Springer Tracts in Mechanical Engineering

Georg-Peter Ostermeyer
Valentin L. Popov
Evgeny V. Shilko
Olga S. Vasiljeva *Editors*

Multiscale Biomechanics and Tribology of Inorganic and Organic Systems

In Memory of Professor Sergey Psakhie

OPEN ACCESS

Springer

Springer Tracts in Mechanical Engineering

Springer Tracts in Mechanical Engineering (STME) publishes the latest developments in Mechanical Engineering - quickly, informally and with high quality. The intent is to cover all the main branches of mechanical engineering, both theoretical and applied, including:

- Engineering Design
- Machinery and Machine Elements
- Mechanical Structures and Stress Analysis
- Automotive Engineering
- Engine Technology
- Aerospace Technology and Astronautics
- Nanotechnology and Microengineering
- Control, Robotics, Mechatronics
- MEMS
- Theoretical and Applied Mechanics
- Dynamical Systems, Control
- Fluids Mechanics
- Engineering Thermodynamics, Heat and Mass Transfer
- Manufacturing
- Precision Engineering, Instrumentation, Measurement
- Materials Engineering
- Tribology and Surface Technology

Within the scope of the series are monographs, professional books or graduate textbooks, edited volumes as well as outstanding PhD theses and books purposely devoted to support education in mechanical engineering at graduate and post-graduate levels.

Indexed by SCOPUS, zbMATH, SCImago.

Please check our Lecture Notes in Mechanical Engineering at http://www.springer.com/series/11236 if you are interested in conference proceedings.

To submit a proposal or for further inquiries, please contact the Springer Editor **in your country**:

Dr. Mengchu Huang (China)
Email: mengchu.Huang@springer.com
Priya Vyas (India)
Email: priya.vyas@springer.com
Dr. Leontina Di Cecco (All other countries)
Email: leontina.dicecco@springer.com
All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11693

Georg-Peter Ostermeyer · Valentin L. Popov ·
Evgeny V. Shilko · Olga S. Vasiljeva
Editors

Multiscale Biomechanics and Tribology of Inorganic and Organic Systems

In Memory of Professor Sergey Psakhie

Springer

Editors
Georg-Peter Ostermeyer
Institute of Dynamics and Vibrations
Technische Universität Braunschweig
Braunschweig, Niedersachsen, Germany

Evgeny V. Shilko
Institute of Strength Physics and Materials Science
Russian Academy of Sciences
Tomsk, Russia

Valentin L. Popov
Technische Universität Berlin
Berlin, Germany

Olga S. Vasiljeva
Department of Biochemistry and Molecular and Structural Biology
Jozef Stefan Institute
Ljubljana, Slovenia

ISSN 2195-9862 ISSN 2195-9870 (electronic)
Springer Tracts in Mechanical Engineering
ISBN 978-3-030-60123-2 ISBN 978-3-030-60124-9 (eBook)
https://doi.org/10.1007/978-3-030-60124-9

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

Preface

The monograph *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems* is dedicated to the memory of Prof. Sergey Grigorievich Psakhie (02.03.1952–22.12.2018). The topic of the monograph reflects the broad range of scientific focus areas of Prof. Psakhie. In compiling this book, we attempted to represent the various aspects of his multifaceted research interests ranging from theoretical physics research and computer modeling for understanding materials at the atomic scale up to applied science for solving specific problems of the rocket and space industry, medicine or geotectonics. The authors of the monograph are the colleagues of S. G. Psakhie from the Institute, collaborators from his international network, his former students, participants of the Workshops he organized and his friends.

Braunschweig, Germany — Georg-Peter Ostermeyer
Berlin, Germany — Valentin L. Popov
Tomsk, Russia — Evgeny V. Shilko
Ljubljana, Slovenia — Olga S. Vasiljeva
June 2020

Acknowledgements

We thank the authors of individual chapters of the monograph for their contributions. We also would like to express a heartfelt thanks to Dr. Jasminka Starcevic for her immense support during the assembly and editing of this book. She is not among the authors of the book, but her decisive assistance in preparation of the monograph was her contribution to the memory of her colleague and friend, Prof. Psakhie.

We would like to acknowledge Technische Universität Braunschweig (Germany), Technische Universität Berlin (Germany), Institute of Strength Physics and Materials Science (Russia), and Jozef Stefan Institute (Slovenia) for financial support of the Open Access publication.

Braunschweig, Germany — Georg-Peter Ostermeyer
Berlin, Germany — Valentin L. Popov
Tomsk, Russia — Evgeny V. Shilko
Ljubljana, Slovenia — Olga S. Vasiljeva
June 2020

Contents

In Memory of Sergey G. Psakhie

Evgeny V. Shilko, Valentin L. Popov, Olga S. Vasiljeva, and Georg-Peter Ostermeyer

1 Scientific Biography of Professor Sergey Grigorievich Psakhie

E. V. Shilko
Institute of Strength Physics, Materials Science SB RAS, 634055 Tomsk, Russia

V. L. Popov (✉)
Technische Universität Berlin, 10623 Berlin, Germany
e-mail: v.popov@tu-berlin.de

O. S. Vasiljeva
Jozef Stefan Institute, Ljubljana 1000, Slovenia

G.-P. Ostermeyer
Technische Universität Braunschweig, 38106 Brunswick, Germany

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_1

Sergey G. Psakhie was born on March 2, 1952 in Tomsk (the main city of the Tomsk region in Siberia, USSR) in a family of school teachers Grigory A. Psakhie and Nadezhda A. Psakhie.

His father, **Grigory Abramovich Psakhie**, was one of the most famous and respected school teachers in the history of Tomsk Region. He has worked as a physics and astronomy teacher for more than half a century and was a school principal in Tomsk and Tomsk Region for about 40 years. Grigory Abramovich is rightfully called the Innovative Teacher. He was one of the first in the USSR to introduce the most advanced system of teaching physics and astronomy at that time, the Shatalov system, and has introduced a large number of other innovations in the school education system in the Tomsk Region. For many years of fruitful educational and enlightenment activities, Grigory Psakhie was awarded numerous orders, medals, and diplomas. The most valuable for him were the Order of the October Revolution and the Janusz Korczak medal "To a teacher who raised students" (the latter had a cult status among the teachers in the USSR). The Academic Lyceum in Tomsk is named after Grigory Psakhie.

Sergey's mother, **Nadezhda Alekseevna Psakhie**, devoted her whole life to preschool and primary school education for children. For more than half a century in the profession, Nadezhda Psakhie worked as a kindergarten teacher, head of a kindergarten, primary school teacher and educator. At all places of work, she was considered an exemplary teacher and educator. She was awarded numerous diplomas of the Departments of public education of Tomsk and the Tomsk Region. Grigory and Nadezhda Psakhie lived together for 65 years and were awarded a special certificate from the President of the Russian Federation Vladimir Putin as a family who lived a long, happy life together.

Grigory and Nadezhda Psakhie

Sergey Psakhie raised and taught two daughters (**Olga** and **Natalia**), and son **Ivan**. Olga (married name Vasiljeva), a Professor of Biochemistry at Jozef Stefan Institute (Slovenia) and an expert in protease biology and oncology, has been working with her father, Sergey Psakhie, on several interdisciplinary projects. Natalia Psakhie is a product manager working at top Silicon Valley (California) tech companies. Ivan Psakhie became a scientist in the molecular biology field.

Sergey Psakhie spent a significant period of his childhood in small villages New Vasyugan, Middle Vasyugan, Kargasok in the north of the Tomsk Region. In some of these villages there were no kindergartens, so Sergey became independent very early and learned to read as early as 3 years old. His parents devoted a lot of time to raising and educating their son, so when Sergey entered the school, the school administration offered to immediately transfer Sergey to the 2nd grade. Already at school, Sergey Psakhie manifested diverse interests, high intelligence, and leadership. He was seriously interested in physics and chess, and at the same time passionately engaged in sports and dances, participated in the performances of the Greek school theater, and wrote talented poems. At the same time, Sergey demonstrated excellent performance in almost all school subjects. In high school, Sergey dreamed of entering a flight school, but in the end, his father-instilled love of physics won.

Upon school graduation, Sergey Psakhie entered the **Physics Department of Tomsk State University**, from which he **graduated in 1976**.

In 1976–1979, Sergey Psakhie studied at the graduate school of Tomsk State University (TSU) and carried out scientific research under the guidance of Professor Viktor E. Panin. Viktor Panin became for Sergey not just a scientific supervisor, but a teacher, colleague, and like-minded person for many subsequent decades. In the graduate school of TSU, Sergey Psakhie chose as a scientific specialization a new, and only emerging at that time, direction: computer modeling of processes and phenomena in solids at the atomic level. This direction would later become one of the leading and most successful scientific areas of his future scientific school.

In 1981, Sergey Psakhie successfully defended his thesis for the degree of candidate of physical and mathematical sciences. The theme of his dissertation was "Investigation of the interaction between atoms of alloying elements and vacancies in diluted aluminum-based alloys".

From 1980 to 1984, Sergey Psakhie was a junior researcher at the Institute of Atmospheric Optics of the Siberian Branch of the Academy of Sciences of the USSR (now V. E. Zuev Institute of Atmospheric Optics of the Siberian Branch of the Russian Academy of Sciences) in the Department of solid-state physics and materials science. This department was headed by his scientific adviser Viktor Panin. The best qualities of Sergey Psakhie as a great scientist, such as scientific instinct, understanding of the physical nature of phenomena, the ability to formulate ideas and results, and extraordinary organizational talent, were clearly manifested already during this period.

In 1984, a new research institute was opened in Tomsk—the Institute of Strength Physics and Materials Science of the Siberian Branch of the Academy of Sciences of the USSR (after the collapse of the USSR in 1992 it was renamed to the Institute of Strength Physics and Materials Science of the Siberian Branch of the Russian

Academy of Sciences, ISPMS SB RAS). Professor Viktor E. Panin was the organizer and first director of the Institute. Together with him and a group of ambitious young scientists, Sergey Psakhie came to the new institute as a senior researcher. The Institute became for Sergey Grigorievich a scientific home for the rest of his life. ISPMS SB RAS is associated with all of his key scientific and career achievements.

I**n 1985**, young ambitious scientist Sergey Psakhie founded a new Laboratory of automation (from 1998, the Laboratory of computer-aided design of materials) and became its head. The scientific formation of his best students have been working in this laboratory. Among them are Dr. Konstantin P. Zolnikov, Dr. Sergey Yu. Korostelev, Dr. Alexey Yu. Smolin, Dr. Andrei I. Dmitriev, Dr. Evgeny V. Shilko, Dr. Andrey V. Dimaki, and many others.

Sergey Psakhie (second from the left) with visitors in the Laboratory of computer-aided design of materials (ca. 1990)

In 1990, Sergey Psakhie successfully defended his dissertation for the degree of Doctor of Physical and Mathematical Sciences at the age of only 38 years and in **1991** received a Doctor of Science degree. The theme of his dissertation was “Interparticle interactions and nonlinear properties of metals under mechanical stress”.

In 1991, Sergey Psakhie took the post of deputy director of the Institute for Research. In this position, he worked until 1993.

In 1994, Sergey Psakhie worked as a visiting professor at North Carolina State University (USA) by the invitation of one of the leading experts in high-rate processes and phenomena in solids, professor Yasuyuki (Yuki) Horie. Work at NCSU significantly changed the worldview of Sergey Psakhie. He fully realized that world-class science can only be successfully developed in close international cooperation. The

foundations of such cooperation were laid in the American period of his career, and Sergey Psakhie was engaged in its development throughout his subsequent scientific career at ISPMS SB RAS. Close collaboration of Sergey Psakhie and his research team with leading American, European and Chinese scientists enabled the achievement of advanced fundamental and applied results in materials science, and led to international recognition in the form of highly cited papers in high ranked journals, international patents, and scientific contracts with leading European universities and international industrial corporations.

The period of work at the NCSU has become extremely fruitful scientifically. In collaboration with Prof. Yasuyuki Horie, Sergey Psakhie developed a new particle method, namely, movable cellular automaton (MCA) method. This method was originally developed to model mechanically activated chemical reactions in powder mixtures. In fact, it was a hybrid numerical technique, which combines the formalisms of discrete element and cellular automaton methods. Later, the formalism of this method was used to create an advanced implementation of the discrete element method, namely of the method of homogeneously deformable discrete elements.

In the same year 1994, Sergey Psakhie became a member of the New York Academy of Sciences and American Ceramic Society.

From 1995 to 2002, Sergey Psakhie continued his scientific work at ISPMS SB RAS as the head of the Laboratory of computer-aided design of materials in close international cooperation with Prof. Yuki Horie, Prof. Zongguang Wang (Institute of Metal Research CAS, China), Dr. Stanko Blatnik (Jožef Stefan Institute and INOVA d.o.o., Slovenia), Dr. Simon Zavsek (Velenje Coal Mine, Slovenia), Prof. Jože Pezdič (University of Ljubljana), Prof. Georg-Peter Ostermeyer (Technische Universität Berlin, Technische Universität Braunschweig after 2002) and Prof. Valentin L. Popov (Universität Paderborn, Technische Universität Berlin after 2002). Sergey Psakhie was closely related to these scientists not only by joint research, but also by many years of friendship.

Thanks to broad scientific collaboration, several new areas of the Sergey Psakhie scientific school were created during this period, and later became extremely successful:

- development of the formalism of hybrid cellular automaton method (the coupled discrete element–finite difference numerical technique) for the study of gas-saturated and fluid-saturated porous materials including coal, sandstone, bone tissues and so on;
- computational study of friction and wear in technical and natural tribounits at different scales using a particle-based approach;
- development of an approach to stress state prognosis and managing the displacement mode and seismic activity of tectonic faults.

The latter approach was co-developed in collaboration with prominent experts in geology and geophysics, Prof. Sergey V. Goldin and Dr. Valery V. Ruzhich. In addition to computational modeling, a key component of these studies was the long-term field research in the dynamics of the segments of tectonic faults and the Lake

Baikal ice cover, as a model block-structured medium. These studies included long-term monitoring of displacements and seismicity in combination with managing impacts (explosions, vibrations, fluid injection). An important result of these studies, which Sergey Psakhie was rightly proud of, is the Patent of Russian Federation (2006) "A method for controlling the displacement mode in fragments of seismically active tectonic faults".

Sergey Psakhie's colleague, Prof. Valentin L. Popov, has contributed to the development of the three aforementioned scientific areas. He initiated and participated in a large number of joint projects and is the co-author of dozens of joint papers. For the past two decades, Valentin Popov has been Sergey Psakhie's closest friend.

In 2002, Sergey Psakhie took a new decisive step. He became the director of his native Institute (ISPMS SB RAS), succeeding his teacher Academician Viktor E. Panin in this post. Sergey Psakhie made this decision after much deliberation, but subsequently he never regretted it. In his new post, Sergey Psakhie became one of the most famous and successful directors of research institutes in Russia and was able to execute the most ambitious projects in a wide spectrum of scientific fields from materials science and mechanical engineering to geotectonics and biomedicine. Under the leadership of Sergey Psakhie, the ISPMS SB RAS developed rapidly and in 2017 became one of the top research institutes in Russia. It is a first-rank institute and is included in the TOP-10 Russian scientific institutes based on the number of papers published in journals indexed in Web of Science.

Despite working at the research institute, Sergey Psakhie never lost touch with his Alma Mater. **Since 1992**, he worked as a professor at the Department of Strength and Design of Tomsk State University, and **in 2005** he took the post of head of the Department of High Technology Physics in Mechanical Engineering at Tomsk Polytechnic University. Working both at the research institute and at leading Tomsk universities, Sergey Psakhie made an invaluable contribution to the formation of Tomsk Consortium of Scientific and Educational Organizations enabling integration of Tomsk's educational and scientific systems.

Sergey Psakhie believed that the presence of scientific schools in Russia, including strong Tomsk scientific schools in the fields of theoretical physics, material science, and advanced medicine, is a key advantage of the education system in the Russian Federation. He always considered Tomsk as a place of attraction for the implementation of joint projects with world leading scientific organizations and did a lot to form such projects in the field of materials science. Examples are the project aimed to develop and create the production of advanced wound healing materials "VitaVallis", development of a new multi-beam electron beam technology for high-performance additive production of large-size metal products and structures for key industries in Russian Federation.

The high reputation of Sergey Psakhie as an outstanding organizer of science, determined his active and multifaceted work in the Russian Academy of Sciences.

From January **2004**, he combined presiding the ISPMS SB RAS with the post of the Deputy Chairman of the Presidium of the Tomsk Scientific Center of the SB RAS, and from **2006 to 2013** with the post of the Chairman of the Presidium of the Tomsk Scientific Center of the SB RAS.

Sergey Psakhie was one of the key leaders in the Siberian Branch of the Russian Academy of Sciences, a member of the Presidium of the SB RAS, and Deputy Chairman of the SB RAS contributing to innovative activities and the development of integrated scientific and educational systems in Siberia.

The merited recognition of Sergey Psakhie's high scientific and organizational achievements and scientific reputation led to his election as a Corresponding Member of the Russian Academy of Sciences **in 2011**.

Sergey Psakhie has contributed to scientific and organizational work in multiple governmental, municipal, and public organizations. He was a member of the Russian Foundation for Basic Research (RFBR) Council, the Russian National Committee on Theoretical and Applied Mechanics, the Russian National Committee on Tribology, the Interstate Coordination Council on physics of strength and plasticity of materials, the Innovation Council of the Siberian Branch of the Russian Academy of Sciences, multiple coordination councils (including international), the expert committee of the Russian Youth Prize in Nano-industry, the committee for the competition of technological projects, the Tomsk Region Administration, a number of councils for awarding prizes in the Tomsk Region, etc. Over the years, he was a member and co-chair of the organizing committees of many international conferences. He was a member of the editorial boards of three international journals.

For his tenure and achievements in science and the organization of science and education in Russia, Sergey Psakhie was awarded a large number of honors including the title of "Honorary Worker of Science and Technology of the Russian Federation", the medal of the Russian Cosmonautics Federation "For Merits", honorary badges of the SB RAS "Silver Sigma", the medal "Honored Veteran of the SB RAS", and the Order of Friendship, among others.

Despite the high workload of organizational activities, Sergey Psakhie always considered science as his main priority. His great fundamental knowledge, unusually wide erudition, sharp mind, and talent made it possible to achieve great results in various scientific fields. Sergey Psakhie has always actively supported interdisciplinary research conducted at the intersection of different sciences. He saw the prospect of such research in the fact that well-established methods and approaches from one field of science, after some modification, can be efficiently used to obtain breakthrough results in another scientific field. The range of scientific interests of Sergey Psakhie was unusually broad and covered the problems of theoretical research and computer modeling of complex nonlinear processes in technical and natural (biological and geological) materials.

In the last decade, Sergey Psakhie's main research interests have been related to the study of the interactions between hard matter and soft matter in multiphase contrast materials, the study of the features and anomalies of the behavior of solids under mechanical confinement (confined matter), the study of dusty plasma, etc. In particular, implementation of the method of particles (movable cellular automaton method) which he developed in cooperation with Prof. Y. Horie and Prof. V. Popov has been broadly used to solve fundamental and important practical problems in the field of mechanics of solids, fracture mechanics, mechanics and physics of friction and wear, geomechanics and tectonophysics, biomedicine.

In 2007, he initiated a new scientific direction connected with the theoretical study of plasma-dust crystals. The study of dusty plasma, which is often called the new state of matter, is one of the most promising scientific fields and is developed by collaborations of leading scientific groups around the world.

Since his student years and throughout his scientific career, Sergey Psakhie was actively involved in radiation materials science. His first publications in 1980s were dedicated to this topic. In the 2000s, Sergey Psakhie initiated scientific cooperation with A. A. Bochvar High-Technology Scientific Research Institute for Inorganic Materials (Moscow, Russia), the lead institution for the development of promising structural materials for nuclear power engineering. This extremely fruitful collaboration brought together several leading Russian scientific teams and continues to this day. The results of theoretical studies of Sergey Psakhie's scientific group are unique and important for the development of new materials for nuclear energy, as they shed light on the particular behavior of solids under conditions of high temperatures, pressures and radiation exposure that materials experience in the core of nuclear reactors.

A large place in the professional activity of Sergey Psakhie in recent years has been devoted to scientific and technical cooperation with S.P. Korolev Rocket and Space Corporation Energia (Korolev, Russia). Joint research was aimed at solving specific problems of the rocket and space industry. Among the largest projects implemented was the development of methods and equipment for non-destructive testing of welded joints, the efficient friction stir welding technology, 3D-printing technology of products from polymeric materials under zero-gravity conditions, a new method and equipment for repair and restoration of the surface of glass illuminators damaged as a result of the impact of micrometeorites, and others. A huge role in the implementation of these ambitious projects was played by Alexander G. Chernyavsky, Deputy Chief Designer of S. P. Korolev Rocket and Space Corporation Energia, who became not only a like-minded person, but also a friend of Sergey Psakhie.

Sergey Psakhie initiated and implemented innovative scientific direction related to the physics of nanoscale states of substances. Since the mid-2000s, the ISPMS SB RAS has become one of the leading nanotechnology innovation centers in Russia.

Sergey Psakhie made a great contribution to the study of the biological effect of low-dimensional metal oxide nanostructures and their use for biomedical applications, including the creation of new cancer treatment strategies. Latter studies have been performed in collaboration with his daughter, Prof. Olga Vasiljeva (Jozef Stefan Institute, Slovenia). As such, his publication with multidisciplinary international team "Ferri-liposomes as an MRI-visible drug-delivery system for targeting tumors and their microenvironment" published in Nature Nanotechnology journal, has more than 300 citations. In 2018, Sergey Psakhie and his team received a US patent for the use of such materials to suppress tumor growth (US Patent 10105318 "Low-dimensional structures of organic and/or inorganic substances and use thereof").

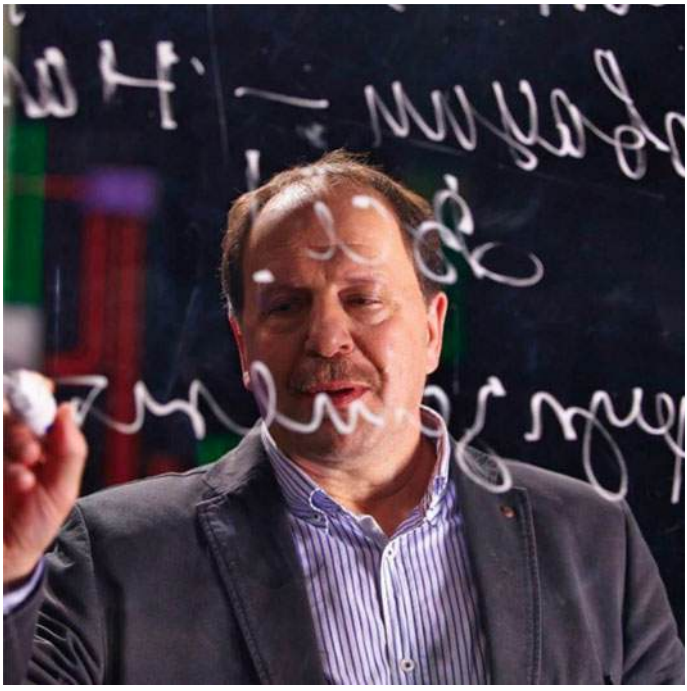
The staff of his laboratory and chair treated Sergey Psakhie not just as head and colleague, but as a Teacher and Scientist. Over the years, he mentored 5 Doctors of Sciences and about 20 Candidates of Science (Ph.D.), and published with co-authors

more than 300 papers. Nowadays, students of his students are successfully defending scientific dissertations.

