of the USA **92** (1995) 6653-6654. http://www.pnas.org/cgi/reprint/92/15/6653
[23] Robert Laughlin. A Different Universe: Reinventing Physics from the bottom down. Basic Books (2005).
[24] R.B. Laughlin and David Pines. The Science of Everything. Proceedings of the National Academy of Sciences of the USA **97** (2000) 28-3.http://www.pnas.org/cgi/reprint/97/1/28
[25] Tamas Vicsek. The bigger picture. Nature **418** (2002) 131. https://arxiv.org/abs/1006.5944
[26] More information on the Hodking-Huxley model and links to its articles can be found at: https://en.wikipedia.org/wiki/Hodgkin%E2%80%93Huxley_model

Talks not presented to the Proceedings

1. Ivana Atanasovska „The introduction of multi-body contact in non-linear dynamics”
2. Ivan Dimitrijević „Cosmological solutions of a new nonlocal gravity model”
3. Goran Đorđević „Nonlinear effects in in RSII cosmological models with tachyon field”
4. Katica (Stevanović) Hedrih „Once again about nonlinear functional coordinate systems and rolling a heavy ball along their coordinate surfaces”
5. Deszo Horvath „Gravity current behind precipitation patterns”
6. Božidar Jovanović „Symmetries, invariant manifolds and integrability of natural mechanical systems on Stiefel varieties”
7. Ljiljana Kolar-Anić „Formation of differential equations to describe the dynamics of a reaction system”
8. Wieslaw Krolikowsky „Nonlocal solitons”
9. Stevan Maćešić „Analysis of reaction fronts observed in autocatalytic systems with reversible reactions”
10. Papp Paszkal „Controlling pattern formation of cadmium-hydroxide system”
11. Marko Stojanović „Linear and nonlinear effects in tachyon inflation in the holographic braneworld”
12. Anatoly Zagorodny „Bogolyubov-Born-Green-Kirkwood-Yvon-hierarchy and effective grain potentials for dusty plasmas”

List of Participants

1. Najdan Aleksić (Belgrade, Serbia)
2. Slobodan Anić (Belgrade, Serbia)
3. Ivana Atanasovska (Belgrade, Serbia)
4. Milivoj Belić (Doha Qatar)
5. Livija Cvetićanin (Novi Sad, Serbia)
6. Željko Čupić (Belgrade, Serbia)
7. Ivan Dimitrijević (Belgrade, Serbia)
8. Milan Dimitrijević (Belgrade, Serbia)
9. Goran Đorđević (Niš, Serbia)
10. Vladan Đorđević (Belgrade, Serbia)
11. Branko Dragovich (Belgrade, Serbia)
12. Katica (Stevanović) Hedrih (Niš, Serbia)
13. Deszo Horvath (Szeged, Hungary)
14. Nenad Filipović (Kragujevac, Serbia)
15. Božidar Jovanović (Belgrade, Serbia)
16. Ljiljana Kolar-Anić (Belgrade, Serbia)
17. Wieslaw Krolikowsky (Doha, Qatar)
18. Stevan Maćešić (Szeged, Hungary)
19. Nataša Mišić (Belgrade, Serbia)
20. Aleksandar Obradović (Belgrade, Serbia)
21. Zoran Ognjanović (Belgrade, Serbia)
22. Nikolaj Ostrovski (Odžaci, Serbia)
23. Papp Paszkal (Szeged, Hungary)
24. Miguel A. F. Sanjuan (Madrid, Spain)
25. Branislav Sazdović (Belgrade, Serbia)
26. Vladimir Škarka (Angers, France)
27. Marko Stojanović (Niš, Serbia)
28. Branko Urošević (Belgrade, Serbia)
29. Slobodan Vuković (Belgrade, Serbia)
30. Anatoly Zagorodny (Kiev, Ukraine)
31. Volodymyr Zasenko (Kiev, Ukraine)
32. Slobodan Zdravković (Belgrade, Serbia)

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

530.182(082)
51-7:53(082)

CONFERENCE on Nonlinearity (1 ; 2019 ; Beograd)
Proceedings of the 1st Conference on Nonlinearity, (October 11-12, 2019, Belgrade, Serbia) / editors B. [Branko] Dragovich, Ž. [Željko] Čupić ; [organizer] Serbian Academy of Nonlinear Sciences [(SANS)], Belgrade ; [Coorganizers Mathematical Institute, Serbian Academy of Sciences and Arts, Belgrade [and] SEENET-MTP (Southeastern European Network in Mathematical and Theoretical Physics, Niš)]. - Belgrade : Serbian Academy of Nonlinear Sciences = Beograd : Srpska akademija nelinearnih nauka, 2020 (Beograd : Skripta internacional). - VIII, 242 str. : ilustr. ; 24 cm

Nasl. u kolofonu: Prva konferencija o nelinearnosti. - Str. V-VI: Preface / Branko Dragovich, Zeljko Cupic. - Napomene uz radove. - Bibliografija uz većinu radova. - Registar.

ISBN 978-86-905633-4-0

а) Нелинеарне теорије -- Зборници б) Математичка физика -- Зборници

COBISS.SR-ID 15860233