Sergey Psakhie suddenly passed away on **December 22, 2018**. On the morning of the day of his death, he met with colleagues and had a long list of planned events. The death of a talented Russian physicist, an outstanding organizer of science, a man who devoted himself to serving and promoting science, shocked the entire Tomsk scientific and educational community and the large number of his Russian and international colleagues and friends.

The loss of such a great scientist and person is shocking and irreparable and will remain so for many more years. Nonetheless, the principles of scientific activities created and applied by Sergey Psakhie, his scientific ideas and undertakings would allow the ISPMS SB RAS, Tomsk and Russian science to successfully develop further. Multiple colleagues, collaborators and mentees of Sergey Psakhie will continue to implement the initiatives of Sergey Psakhie and carry on his innovative work.

The name **Sergey G. Psakhie** is forever inscribed in gold letters in the history of Russian and international science.



Sergey Grigorievich Psakhie (02.03.1952–22.12.2018)

2 Georg-Peter Ostermeyer: Twenty Years of Friendship with Sergey Psakhie

G.-P. Ostermeyer is director of the Institute of Dynamics and Vibrations at the University of Braunschweig, Germany.

My wife Ulrike and I first met Professor Psakhie in 1999 in Berlin. There, at the TU Berlin, I established a Collaborative Research Center on friction and invited Prof. Psakhie several times in this context. Besides scientific discussions, we also explored Berlin together and went on extensive hikes in the nature around Berlin. We became friends very soon and he remained our very good friend until the end.

This was also because we were both interested in the philosophy of discrete methods in tribology. We both shared the opinion that mesoscopic simulation methods in particular are essential for tribology. We were in contact all the time and have developed together many ideas. One of these ideas was the simulation of the thermalization process on the mesoscale based on the Shannon theorem. The method of meso particles, which I developed in the late 1990s, and the method of movable cellular automata, developed by Sergei Psakhie, solved this problem in different ways, and the combination of both approaches was a very attractive idea.

It was always a pleasure to have discussions with Sergey Psakhie. With his very broad range of interests and his extensive experience in discrete methods, we would often return to the topic of numerical modeling of multiscale phenomena. Examples of such are material textures in friction processes, which react on slow time scales with surprising motion dynamics, to friction loads, or chemical reactions in the friction boundary layer, which are found on completely different time and size scales. We were fascinated by the idea of treating multiscale effects with different abstract, i.e. scale-independent, methods, whereby the interconnection of these methods alone represents the scale spread. My wife Ulrike, a mathematician, was always fascinated by the mathematical depth of his argumentation even on topics, which were quite far from his fields of research, and Sergey Psakhie enjoyed these inspiring discussions with her.

The last time we met personally was during the International Workshop “Advances in Tribology: Science, Technology and Education” in Karlovy Vary in 2015, organized by Sergey Psakhie together with Valentin Popov and myself. This meeting was like a throwback to the early days of our friendship. Once again, we were walking (and working) together in the nature. One time the three of us took a shortcut on the way back to the hotel - but the intensive discussions made sure that we didn’t arrive any earlier. The plan to hold a joint conference in Tomsk was born and we made plans to try to organize this conference thematically and chronologically.

Unfortunately, these plans could not be realized because Sergei Psakhie left us suddenly and completely unexpectedly in December 2018. This was a great shock for all of us.

I am very glad that I can contribute to the memory of the great scientist and man Sergei Psakhie by co-editing this monograph and co-organizing the commemorative workshop on “Multiscale Biomechanics and Tribology of Inorganic and Organic

Systems" in Tomsk in October 2019. My wife and I experienced and got to know Tomsk, the home city of Sergey, up close for the first time. We visited his place of work, and also the nature around Tomsk, with the feeling that he was still very much with us.

At the German-Russian Workshop "Advances in Tribology: Science, Technology and Education", Karlovy Vary, Czech Republic, March 2, 2015. First row left: S. G Psakhie, rigth: G. P. Ostermeyer, behind them: Ulrike Ketterl-Ostermeyer. Other participants (from the left): E. Shilko, V. Aleshin, M. Popov, V. Popov, A. Smolin, A. Dmitriev, R. Pohrt, J. Starcevic, E. Kolubaev, A. Korsunsky, A. Dimaki

3 Valentin L. Popov: A Word of Sergey Psakhie

V. L. Popov is the head of Department of System Dynamics and Friction Physics at the Technische Universität Berlin.

My collaboration and close friendship with Sergey Grigorievich Psakhie started in 1997. Together with Prof. E. Santner, at that time the head of tribology department of the Federal Institute of Materials Research and Testing (BAM), we submitted a joint project devoted to simulation of tribological processes with the Method of Movable Cellular Automaton. This project became a great success and established for many years one of research directions at the BAM.

Sergey Psakhie was the most important partner of the Department of System Dynamics and Friction Physics of Technische Universität Berlin, which I headed since 2002. Our cooperation covered diverse areas to which tribology can be applied: molecular motors and earth tectonics, problems of material wear and damping of aerospace structures, active control of friction and, of course, numerical modeling methods in tribology. Most important for us was scientific collaboration in the field of

application of mesoparticle approach, in particular Cellular Automata, to tribological problems. Among a large number of papers, I can highlight our very fundamental and highly cited programmatic joint paper, "Numerical Simulation Methods in Tribology", appeared in 2007, which contained the most important future developments in this field such as Method of Dimensionality Reduction, applications of MCA and stochastic differential equations.

The forms of cooperation were correspondingly manifold: annual German-Russian workshops (in fact international conferences), seismological field expeditions, multiple joint projects, invited professorships, international laboratories, visits of student groups, academic exchanges and finally the Double Degree Program with the Polytechnic University initiated by Prof. Psakhie in 2006. The program ran on the base of the Department of High Technology Physics in Mechanical Engineering headed by Sergey Psakhie. Since that time, about 100 German students studied in Tomsk thanks to these programs.

Sergey Psakhie was director of an academic institute. And he suited this position very well. He was a surprisingly harmonious and versatile person. Despite the fact that he devoted his life to materials, he was always attracted to biological topics. He believed that the twenty-first century is the century of biology, especially molecular biology. He was an expert in both fields. And not just an expert. He was a visionary, theorist and practitioner in one. Sergey Psakhie was a very erudite and well-read man. My wife Elena, a philologist with a double education (graduated from the Tomsk State University in Russian philology and from Paderborn University in German philology) had many times to urgently re-read works, heroes or conflicts of which Sergey mentioned.

In October of 2018, Sergey Psakhie and I, together with the rector of TSU, Eduard Vladimirovich Galazhinsky, discussed a new scientific direction - active biocontact mechanics, the purpose of which is the natural restoration of joints controlled by mechanical and medical means, instead of surgical replacement. Already in December 2018 the preparation of a joint multilateral project was in full swing. But when I received the sad news on Saturday morning, 22. December 2018, my thoughts were not about cooperation and not about laboratories. Because Sergey Psakhie was not only my colleague, but also a close friend of my whole family. The memory machine has started and cannot stop until now…

In 2005, we celebrated Christmas at the Baltic Sea, in Warnemünde. It is easy to imagine the weather in North Germany at the end of December. Nevertheless, Sergey jokingly suggested taking a swim next morning. Indeed, at 6.00 am in the morning, still in the dark, my sons, Nikita and Misha woke him up and pulled him to go swimming. There is a historical photograph capturing Sergey Psakhie with Nikita and Misha swimming in fur hats in the Baltic Sea on the 25th of December.

We often took vacations together and Sergey supplied us with ideas for movies to watch. One of these movies was "The Discreet Charm of the Bourgeoisie" by Luis Bunuel, which we watched many times.

Like many parents, I sometimes suffer from a critical attitude towards my children. Sergey constantly corrected me and told me how talented they are and what they fascinate him with.

Sergey introduced Russia to my German colleagues. I would like to mention just two names: Jasminka Starcevic, who was a devoted close friend of Sergey Psakhie; and, of course, Professor Ostermeyer with his wife Ulrike, with whom Sergey Psakhie had a family friendship.

Sergey Psakhie was a true friend who we regard as a member of our family. He entered our lives a long time ago and will remain in them forever.

Sergey Psakhie and Valentin Popov at the German-Russian Workshop on "Numerical simulation methods in tribology: possibilities and limitations" (Technische Universität Berlin, March 2005)

4 Lev B. Zuev: From My Memories of Sergey G. Psakhie

L. B. Zuev is the head of Laboratory of Strength Physics at the Institute of Strength Physics and Materials Science, SB RAS.

I met Sergey Grigorievich Psakhie in 1983. At that time, I headed the Department of Physics at the Siberian Metallurgical Institute in Novokuznetsk, but my moving to the Institute of Strength Physics and Materials Science in Tomsk (which had just formed) was under discussion. The future director of the institute, V. E. Panin, invited me to participate in a Session of the Scientific Council of the USSR Academy of Sciences on strength and ductility. The session was held in a pioneer camp in the village of Zavarzino, located close to the Tomsk Academic Township. Sergey Psakhie was responsible for organization of this Session. It was the fall of 1983 - four months before the official opening of the Institute.

I worked with Sergey Psakhie as his deputy for ten years, from 2002 to 2012. In these years, on his initiative, the research topics and the structure of the Institute have been substantially modernized. New laboratories and new research areas opened, often quite distant from the physics of strength, but ideally fitting into materials science in a broader sense. The technological base of the Institute has been significantly expanded and improved. This allowed rapid progress in our research at the highest international level.

I liked the style of the meetings Sergey Psakhie ran at the Institute. They were relatively brief. He usually listened to a number of points of view, which were presented by the participants, and then quickly made a decision. I want to note that the decision was made at the right time, not too early, and not too late. This contributed to the formation of an optimal point of view, but excluded a long debate that almost never leads to a useful result.

It seems important to me that Sergey Psakhie had a sense of what is now called "growth points", an ability to notice teams and people who have the potential for growth and development. He always considered it his duty to promote their development, which, in turn, contributed to the progress of the Institute as a whole.

I think that even organizational work gave Sergey Psakhie pleasure. He clearly sought to concentrate in his hands the work on opening new areas of research, networking, negotiations, making key decisions, choosing partners, etc.

I was twelve years older than Sergey Psakhie, and with such a difference in age we have not been friends. However, closer, what is called "human" contacts with him developed due to my health problems. His enormous decisive help in this respect did not belong to the scope of duties of director. When a medical treatment was necessary, he immediately picked up the phone, called medical doctors he knew, and even accompanied me to the hospital, which was not necessary at all. Sergey Psakhie considered such help a person's duty.

I am thankful to Sergey Psakhie for a lot, and especially for the editing of my book "Autowave plasticity. Localization and collective modes", which was published in 2018. He proposed the concise title of this monograph and wrote a short introduction to it. I still managed to present him a copy of the book…

I think that Sergey Psakhie was a man of integrity, and such people are always complex. Nevertheless, I am happy that I had the opportunity to work with him for a long time…

Meeting of the director board of ISPMS SB RAS (ca. 1985). Third from the left: S. G. Psakhie, second from the left L. B. Zuev. The meeting is chaired by the founder and first director of ISPMS SB RAS, V.E. Panin

5 Valery V. Ruzhich: On the "Earthquake Vaccine" Project of Sergey Psakhie

V. V. Ruzhich is the Principal Researcher of the laboratorium of Sesimology at the Institute of Earth's Crust, SB RAS, Irkutsk.

The eternal question of how to reduce the destructive consequences of strong earthquakes remains unresolved for many thousands of years. Humanity does not have means to affect the energetics of the deep tectonic processes leading to earthquakes. However, it is legitimate to pose the question: is it possible to find ways to reduce, disperse, or "streamline" the rampant play of the underground elements, manifested in the form of seismo-tectonic catastrophes? This issue was at the center of the project that emerged in the late 1990s and was realized under guidance of Sergey Psakhie. Journalists called the methods developed in the framework of this project "vaccination" from earthquakes.

Our first meeting with Sergey Psakhie took place in 1999 at the initiative of Academician Sergei Vasilievich Goldin at one of the interdisciplinary seminars. The ambitious idea of creating methods of controlled technogenic impacts on faults to mitigate and dissipate the destructive energy stored in the crust interested Sergey Psakhie, although it was very far from the subject of his main research. He belonged

to a rare category of researchers with a huge arsenal of knowledge and own scientific developments, but at the same time very sensitive to innovation. He initiated a project with participation of specialists from 7 scientific organizations of the Russian Federation and Technische Universität Berlin. By 2003, at the Listvyanka landfill in the Angarsk seismically active fault zone, wells were drilled, strain and seismic regime monitoring was organized, and the seismic zone was modeled theoretically. In August 2004, original field experiments were conducted on a real seismic hazardous fault with vibrations, shocks and explosions in combination with injecting water solutions through wells into particular fault fragments. With this complex action, it was possible to induce an accelerated creep mode and to shift the fault banks on the length of about 100 m by 8–10 mm. In 2006, these unique results were recognized by the Siberian Branch of the Russian Academy of Sciences as a breakthrough in seismic research. In the same year, the authors obtained the patent of the Russian Federation "A method for controlling the displacement mode in fragments of seismically active tectonic faults".

Ice chronicle on Baikal. In 2005 Sergey Psakhie proposed a unique scientific project to study the conditions of deformation and dynamic fracture of the ice cover of Lake Baikal, related to the seismically active Baikal rift basin. Annually, many kilometers long main cracks occur in the ice cover, which are accompanied by dynamic phenomena called ice impacts - registered in the form of seismic tremors with an energy comparable to weak tectonic earthquakes. It was assumed that their study might facilitate understanding of similar processers in the lithosphere. An interdisciplinary team of researchers from several scientific institutions, under the guidance of Sergey Psakhie, collected information that contributed to improvement of methods for controlling the regimes of seismic emission generation.

Our last meeting with Sergey Psakhie took place in Tomsk in the fall of 2018, during the annual international conference. With his extensive scientific and organizational activities, Sergey subjected his health to extreme loads, reaching far beyond the functions of the director of a research institute. His regular trips to Moscow and Novosibirsk for solving numerous organizational and funding issues did not leave him any time for recreation. At our last lunch in the cafe of the Academic Township in Tomsk, we discussed the plan for a joint winter expedition in 2019 on the ice of Lake Baikal. When parting, I asked him about his health—he looked tired, but habitually lively. Sitting behind the steering wheel of his SUV, he cheerfully replied: "Fine, only the left lower leg aches a little…"

Two months later, on the day of my birth, friends from Tomsk sent a deafeningly deplorable message about the sudden tragic departure of Sergey Grigorievich Psakhie…

As a tribute of deep respect and in memory of Sergey Psakhie, it is necessary for all of us, his friends and followers to continue research on improving the "vaccine" that is in great demand by the international community against imminent seismic disasters.

On the ice of Lake Baikal. Prof. Ruzhich is the right in the front row. Prof. Psakhie is behind him

6 Most Important Publications of Prof. Sergey Grigorievich Psakhie

I. ***Particle-based and continuum approaches to computational modeling of materials***

Development of particle-based and continuum approaches to computational modeling of the mechanical behavior of materials under complex loading conditions.

1. Grigoriev, A. S., Shilko, E. V., Skripnyak, V. A., **Psakhie, S. G.** (2019). Kinetic approach to the development of computational dynamic models for brittle solids, *International Journal of Impact Engineering*, 123, 4–25.
2. Grinyaev, Yu. V., Chertova, N. V., Shilko, E. V., **Psakhie, S. G.** (2018). The continuum approach to the description of semi-crystalline polymers deformation regimes: the role of dynamic and translational defects, *Polymers*, 10, 1155.
3. **Psakhie, S. G.**, Dimaki, A. V., Shilko, E. V., Astafurov, S. V. (2016). A coupled discrete element-finite difference approach for modeling mechanical response of fluid-saturated porous materials, *International Journal for Numerical Methods in Engineering*, 106, 623–643.
4. Shilko, E. V., **Psakhie, S. G.**, Schmauder, S., Popov, V. L., Astafurov, S. V., Smolin, A. Yu. (2015). Overcoming the limitations of distinct element method

for multiscale modeling of materials with multimodal internal structure, *Computational Materials Science*, 102, 267–285.

5. **Psakhie, S. G.**, Shilko, E. V., Grigoriev, A. S., Astafurov, S. V., Dimaki, A. V., Smolin, A. Yu. (2014). A mathematical model of particle–particle interaction for discrete element based modeling of deformation and fracture of heterogeneous elastic–plastic materials, *Engineering Fracture Mechanics*, 130, 96–115.
6. **Psakhie, S. G.**, Shilko, E. V., Smolin, A. Yu., Dimaki, A. V., Dmitriev, A. I., Konovalenko, Ig. S., Astafurov, S. V., Zavsek, S. (2011). Approach to simulation of deformation and fracture of hierarchically organized heterogeneous media, including contrast media, *Physical Mesomechanics,* 14, 224–248.
7. Grinyaev, Yu. V., **Psakhie, S. G.**, Chertova, N. V. (2008). Phase space of solids under deformation, *Physical Mesomechanics,* 11, 228–232.
8. **Psakhie, S. G.**, Smolin, A. Yu., Stefanov, Yu. P., Makarov, P. V., Chertov, M. A. (2004). Modeling the behavior of complex media by jointly using discrete and continuum approaches, *Technical Physics Letters*, 30, 712–714.
9. **Psakhie, S. G.**, Horie Y., Ostermeyer G.-P., Korostelev, S. Yu., Smolin, A. Yu., Shilko, E. V., Dmitriev, A. I., Blatnik, S., Spegel, M., Zavsek, S. (2001). Movable cellular automata method for simulating materials with mesostructured, *Theoretical and Applied Fracture Mechanics*, 37, 311–334.
10. Popov, V. L., **Psakhie, S. G.** (2001). Theoretical principles of modeling elastoplastic media by movable cellular automata method. I. Homogeneous media, *Physical Mesomechanics*, 4, 15–25.
11. **Psakhie, S. G.**, Horie, Y., Korostelev, S. Yu., Smolin, A. Yu., Dmitriev, A. I., Shilko, E. V., Alekseev, S. V. (1995). Method of movable cellular automata as a tool for simulation within the framework of mesomechanics, *Russian Physics Journal,* 38, 1157–1168.

II. ***Contact interaction, friction and wear***

1. Dimaki, A. V., Dudkin, I. V., Shilko, E. V., **Psakhie, S. G.**, Popov, V. L. (2020). Role of Adhesion Stress in Controlling Transition between Plastic, Grinding and Breakaway Regimes of Adhesive Wear, *Scientific Reports*, 10, 1585.
2. Willert, E., Dmitriev, A. I., **Psakhie, S. G.**, Popov, V. L. (2019). Effect of elastic grading on fretting wear, *Scientific Reports*, 9, 7791.
3. Popov, V. L., Dimaki, A. V., **Psakhie, S. G.**, Popov, M. V. (2015). On the role of scales in contact mechanics and friction between elastomers and randomly rough self-affine surfaces, *Scientific Reports*, 5, 11139.
4. Li, Q., Dimaki, A., Popov, M., **Psakhie, S. G.**, Popov, V. L. (2014). Kinetics of the coefficient of friction of elastomers, *Scientific Reports*, 4, 5795.
5. **Psakhie, S. G.**, Popov, V. L., Shilko, E. V., Smolin, A. Yu., Dmitriev, A. I. (2009). Spectral analysis of the behavior and properties of solid surface layers. Nanotribospectroscopy, *Physical Mesomechanics*, 12, 221–234.
6. Popov, V. L., **Psakhie, S. G.** (2007). Numerical simulation methods in tribology, *Tribology International*, 40, 916–923.

7. Bucher, F., Dmitriev, A. I., Ertz, M., Knothe, K., Popov, V. L., **Psakhie, S. G.**, Shilko, E. V. (2006). Multiscale simulation of dry friction in wheel/rail contact, *Wear*, 261, 874–884.
8. Popov, V. L., **Psakhie, S. G.**, Dmitriev, A. I., Shilko, E. V. (2003). Quasi-fluid nano-layers at the interface between rubbing bodies: simulations by movable cellular automata, *Wear*, 254, 901–906.
9. Popov, V. L., **Psakhie, S. G.**, Shilko, E. V., Dmitriev, A. I., Knothe, K., Bucher, F., Ertz, M. (2002). Friction coefficient in "rail-wheel" contacts as a function of material and loading parameters, *Physical Mesomechanics*, 5(3-4), 17–24.

III. ***Seismic activity of geological media and controlling displacement modes in fault zones***

1. Shilko, E. V., Dimaki, A. V., **Psakhie, S. G.** (2018). Strength of shear bands in fluid-saturated rocks: a nonlinear effect of competition between dilation and fluid flow, *Scientific Reports*, 8, 1428.
2. Ruzhich, V. V., **Psakhie, S. G.**, Chernykh, E. N., Shilko, E. V., Levina, E. A., Dimaki, A. V. (2018). Baikal ice cover as a representative block medium for research in lithospheric geodynamics, *Physical Mesomechanics*, 21, 223–233.
3. Grigoriev, A. S., Shilko, E. V., Astafurov, S. V., Dimaki, A. V., Vysotsky, E. M., **Psakhie, S. G.** (2016). Effect of dynamic stress state perturbation on irreversible strain accumulation at interfaces in block-structured media, *Physical Mesomechanics*, 19, 136–148.
4. **Psakhie, S. G.**, Dobretsov, N. L., Shilko, E. V., Astafurov, S. V., Dimaki, A. V., Ruzhich, V. V. (2009). Model study of the formation of deformation-induced structures of subduction type in block-structured media. Ice cover of Lake Baikal as a model medium, *Tectonophysics*, 465, 204–211.
5. **Psakhie, S. G.**, Shilko, E. V., Astafurov, S. V., Dimaki, A. V., Ruzhich, V. V., Panchenko, A. Yu. (2008). Model study of the formation and evolution of deformation induced structures of the subduction type in the ice cover of Lake Baikal, *Physical Mesomechanics*, 11, 55–65.
6. **Psakhie, S. G.**, Ruzhich, V. V., Shilko, E. V., Popov, V. L., Astafurov, S. V. (2007). A new way to manage displacements in zones of active faults, *Tribology International*, 40, 995–1003.
7. Dobretsov, N. L., **Psakhie, S. G.**, Ruzhich, V. V., Popov, V. L., Shilko, E. V., Granin, N. G., Timofeev, V. Yu., Astafurov, S. V., Dimaki, A. V., Starcevic, Ya. (2007). Ice cover of Lake Baikal as a model for studying tectonic processes in the Earth's crust, *Doklady Earth Sciences*, 413, 155–159.
8. **Psakhie, S. G.**, Ruzhich, V. V., Shilko, E. V., Popov, V. L., Dimaki, A. V., Astafurov, S. V., Lopatin, V. V. (2005). Influence of the state of interfaces on the character of local displacements in fault-block and interfacial media, *Technical Physics Letters*, 31, 712–715.
9. **Psakhie, S. G.**, Ruzhich, V. V., Smekalin, O. P., Shilko, E. V. (2001). Response of the geological media to dynamic loading, *Physical Mesomechanics*, 4(1), 63–66.

IV. ***Vortices as a fundamental mechanism of material response to loading***

1. Shilko, E. V., Astafurov, S. V., Grigoriev, A. S., Smolin, A. Yu., **Psakhie, S. G.** (2018). The fundamental regularities of the evolution of elastic vortices generated in the surface layers of solids under tangential contact loading, *Lubricants*, 6, 51.
2. Shilko, E. V., Grinyaev, Yu. V., Popov, M. V., Popov, V. L., **Psakhie, S. G.** (2016). Nonlinear effect of elastic vortexlike motion on the dynamic stress state of solids, *Physical Review. E,* 93, 053005.
3. **Psakhie, S. G.**, Shilko, E. V., Popov, M. V., Popov, V. L. (2015). The key role of elastic vortices in the initiation of intersonic shear cracks, *Physical Review. E*, 91, 063302.
4. **Psakhie, S. G.**, Zolnikov, K. P., Dmitriev, A. I., Smolin, A. Yu., Shilko, E. V. (2014). Dynamic vortex defects in deformed material, *Physical Mesomechanics*, 17, 15–22.
5. **Psakhie, S. G.**, Smolin, A. Yu., Shilko, E. V., Korostelev, S. Yu., Dmitriev, A. I., Alekseev, S. V. (1997). About the features of transient to steady state deformation of solids, *Journal of Materials Science and Technology*, 13, 69–72.
6. **Psakhie, S. G.**, Smolin, A. Yu., Korostelev, S. Yu., Dmitriev, A. I., Shilko, E. V., Alekseev, S. V. (1995). The study of establishing the steady mode of deformation of solids by the method of movable cellular automata, *Pis'ma Zh. Tech. Phys,* 21(20), 72.

V. ***Atomic mechanisms of inelastic deformation of crystalline solids***

1. **Psakhie, S. G.**, Zolnikov, K. P., Kryzhevich, D. S., Korchuganov, A. V. (2019). Key role of excess atomic volume in structural rearrangements at the front of moving partial dislocations in copper nanocrystals, *Scientific Reports*, 9, 3867.
2. Korchuganov, A. V., Tyumentsev, A. N., Zolnikov, K. P., Litovchenko, I. Yu., Kryzhevich, D. S., Gutmanas, E., Li, S. X., Wang, Z. G., **Psakhie, S. G.** (2019). Nucleation of dislocations and twins in fcc nanocrystals: Dynamics of structural transformations, *Journal of Materials Science and Technology*, 35, 201–206.
3. Zolnikov, K. P., Korchuganov, A. V., Kryzhevich, D. S., Chernov, V. M., **Psakhie, S. G.** (2019). Formation of Point Defect Clusters in Metals with Grain Boundaries under Irradiation, *Physical Mesomechanics*, 22(5), 355–364.
4. Korchuganov, A. V., Zolnikov, K. P., Kryzhevich, D. S., **Psakhie, S. G.** (2017). Primary Ion-Irradiation Damage of BCC-Iron Surfaces, *Russian Physics Journal*, 60, 170–174.
5. Korchuganov, A. V., Zolnikov, K. P., Kryzhevich, D. S., Chernov, V. M., **Psakhie, S. G.** (2016). MD simulation of plastic deformation nucleation in stressed crystallites under irradiation, *Physics of Atomic Nuclei*, 79(7), 1193–1198.
6. Zolnikov, K. P., Korchuganov, A. V., Kryzhevich, D. S., Chernov, V. M., **Psakhie, S. G.** (2015). Structural changes in elastically stressed crystallites under irradiation, *Nuclear Instruments & Methods in Physics Research Section B-Beam Interactions with Materials and Atoms*, 352, 43–46.

7. Dmitriev, A. I., Nikonov, A. Yu., **Psakhie, S. G.** (2011). Atomistic mechanism of grain boundary sliding with the example of a large-angle boundary $2 = 5$. Molecular dynamics calculation, *Physical Mesomechanics*, 14, 24–31.
8. **Psakhie, S. G.**, Zolnikov, K. P., Dmitriev, A. I., Konovalenko, Iv. S. (2009). Kinematic properties of nanostructures based on bilayer nanocrystalline films, *Physical Mesomechanics*, 12, 112–116.
9. **Psakhie, S. G.**, Zolnikov, K. P., Kryzhevich, D. S. (2007). Elementary atomistic mechanism of crystal plasticity, *Physics Letters A*, 367, 250–253.
10. **Psakhie, S. G.**, Zolnikov, K. P., Kryzhevich, D. S., Lipnitskii, A. G. (2006). On structural defect generation induced by thermal fluctuations in materials with a perfect lattice under dynamic loading, *Physics Letters A*, 349, 509–512.
11. **Psakhie, S. G.**, Zolnikov, K. P. (1998). Possibility of a vortex mechanism of displacement of the grain boundaries under high-rate shear loading, *Combustion Explosion and Shock Waves*, 34(3), 366–368.
12. **Psakhie, S. G.**, Zolnikov, K. P. (1997). Anomalously high rate of grain boundary displacement under fast shear loading, *Technical Physics Letters*, 23, 555–556.
13. **Psakhie, S. G.**, Korostelev, S. Y., Negreskul, S. I., Zolnikov, K. P., Wang, Z. G., Li, S. X. (1993). Vortex mechanism of plastic-deformation of grain-boundaries—computer-simulation, *Physica Status Solidi B-Basic research*, 176(2), K41–K44
14. Zolnikov, K. P., **Psakhie, S. G.**, Panin, V. E. (1986). Alloy phase-diagrams using temperature, concentration and density as variables, *Journal of Physics F—Metal Physics*, 16, 1145–1152.
15. **Psakhie, S. G.**, Panin, V. E., Chulkov, E. V., Zhorovkov, M. F. (1980). Pseudopotential theory calculation of bounding energy of zinc atom with vacancy in aluminum, *Fizika Metallov i Metallovedenie*, 50, 620–622.
16. **Psakhie, S. G.**, Panin, V. E., Chulkov, E. V., Zhorovkov, M. F. (1980). Calculation of the bond-energy of Mg and Zn impurities with vacancies in Al-alloys, *Izvestiya Vysshikh Uchebnykh Zavedenii Fizika*, 8, 99–104.

VI. ***Biological and medical applications***

1. Sharipova, A., Gotman, I., **Psakhie, S. G.**, Gutmanas, E. Y. (2019). Biodegradable nanocomposite Fe-Ag load-bearing scaffolds for bone healing, *Journal of the Mechanical Behavior of Biomedical Materials*, 98, 246–254.
2. Lerner, M. I., Mikhaylov, G., Tsukanov, A. A., Lozhkomoev, A. S., Gutmanas, E., Gotman, I., Bratovs, A., Turk, V., Turk, B., **Psakhie, S. G.**, Vasiljeva, O. (2018). Crumpled aluminum hydroxide nanostructures as a microenvironment dysregulation agent for cancer treatment, *Nano Letters*, 18, 5401–5410.
3. Sharipova, A., Swain, S. K., Gotman, I., Starosvetsky, D., **Psakhie, S. G.**, Unger, R., Gutmanas, E. Y. (2018). Mechanical, degradation and drug-release behavior of nano-grained Fe-Ag composites for biomedical applications, *Journal of the Mechanical Behavior of Biomedical Materials*, 86, 240–249.
4. Tsukanov, A. A., **Psakhie, S. G.** (2017). From the soft matter-hard matter interface to bio-self-organization and hybrid systems, *Physical Mesomechanics*, 20, 43–54.

5. Tsukanov, A. A., **Psakhie, S. G.** (2016). Energy and structure of bonds in the interaction of organic anions with layered double hydroxide nanosheets: A molecular dynamics study, *Scientific Reports*, 6, 19986.
6. Lozhkomoev, A. S., Glazkova, E. A., Bakina, O. V., Lerner, M. I., Gotman, I., Gutmanas, E. Y., Kazantsev, S. O., **Psakhie, S. G.** (2016). Synthesis of core-shell AlOOH hollow nanospheres by reacting Al nanoparticles with water, *Nanotechnology*, 27, 205603.
7. Bakina, O. V., Glazkova, E. A., Svarovskaya, N. V., Lozhkomoev, A. S., Lerner, M. I., **Psakhie, S. G.** (2015). The influence of precursor disaggregation during synthesis of low-dimensional AlOOH structures on their morphology, *Russian Physics Journal*, 57, 1669–1675.
8. Mishnaevsky, L. Jr., Levashov, E., Valiev, R. Z., Segurado, J., Sabirov, I., Enikeev, N., Prokoshkin, S., Solov'yov, A. V., Korotitskiy, A., Gutmanas, E., Gotman, I., Rabkin, E., **Psakhie, S. G.**, Dluhoš, L., Seefeldt, M., Smolin, A. (2014). Nanostructured titanium-based materials for medical implants: Modeling and development, *Materials Science and Engineering: R: Reports*, 81, 1–19.
9. Mikhaylov, G., Mikac, U., Magaeva, A. A., Itin, V. I., Naiden, E. P., Psakhye, I., Babes, L., Reinheckel, T., Peters, C., Zeiser, R., Bogyo, M., Turk, V., **Psakhie, S. G.**, Turk, B., Vasiljeva, O. (2011). Ferri-liposomes as an MRI-visible drug-delivery system for targeting tumours and their microenvironment, *Nature Nanotechnology*, 6, 594–602.

VII. ***Dusty plasma as a special state of particulate matter***

1. **Psakhie, S. G.**, Zolnikov, K. P., Abdrashitov, A. V. (2010). On the formation of structural states in dusty plasmas & *Physical Mesomechanics*, 13(5–6), 275–282.
2. Abdrashitov, A. V., Zolnikov, K. P., **Psakhie, S. G.** (2010). Effect of the Anisotropy of Confining Field on the Structure of Dusty Plasma Clusters, *Technical Physics Letters*, 36(10), 910–913.
3. **Psakhie, S. G.**, Zolnikov, K. P., Abdrashitov, A. V. (2009). Studying the response of a Coulomb ball of charged dust particles to external pulsed loads &*Technical Physics Letters*, 35(2), 120–122.
4. **Psakhie, S. G.**, Zolnikov, K. P., Skorentsev, L. F., Kryzhevich, D. S., Abdrashitov, A. V. (2008). Structural features of bicomponent dust Coulomb balls formed by the superposition of fields of different origin in plasma, *Physics of Plasmas,* 15, 053701.
5. **Psakhie, S. G.**, Zolnikov, K. P. (2008). Structure of binary dust Coulomb balls in confining fields of different origin, *Physical Mesomechanics*, 11(3-4), 144–148.
6. **Psakhie, S. G.**, Zolnikov, K. P., Skorentsev, L. F., Kryzhevich, D. S., Abdrashitov, A. V. (2008). Structural features of two-component dusty plasma Coulomb balls, *Technical Physics Letters*, 34(4), 319–322.

Biomechanical and Tribological Aspects of Orthopaedic Implants

Irena Gotman

Abstract Orthopaedic and dental implant treatments have allowed to enhance the quality of life of millions of patients. Total hip/knee arthroplasty is a surgical replacement of the hip/knee joint with an artificial prosthesis. The aim of joint replacement surgery is to relieve pain improve function, often for sufferers of osteoarthritis, which affects around a third of people aged over fifty. Nowadays, total hip and knee replacement (THR) surgeries are considered routine procedures with generally excellent outcomes. Given the increasing life expectancy of the world population, however, many patients will require revision or removal of the artificial joint during their lifetime. The most common cause of failure of hip and knee replacements is mechanical instability secondary to wear of the articulating components. Thus, tribological and biomechanical aspects of joint arthroplasty are of specific interest in addressing the needs of younger, more active patients. The most significant improvements in the longevity of artificial joints have been achieved through the introduction of more wear resistant bearing surfaces. These innovations, however, brought about new tribocorrosion phenomena, such as fretting corrosion at the modular junctions of hip implants. Stiffness mismatch between the prosthesis components, non-physiological stress transfer and uneven implant-bone stress distribution are all involved in premature failure of hip arthroplasty. The development of more durable hip and knee prostheses requires a comprehensive understanding of biomechanics and tribocorrosion of implant materials. Some of these insights can also be applied to the design and development of dental implants.

Keywords Total joint replacements · Low friction arthroplasty · Stress shielding · Wear of polyethylene · Hard-on-hard articulations · Fretting wear · Tribocorrosion

I. Gotman (✉)
Department of Mechanical Engineering, ORT Braude College, Karmiel, Israel
e-mail: irenag@braude.ac.il

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_2

1 Introduction to Orthopaedic and Dental Implantable Devices

Orthopaedic and dental implants are surgical components that replace or interface with the bone. The most important implantable orthopaedic devices are total joint replacements, primarily total hip (THR) and total knee replacements (TKR), Fig. 1(left).

The aim of joint replacement surgical procedure (arthroplasty) is to relieve pain, improve function, and enhance quality of life, often for sufferers of osteoarthritis, which affects around a third of people aged over 50. Worldwide, more than one million THR surgeries and about two million TKR surgeries are performed every year, approximately 50% of which are done in the US. The rapidly aging population and a high prevalence of degenerative bone conditions in the elderly drive the demand for joint replacements even higher. According to American Academy of Orthopaedic Surgeons, the number of primary THRs in the US is projected to reach 635,000 in the year 2030 (171% increase vs. year 2014) and 1.23 million in the year 2060 (330% increase). Similarly, the projections for primary TKR are 1.28 million in the year 2030 (189% increase vs. year 2014) and 60 million in the year 2060 (382% increase).

Total joint arthroplasty is considered one of the most successful surgical interventions performed today. The reported survivorship after 15 years is above 90% for total hip replacements [1] and ranges between 82 and 98% for total knee replacements [2]. Despite this success, failures of joint replacements do occur, in which case the patient is required to undergo a revision surgery to replace the failed implant. In fact, the number of revision procedures is increasing faster than the number of primary arthroplasties. The projections of American Academy of Orthopaedic Surgeons for the US are 72,000 revision THRs (142% increase vs. year 2014) and 120,000 revision TKRs (190% increase) in the year 2030 and 110,000 revision THRs (219% increase vs. year 2014) and 253,000 revision TKRs (400% rise) in the year 2060. The rising rates

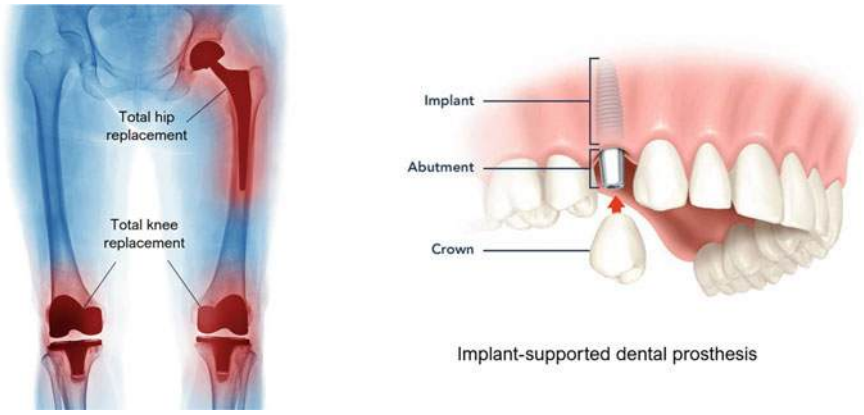


Fig. 1 Total joint replacements (left) and implant-supported dental prosthesis (right)

of failure are not surprising given the fact that patients currently receiving THR/TKR are ~20% heavier, more physically active, and live more than 25% longer compared with several decades ago. According to the latest study, approximately 42% of THRs will not last more than 25 years [3, 4]. This means that some artificial joints will need to be replaced at least once during a patient's lifetime. For younger patients (under 60 years of age), the lifetime risk of revision is around 30%, with the highest revision rates reported for males between the ages of 50–55 [5]. These clinical data indicate that the ultimate goal of joint arthroplasty—long-term pain-free function for the rest of the patient's life—has not yet been achieved.

Another important class of bone-interfacing prostheses are dental implants, Fig. 1(right). A dental (or endosseous) implant is a surgical replacement of the tooth root that interfaces with the with the bone of the jaw or skull to support a dental prosthetic (e.g. crown or bridge). The main objective of implant supported dental restoration is the permanent replacement of missing teeth. Every year, more than 800,000 dental implants are placed in the United States and more than 1.8 million in the European Union. As population is aging, the demand for dental implants will rise significantly in the next decade. Projection models suggest that among the US adults missing teeth, the prevalence of dental implant use could reach as high as 23% by the year 2026 [6]. Dental implants are designed to last a lifetime and have, according to different sources, a success rate of over 90% at 10–15 years follow-up [7]. However, failures do happen, especially in medically compromised patient population. It has been suggested that failure patterns and mechanisms behind bone loss around dental implants have much in common with joint replacements, and that biomechanical under- or overloading and synergy between friction, wear and corrosion are responsible for the majority of dental and orthopaedic implant system failures [8–10]. Therefore, a total hip replacement will be used throughout this paper as it is a representative example of an orthopaedic implant and because the knowledge from the discipline of orthopedics can be applied to oral implants.

2 Tribology of Total Hip Replacement

2.1 Charnley Low Friction Arthroplasty

The hip is one of the body's largest weight-bearing joints. This geometrically simple "ball-in-socket" joint consists of the head (top of the thigh bone, or femur) articulating inside the acetabular socket of the pelvis. The layer of articular cartilage covering the bone surfaces lubricated by the synovial fluid provides the joint with exceptional tribological properties. Total hip arthroplasty consists of replacing both the acetabulum and the femoral head with artificial components. The first total hip replacements were performed by Wiles (1938) and McKee (1951) [11]. In those early designs, both bearing surfaces (acetabular cup and femoral head) were made of stainless steel and were fixed to the bone with screws and bolts. These historical implants experienced

high incidence of early failure associated with the component loosening, typically of the acetabular cup. The unsatisfactory clinical performance was primarily due to elevated friction, jamming and wear within the bearings. It turned out that the main limitation of the early THR designs was that they mimicked normal hip joint anatomy. Large femoral heads coupled with inconsistent manufacturing tolerances generated high frictional torque (turning force) on the articulating surfaces leading to high shear stresses and loosening at the acetabular cup-bone interface.

Realization of the tribological nature of failure of early hip replacement designs brought Sir John Charnley to introduce his revolutionary concept of "low-friction arthroplasty" (LFA) [12]. LFA follows the principle of low-frictional torque based on the largest possible difference between the radius of the femoral head and that of the outer aspect of the acetabular component [13]. Charnley and his colleagues concluded that in order to minimize frictional torque and protect the cup-bone interface, the head diameter should be not greater than half of the external diameter of the cup. Consequently, the head diameter of THR was reduced from the earlier used 41.5–22.2 mm, less than half anatomical femoral head diameter (48–55 mm on the average). Charnley also recognized that in addition to the low frictional torque design, it is important that the acetabular cup material had a low friction coefficient against the material of the femoral head (stainless steel in Charnley prosthesis). The first material used was a self-lubricating polymer polytetrafluoroethylene (PTFE). PTFE sockets, however, wore out disappointingly fast causing "intense foreign body reaction" to wear debris and gross destruction of bone. The next polymeric material used—ultra-high molecular weight polyethylene (UHMWPE) [14] proved much more successful. UHMWPE had excellent wear resistance, low friction and high impact strength, and no problems were observed with metal-on-UHMWPE (M-PE) bearings in the early years post-implantation. Thus, the biomechanical concept of low friction arthroplasty combined with the use of a low-wear acetabular cup material (UHMW polyethylene) started a new era in joint replacement surgery. Very soon, a more biocompatible and corrosion resistant cobalt-chromium alloy (CoCr) came to replace stainless steel in the femoral component of Charnley prostheses. From then on, arthroplasty has known considerable evolution, but metal-on-UHMWPE (CoCr-PE) articulation remains the gold standard for artificial hips and other artificial joints, including the knee and shoulder. It is definitely Sir John Charnley to be credited with advancing our understanding of tribological effects as they apply in orthopaedics, and the significance of friction, wear and lubrication of implant materials for their longevity and function, and particularly the body's reaction to the particulate debris produced as a result of implant wear [15].

Another breakthrough made by John Charnley was the introduction of a self-curing acrylic resin (bone cement) as a grouting agent to secure the implant components to bone. The cement is injected as a dough-like mass and hardens around the implant to ensure its anchorage in the bone [16]. The use of bone cement allowed for the firm fixation of hip replacements unachievable with the previously used screws and bolts. Charnley cemented THA rapidly gained widespread popularity and became one of the most successful orthopaedic procedures with reported survivorship rates greater than 90% at 15–20 years. In young, physically active patients, however, failure

rates were significantly higher. The main mode of failure was aseptic loosening (loss of fixation) of one or both implant components, secondary to periprosthetic osteolysis—resorption of bone surrounding the implant [17, 18]. At the time, osteolysis was thought to be caused by biological reaction to bone cement described as "cement disease". This led to the erroneous conclusion that the problem of aseptic loosening can be solved if the use of bone cement is avoided. Consequently, the innovative concept of cementless fixation was developed. In the cementless approach, implant components are stabilized within the bone by bone ingrowth into the porous surface layer or by bone ongrowth onto the textured surface. The material of choice for cementless prostheses are titanium alloys (mostly Ti6Al4V) due to their superior biomechanical compatibility: Ti6Al4V is capable of osseointegration (establishing direct contact with bone) and has a low modulus of elasticity, half that of CoCr alloy (110 vs. ~230 GPa) [19]. The latter is important for minimizing periprosthetic bone resorption due to stress shielding caused by stiffness mismatch between the implant and the bone (elastic modulus of ~20 GPa) [20]. Against early expectations, cementless fixation did not eliminate the problem of aseptic loosening of hip replacements and the outcomes were no better than with cemented THA.

2.2 *Wear of Polyethylene—The Main Culprit of Aseptic Loosening*

Once it became clear that not "cement disease" was the problem, the proposition that aseptic loosening is related to particulate wear debris from the UHMWPE acetabular component was put forward. The hard metal femoral head can produce wear of the polyethylene surface during articulation through both abrasive and adhesive mechanisms. Positive asperities on the hard counterface can abrade the polyethylene surface, which is relatively softer, producing abrasive wear debris. Friction between the articulating surfaces shears off particles, producing adhesive wear debris. Hard particles present in the joint space (cement, metal, bone) can enter between the articulating surfaces, embed in the polyethylene and abrade the metallic counterface—"third body" wear mechanism. In total knee replacements, the dominant form of wear is delamination of polyethylene which occurs as a result of cyclic compressive-tensile loading that leads to subsurface cracking. Wear particles migrate into tissues and are phagocytosed by macrophages which become activated and release pro-inflammatory cytokines that stimulate bone resorption (osteolysis) around the implant leading to prosthesis loosening [21–23]. Typically, billions of submicron UHMWPE wear particles (average diameter of 0.3–0.5 μm) per year are released into periprosthetic fluids. Two main factors affecting the volumetric wear of UHMW polyethylene are diameter and material of the femoral head that articulates against the polyethylene. CoCr alloy is a metal traditionally used for the femoral head. Meanwhile, titanium alloys exhibit poor tribological behavior under abrasive and adhesive wear and should not be used for manufacturing femoral heads. Titanium is much softer than CoCr and is

easily scratched by hard "third body" particles that intrude between the articulating surfaces which results in increased friction and abrasive wear of the polyethylene. For CoCr heads articulating against UHMWPE, larger heads are associated with greater volumetric wear of polyethylene and high revision rates. The popular 28 mm diameter is a compromise between wear performance and risk for dislocation of the implant.

3 Alternative Bearing Surfaces

Since the inflammatory response to wear debris was established as the main cause of aseptic loosening, efforts at extending joint replacement longevity have focused primarily on development of more wear resistant bearing surfaces. Two major directions included (*i*) improving the quality of UHMWPE and (*ii*) avoiding the use of polyethylene bearing altogether. Alternatively, attempts were made to improve the wear resistance of metallic components by providing them with a hard, wear resistant surface.

Prior to being introduced into clinical practice, new artificial joint materials must be submitted to realistic preclinical tests. The tribological performance of novel THA bearing couples is tested in hip joint simulators designed to mimic the biomechanics of hip joint in a simulated physiological environment. Despite the reported discrepancies between in vitro simulation results and wear data from explanted devices, joint simulators are instrumental in predicting clinical wear performance of new bearing surfaces and identifying the risk of clinically relevant wear [24].

3.1 Cross-Linked Polyethylene

Radiation cross-linking significantly decreased the wear rate of UHMWPE against CoCr in simulation studies, Fig. 2.

In total hip arthroplasty, this has translated into better long-term outcomes and a significant reduction in the rate of revision for younger patients [25]. Furthermore, highly cross-linked polyethylene may allow use of large-diameter femoral heads without concern about increased polyethylene wear [26]. The biomechanical rationale for using large-diameter femoral heads is that they allow for a greater range of motion and limit the risk of dislocation by increasing jump distance—distance the head has to "jump" before leaving the acetabular cup [27]. For total knee arthroplasty, however, clinical evidence has been inconclusive, and no distinctive improvement was observed when highly cross-linked polyethylene (HXLPE) was used compared to conventional UHMWPE [28–30]. This could be attributed to distinct biomechanical environments and different relative contributions of polyethylene wear mechanisms in the two types of joints: adhesive and abrasive wear in the hip versus fatigue wear and delamination in the knee. In addition, the reduced fatigue strength and toughness

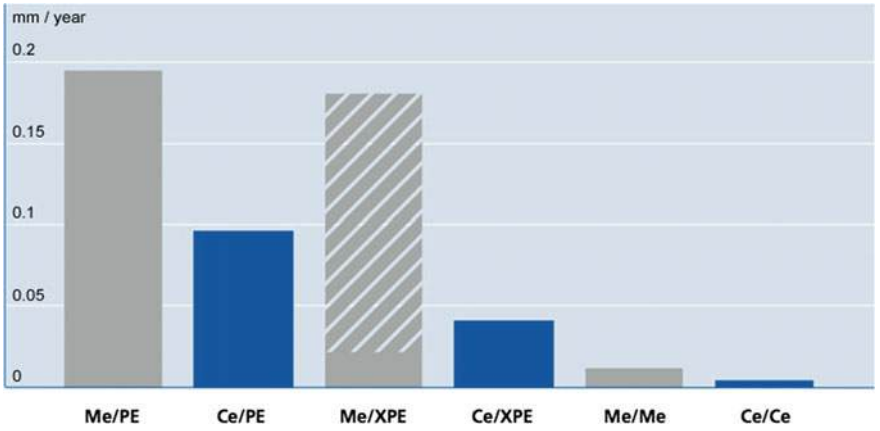


Fig. 2 In vivo linear wear rates of different bearing couples.: Me—metal, PU—UHMW polyethylene, Ce—ceramic, XPE—cross-linked polyethylene. https://www.ceramtec.com/biolox/clinical-experience/wear-osteolysis/

of cross-linked polyethylene may cause the cracking of HXLPE knee replacement components.

3.2 Ceramic-on-Polyethylene Articulation

The wear of polyethylene acetabular components has been further reduced by replacing the femoral head material from the CoCr metal to a ceramic. The first ceramic used was aluminum oxide (alumina). Alumina is oxidation resistant, biocompatible, very hard and scratch resistant. Furthermore, the fine grain structure allows alumina to be polished to a very low surface roughness ($Ra < 0.05\ \mu m$) resulting in reduced wear of the polyethylene countersurface. In addition, the high wettability of alumina positively affects the lubricating film thus decreasing the coefficient of friction [31]. Due to the brittle nature of alumina, however, a few (but potentially devastating) fracture failures of the early ceramic heads (Biolox® forte, CeramTec, GmbH) were reported. To reduce the risk of brittle fracture of femoral heads, an alternative ceramic material—yttria stabilized zirconia (Y-TZP) was introduced. Y-TZP is a metastable material that exhibits an extremely high (for a ceramic) fracture toughness due to a unique transformation toughening mechanism. High fracture toughness combined with excellent tribological behavior against polyethylene made zirconia femoral heads very popular in the last decade of the past century. It turned out, however, that when exposed to body fluids, the metastable tetragonal phase may transform to the stable monoclinic structure [32, 33]. This aging process occurs in vivo on the surface of zirconia heads, leading to their roughening and microcracking. The problem became apparent in the year 2000 when an unusually large

amount of failures was reported following a change in the manufacturing process of zirconia heads. In 2001, the company St. Gobain Desmarquest (Vincennes, France) issued a voluntary recall, and the use of zirconia femoral heads in hip arthroplasty came to an end.

An important step towards enhancing the resistance of ceramic femoral heads to brittle fracture was the introduction of zirconia-toughened alumina—a composite ceramic (Biolox delta) whose fracture toughness is twice that of Biolox forte [34]. The volumetric wear of polyethylene acetabular cups articulating against ceramic femoral heads is several times lower than that of their metal-on-polyethylene counterparts, Fig. 2. The absence of allergic reaction to alumina makes ceramic-on-polyethylene articulation especially suitable for patients suffering from immune hypersensitivity to metals such as nickel, chromium and cobalt.

3.3 Hard-on-Hard Articulations

Notwithstanding the improved wear resistance of cross-linked polyethylene, the most spectacular reductions in volumetric wear of articulating joints (by one-two orders of magnitude) are achieved with hard-on-hard (metal-on-metal and ceramic-on-ceramic) bearings, eliminating altogether the soft polyethylene component, Fig. 2.

3.3.1 Ceramic-on-Ceramic Bearings

Ceramic-on-ceramic (CoC) articulations exhibit by far the lowest wear rates and are considered a viable option for young, active patients. Excellent mid-term clinical outcomes are reported for modern CoC hip replacements using zirconia-toughened alumina (Biolox delta), and component fractures are extremely rare [35, 36]. One well-recognized and annoying complication of ceramic-on-ceramic THA is squeaking—a high pitched, audible sound that occurs during movement, often related to a specific activity [37, 38]. The reported incidence of squeaking in CoC THA lies between 0.5% and >20%. A likely cause of squeaking is adverse tribological conditions caused by the loss of fluid film lubrication and high friction between the ceramic components. Friction generates forced vibrations that cause the metallic parts to resonate and convert vibrational energy into an audible noise. The incidence of squeaking is strongly affected by implant- and patient-specific factors.

3.3.2 Metal-on-Metal Bearings

Metal-on-metal (MoM) total hip arthroplasty almost totally abandoned in the mid-1970's in favor of Charnley's metal-on-polyethylene THA, made its comeback in the very beginning of the twenty-first century [39–41]. By that time, it became clear that

the failures of fist generation metal-on-metal implants were not due to the bearing surface material but were mainly caused by design errors and inadequate manufacturing. Survivorship analysis of hip replacements implanted between 1965 and 1973 revealed a surprisingly great longevity among some of the original all-metal designs. At long-term follow-up, the wear of the long-lived metal-on-metal McKee–Farrar prostheses was by at least one order of magnitude smaller compared to the metal-on-polyethylene Charnley prostheses. Novel CoCr-on-CoCr devices designed with the standard small femoral head diameter (22 and 28 mm) exhibited very low wear rates in hip simulator tests, only slightly higher than those of ceramic-on-ceramic couples. At the same time, they possessed an obvious advantage of not being brittle. An important observation from the simulation tests was that more effective fluid film lubrication and correspondingly low wear rates were achieved with larger diameter femoral heads. Based on these results, large diameter ($\geq$36 mm) MoM articulations were developed and quickly gained a big share of the market, both in the USA (approx. 30% in 2006–2007) and worldwide. It took only a few years to realize that MoM THA was associated with higher revision rates and lower patient satisfaction. In addition to the well-known phenomenon of aseptic loosening), a new mode of failure was observed—adverse local tissue reaction (ALTR) [42, 43]. ALTR included periprosthetic soft tissue inflammation, soft tissue necrosis, and pseudotumor formation. Some patients were asymptomatic but those presenting with pain and elevated metal blood levels had to be revised.

The final blow for metal-on-metal designs came in 2010 when DePuy, J&J's orthopaedic branch voluntarily recalled its ASR MoM hip system due to an unacceptably high failure rate (~13% after 5 years) [44]. The use of metal-on-metal devices declined rapidly to less than 1% of all the THR systems being implanted today. The analysis of failed implants revealed that wear particles from MoM articulations are approximately 50 nm in size, much smaller than the 0.3–0.5 μm UHMWPE debris particles. Thus, despite the low volumetric wear of MoM bearings, the actual number of released particles is considerably higher than for conventional metal-on-PE bearings. Moreover, these metallic nanoparticles are more biologically active and corrode rapidly in the body fluids releasing large amounts of potentially toxic cobalt and chromium ions. The unfavorable outcomes of large diameter metal-on-metal hip replacements are the result of their complex and not well-understood tribology. From the biomechanical and tribological point of view, MoM articulations were found to be extremely unforgiving to positioning and manufacturing mistakes: slight deviations from the optimal alignment, sphericity and radial clearance could lead to adverse lubrication conditions and excessive wear. To date, there are no FDA-approved metal-on-metal total hip replacement devices marketed for use in the US, the only available options being ceramic-on-ceramic, ceramic-on-polyethylene and metal-on-polyethylene bearings.

4 Bearing Materials in Total Knee Replacement

In total knee arthroplasty (TKR), bearing surface options have been much more limited. Practically all state-of-the-art knee replacements use a CoCr alloy femoral component articulating on polyethylene. The knee joint has a complex nonconformal geometry and is subjected to high contact stresses. Polyethylene is sufficiently compliant to accommodate stress concentration caused by misalignment or surface-to-surface contact of asperities. The rigid nature of ceramics makes ceramic-on-ceramic articulation much less forgiving of surface irregularities and slight malposition thus leading to increased risk of brittle fracture. Therefore, transferring the benefits of excellent tribological properties of ceramics to the complex geometry of knee prostheses remains challenging. At the present time, all-ceramic knee endoprosthesis is not a feasible option. Several TKA designs having a ceramic femoral component articulating on polyethylene are available for clinical use or are under clinical trial [45–47]. These devices represent a promising alternative for patients with a known hypersensitivity to metals, but it is still early to draw conclusions regarding their long-term outcomes in terms of longevity, wear damage and incidence of brittle fracture.

5 Surface-Modified Bearing Materials

The desired alternative to existing articulating materials for joint implants would combine the fracture toughness of metals with the wear performance of ceramics [48]. One approach for achieving this is to deposit or overlay a ceramic coating onto a metallic substrate. The bond created between the deposited coating and the substrate is only physical (rather than chemical) resulting in relatively weak coating adhesion. Given the significantly different mechanical properties of the ceramic film and the underlying metal, adhesive failure between the two materials occur under load or during articulation. Diamond like carbon (DLC) and titanium nitride (TiN) are the most extensively studied wear resistant coatings for artificial joints. Despite their high hardness and excellent biocompatibility, delamination and spalling of such coatings has been observed in clinical trials and some wear simulation tests. Insufficient adhesion and inadequate load bearing capacity of the underlying softer metallic substrate are believed to be the major obstacles on the way to successful implementation of hard coatings into clinical practice [48]. It is believed that these shortcomings of externally applied ceramic layers can be alleviated by diffusional surface hardening—reactively diffusing a non-metallic element into the substrate at elevated temperatures thus transforming the surface from metal to ceramic.

Oxidized zirconium (OXINIUM, Smith & Nephew Orthopaedics) was developed for orthopedic applications to provide improvements over CoCr alloy for resistance to roughening, frictional behavior, and biocompatibility without the mechanical limitations of brittle monolithic ceramics [49]. The ceramic surface is formed

by heating a zirconium alloy in air to allow oxygen to diffuse into the substrate and to transform the metal surface to zirconium oxide (zirconia) ceramic. Despite the consistently lower wear of polyethylene components articulating against the ceramic surfaced OXINIUM in knee and hip joint simulator tests, clinical studies have shown no statistically significant differences in mid-term implant survivorship between OXINIUM and CoCr components [50–52]. Additional research is needed for the clinical performance of OXINIUM to better understand long-term outcomes. Meanwhile, OXINIUM contains no detectable nickel or chromium, which makes such implants a safer choice for patients with metal allergies.

6 Fretting Wear Damage of Total Joint Replacements

6.1 Modular Connections of Hip Prostheses

In early hip replacement devices, the femoral stem and head were produced as a single-piece, monolithic component—a so-called monobloc design, Fig. 3a. Nowadays, almost all hip joints are modular and consist of a separate femoral head that fits on the stem, Fig. 3b, c. The reliable joining of modular components of total joint replacements is based on the concept of a Morse taper, i.e. that of the cone in the cone [53, 54]. The two components of the Morse taper form a firm fit that relies on friction and mechanical interlocking. Modularity provides many advantages, such as greater intraoperative flexibility allowing the surgeon to restore the patient's anatomy and

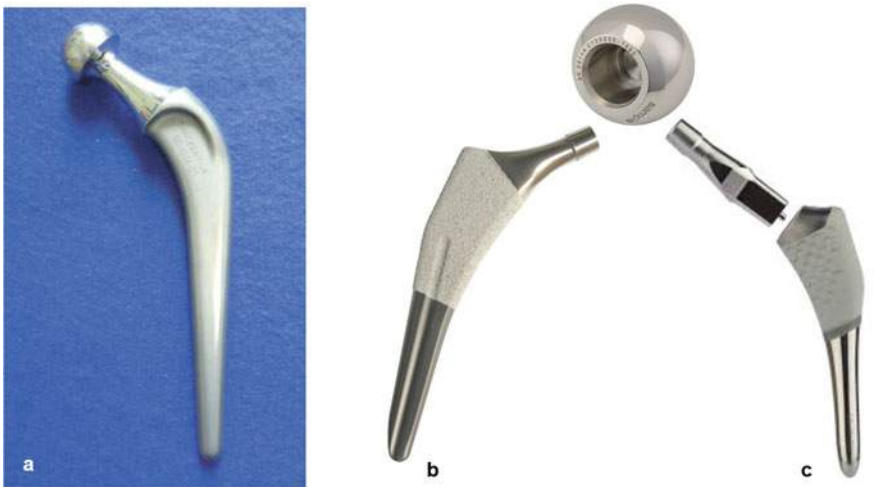


Fig. 3 Different designs of femoral components of total hip replacement: monoblock (Charnley) (**a**) and modular (**b**, **c**): b—monolithic stem, c—dual-taper stem

to adjust leg length, decreased implant inventory and potential ease of revision by exchanging only the failed component. Furthermore, modularity allows the combination of head and stem made of different materials with specific properties thus optimizing the clinical performance of the whole assembly. For example, a stem from titanium alloy that is most suitable for cementless fixation but has an inadequate wear resistance can be combined with a hard, wear resistant Co alloy or ceramic head.

Despite its benefits, the modular design has been associated with higher revision rates due to adverse tissue reaction, neck fracture and femoral head disassociation. Modularity creates additional mechanical junctions (neck-head and neck stem interfaces having a crevice-like geometry) that become weak points where micromotion and wear can occur [55]. The hip joint is subject to cyclic stresses from gait loading amounting to more than one million cycles a year. As a result of cyclic loading, a low amplitude oscillating relative motion occurs at the taper junction of femoral components made of dissimilar materials and having different rigidity. This leads to the tribological process of fretting causing surface damage of the fitting contact surfaces. The process is often referred to as "mechanically assisted crevice corrosion"(MACC) and can be briefly described as follows [56–59], Fig. 4a–d. Rubbing between the taper surfaces under stress leads to mechanical disruption of the protective oxide film and corrosion followed by rapid regeneration of the oxide layer (repassivation). This is accompanied by oxygen consumption, metal ion release and hydrolysis, and voltage drop. As the mechanical damage to the oxide is continuously repeated, oxygen in the crevice is depleted while the liberated hydrogen ions acidify the fluid to the point where repassivation becomes impossible. Given that the corrosion resistance

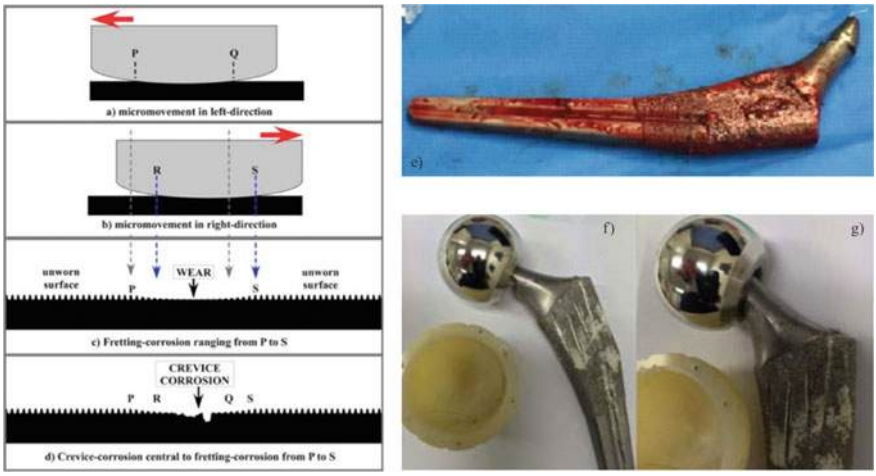


Fig. 4 **a**–**d** Schematic of mechanically assisted crevice corrosion (MACC); **e** retrieved Accolade stem with marked fretting wear of the neck taper; **f** retrieved Accolade stem with instability and dissociation of the head-neck junction. Reproduced from P. Walker et al., Reconstructive Review 2016, September; 6(3):13–18

of both cobalt and titanium alloys relies on passivity, the inability to rebuild the passive oxide film results in active corrosion. An additional feature of MACC is hydrogen gas accumulation below the modular neck and hydrogen embrittlement on the surface of titanium components [60, 61]. MACC at the femoral head-neck interface of modular hip replacements is commonly referred to as trunnionosis.

Wear debris and metal ions produced at taper junctions can lead to elevated blood metal ion concentrations and adverse local tissue reactions. The histological appearance of periprosthetic tissues surrounding corroded trunnions is similar to tissues surrounding failed hip replacements with metal-on-metal bearings. Some patients will remain asymptomatic; others will develop adverse clinical symptoms that require revision: necrosis, pseudotumors, pain, etc. In heavy patients, cases of catastrophic fatigue fracture of titanium male stem tapers initiated at notch-like irregularities of the fretted surface were reported [62, 63].

Several modular hip systems are available on the market, differing in design and femoral stem/neck/head material. As discussed above, the head is typically made of the wear-resistant cobalt-chromium alloy or ceramic whereas the stem can be either cobalt-chromium or titanium alloy. The stem can be monolithic (one-piece design), Fig. 3b, or can have an exchangeable neck (dual-taper stem), Fig. 3c. In the latter case, an additional stem/neck interface is introduced which only increases the concern regarding the occurrence of mechanically assisted crevice corrosion. Ti alloys are the common material choice for cementless stems due to their superior osseoconductivity and low stiffness preventing bone resorption secondary to stress-shielding. In this respect, beta-titanium alloys whose elastic modulus is significantly lower than that of the standard Ti-6Al-4V alloy are most favorable candidate materials [64]. However, combining beta-titanium with the high-modulus Co-Cr alloy components has proven disastrous and led to several major Hip Replacement recalls. The beta-titanium involved is a proprietary Ti-12Mo-6Zr-2Fe (TMZF®) alloy having the elastic modulus of around 75 GPa that was developed by Stryker Orthopaedics.

When a monolithic TMZF stem (Accolade I stem design) was used in combination with a Co-Cr alloy head (LFIT V40) supposed to minimize the risk of dislocation, high incidence of failures due to taper wear and adverse local tissue reaction was reported [65–68]. The patients experienced pain and restricted motion requiring a revision surgery. In some cases, tapers were damaged to such a significant level that head dissociation from the stem occurred. Examples of such gross stem taper failure (GTF) and head disassociation are shown in Fig. 4e–g. It is hypothesized that micromotion and fretting corrosion at the taper junction results in the widening of the gap between the head and the neck which allows the head to turn and move on the stem. The harder Co-Cr head abrades the softer titanium alloy neck to such an extent that the head can easily disassociate from the stem. Importantly, GTF has occurred recurrently with stems made of the low-modulus TMZF alloy, very rarely with stems from the standard Ti-6Al-4V alloy and never with the high modulus Co-Cr alloy stems suggesting the influence of the material stiffness. Indeed, numerical modelling has shown that the deformation and micromotion at the Co-Cr head-stem taper interface was significantly larger for the TMZF-alloy stem compared to the CoCr and even to Ti-6Al-4V alloy stems. The phenomenon of head disassociation

was most often observed with large diameter Co-Cr heads (36 mm and larger) and these were voluntarily recalled by the manufacturer in 2016. Other femoral head sizes, as well as ceramic heads remain on the market however problems with the device continue to be of concern as failures in implant sizes outside of the recall are being reported. The Accolade I stem made of the low-modulus TMZF beta titanium alloy was never recalled but its use declined rapidly. In 2012 Stryker replaced Accolade I stem with the standard Ti-6Al-4V alloy stem (Accolade II).

The most well-known case of tribocorrosion of modular hip replacements is the failure of dual-taper Rejuvenate and ABG II stems launched by Stryker Orthopaedics in 2009 [69–72]. Both designs combined a low-modulus TMZF femoral stem and an exchangeable cobalt-chromium alloy neck. Since Co-Cr is harder and stiffer than titanium, it was suggested that this would allow safer and long-term use of the modular neck. The femoral heads were either Co-Cr or ceramic articulating on a UHMW polyethylene acetabular cup. Extremely high revisions rates secondary to tribocorrosion at the taper connection were reported for both designs, reaching, for Rejuvenate stem, 65% three years post-implantation. Due to these unacceptable failure rates and the ensuing FDA investigation, Stryker was forced to issue a voluntary recall of both products in 2012. Similarly to the case of Accolade II, the culprit in the failure of the dual-taper mixed-metal stem was the low elastic modulus of the TMZF alloy. Following the recall of the Rejuvenate and ABG II, Stryker discontinued the use of the low-stiffness beta-titanium and replaced it with the standard Ti-6Al-4V alloy. Current recommendations regarding modular hip replacements include avoiding femoral stems with low flexural rigidity and reducing the number of modular junctions, e.g. by using fixed neck stems. Also, substituting Co-Cr alloy heads with the chemically inert ceramic heads is expected to reduce MACC since corrosion processes will only occur on the metallic stem taper and will not be accelerated by galvanic coupling with a dissimilar metal [73, 74]. From the biomechanical point of view, smaller diameter heads are typically reported to produce less fretting damage since they generate smaller head-neck moment arm and correspondingly smaller torsional forces at the head-neck taper junction [75, 76]. The use of small diameter heads (36 mm and less), however, is associated with an increased risk of dislocation of hip replacement and is not willingly accepted by orthopaedic surgeons.

It follows from the above discussion that even if all the guidelines are followed, tribocorrosion at modular junctions of orthopaedic and dental implants exposed to cyclic loading cannot be fully eliminated. To overcome the problem of head-neck taper degradation, different surface engineering approaches are being investigated. One possibility is to coat a titanium alloy stem taper with a hard, wear-resistant film. The results of in vitro evaluation of TiN and TiN/AlN coatings suggest that these coatings provide superior fretting and fretting corrosion resistance to the tapered interfaces of the Co-Cr-Mo and Ti-6Al-4V alloy components [77–79]. Increasing the interfacial bond strength between the coating and the substrate could improve the fretting and corrosion resistance even more.

6.2 Stem-Cement Interface

Another joint replacement zone prone to tribocorrosion is the stem-cement interface of cemented prostheses [80–83]. Under physiological loading, this interface experiences a low amplitude oscillatory micromotion. During relative sliding, hard radiopacifier particles (e.g., ZrO_2) within the cement abrade the polished surface of the femoral stem and induce a tribocorrosive interaction. The effect is most pronounced for titanium alloy stems that experience larger flexural deflections and are more easily abraded than the stiffer and harder CoCr stems [84]. Fretting wear damage results in the formation of gaps/crevices between the cement mantle and the titanium stem leading to crevice corrosion of the metallic surface. Both the surface damage and the immunological reaction to released particles and ions compromise the stem stability and may lead to premature failure of the cemented joint prosthesis. Similarly to the taper junctions of modular implants, the low elastic modulus of titanium here is a drawback rather than an advantage. Flexural deflections of femoral stems lead to the cracking of the cement mantle and debonding at cement-stem interface, and both phenomena are much more pronounced for the low-stiffness titanium stems. These biomechanical and tribocorrosion problems make Ti alloy stems a much less popular option for cemented hip replacements. Although cemented Ti alloy stems are still available on the market, it is believed by many that the use of titanium stems in cemented THA should be abandoned [85].

7 Tribocorrosion in Dental Implants

Despite high success rates, 5–11% of implant-supported dental restorations fail within 10–15 years and must be removed [86]. The dominant failure mode of dental implants is peri-implantitis (inflammation of tissues surrounding the implant) and the associated loss of supporting bone. Tribocorrosion at internal connections between the prosthesis parts (implant, abutment and crown) and at implant-bone interface is among the major contributors to peri-implantitis [87–90]. The implant and the abutment are made of titanium or titanium alloy, whereas the crown is usually made of porcelain or zirconia ceramic. Despite very precise machining, there always is a microgap between the implant and the prosthetic connector. During mastication, micromotion occurs at the implant-abutment and abutment-crown interface leading to fretting and crevice corrosion. The situation is obviously similar to mechanically assisted crevice corrosion at taper connections of total hip replacements. The released degradation products (metal ions and metal oxide particles) initiate inflammatory tissue response that can eventually result in peri-implant bone loss. Furthermore, material loss enlarges the microgap between the abutment and the implant allowing for rapid gap colonization by oral microorganisms and subsequent bacterial infection. The combined action of microbiological, mechanical and tribocorrosion factors

promotes peri-implantitis and the associated bone loss eventually leading to implant failure.

8 Summary

With the growing demand for orthopaedic and dental implants and expectations of longer device lifetimes for the younger patients, wear and corrosion of articulating surfaces and modular junctions of these implantable devices are a prime concern. Aside from the patient's activity and physiological state, the biotribological performance of a prosthesis depends on the mechanical design and the materials used (metals, ceramics, polymers). Despite the substantial improvements achieved in both directions, the ultimate goal of total joint replacements—long-term pain-free function for the rest of the patient's life—has not yet been achieved. Currently there are gaps in our understanding of biomechanics and wear behavior of medical implants. Closing these gaps will help guide future research in this field and improve the longevity of orthopaedic and dental implantable devices.

References

1. Ferguson RJ, Palmer AJR, Taylor A, Porter ML, Malchau H, Glyn-Jones S (2018) Hip replacement. Lancet 392:1662–1671
2. Price AJ, Alvand A, Troelsen A, Katz JN, Hooper G, Gray A, Carr A, Beard D (2018) Knee replacement. Lancet 392:1672–1682
3. Cook R, Davidson P, Martin R (on behalf of NIHR Dissemination Centre) (2019) More than 50% of hip replacements appear to last 25 years. Journal BMJ 367:l5681
4. Evans JT, Evans JP, Walker RW, Blom AW, Whitehouse MR, Sayers A (2019) How long does a hip replacement last? A systematic review and meta-analysis of case series and national registry reports with more than 15 years of follow-up. Lancet 393:647–654
5. Bayliss LE, Culliford D, Monk AP, Glyn-Jones S, Prieto-Alhambra D, Judge A, Cooper C, Carr AJ, Arden NK, Beard DJ, Price AJ (2017) The effect of patient age at intervention on risk of implant revision after total replacement of the hip or knee: a population-based cohort study. Lancet 389:1424–1430
6. Elani HW, Starr JR, Da Silva JD, Gallucci GO (2018) Trends in dental implant use in the U.S., 1999–2016, and projections to 2026. J Dental Res 97(13):1424–1430
7. Moraschini V, da C. Poubel LA, Ferreira VF, dos SP Barboza E (2015) Evaluation of survival and success rates of dental implants reported in longitudinal studies with a follow-up period of at least 10 years: a systematic review. Int J Oral Maxillofac Surg 44:377–388
8. Albrektsson T, Becker W, Coli P, Jemt T, Mölne J, Sennerby L (2019) Bone loss around oral and orthopedic implants: an immunologically based condition. Clin Implant Dentistry Related Res 21(4):786–795
9. Coli P, Sennerby L (2019) Is peri-implant probing causing over-diagnosis and over-treatment of dental implants? J Clin Med 8(8):1123. https://doi.org/10.3390/jcm8081123
10. Souza JCM, Henriques M, Teughels W, Ponthiaux P, Celis JP, Rocha LA (2015) Wear and corrosion interactions on titanium in oral environment: literature review. J Bio- Tribo-Corrosion 1(13). https://doi.org/10.1007/s40735-015-0013-0

11. Gomez PF, Morcuende JA (2005) Early attempts at hip arthroplasty–1700s to 1950s. Iowa Orthop J 25:25–29
12. Wroblewski BM, Siney PD, Fleming PA (2006) The Charnley Hip replacement—43 years of clinical success. Acta Chirurgiae Orthopaedicae et Traumatologiae Cechoslovaca 73(1):6–9
13. Wroblewski BM, Siney PD, Fleming PA (2009) The principle of low frictional torque in the Charnley total hip replacement. J Bone Joint Sur Br Vol 91-B:855–858
14. Sobieraj MC, Rimnac CM (2009) Ultra high molecular weight polyethylene: mechanics, morphology, and clinical behavior. J Mech Behav Biomed Mater 2(5):433–443
15. Sonntag R, Beckmann NA, Reinders J, Kretzer JP (2015) Materials for total joint arthroplasty—biotribology of potential bearings. Imperial College Press, London, pp 15–40
16. Vaishya R, Chauhan M, Vaish A (2013) Bone Cement. J Clin Orthop Trauma 4(4):157–163
17. Iannotti JP, Balderston RA, Booth RE, Rothmann RH, Cohn JC, Pickens G (1986) Aseptic loosening after total hip arthroplasty. J Arthroplasty 1:99–103
18. Sundfeldt M, Carlsson LV, Johansson CB, Thomsen P, Gretzer C (2006) Aseptic loosening, not only a question of wear: a review of different theories. J Acta Orthop 77(2):177–197
19. Geetha M, Singh AK, Asokamani R, Gogia AK (2009) Ti based biomaterials, the ultimate choice for orthopaedic implants—a review. Prog Mater Sci 54:397–425
20. Sumner DR (2015) Long-term implant fixation and stress-shielding in total hip replacement. J Biomech 48(5):797–800
21. Abu-Amer Y, Darwech I, Clohisy JC (2007) Aseptic loosening of total joint replacements: mechanisms underlying osteolysis and potential therapies. Arthritis Res Therapy 9(1):6
22. Harris WH (1994) Osteolysis and particle disease in hip replacement. A review. Acta Orthop Scand 65:113–123
23. Wooley PH, Schwarz EM (2004) Aseptic loosening. Gene Ther 11:402–407
24. Medley JB (2016) Can physical joint simulators be used to anticipate clinical wear problems of new joint replacement implants prior to market release? Proc Inst Mech Eng Part H J Eng Med 230(5):347–358
25. Dumbleton JH, D'Antonio JA, Manley MT, Capello WN, Wang A (2006) The basis for a second-generation highly cross-linked UHMWPE. Clin Orthop Relat Res 453:265–271
26. Allepuz A, Havelin L, Barber T, Sedrakyan A, Graves S, Bordini B, Hoeffel D, Cafri G, Paxton E (2014) Effect of femoral head size on metal-on-HXLPE hip arthroplasty outcome in a combined analysis of six national and regional registries. J Bone Joint Surg 96(1):12–18
27. Banerjee S, Pivec R, Issa K, Kapadia BH, Khanuja HS, Mont MA (2014) Large-diameter femoral heads in total hip arthroplasty: an evidence-based review. Am J Orthop (Belle Mead, N.J.) 43(11):506–512
28. Brown TA, van Citters DW, Berry DJ, Abdel M (2017) The use of highly crosslinked polyethylene in total knee arthroplasty. Bone Joint J 99B(8):996–1002
29. Wilhelm SK, Henrichsen JL, Siljander M, Moore D, Karadsheh M (2018) Polyethylene in total knee arthroplasty: where are we now? J Orthop Surg 23(6):1–7
30. de Steiger RN, Muratoglu O, Lorimer M, Cuthbert AR, Graves SE (2015) Lower prosthesis-specific 10-year revision rate with crosslinked than with non-crosslinked polyethylene in primary total knee arthroplasty. Acta Orthop 86(6):721–727
31. Cash D, Khanduja V (2014) The case for ceramic-on-polyethylene as the preferred bearing for a young adult hip replacement. HIP Int 24(5):421–427
32. Habermann B, Ewald W, Rauschmann M, Zichner L, Kurth AA (2006) Fracture of ceramic heads in total hip replacement. Arch Orthop Trauma Surg 126:464–470
33. Chevalier J (2006) What future for zirconia as a biomaterial? Biomaterials 27(4):535–543
34. Kurtz SM, Kocagöz S, Arnholt C, Huet R, Ueno M, Walter WL (2014) Advances in zirconia toughened alumina biomaterials for total joint replacement. J Mech Behav Biomed Mater 31:107–116
35. Reuven A, Manoudis G, Aoude A, Huk OL, Zukor D, Antoniou J (2014) Clinical and radiological outcome of the newest generation of ceramic-on-ceramic hip arthroplasty in young patients. Adv Orthop Surg. https://doi.org/10.1155/2014/863748

36. Kuntz M, Usbeck S, Pandorf T, Heros R (2011) Tribology in total hip arthroplasty (Knahr K, ed). Springer, Berlin, pp 25–40
37. Brockett CL, Williams S, Zhongmin J, Isaac GH, Fisher J (2013) Squeaking hip arthroplasties. J Arthrop 28(1):90–97
38. Mai K, Verioti C, Ezzet KA, Copp SN, Walker RH, Colwell Jr CW (2010) Incidence of 'squeaking' after ceramic-on-ceramic total hip arthroplasty. Clin Orthop Relat Res 468(2):413–417
39. Triclot P (2011) Metal-on-metal: history, state of the art (2010). Int Orthop 35:201–206
40. Drummond J, Tran P, Fary C (2015) Metal-on-metal hip arthroplasty: a review of adverse reactions and patient management. J Funct Biomater 6(3):486–499
41. Delaunay C, Petit I, Learmonth ID, Oger P, Vendittoli PA (2010) Metal-on-metal bearings total hip arthroplasty: the cobalt and chromium ions release concern. Orthoaed Traumatol Surg Res OTSR 96(8):894–904
42. Matharu GS, Eskelinen A, Judge A, Pandit HG, Murray DW (2018) Acta Orthop 89(3):278–288
43. Engh CA, MacDonald SJ, Sritulanondh S, Korczak A, Naudie D, Engh C (2014) Metal ion levels after metal-on-metal total hip arthroplasty. J Bone Joint Surg 96(6):448–455
44. Fernández-Valencia J, Gallart X, Bori G, Ramiro SG, Combalía A, Riba J (2014) Assessment of patients with a DePuy ASR metal-on-metal hip replacement: results of applying the guidelines of the Spanish Society of Hip surgery in a tertiary referral hospital. Adv Orthop. https://doi.org/10.1155/2014/982523
45. Bergschmidt P, Bader R, Kluess D, Zietz C, Mittelmeier W (2012) The all-ceramic Knee endoprosthesis—the gap between expectation and experience with ceramic implants. Seminars Arthrop 23:262–267
46. Meier E, Gelse K, Trieb K, Pachowsky M, Henning FF, Mauerer A (2016) First clinical study of a novel complete metal-free ceramic total knee replacement system. J Orthop Surg Res 11(21). https://doi.org/10.1186/s13018-016-0352-7
47. Solarino G, Piconi C, De Santis V, Piazzolla A, Moretti B (2017) Ceramic total knee arthroplasty: ready to go? Joints 5(4):224–228
48. Gotman I, Gutmanas EY, Hunter G (2017) Comprehensive biomaterials II (Ducheyne P, ed). Elsevier, Amsterdam, pp 165–203
49. Hunter G, Jones WM, Spector M (2005) Total knee arthroplasty (Bellemans J, Ries MD, Victor JMK, ed). Springer-Verlag, Heidelberg, pp 371–377
50. Vertullo CJ, Lewis PL, Graves S, Kelly L, Lorimer M, Myers P (2017) Twelve-year outcomes of an oxinium total knee replacement compared with the same cobalt-chromium design: an analysis of 17,577 prostheses from the Australian Orthopaedic Association National Joint Replacement Registry. J Bone Joint Surg Am 99(4):275–283
51. Heyse T, Haas SB, Efe T (2012) The use of oxidized zirconium alloy in knee arthroplasty. J Expert Rev Med Dev 9(4):409–421
52. Piconi C, De Santis V, Maccauro G (2017) Clinical outcomes of ceramicized ball heads in total hip replacement bearings: a literature review. J Appl Biomater Funct Mater 15(1):e1–e9
53. Vierra BM, Blumenthal SR, Amanatullah DF (2017) Modularity in total hip arthroplasty: benefits, risks, mechanisms, diagnosis, and management. Orthopedics 40(6):355–366
54. Hernigou P, Queinnec S, Henri C, Lachaniette F (2013) One hundred and fifty years of history of the Morse taper: from Stephen A. Morse in 1864 to complications related to modularity in hip arthroplasty. Int Orthop 37(10):2081–2088
55. Falkenberg A, Drummen P, Morlock MM, Huber G (2019) Determination of local micromotion at the stem-neck taper junction of a bi-modular total hip prosthesis design. Med Eng Phys 65:31–38
56. Urish KL, Giori NJ, Lemons JE, Mihalko WM, Hallab N (2019) Trunnion corrosion in total hip arthroplasty-basic concepts. Orthop Clin North Am 50:281–288
57. Weiser MC, Lavernia CJ (2017) Trunnionosis in total hip arthroplasty. J Bone Joint Surg Am 99:1489–1501
58. Pourzal R, Lundberg HJ, Hall DJ, Jacobs JJ (2018) What factors drive taper corrosion? J Arthrop 33(9):2707–2711

59. Mistry JB, Chughtai M, Elmallah RK, Diedrich A, Le S, Thomas M, Mont MA (2016) Trunnionosis in total hip arthroplasty: a review. J Orthop Traumatol 17(1):1–6
60. Rodrigues DC, Urban RM, Jacobs JJ, Gilbert JL (2009) In vivo severe corrosion and hydrogen embrittlement of retrieved modular body titanium alloy hip-implants. J Biomed Mater Res B Appl Biomater 88(1):206–219
61. Weber AE, Skendzel JG, Waxman DL, Blaha JD (2013) Symptomatic aseptic hydrogen pneumarthrosis as a sign of crevice corrosion following total hip arthroplasty with a modular neck: a case report. JBJS Case Connector 3(3):e76
62. Morlock MM, Bünte D, Ettema H, Verheyen CC, Hamberg A, Gilbert J (2016) Primary hip replacement stem taper fracture due to corrosion in 3 patients. J Acta Orthop 87(2):189–192
63. Fokter SK, Rudolf R, Molicnik A (2014) Titanium alloy femoral neck fracture—clinical and metallurgical analysis in 6 cases. Acta Orthop 87(2):197–202
64. Niinomi M, Nakai M (2011) Titanium-based biomaterials for preventing stress shielding between implant devices and bone. Int J Biomater, Article ID 836587. https://doi.org/10.1155/2011/836587
65. Swann RP, Webb JE, Cass JR, Van Citters DW, Lewallen DG (2015) Catastrophic head-neck dissociation of a modular cementless femoral component. JBJS Case Connector 5(3):e71. https://doi.org/10.2106/JBJS.CC.N.00179
66. Urish KL, Hamlin BR, Plakseychuk AY, Levison TJ, Kurtz S, DiGioia AM (2017) Letter to the Editor on "Trunnion failure of the recalled low friction ion treatment cobalt chromium alloy femoral head." J Arthrop 32(9):2857–2863
67. Martin AJ, Jenkins DR, Van Citters DW (2018) Role of corrosion in taper failure and head disassociation in total hip arthroplasty of a single design. J Orthop Res 36(11):2996–3003
68. Matsen Ko L, Chen AF, Deirmengian GK, Hozack WJ, Sharkey PF (2016) The Journal of bone and joint surgery. American 98(16):1400–1404
69. Molloy DO, Munir S, Jack CM, Cross MB, Walter WL, Walter WK (2014) Fretting and corrosion in modular-neck total hip arthroplasty femoral stems. J Bone Joint Surg 96(6):488–493
70. Walsh CP, Hubbard JC, Nessler JP, Markel DC (2015) Revision of recalled modular neck rejuvenate and ABG femoral implants. J Arthrop 30(5):822–826
71. Nawabi DH, Do HT, Ruel A, Lurie B, Elpers ME, Wright T, Potter HG, Westrich GH (2016) Comprehensive analysis of a recalled modular total hip system and recommendations for management. J Bone Joint Surg 98:40–47
72. Di Laura A, Hothi HS, Henckel J, Kwon YM, Skinner JA, Hart AJ (2018) Retrieval findings of recalled dual-taper hips. J Bone Joint Surg Am 100(19):1661–1672
73. Kurtz SM, Kocagöz SB, Hanzlik JA, Underwood RJ, Gilbert JL, MacDonald DW, Lee GW, Mont MA, Kraay MJ, Klein GR, Parvizi J (2013) Do ceramic femoral heads reduce taper fretting corrosion in hip arthroplasty? A retrieval study. Clin Orthop Relat Res 471(10):3270–3282
74. Bull SJ, Moharrami N, Langton D (2017) Mechanistic study of the wear of ceramic heads by metallic stems in modular implants. J Bio Tribo-Corrosion 3:7
75. Triantafyllopoulos GK, Elpers ME, Burket JC, Esposito CI, Padgett DE, Wright TM (2016) Otto Aufranc award: large heads do not increase damage at the head-neck taper of metal-on-polyethylene total hip arthroplasties. Clin Orthop Relat Res 474(2):330–338
76. Morlock MM (2015) The taper disaster—how could it happen? HIP Int 25(4):339–346
77. Goldberg JR, Gilbert JL (2003) In vitro corrosion testing of modular hip tapers. J Biomed Mater Res Part B Appl Biomater 64(2):78–93
78. Zachary JC, Coury JG, Cohen J (2017) Taper technology in total hip arthroplasty. JBJS Rev 5(6):e2
79. Shenhar A, Gotman I, Radin S, Ducheyne P, Gutmanas EY (2000) Titanium nitride coatings on surgical titanium alloys produced by a powder immersion reaction assisted coating method: residual stresses and fretting behavior. Surf Coat Technol 126(2–3):210–218
80. Shearwood-Porter N, Browne M, Milton JA, Cooper MJ, Palmer MR, Latham JM, Wood RJK, Cook RB (2017) Damage mechanisms at the cement–implant interface of polished cemented femoral stems. J Biomed Mater Res Part B Appl Biomater 105(7):2027–2033

81. Choi D, Park Y, Yoon Y-S, Masri BA (2010) In vitro measurement of interface micromotion and crack in cemented total hip arthroplasty systems with different surface roughness. Clin Biomech 25:50–55
82. Zhang H, Brown LT, Blunt LA, Jiang X, Barrans SM (2009) Understanding initiation and propagation of fretting wear on the femoral stem in total hip replacement. Wear 266(5–6):566–569
83. Bader R, Steinhauser E, Holzwarth U, Schmitt M, Mittelmeier W (2004) A novel method for evaluation of the abrasive wear behaviour of total hip stems at the interface between implant surface and bone cement. Proc Inst Mech Eng, Part H 218(4):223–230
84. Thomas SR, Shukla D, Latham PD (2004) Corrosion of cemented titanium femoral stems. J Bone Joint Surg (BR) 86B:974–978
85. Hallan G, Espehaug B, Furnes O, Wangen H, Hol PJ, Ellison P, Havelin LI (2012) Is there still a place for the cemented titanium femoral stem? Acta Orthop 83(1):1–6
86. Noronha Oliveira M, Schunemann WVH, Mathew MT, Henriques B, Magini RS, Teughels W, Souza JCM (2018) Can degradation products released from dental implants affect peri-implant tissues? J Periodontal Res 53:1–11
87. Cruz HV, Souza JCM, Henriques M, Rocha LA (2011) Biomedical tribology (Paulo Davim J, ed). Nova Science Publishers, Hauppauge (New York), pp 2–33
88. Corne P, De March P, Cleymand F, Geringer J (2019) Fretting-corrosion behavior on dental implant connection in human saliva. J Mech Behav Biomed Mater 94:86–92
89. Alrabeah GO, Knowles JC, Petridis H (2018) Reduction of tribocorrosion products when using the platform-switching concept. J Dent Res 97:995–1002
90. Fransson C, Wennström J, Tomasi C, Berglundh T (2009) Extent of peri-implantitis-associated bone loss. J Clin Periodontol 36:357–363

A New Method for Seismically Safe Managing of Seismotectonic Deformations in Fault Zones

Valery V. Ruzhich and Evgeny V. Shilko

Abstract The authors outline the results of long-term interdisciplinary research aimed at identifying the possibility and the methods of controlling tangential displacements in seismically dangerous faults to reduce the seismic risk of potential earthquakes. The studies include full-scale physical and numerical modeling of P-T conditions in the earth's crust contributing to the initiation of displacement in the stick-slip regime and associated seismic radiation. A cooperation of specialists in physical mesomechanics, seismogeology, geomechanics, and tribology made it possible to combine and generalize data on the mechanisms for the formation of the sources of dangerous earthquakes in the highly stressed segments of faults. We consider the prospect of man-caused actions on the deep horizons of fault zones using powerful shocks or vibrations in combination with injecting aqueous solutions through deep wells to manage the slip mode. We show that such actions contribute to a decrease in the coseismic slip velocity in the fault zone, and, therefore, cause a decrease in the amplitude and energy of seismic vibrations. In conclusion, we substantiate the efficiency of the use of combined impacts on potentially seismically hazardous segments of fault zones identified in the medium-term seismic prognosis. Finally, we discuss the importance of the full-scale validation of the proposed approach to managing the displacement regime in highly-stressed segments of fault zones. Validation should be based on large-scale tests involving advanced technologies for drilling deep multidirectional wells, injection of complex fluids, and localized vibrational or pulse impacts on deep horizons.

Keywords Seismically active fault · Earthquake · Friction · Healing · Field experiment · Vibro-pulse impact · Deep drilling · Fluid injection · Shear stress relaxation · Slip control

V. V. Ruzhich (✉)
Institute of the Earth's Crust, SB RAS, 664033 Irkutsk, Russia
e-mail: ruzhich@crust.irk.ru

E. V. Shilko
Institute of Strength Physics and Materials Science, SB RAS, 634055 Tomsk, Russia
e-mail: shilko@ispms.tsc.ru

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_3

1 Introduction

The methods developed by the world community for countering natural seismic disasters are insufficient since they do not allow an efficient reduction of the almost annual losses caused by strong earthquakes [1–3]. To date, one has come to understand that difficulties in finding solutions to this problem are concerned with the lack of reliable information on the deep tribophysical and geochemical characteristics of earthquake preparation processes in fault zones. These geological processes are hidden at inaccessible seismic focal depths of the lithosphere (7–30 km) and occur at elevated temperatures and pressures of hundreds of megapascals. An important role in the preparation of earthquakes belongs to the migration of fluids of various compositions [4, 5]. The results of a long- and medium-term prognosis make it possible to identify foci with indications of the final stage of the preparation of strong earthquakes with a probability of the order of $P = 0.6$–0.7. In most cases, these are the same segments of the zones of interplate and intracontinental seismic faults, where "hot spots" of dangerous or catastrophic events have already arisen for many tens or hundreds of thousands of years. The modern possibilities of making a short-term forecast of strong and catastrophic earthquakes are still limited. Moreover, such a forecast cannot efficiently ensure seismic safety, since it does not prevent the large-scale negative consequences of destruction [6]. The world scientific community recognizes the need to search for new ways to more reliably ensure seismic safety.

The chapter is devoted to a review of the results of a joint 20-year interdisciplinary research led by Professor Sergey Grigorievich Psakhie, the corresponding member of the Russian Academy of Sciences. These studies were aimed at exploring the safe methods of managing deformations (displacements) in the seismically dangerous segments of active faults [7, 8]. Specialists from various scientific institutions were involved, including the Institute of the Earth's Crust SB RAS (IEC SB RAS), the Institute of Strength Physics and Materials Science SB RAS (ISPMS SB RAS), the Institute of Geosphere Dynamics RAS (IDG RAS), Technische Universität Berlin (TU Berlin), as well as other scientific institutions from Ukraine, Mongolia, and China. The success achieved was largely related to the use of diverse scientific approaches and methods including computer simulation, laboratory physical modeling, and field experiments in rock massifs and regions of the block-structured ice cover of Lake Baikal. Such a variety of techniques and objects of study made it possible to take a fresh look at the nature and mechanisms of destruction processes in the earth's crust and reveal a set of criteria governing the occurrence of seismically dangerous dynamic phenomena in zones of deep faults.

In the chapter, we outline key aspects of this multidisciplinary research.

In particular, we describe the results of a detailed geological, geophysical, and petrological study of the structure and physicochemical transformations of rocks in earthquake foci in the deep horizons of faults. Such information was obtained through examining the deeply denudated blocks of rocks in the fault zones, which were exhumed to the earth's surface from depths of 10–20 km. Using new methods of

analysis, a number of representative deep segments of fault zones (deep paleoseismic dislocations) were studied in detail. The main attention was paid to segments, which were the centers of earthquakes in past epochs of geological evolution.

Significant efforts of the collaboration participants were directed at conducting field experiments to study the complex mechanisms of the formation of seismic radiation sources. The advantages of such experiments over laboratory tests on small samples of artificial and natural materials are the large size of the studied objects (segments of fault zones tens or hundreds of meters long), the interrelation between the studied objects and the surrounding rocks, the possibility of using broadband seismic stations and strain gages, as well as the possibility of modeling various impacts using high-power jacks, pile drivers and weak explosions. Emphasis was placed on revealing the effect of the impacts on the features of changing the displacement mode in the studied fault segments from slow creep to accelerated creep and coseismic slip. These features largely determine the generation of seismic oscillations of a wide amplitude-frequency range.

Based on a comprehensive analysis of the results of geological and geophysical studies of deep paleoseismic dislocations, field experiments, and numerical modeling, we formulated the concept of the response of faults to external actions and the approach to controlling and managing deformations in fault segments using man-made vibro-pulse and hydraulic wave impacts. Due to the limited effectiveness of numerical modeling and low-scale field experiments in the near-surface part of the earth's crust, the authors cooperated with deep drilling specialists, which have high competence in the exploration and exploitation of hydrocarbon deposits in the Late Proterozoic crystalline formations of the Siberian Craton [9, 10]. The combination of dynamic mechanical impacts with the injection of solutions into fault segments through deep multidirectional wells significantly expands the possibilities of the developed approach to managing displacements in seismically dangerous fault segments. Here we also discuss the prospects of hydraulic wave impacts on deep seismically active segments by means of drilling horizontal, inclined, and vertical wells up to 5–7 km deep, hydraulic fracturing of rock massifs, and injection of solutions of various compositions.

Finally, we consider the difficulties of implementing a complex impact on the deep fault segments and provide a brief overview of current and planned studies by different scientific groups.

2 Methodological Basis

Of the many earthquake source focal models known in the world, the "stick–slip" model is considered to be the optimal and most general model [11, 12]. It is based on the fundamental tribological laws of sliding friction in seismic-generating fault zones, which have a textured roughness, specific mineralization, and fluid saturation. This model uses the main achievements of tribology and, in particular, takes into account the influence of such frictional parameters as the ratio of shear and normal

resistance forces (characterized by stiffness ratio), roughness, the variation of sliding velocity, the presence of lubricant, wear of the sliding surfaces, vibrations, etc. [13]. Detailed geological and experimental study of fracture processes in rock massifs has shown that the fundamental tribological patterns of creep or stick-slip sliding can be successfully used to determine the causes and mechanisms of strong earthquakes in fault segments [14, 15].

To clarify the conditions of earthquake preparation, we studied the zones of seismic dislocations that formed not only with modern earthquakes but also with paleo-earthquakes. Traces of such ancient earthquakes are captured in the seismic focal horizons of the denudated crust. During the long geodynamic evolution of the lithospheric shell of the Earth, these horizons were on the Earth's surface in the orogeny regions after kilometers-long denudation. They are currently available for visual geological exploration. To obtain a more in-depth understanding of the real structure of the deep segments of the sources of strong earthquakes, these studies were conducted in the Baikal rift zone and the collision zones in Mongolia. The results of these studies are discussed below.

From the viewpoint of practical geodynamic applications of the methods of physical mesomechanics, it is advisable to consider the lithospheric shell of our planet as the upper hierarchically organized geosphere in which the seismotectonic destruction process occurs at various scales according to well-known geomechanical laws. In particular, various tribochemical and geomechanical processes manifest themselves at various hierarchical scales of the earth's crust. According to the well-known kinetic concept of the strength of solids by S. N. Zhurkov and his followers [16, 17], multiscale fracture-related processes manifest themselves in stage-by-stage accumulation of small discontinuities and their coalescence into longer ones due to the fracture of the separating "jumpers" (structural barriers). This can be observed in geological destructive processes. Hierarchical transitions within the self-organized process of destruction of the earth's crust reach the highest level in the form of the formation of intraplate and interplate seismic faults with a length of hundreds to thousands of kilometers.

The basic geomechanical model of a seismically active fault is based on the idea that a focus of a very strong earthquake with $M = 7.0$–9.0 is formed as a result of successive long and multi-stage coalescence of faults ranging from the first tens of kilometers to longer ones (70–120 km). This is clearly illustrated by many recent strong earthquakes, such as the Gobi-Altai earthquake of 1957 ($M = 7.9$). To find out the conditions for the formation of different-scale foci of earthquakes (including foci of super-earthquakes) and mechanisms of fracture of these interface zones, the authors use a multiscale approach. It is based on the data of many years of field experiments, geological and structural studies, and numerical modeling using the method of movable cellular automata. On each scale, specific tribochemical and geomechanical factors that determine the features of contact interactions, including the frictional regimes, can be distinguished. The friction features, in turn, affect the amplitude-frequency and energy parameters of the generation of seismic vibrations

[18]. Further development of this concept is associated with an analysis of the geomechanical consequences of fracture of structural barriers and contact interactions of different scale asperities on the glide planes of the fault limbs.

Pore fluids play an important and, in some cases, determining role in petrochemical processes in zones of highly stressed and seismically dangerous faults. Interstitial fluid largely determines the slip modes, and therefore, the modes and features of the generation of seismic vibrations [19, 20]. Traditionally, the term "fluids" include liquids and gases that fill void spaces arising from deformation in rock masses. Fluids are also magmatic melts, which themselves are sources of hydrothermal fluids ascending to the earth's surface through conducting faults [21]. The migration of fluid solutions affects the shear resistance of the faults through the lubrication and hydraulic effects on the walls of the fractured rock. Therefore, the pore fluid should be taken into account among key factors that control the nucleation and dynamics of coseismic displacements in discontinuities of different hierarchical levels.

3 Geological Study of Exhumed Seismic Dislocations of Paleo-Earthquakes in the Southeastern Boundary of the Siberian Craton

To determine the deep-seated conditions for the modern earthquakes, a collaboration of scientists from IEC SB RAS and IDG RAS carried out a geological and geophysical study of the exhumed deep segments of the Primorsky section of the marginal zone on the southeastern marginal seam of the ancient Siberian craton. Due to natural exhumation from the middle depths of the earth's crust, deep fault segments are now available for study. Here, the traces of different age foci of paleo-earthquakes (ancient coseismic slips) formed at elevated temperatures and pressures are preserved. We note that similar P-T conditions currently exist at depths of 10–25 km, where the seismic focal horizon is located and most of the modern earthquake foci of various energy scales occur (including the strongest events with $M = 7.0$–7.9). Many of them are confined to the zone of the Primorsky riftogenic fault with a length of about 200 km. This fault is located on the northwestern coast of the Baikal rift depression (Fig. 1) and was formed for 65 Ma. It is important to note that the deeply denudated (exhumed) Pribaikalsky segment of the fault allows the visual study of the structure and P-T conditions for the formation of ancient earthquakes of different ages (including those with an age of the order of 700–450 Ma) at activation depths of 12–18 km [22].

Ancient sliding mirrors, formed during the dynamic interaction of irregularities, were sampled in this fault segment (Fig. 2). The samples were subjected to petrological studies using the polished sections. Tectonites (rocks transformed by the friction of rocks) taken in the zone of later earthquakes have indications of gneissing, cataclase, and milonitization in the Early Paleozoic period. To elucidate the natural conditions of tribochemical processes in the deep foci of paleo-earthquakes and

Fig. 1 Morphological expression of the Primorsky riftogenic zone: view from space (**a**) and one of the sections on the shore of the Small Sea (**b**). The arrows in (a) indicate the test sites where a comprehensive study of ancient coseismic displacements was carried out. These displacements occurred hundreds of millions of years ago in the centers of paleo-earthquakes at depths of about 15–20 km, at high temperatures and pressures

the structure of contact spots with traces of coseismic slip, we carried out a petrological study of changes in the material composition of the mineral coating on slip planes. Coseismic slip and the structure of the surface layers were analyzed by sliding mirrors containing indications of high-rate displacements, including pseudotachylyte (products of frictional heating in faults under stress metamorphism at elevated

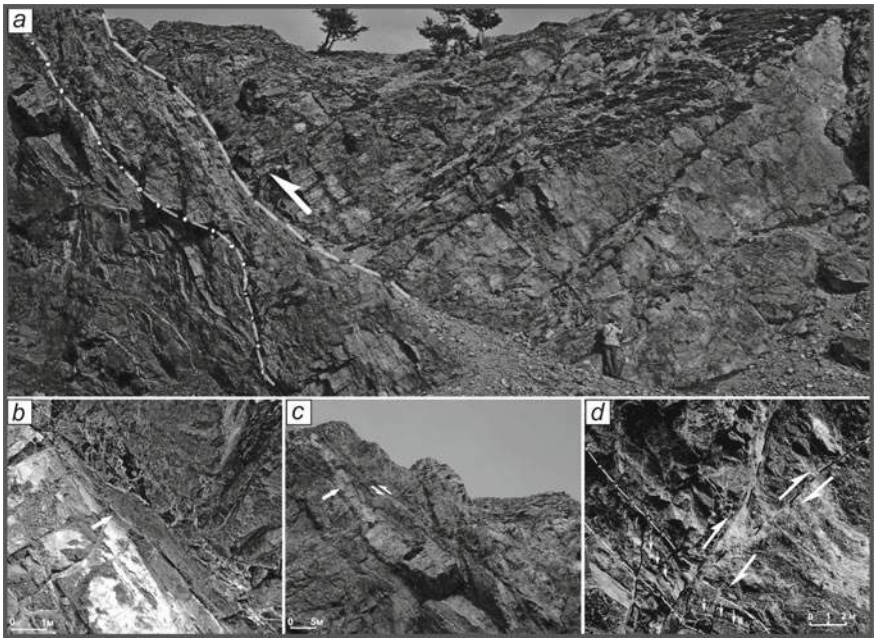


Fig. 2 Coseismic fractures of different ages containing multiscale asperities in the centers of paleoearthquakes at seismic focal depths from 5 km to 12–18 km: **a** the oldest coseismic slip zone at a depth of about 18 km; **b** low-scale roughness in a small coseismic rupture; **c** longer coseismic rupture and "macroasperity" shown by the arrow; **d** intersecting coseismic ruptures that occurred at different times and different depths and PT conditions, as shown by minerals deposited from fluids. The dash-dot line in (a) shows the oldest coseismic slip zone at a depth of about 18 km. The dashed line and the arrow to the right (slip direction) denote the younger zone of coseismic slip. Slip mirrors and newly formed mineralization with traces of thermal heating during friction were taken from this zone

temperatures, pressures, and the participation of fluids). Recently, G.A. Sobolev and his colleagues demonstrated the possibility of efficient application of modern physical methods for determining strains and stresses in fine ground nanocrystals of various minerals, including quartz, albite, chlorite, epidote, and other varieties found on sliding mirrors from the exhumed segment of the Vilyui deep fault [23, 24]. Consequently, tectonic glide mirrors identified at fault exhumation sites, as well as pseudotachylyte, can be used to recognize and analyze coseismic planes. This makes it possible to obtain important information about the kinematics and tribophysical processes in the deep fault segments at the moments of the generation of seismic events by the examples of past tectonic eras.

The analysis of the slip mirrors showed the finest brittle attrition of minerals. This is typical for high-velocity sliding. The important role of fluids in eliminating (healing) the consequences of fracture is illustrated through the presence of fine-grained aggregates of quartz grains and feldspars with a subordinate amount of sericite, chlorite, and iron ore minerals. It is important to note that such mineral

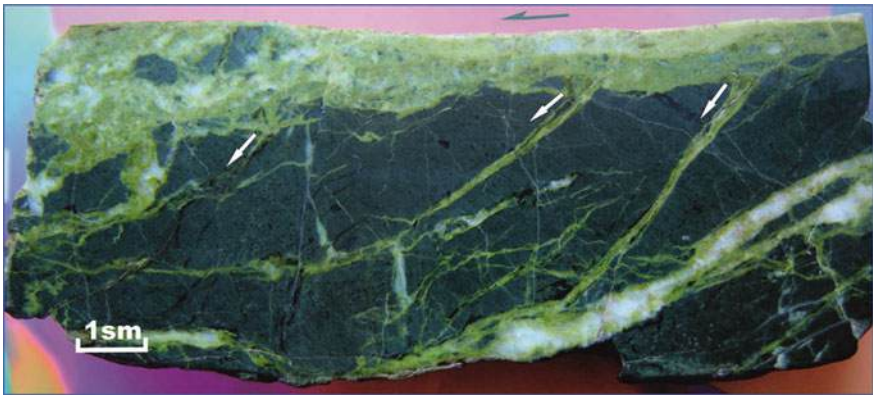


Fig. 3 The cross-section of a sample of an ancient sliding mirror taken from the zone of the exhumed segment of the Dolinozersky Fault (still highly seismic hazardous) in southern Mongolia. It captures the consequences of two acts of coseismic slip (the arrow at the upper edge) with the "healing" by the epidote and subsequent quartz mineralization

neoformations have a reduced shear resistance and can play the role of a mineral lubricant that facilitates subsequent sliding acts and prevents reverse closure of crack walls. Sliding of rough walls of the fault is accompanied by crack opening (dilatancy), and mineralized fluids are sucked into the vacuum gap. This ensures filling the cavities with new minerals, including ore. This is the hydrothermal mechanism of mineral "healing" of fault segments in the upper part of the earth's crust at relatively low pressures and temperatures [25].

The collected information on the dating of identified coseismic planes and the P-T-f conditions for the formation of slip mirrors and pseudotachylyte was used to make a qualitative and quantitative estimation of the exhumation of the Primorsky section of the margin, as well as of the structure of the seismic focal horizon. In particular, the oldest absolute dating established by the $^{40}Ar/^{39}Ar$ method is 673 ± 4.8 Ma. This corresponds to one of the stages of the Neoproterozoic era, i.e., the time of break-up of the Rodinia supercontinent and the separation of the Siberian craton along the margin. 415.4 ± 4.1 Ma refers to a later stage in the evolution of the margin during the collision of the Olkhon terrane with the Siberian craton [22]. The obtained dating made it possible to roughly estimate the occurrence depths of the identified strong paleo-earthquakes. They amounted to about 18 km in the earlier (Neoproterozoic) period, and about 12 km in the later (Middle Paleozoic) stage [22]. Similar results were obtained in the neighboring section of the Primorsky fault (5 km to the northeast). In the coseismic slip plane, the temperature of frictional heating during the occurrence of pseudotachylyte reached approximately 900 ºC, and the pressure reached 700–800 MPa. After dynamic slip, the pressure rapidly decreased to a "natural" (lithostatic) value, which was estimated as ≤150–200 MPa in the considered case. This corresponds to depths of the order of 5–7 km [26].

Figure 3 shows a photo of a very ancient sliding mirror taken from the exhumed zone of the Dolinozersky Fault in southern Mongolia. A devastating Gobi-Altai earthquake ($M = 7.9$) occurred here in December 1957.

The polished section shows the internal structure of the coseismic rupture zone with the shearing of asperities on the slip plane and the transfer of fragments of basaltic anorthosite from an ancient volcano. Large and small fragments of the parent rock move within the mass of newly formed chlorite and calcite minerals deposited from the flowing fluid. These minerals healed the fracture cavity. Mineral "healing" products also fill wing cracks that formed under the condition of large friction in combination with the fracture of asperities. Wear debris carries over when sliding in the direction of the white arrows. This is the widespread mechanism of the formation of sources of strong earthquakes accompanied by the occurrence of wing cracks.

Therefore, the geological study of deep segments of fault zones opens up a unique opportunity to reconstruct the conditions for the nucleation of coseismic slips in the areas of the most severe contact interaction of different scale asperities in fault zones with the participation of fluids. The mineralization of the fluid determines the mineral composition of crack fillers. Mineral composition in many respects affects the frictional sliding of the fault walls during subsequent activations. Although the processes of stress-metamorphic transformation of rocks in the considered fault zones at seismic focal depths occurred in past eras of seismotectonic activation, it is clear that the same mechanisms act in modern earthquake foci. Therefore, the database of ancient earthquakes is fundamentally important for creating adequate models of modern earthquake sources [14].

4 Features of the Response of Tectonic Fault Segments to Man-Caused Impacts

Below we describe the most important results of field experiments carried out in different years on the segments of the Angarsky and Primorsky faults in the Baikal rift zone.

The first one was carried out in 2004–2006 on the segment of the Angarsky fault zone near the Listvyanka village (southwestern coast of Lake Baikal), Irkutsk Region. The Angarsky fault has a northwestern strike and is characterized by faults of the reverse-strike-slip type. The studies were conducted with the participation of scientific groups of several institutes of the Russian Academy of Sciences, as well as the Technische Universität Berlin. At this test site, the effect of high strain sensitivity of the principal slip zone to natural and man-made impacts was revealed [27]. The most important result of the research is related to the demonstration of the ability to manage the shear displacement mode in the fault zone. In order to carry out man-made impacts on a selected fault segment 100 m long, several wells were drilled to a depth of 30 m for 14 days, and water solutions were injected into the fault zone (Fig. 4a). Then local dynamic impacts with weak explosions at a depth of about 15 m

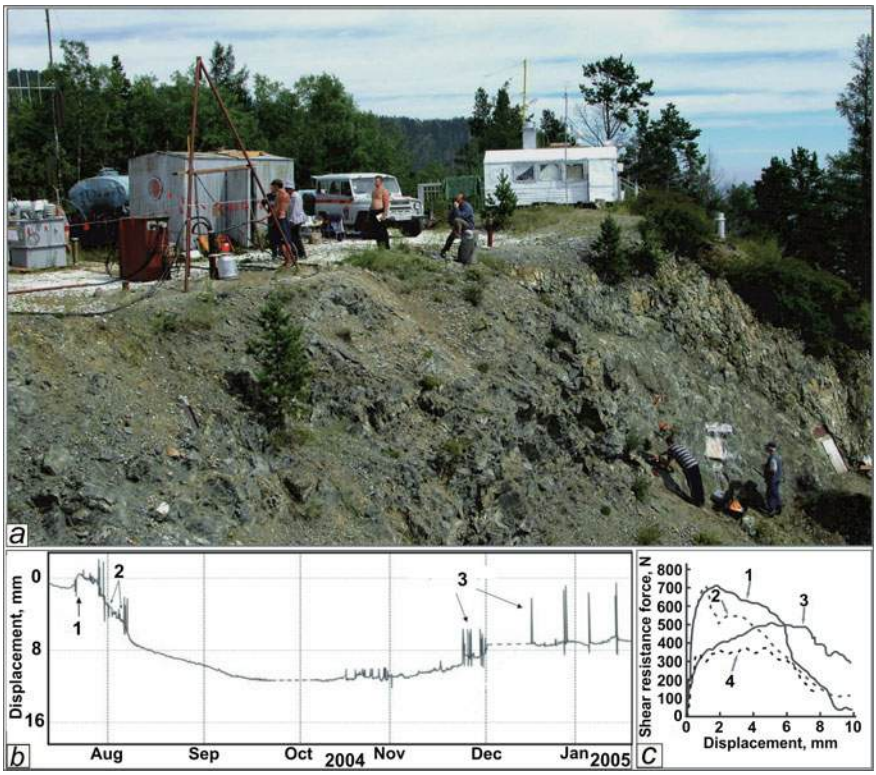


Fig. 4 General view of the test site in the Angarsky fault zone section (**a**), a graph of shear displacements in the fault zone at different stages of the experiment (**b**) and results of numerical modeling [7] showing the dependence of the shear resistance on the relative tangential displacement at a constant displacement velocity and various additional actions. Figure (a) shows a part of the test site with three wells for water injection and explosive actions, a platform for a falling pile and locations for strain gages and seismic sensors. The numbers in Figure (b) are as follows: stress tests with a falling pile before the experiment (1); well drilling, water injection, and explosions in the well (2); displacement "jumps" at the stage of return deformation (3). The numbers in Figure (c) are: "dry" fault zone, no additional impacts (1); "dry" fault zone, additional vibration (2); watered fault zone, no additional impacts (3); watered fault zone, additional vibration (4)

in watered wells were carried out at the next stage. To study the effect of the impacts on the stress state of the fault segment, we performed mechanical stress tests on the surface with a falling pile weighing 100 kg. Such tests were performed before and during the whole course of the experiment.

The fault responded in the form of initiated tangential displacements in the principal slip zone. This was recorded by strain and seismic sensors. The vibrations during drilling and subsequent watering of the fault zone, accompanied by weak cumulative explosions in watered wells, have led to a qualitative change in the displacement mode. In particular, the creep velocity increased by several orders of magnitude: from "natural" values of 0.65–3.3 μm/day (0.24 mm/year ÷ 1.2 mm/year) to 0.5 mm/day.

Although the creep acceleration was of a short-term nature (less than a month), the total amplitude of the initiated tangential displacements in the principal slip zone reached 8–10 mm (Fig. 4b). This corresponds to 10 years of accumulation of displacements in the natural mode. Note that in the months after the experiment, a partial recovery (reverse displacement) occurred in the slip plane, and the residual displacement amplitude decreased to 5–6 mm. This is the result of the elastic "return" of the surrounding rock mass. It is noteworthy that the return motion was accompanied by seismic pulses with amplitudes comparable with seismic amplitudes during initiated (accelerated) creep. A similar elastic-plastic partial return with significant "residual" displacements is also observed in seismic dislocations after strong earthquakes. This indicates the generality of the described effect for the zones of tectonic faults.

Field experiments and numerical simulations using the method of movable cellular automata [7] showed that the injection of aqueous solutions into the fault zone in combination with vibroimpulse impact trigger local relaxation of "excess" shear stresses in the fault segment (Fig. 4c). Stress relaxation is manifested particularly in the transition from natural to accelerated creep, followed by its gradual deceleration to natural values determined by regional tectonic motion. Displacement monitoring on the studied segment of the Angarsky fault over 14 years shows slow creep and extremely weak microseismicity. Moreover, the magnitude of deformation and seismic response of the fault segment to the standard dynamic test impacts of the falling pile is one order of magnitude lower than before the experiment. This confirms a considerable decrease in the level of shear stresses. Thus, the field experiment confirmed the statement about the possibility of a controlled change in the slip mode in the segments of fault zones to safely reduce the level of shear stresses and seismic activity [7, 28]. The developed method of shear stress relaxation in fault segments is protected by a patent of the Russian Federation [29].

From 2015 to 2018 joint research was conducted by the IEC RAS and IDG RAS research teams in the coastal segment of the Primorsky Fault zone on the Baikal coast (near Olkhon Island). The studies included geophysical observation and field experiments. Figure 5 shows a general view of the testing ground. The purpose of the experiments was to study the influence of quasistatic (carried out by jacks) and dynamic (explosive) effects on the displacement mode during contact interaction with a single artificial asperity. A hydraulic flat jack was installed in a large crack in the central part of the fault zone. The crack walls were pulled apart with a force of 43 tons. A 5 cm thick granite gneiss ore was inserted into the cavity to study the fracture of the rock under conditions of high local compressive stress and additional (man-made) impacts. After the jack was removed, the ore was compressed by the crack walls with the force of the crack extension (~ 43 tons).

The processes in the studied segment of the fault zone after the insertion of a single macroasperity were recorded by strain gages and seismometers. After several days of monitoring the natural deformations and the seismic background, a series of weak explosions were performed (the mass of one powder charge was 0.1 kg). The explosion chamber was located in the large plate within the fault segment at a distance of 3 m from the seismic sensors. The recorded amplitude of seismic

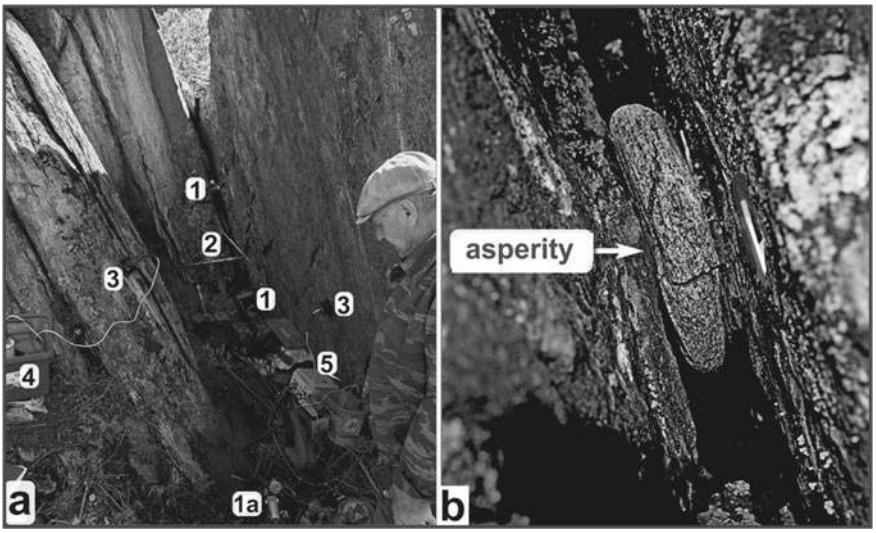


Fig. 5 Studied segment of the Primorsky fault zone with measuring equipment (**a**) and artificial asperity (granite gneiss ore) with cracks resulting from experiments (**b**). The numbers in Figure (a) are as follows: jacks (1); pump (1a); displacement sensors (2); seismic sensors (3). The inclined arrow in (b) shows the direction of the natural normal faulting displacement

acceleration during the explosion was 219 cm/s^2, and the frequency was about 70–75 Hz. Figure 6 shows a seismic recording for 330 min after the explosion. It reflects the result of contact interaction during the explosion, in particular, the initiated slip

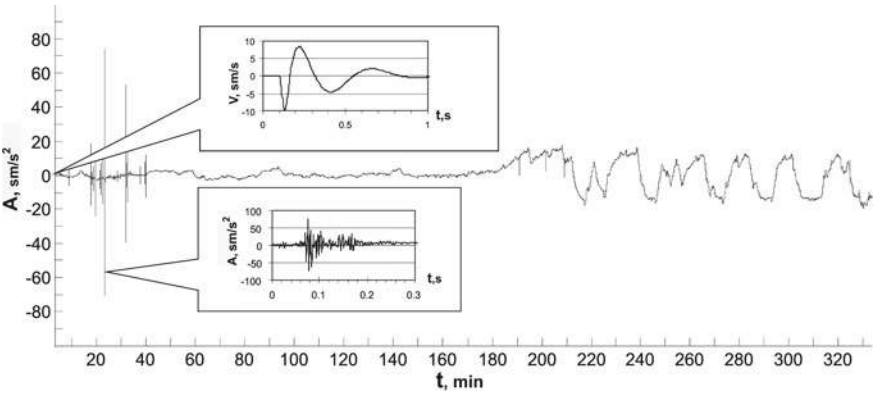


Fig. 6 Response to explosive action in one of the cracks in the Primorsky fault zone. The main figure shows the general view of the seismic record within 330 min after the explosion. The upper inset shows the oscillation velocity of the rock mass for 1 s and its upper value of 0.7 mm/s. The bottom inset shows the record of the first induced seismic response (the largest seismic pulse) that occurred 22 min after the explosion (amplitude 70 cm/s^2, frequency 160–175 Hz)

in the contact pair "crack wall—ore". Within 40 min after the explosion (with some delay) a series of strong seismic pulses (response to the explosion) were formed. Pulse amplitudes reached several tens of percent of the acceleration amplitudes during the explosion.

The measurements and visual observation showed that stress relaxation and redistribution in the surrounding massif after the explosion promote reaching the critical stresses at the contacts of the artificial asperity with the crack walls and cracking of the ore after some time (tens of minutes) after the explosion. Cracks propagated in the ore in the dynamic regime and were accompanied by the generation of a series of large seismic pulses. In Fig. 6, these pulses are shown on the left side (time interval of 15–40 min after the moment of explosion).

An analysis of the record of displacements in the main crack showed moderate compressive deformations of the crack with characteristic amplitudes of 25–35 μm in combination with reverse faulting during the period of the subsequent "calm". The largest seismic impulse was associated with the fracture of the artificial asperity and characterized by the following parameters: amplitude 70 cm/s^2, frequency 160–175 Hz. The next phase of the response began 190 min after the explosion and lasted about 140 min. It is characterized by low-frequency "pulses" with an average period of about 18 min and acceleration amplitudes of 15–20 cm/s^2. The generation of such extremely slow seismic pulses is a deformation response of the rock massif to a dynamic disturbance of a local stress state. The response manifests itself in the form of a return slip and can be described by viscoelastic models.

This example demonstrates the generation of high-velocity elastic and subsequent slowed-down oscillations of the rock massif caused by the intermittent slip mode at the contact with the large-scale asperity. It is noteworthy that the unloading of the surrounding rock mass manifested itself in the form of intermittent slip described by the stick-slip model, with the generation of slow deformation waves. In subsequent similar tests, the contact area of the ore and the crack walls was wetted. This led to the initiation of small cracks in the artificial asperity at the contact point accompanied by numerous packets of audible seismic-acoustic pulses. Over the subsequent months of application of natural static pressure by the rock mass, the ore gradually split into several large fragments.

The described experiment showed that an explosive (high-velocity) impact causes a strong viscoelastic-plastic return accompanied by high-amplitude seismic pulses and subsequent slow motions. These slow oscillations can presumably be regarded as being analog to slow waves associated with intermittent sliding of contacting surfaces in interplate faults [30, 31]. In this case, the observed propagation phenomena of slow waves are apparently associated with the phenomena of self-organization of ensembles of cracks in the rock mass (cracks are highly sensitive to the rate and amplitudes of external dynamic impacts). Note that slow motions were clearly recorded in other (larger-scale) experiments including technological explosions with multi-ton explosive charges in the quarry of the Udachnaya diamond pipe.

Based on the results of the above field experiments, as well as many other tests, we can draw an important conclusion. Safe stress relaxation in highly stressed (seismically hazardous) segments of tectonic fault zones can be achieved using local

impacts with low impact velocities and energies [32]. An additional and extremely effective relaxation stimulation is the saturation of contact slip surfaces with aqueous solutions, especially when the physicochemical properties of the liquid determine the important role of the Rehbinder effect [33, 34].

5 Prospects for the Implementation of Controlled Impacts on Fault Segments Through Deep Wells

The examples discussed above consider methods of managing impacts on small near-surface areas of seismically active faults. The question remains whether these experiments can be extended to the scales of the long fault segments and deep regions characterized by high pressures and temperatures, and complex fluid circulation. To answer this question, we consider the possibility of implementing control actions on segments of deep sections of fault zones through wells up to 3–5 km deep. Here we can take the experience of drilling such wells within the ancient Siberian platform for injecting solutions and hydraulic fracturing. Such work is carried out, particularly, at hydrocarbon deposits in the Republic of Yakutia (Russia), where hydrocarbon reservoirs are at relatively shallow depths of about 2 km.

Deep drilling and hydraulic fracturing allow to open the natural fluid fracture-pore reservoirs containing oil, gas, and saltwater. It should be noted that the magnitude of the pressure in such reservoirs is often close to rock pressure. The application of this technology allows the use of fluid injection at a given pressure through a network of inclined wells as a hydraulic action on the walls of the principal slip plane, ensuring their opening or closing [35]. The hydraulic effect of the injected fluid can be compared to a first approximation with the action of a powerful hydraulic jack (the role of the jack is played by the fluid pumped under pressure). The effect of fluid saturation in combination with hydraulic fracturing reduces friction and changes the slip mode of the fault limbs at the injection sites.

Man-made pressure effects transmitted through the reservoir fluid system allow the fractured reservoir (within the hydraulic influence region) to be transferred to another stress state. These effects include the influence on the value of shear resistance in the segments of the drilled sections of the fault zones. When the seam pressure is reduced to a certain level, the permeability of the rock decreases, cracks close, and, therefore, friction and shear resistance increase. To re-open the cracks, it is necessary to re-increase the pressure of the fluid system to an appropriate level. This makes it possible to control the stress-strain state of fault segments and the mode of their seismic activity [36]. It is also important to take into account the natural ratio of the angles of inclination of fault zones with the vertical direction of gravity.

It is known that vertical cracks in the strike-slip fault zones have greater openness than inclined ones since the rock pressure on the walls of vertical cracks is much less than on the walls of inclined (and especially mildly sloping) cracks. Accordingly, the shear resistance of vertical cracks is also significantly less. In particular, during

the drilling of reservoirs within the zones of subvertical faults, seam pressure can sharply drop to the value of pore pressure in the surrounding rock mass [37]. Another situation occurs with the subhorizontal inclination of the principal slip zone of a thrust fault. Here, the natural fluid system can be at extremely high pressures due to push-up conditions. In such cases, the use of standard drilling technologies can lead to strong dynamic responses of fault zones, especially in the absence of pore fluids. Injection of drilling fluid into the zones of subhorizontal fractures can lead to a sharp decrease in shear resistance and activation of coseismic displacements.

Injection of solutions into the fault at a controlled rate allows managing pressure drop and the frictional resistance of the fault zone, as was shown by experimental studies. Such technological impacts on highly stressed fault segments can be supplemented by the use of proppants to change the slip mode. Natural proppant-fixed cracks are colmataged, which prevents them from closing. Another technique is the use of additives in drilling fluids, for example, granules of copolymers (alpine drill beads). Such additives contribute to a decrease in frictional resistance in the contact spots of asperities in the fault zone and to a decrease in the critical density of accumulated elastic strain energy at which seismic vibrations are generated. The latter leads to a decrease in the amplitude of seismic oscillations.

Local hydrodynamic and tectonic conditions must be taken into account during the drilling process. For example, natural filtering fracture simplifies fluid injection and pellet delivery into the fault through drilled fan lateral branches in horizontal fishbones. At a subhorizontal slope of the principal slip zone of the thrust fault, the natural fluid system can be at extremely high pressures due to push-up conditions. This contributes to the creation of frictional instabilities in the fault zone. Upon identifying such features by geological and geophysical data, special measures to reduce seismic risk have to be developed.

Deep drilling of injection wells to depths of 5–7 km in the fault zones shows that the injection pressure and fluid saturation are able to propagate to much deeper horizons due to hydraulic connection with existing natural fluid systems. This indicates a real possibility of man-made impacts on seismically hazardous deep fault segments through deep wells since such impacts can reach the level of the seismic focal layer in the continental crust.

The example shown below is one of the options for drilling a well-studied segment of the Tunka fault in the Baikal rift zone near the Arshan settlement (Irkutsk Region, Russia). Some volcanic apparatuses and coseismic fractures of 4 destructive earthquakes ($M_w \sim 7.5$, the average recurrence period is 3.9 ± 0.6 thousand years) were found in the vicinity of this segment. The latest paleoseismic dislocation in this area is about 1 thousand years old. We also note the recent strong earthquake of August 22, 1814, in the neighboring segment of the Tunka fault zone. The energy of the tremors was about 9 balls. Currently, there is a seismic gap in the Arshan segment of the fault, which is a warning sign of the preparation of the next stage of activation in the coming years or the first decades. Based on these assumptions, the seismic situation in the Arshan segment of the Tunka fault can be assessed as potentially dangerous.

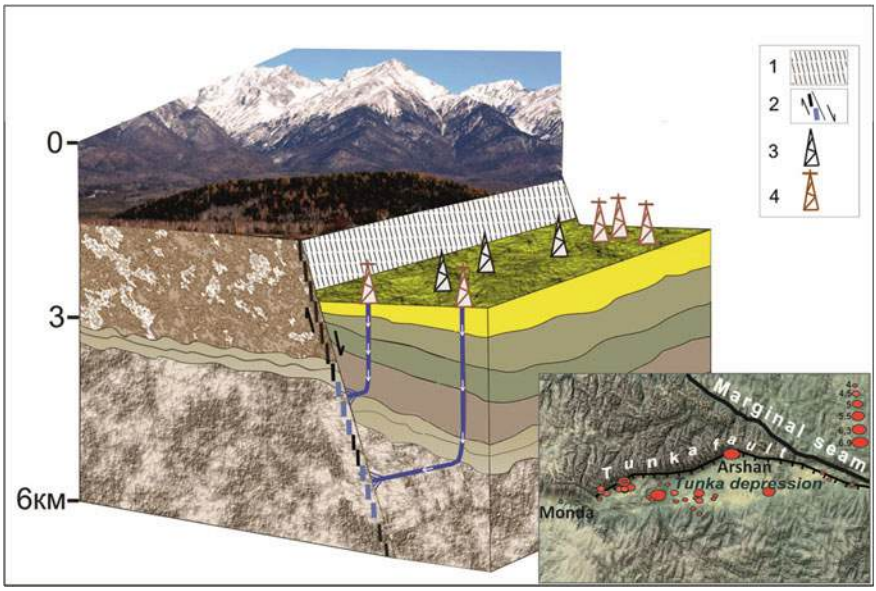


Fig. 7 The scheme of the proposed placement of two groups of inclined injection and monitoring wells for sequential hydrodynamic vibro-pulse impacts on the Arshan segment of the Tunka fault. The inset shows the central part of the Tunka rift basin, the Arshan segment of the Tunka fault zone, and the epicenters of earthquakes $M_w = 4.0–6.9$ over the past 70 years

Therefore, this segment is an appropriate place to validate the method of preventive relaxation of shear stresses using a hydraulic injection of fluids with controlled pressure into the fracture zone through a network of deep inclined or horizontal multilateral wells [38, 39]. Below is a simplified block diagram of the planned test site in the Arshan segment of the Tunka rift faulting zone (Fig. 7).

The advantages of using modern technology for driving deep wells into fault zones are determined by the possibility to control the volume and modes of injection of different solutions. Nevertheless, it is necessary to take into account the complex additional factors (geophysical, seismotectonic and geostructural), including the spatial orientation of principal slip zones with respect to the main axes of tectonic stresses, the length and thickness of fault zones, the values of dip angles, triple junctions, etc. It is also important to conduct preliminary test actions to assess the features of the "macroscopic" response of the fault zone.

6 Discussion

When discussing the results of the described interdisciplinary studies aimed at developing an approach to the man-made management of displacement modes in fault

zones, it is important to consider the similar experience of foreign scientific groups. In particular, it is necessary to take into account the known results of studying the consequences of injecting solutions through wells 3–5 km deep. The watering of fault zones sometimes led to the initiation of many weak earthquakes. Hypocenters of earthquakes extended in depth to distances commensurate with the length of the wells themselves. This indicates fluid saturation of rock masses in much deeper horizons. The authors of the monograph [40] studied indirect seismological evidence of high and rapidly changing permeability of the middle and lower crust from data on the systematic emergence of hypocenters of aftershock sequences of strong earthquakes. Such a notion can be considered as an indication of the presence of a hydraulic connection between fluids of the lower and middle crust with hydrological systems in the surface layers of the earth's crust. Thus, fluid injected through insufficiently deep wells (3–5 km) can hydraulically act on deeper fluid-saturated layers of the earth's crust down to the depths of preparation of the sources of strong earthquakes. This way is assumed in the framework of the proposed approach.

Laboratory tribological studies indicate that many classical laws of friction are valid for zones of active tectonic faults and determine the regimes of seismic activity in deep fault horizons [13, 41, 42]. This means that the regimes of radiation of seismic vibrations during tectonic creep or coseismic slip should be considered as the results of visco-elastic–plastic contact interactions between randomly rough sliding surfaces. Classical tribological laws also hold upon transitions to higher hierarchical levels of seismotectonic destruction in the lithosphere. This refers, for example, to transient frictional slip regimes from aseismic to accelerated ones, or dynamic failure in the final stage of earthquake preparation. These regimes, in turn, affect the amplitude-frequency and energy parameters of the radiated seismic vibrations. The injection of aqueous solutions into fault zones not only reduces the strength characteristics of rocks and shear resistance in slip planes but also promotes a hydraulic (wedging apart) effect. The latter causes hydraulic fracturing or slow opening of natural cracks and hence reduces the contact area and the friction coefficient. The experience gained during the operation of underground hydrocarbon storage, the extraction of geothermal resources, and the development of gas and oil fields in seismically active regions show that aqueous solutions injected through wells under high pressure can reactivate existing deep fault segments and provoke dangerous earthquakes with $M_w \geq 4.0$–5.2 [43–46]. Many authors are of the opinion that the unintentional activation of faults caused by fluid injection should be used instead to manage the relaxation of excess stresses in fault zones. To implement this in practice, it is necessary to identify and use tribological mechanisms that control slip modes and generation of seismic vibrations in fault zones [47]. It was shown above that such knowledge helps to exclude the acts of strong seismic activation of fault segments at significant distances from injection wells.

Many representatives of the scientific community believe that even energetically moderate technogenic impacts including hydraulic fracturing by injecting drilling fluids can induce strong earthquakes even in aseismic areas. Examples are geological exploration and exploitation of underground geothermal water resources in Texas

(USA) and Switzerland [48]. However, experience of deep drilling in the exploration and development of hydrocarbon deposits in aseismic regions (including the Siberian platform), indicates the absence of such dangerous phenomena. Note that even underground nuclear explosions with megaton charges were not able to trigger multiscale transitions to merge small faults into a long seismically active fault and initiate destructive earthquakes with $M_w \geq 7$ [49]. Nevertheless, we cannot exclude induced low-energy seismicity in the form of a series of moderate earthquakes. The most striking example is the recent Bachat man-provoked seismic event in Kuzbass (Russia) on 18.06.2013 ($M_w = 6.1$) with a hypocenter depth of 4 km [50]. It was initiated after long-term mining operations in coal mining at depths of up to 2 km. This extraordinary earthquake is the result of regular powerful explosions, as well as the excavation of large volumes of rock mass over an area of 100 km^2.

The magnitude of the most well-known (and rather rare) technologically induced moderately hazardous earthquakes induced by mining operations ranges within 2.8–6.1. There are some examples of the initiation of strong earthquakes with $M_w = 7$ as a result of the uncontrolled long-term withdrawal of huge volumes of oil and gas from deep-seated deposits. This kind of phenomena took place near the city Gazli or the city Neftegorsk. A huge number of such examples show that the increasing power of anthropogenic impact on the upper layers of the earth's crust leads to the unintentional provocation of dangerous seismodynamic phenomena. This makes it necessary to mobilize the efforts of the scientific community to prevent such kind of phenomena. The proposed approach can be efficiently used to solve this problem.

7 Conclusion

Uncontrolled long-term high-energy man-made impacts on highly stressed areas of the earth's crust are increasingly leading to unintended seismic disasters. We showed the possibility to prevent or neutralize such disasters by implementing proactive measures including controlled vibration-pulse actions in combination with controlled injections of solutions through deep wells. Such kind of technology is based on the approach, which integrates advanced geological and geophysical methods in combination with advanced deep drilling technologies and numerical and physical modeling. This will make it possible, in the near future, to provide a more reliable solution to the global problem of ensuring seismic safety. The results considered above provide the basis for conclusions about the feasibility of projects aimed at effectively reducing seismic risk [51].

Professor S. G. Psakhie was among the drivers of the development of the approach to managing the displacement mode and seismic activity of highly stressed fault zones. At various seminars, we discussed the prospects of the implementation of large-scale field tests using vibro-pulse and hydraulic methods of influencing segments of seismically dangerous fault zones through drilled deep wells. The most challenging problems are appropriate funding, government approvals, and the preparation of test sites in areas, which are at a safe distance from settlements. Such test

sites are available, for example, in the poorly populated territory of the Mongolian People's Republic, where seismic and geological studies have revealed a system of deep seismic hazardous faults. To create a technology aimed at solving the described multidisciplinary problem, combined efforts of the world community and the support of governments are required. The next immediate task for the implementation of such an important project is to organize a series of large-scale field tests in the segments of seismically dangerous faults using advanced deep drilling technologies.

Acknowledgements The authors are grateful to the great number of participants in the many years of research. The authors express especially grateful memory and deep gratitude to Professor S. G. Psakhie as the organizer, teacher, author, and mastermind of the studies aimed at solving the global problem of ensuring the seismic safety of the human community. The work was performed according to the Government research assignments for IEC SB RAS (V.V.R.), and for ISPMS SB RAS (E.V.S.).

References

1. Ogata Y (2017) Statistics of earthquake activity: models and methods for earthquake predictability studies. Annu Rev Earth Planet Sci 45:497–527. https://doi.org/10.1146/annurev-earth-063016-015918
2. Booth E (2018) Dealing with earthquakes: the practice of seismic engineering 'as if people mattered.' Bull Earthq Eng 6:1661–1724. https://doi.org/10.1007/s10518-017-0302-8
3. Panza FG, Kossobokov VG, Peresan A, Nekrasova K (2014) Earthquake hazard, risk and disasters. Academic Press, Cambridge, UK, pp 309–357 (Why are the Standard Probabilistic Methods of Estimating Seismic Hazard and Risks Too Often Wrong). https://doi.org/10.1016/B978-0-12-394848-9.00012-2
4. Chen X, Nakata N, Pennington C, Haffener J, Chang JC, He X, Zhan Z, Ni S, Walter JI (2017) The Pawnee earthquake as a result of the interplay among injection, faults and foreshocks. Sci Rep 7:4945. https://doi.org/10.1038/s41598-017-04992-z
5. Miller S (2013) The role of fluids in tectonic and earthquake processes. Adv Geophys 54:1–46. https://doi.org/10.1016/B978-0-12-380940-7.00001-9
6. Ruzhich VV, Psakhie SG, Shilko EV, Vakhromeev AG, Levina EA (2019) On the possibility of development of the technology for managing seismotectonic displacements in fault zones. AIP Conf Proc 2051:020261. https://doi.org/10.1063/1.5083504
7. Psakhie SG, Ruzhich VV, Shilko EV, Popov VL, Astafurov SV (2007) A new way to manage displacements in zones of active faults. Tribol Int 40:995–1003. https://doi.org/10.1016/j.triboint.2006.02.021
8. Ruzhich VV, Psakhie SG, Chernykh EN, Shilko EV, Levina EA, Dimaki AV (2018) Baikal ice cover as a representative block medium for research in lithospheric geodynamics. Phys Mesomech 21:223–233. https://doi.org/10.1134/S1029959918030062
9. Vakhromeev AG, Ivanishin VM, Sverkunov SA, Polyakov VN, Razyapov RK (2019) Deep well as a facility for on-line hydraulic studies of the stress state of the rock mass in fluid-saturated fractured reservoirs. Geodyn Tectonophys 10(3):761–778. https://doi.org/10.5800/GT-2019-10-3-0440
10. Pang X-Q, Jia C-Z, Wanf W-Y (2015) Petroleum geology features and research developments of hydrocarbon accumulation in deep petroliferous basins. Petrol Sci 12:1–53. https://doi.org/10.1007/s12182-015-0014-0
11. McGarr A (2012) Relating stick-slip friction experiments to earthquake source parameters. Geophys Res Lett 39:L05303. https://doi.org/10.1029/2011GL050327

12. Mclaskey GC, Yamashita F (2017) Slow and fast ruptures on a laboratory fault controlled by loading characteristics. J Geophys Res Solid Earth 122:3719–3738. https://doi.org/10.1002/2016JB013681
13. Popov VL (2010) Contact mechanics and friction. Physical principles and applications. Springer-Verlag Berlin Heidelberg, Berlin
14. Ruzhich VV, Kocharyan GG (2017) On the structure and formation of earthquake sources in the faults located in the subsurface and deep levels of the crust. Part I. Subsurface level. Geodyn Tectonophys 8(4):1021–1034. https://doi.org/10.5800/GT-2017-8-4-0330
15. Nielsen S (2017) From slow to fast faulting: recent challenges in earthquake fault mechanics. Philos Trans R Soc A Math Phys Eng Sci 375(2103):20160016. https://doi.org/10.1098/rsta.2016.0016
16. Zhurkov SN (1984) Kinetic concept of the strength of solids. Int J Fract 26:295–307. https://doi.org/10.1007/BF0096296.1
17. Petrov YuV, Karihaloo BL, Bratov VV, Bragov AM (2012) Multi-scale dynamic fracture model for quasi-brittle materials. Int J Eng Sci 61:3–9. https://doi.org/10.1016/j.ijengsci.2012.06.004
18. Filippov AE, Popov VL, Psakhie SG, Shilko EV (2006) Converting displacement dynamics into creep in block media. Tech Phys Lett 32:545–549. https://doi.org/10.1134/S1063785006060290
19. Scuderi MM, Collettini C (2016) The role of fluid pressure in induced vs. triggered seismicity: insights from rock deformation experiments on carbonates. Sci Rep 6:24852. https://doi.org/10.1038/srep24852
20. Cornelio C, Spagnuolo E, Di Toro G, Nielsen S, Violay M (2019) Mechanical behaviour of fluid-lubricated faults. Nat Commun 10:1274. https://doi.org/10.1038/s41467-019-09293-9
21. Loreto MF, Düşünür-Doğan D, Üner S, İşcan-Alp Y, Ocakoğlu N, Cocchi L, Muccini F, Giordano P, Ligi M (2019) Fault-controlled deep hydrothermal flow in a back-arc tectonic setting, SE Tyrrhenian Sea. Sci Rep 9:17724. https://doi.org/10.1038/s41598-019-53696-z
22. Ruzhich VV, Kocharyan GG, Travin AV, Saveleva VB, Ostapchuk AA, Rasskazov SV, Yasnygina TA, Yudin DS (2018) Determination of the PT conditions that accompanied a seismogenic slip along a deep segment of the marginal suture of the Siberian Craton. Dokl Earth Sci 481:1017–1020. https://doi.org/10.1134/S1028334X18080081
23. Sobolev GA, Vettegren' VI, Ruzhich VV, Ivanova LA, Mamalimov RI, Shcherbakov IP (2015) A study of nanocrystals and the glide-plane mechanism. J Volcanol Seismol 9:151–161. https://doi.org/10.1134/S0742046315030057
24. Vettegren VI, Ponomarev AV, Sobolev GA, Shcherbakov IP, Mamalimov RI, Kulik VB, Patonin AV (2017) Structural changes in the surface of a heterogeneous nanocrystalline body (sandstone) under the friction. Phys Solid State 59:588–593. https://doi.org/10.1134/S1063783417030313
25. Medvedev VY, Ivanova LA, Lysov BA, Ruzhich VV, Marchuk MV (2014) Experimental study of decompression, permeability and healing of silicate rocks in fault zones. Geodyn Tectonophys 5(4):905–917. https://doi.org/10.5800/GT-2014-5-4-0162
26. Ruzhich VV, Kocharyan GG, Saveleva VB, Travin A (2018) On the structure and formation of earthquake sources in the faults located in the subsurface and deep levels of the crust. Part II. Deep level. Geodyn Tectonophys 9(3):1039–1061. https://doi.org/10.5800/GT-2018-9-3-0383
27. Ruzhich VV, Truskov VA, Chernykh EN, Smekalin OP (1999) Neotectonic movements in fault zones of the Baikal region and their origin mecha-nisms. Russ Geol Geophys 40(3):356–368
28. Astafurov SV, Shilko EV, Psakhie SG, Rhuzich VV (2008) Effect of local stress on the interface response to dynamic loading in faulted crust. Russ Geol Geophys 49(1):52–58. https://doi.org/10.1016/j.rgg.2007.12.007
29. Psakhie SG, Popov VL, Shilko EV et al (2006) A method for controlling the displacement mode in fragments of seismically active tectonic faults. RF Patent 2273035, 27 Mar 2006
30. Bürgmann R (2018) The geophysics, geology and mechanics of slow fault slip. Earth Planet Sci Lett 495:112–134. https://doi.org/10.1016/j.epsl.2018.04.062
31. Michel S, Gualandi A, Avouac J-P (2019) Similar scaling laws for earthquakes and Cascadia slow-slip events. Nature 574:522–526. https://doi.org/10.1038/s41586-019-1673-6

32. Ostapchuk AA, Pavlov DV, Ruzhich VV, Gubanova AE (2019) Seismic-acoustics of a block sliding along a fault. Pure Appl Geophys. https://doi.org/10.1007/s00024-019-02375-1
33. Rebinder PA, Shchukin ED (1973) The surface phenomena in solids during the course of their deformation and failure. Soviet Physics Uspeki 15(5):533–554
34. Traskin VYu (2009) Rehbinder effect in tectonophysics. Izv Phys Solid Earth 45:952. https://doi.org/10.1134/S1069351309110032
35. Vakhromeev AG, Sizykh VI (2006) The role of nappe tectonics in the development of abnormally high formation pressure and economic metalliferous brines: a case study of the southern Siberian craton. Dokl Earth Sci 407:209–212. https://doi.org/10.1134/S1028334X06020115
36. Vakhromeev AG, Sverkunov SA, Ivanishin VM, Razyapov RK, Danilova EM (2017) Geodynamic aspects in the study of complex mining and geological conditions for drilling into oil-and-gas reservoirs in the riphean carbonate rocks: an overview of the problem as exemplified by the deposits in the Baikit petroliferous district. Geodyn Tectonophys 8(4):903–921. https://doi.org/10.5800/GT-2017-8-4-0323
37. Vakhromeev AG, Sverkunov SA, Siraev RU, Razyapov RK, Sotnikov AK, Chernokalov KA (2016) The primary method of drilling a horizontal hole in a fracture type of oil and gas saturated carbonate reservoir under conditions of abnormally low reservoir pressure. RF Patent 2602437, 20 Nov 2016, Bull. No. 32 (in Russian)
38. Ruzhich VV, Psakhie SG, Shilko EV, Vakhromeev AG, Levina EA (2018) On the possibility of development of the technology for managing seismotectonic displacements in fault zones. AIP Conf Proc 2051:020261. https://doi.org/10.1063/1.5083504
39. Ma T, Chen P, Zhao J (2016) Overview on vertical and directional drilling technologies for the exploration and exploitation of deep petroleum resources. Geomech Geophys Geo-Energy and Geo-Resour 2:365–395. https://doi.org/10.1007/s40948-016-0038-y
40. Rodkin MV, Rundquist DV (2017) Geofluidogeodynamics. Application to seismology, tectonics, and processes of ore and oil genesis. Publishing House "Intellect", Dolgoprudny (RU)
41. Popov VL, Grzemba B, Starcevic J, Popov M (2012) Rate and state dependent friction laws and the prediction of earthquakes: what can we learn from laboratory models? Tectonophysics 532–535:291–300. https://doi.org/10.1016/j.tecto.2012.02.020
42. Kocharyan GG, Novikov VA (2016) Experimental study of different modes of block sliding along interface. Part 1. Laboratory experiments. Phys Mesomech 19:189–199. https://doi.org/10.1134/S1029959916020120
43. Sibson RH (1973) Interactions between temperature and pore fluid pressure during an earthquake faulting and a mechanism for partial or total stress relief. Nat Phys Sci 243:66–68. https://doi.org/10.1038/physci243066a0
44. Bachmann CE, Wiemer S, Woessner J, Hainzl S (2011) Statistical analysis of the induced Basel 2006 earthquake sequence: introducing a probability-based monitoring approach for Enhanced Geothermal Systems. Geophys J Int 186(2):793–807. https://doi.org/10.1111/j.1365-246x.2011.05068.x
45. Guglielmi Y, Cappa F, Avouac J-P, Henry P, Elsworth D (2015) Seismicity triggered by fluid injection-induced aseismic slip. Science 348(6240):1224–1226. https://doi.org/10.1126/science.aab0476
46. Weingarten M, Ge S, Godt JW, Bekins BA, Rubinstein JL (2015) High-rate injection is associated with increase in U.S. mid-continent seismicity. Science 348(6241):1336–1340. https://doi.org/10.1126/science.aab1345
47. Rutqvist J, Rinaldi AP, Cappa F, Jeanne P, Mazzoldi A, Urpi L, Guglielmi Y, Vilarrasa V (2016) Fault activation and induced seismicity in geologic carbon storage—lessons learned from recent modeling studies. J Rock Mech Geotech Eng 8(6):789–804. https://doi.org/10.1016/j.jrmge.2016.09.001
48. Rinaldi AP, Rutqvist J, Cappa F (2014) Geomechanical effects on CO_2 leakage through fault zones during large-scale underground injection. Int J Greenhouse Gas Control 20:117–131. https://doi.org/10.1016/j.ijggc.2013.11.001

49. Tarasov NT, Tarasova NV (1995) Earthquakes induced by underground nuclear explosions. Environmental and ecological problems. NATO ASI Series (2. Environment), vol 4. Springer-Verlag Berlin, Heidelberg, Berlin, pp 215–223 (Response of seismoactive medium to nuclear explosions)
50. Emanov AF, Emanov AA, Fateev AV, Leskova EV, Shevkunova EV, Podkorytova VG (2014) Mining-induced seismicity at open pit mines in Kuzbass (Bachatsky earthquake on June 18, 2013). J Min Sci 50:224–228. https://doi.org/10.1134/S1062739114020033
51. Mirzoev K, Nikolaev AV, Lukk AA, Yunga SL (2009) Induced seismicity and the possibilities of controlled relaxation of tectonic stresses in the earth's crust. Izv Phys Solid Earth 45:885–904. https://doi.org/10.1134/S1069351309100061

Particle-Based Approach for Simulation of Nonlinear Material Behavior in Contact Zones

Evgeny V. Shilko, Alexey Yu. Smolin, Andrey V. Dimaki, and Galina M. Eremina

Abstract Methods of particles are now recognized as an effective tool for numerical modeling of dynamic mechanical and coupled processes in solids and liquids. This chapter is devoted to a brief review of recent advances in the development of the popular particle-based discrete element method (DEM). DEM is conventionally considered as a highly specialized technique for modeling the flow of granular media and the fracture of brittle materials at micro- and mesoscopic scales. However, in the last decade, great progress has been made in the development of the formalism of this method. It is largely associated with the works of the scientific group of Professor S. G. Psakhie. The most important achievement of this group is a generalized formulation of the method of homogeneously deformable discrete elements. In the chapter, we describe keystones of this implementation of DEM and a universal approach that allows one to apply various rheological models of materials (including coupled models of porous fluid-saturated solids) to a discrete element. The new formalism makes possible qualitative expansion of the scope of application of the particle-based discrete element technique to materials with various rheological properties and to the range of considered scales form microscopic to macroscopic. The capabilities of this method are especially in demand in the study of the features of contact interaction of materials. To demonstrate these capabilities, we briefly review two recent applications concerning (a) the effect of adhesive interaction on the regime of wear of surface asperities under tangential contact of bodies and (b) the nonmonotonic dependence of the stress concentration in the neck of the human femur on the dynamics of hip joint contact loading.

Keywords Discrete element method · Deformable element · Movable cellular automata · Many-body interaction · Plasticity · Poroelasticity · Surface adhesion · Adhesive wear · Bone tissue · Pore fluid

E. V. Shilko (✉) · A. Yu. Smolin · A. V. Dimaki · G. M. Eremina
Institute of Strength Physics and Materials Science SB RAS, 634055 Tomsk, Russia
e-mail: shilko@ispms.tsc.ru

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_4

1 Introduction

Starting with the classic works of Cauchy and Navier [1, 2], the development of the formalism of the discrete representation of the medium at a continuum ("super-atomic") scale is considered among the fundamental problems for the mechanics of solids. In the framework of this representation, a matter is described by an ensemble of interacting particles. Each particle models a sufficient number of atoms or molecules to describe the state and response of the particle in terms of thermodynamic parameters and classical mechanical models (Fig. 1). A discrete description of solids and liquids was initially considered as a way to fill the gap between molecular mechanics and continuum mechanics [3]. However, the rapid development of the formalism of particle methods in the last two decades has made it possible to apply them to study the mechanical behavior of diverse solids as well as various mechanically assisted or activated processes in the entire spectrum of spatial scales from atomic to macroscopic.

The traditional approach to the numerical study of the behavior of materials on the "above-atomic" spatial scales is based on the methods of continuum mechanics such as finite element, finite difference and boundary element methods (FEM, FDM and BEM) [4–9]. The formalism of these methods allows easy implementation of various linear and nonlinear (including coupled thermomechanical and poroelastic) rheological models. Moreover, advanced implementations of FEM, FDM and BEM include the ability to directly model fracture [10, 11]. Despite the well-known advantages of continuum numerical methods, their fundamental limitation is difficulty in modeling of complex fracture-related problems including development of multiple fractures, contact interaction of the initial and newly formed surfaces, wear of surface layers and a change in surface roughness, flow of granular media, etc.

Mentioned limitation is not inherent in particle-based methods [12–19]. The most relevant and efficient particle-based method for numerical modeling of the above-said complex mechanical processes in solids is the discrete element method (DEM) [18, 19]. The constantly growing interest in this numerical technique is determined by the ability to solve a variety of complex and non-linear contact problems, where the processes of fragmentation and mass transfer of fragments play a key role. This

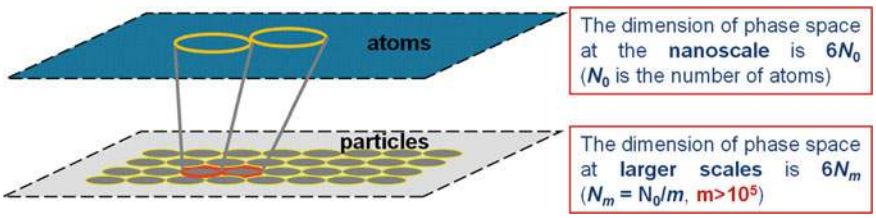


Fig. 1 Representation of the material by ensemble of particles as an extension of atomistic representation to higher spatial (and temporal) scales

chapter is devoted to the analysis of achievements in the development of DEM for computer simulation of the mechanical behavior of consolidated solids.

Creation of this method is attributed to Cundall [20, 21]. In the framework of original implementation of DEM, a discrete element is treated as a finite part of a solid body (or a particle in particulate/granular material) bounded by clearly defined (exact) surface. The latter qualitatively distinguishes this method from the methods of quasiparticles with "fuzzy" surface. Elements can be either chemically bonded if they model a consolidated material, or contact if the contact interaction of the fragments is modeled. Changing the type of bond of elements (chemically bonded $\leftrightarrow$ contact $\leftrightarrow$ noninteracting) is governed by the applied criteria of fracture, bond formation and contact loss [22, 23]. The stress state of a discrete element is determined by the mechanical load of the surrounding elements on the surface of the element. The absence of a constraint in the form of the continuity equation makes the DEM extremely attractive for numerical studying complex processes in contact zones.

The term "discrete element method" is now used as a generic name for a large group of numerical techniques based on these general principles for representing the medium [18, 21, 24, 25]. Various representatives of this group differ in several key features: the principle of local or global force equilibrium (explicit or implicit formulation); approximation of the shape of the volume modeled by a discrete element; approximation to the description of the deformability of a discrete element.

Explicit DEM is the most popular and is widely used to solve fracture- and contact interaction related dynamic problems. It implies the formulation of equations of motion for each discrete element and the parallel solution of the system of these equations with an explicit time marching scheme (Euler, Verlet or other integration algorithm). Since a discrete element simulates a finite volume of material, the mechanical interaction of such finite volumes should lead not only to their translational motion, but also to rotation determined by the moments of interaction forces. The form of the dynamics equations for the rotational degrees of freedom of an element (Euler's equations) is determined by the shape of an element. In the general case, Euler's equations are written in integral form using the inertia tensor (tensor of inertia, in turn, is formulated as an integral) [18]. Integral Euler equation is cumbersome and computationally costly, that is why elements with complex nonequiaxial geometry (polygonal [26], superquadric [27] elements) are used only for modeling granular or fragmented (block-structured) materials. At the same time, a consolidated material can be much more efficiently modelled by an ensemble of bonded equiaxial elements, the shape of which is approximated by equivalent sphere (3D problem) or a disk of a given height (2D problem) [18, 22, 28]. The efficiency of an approximation of an equivalent disk/sphere is determined by the simplicity of the Newton–Euler motion equations for a discrete element:

$$\begin{cases} m_i \frac{d^2\vec{r}_i}{dt^2} = m_i \frac{d\vec{v}_i}{dt} = \vec{F}_i = \sum\limits_{k=1}^{N_i} \vec{F}_{ik} = \sum\limits_{k=1}^{N_i} \left(\vec{F}^c_{ik} + \vec{F}^t_{ik} \right) \\ J_i \frac{d\vec{\omega}_i}{dt} = \vec{M}_i = \sum\limits_{k=1}^{N_i} \vec{M}_{ik} \end{cases}, \tag{1}$$

where i is the number of discrete element, $\vec{r}_i$, $\vec{v}_i$ and $\vec{\omega}_i$ are the radius-vector, velocity vector and angular velocity pseudovector respectively, m_i is the mass of the element i, J_i is moment of inertia of an equivalent disk or sphere, $\vec{F}_i$ and $\vec{M}_i$ are the total force and torque acting on the element i by the neighbors, $\vec{F}_{ik}$ is the force of interaction of the considered element i with the neighbor k, $\vec{F}^c_{ik}$ and $\vec{F}^t_{ik}$ are the central (along the line connecting mass centers of the elements i and k) and tangential (in transverse plane) components of the force $\vec{F}_{ik}$, $\vec{M}_{ik}$ is the moment of interaction forces (includes tangential moment of $\vec{F}^t_{ik}$ and twisting moment [22]), N_i is the number of neighbors of the element i.

One can see that the approximation of the equivalent disk/sphere allows the use of Euler's equations in the most trivial and computationally efficient form. Another important consequence of this approximation is the formal independence of the central and tangential interactions: the force $\vec{F}^c_{ik}$ does not cause acceleration in the plane of the tangential interaction, and the force $\vec{F}^t_{ik}$ does not cause acceleration along the line connecting the centers of mass of the elements.

The forces $\vec{F}^c_{ik}$ and $\vec{F}^t_{ik}$ of interaction of discrete elements are traditionally represented as the sum of the potential ($\vec{F}^{cp}_{ik}$ and $\vec{F}^{tp}_{ik}$) and viscous ($\vec{F}^{cv}_{ik}$ and $\vec{F}^{tv}_{ik}$) constituents [20, 22]. From the point of view of the rheological description, viscous forces have a meaning similar to the physical meaning of the damper in the Kelvin-Voight viscoelastic model. A key component of constructing a discrete-element model of a material is the determination of the structural type and coefficients of the potential interaction forces.

In the framework of the traditional DEM implementation, the central and tangential potential forces of interaction of equivalent balls/disks (F^{cp}_{ik} and F^{tp}_{ik}) are calculated in the pair-wise approximation. From the physical point of view, pair-wise potential corresponds to the approximation of a non-deformable (rigid) element (a system of springs or rods). Here, the term "rigid" is used in the sense that the interaction of the element i with the neighbor k does not change its volume and shape and, therefore, does not cause a change in the forces of interaction of the element i with other neighbors (these neighbors "feel" the result of the interaction in the pair $i-k$ only indirectly, through the motion of the element i). This approximation is widely used for micro- and mesoscopic description of the processes of damage accumulation and fracture of brittle materials, contact interaction of elastic bodies, including the dynamics of block-structured media. In this case, element-element interaction is traditionally modelled using harmonic interaction potentials (such an interaction is schematically represented by connecting the centers of the elements with two springs oriented in the central and tangential directions) [22, 28, 29]. Some models also use nonlinear (elastic–plastic or viscoelastic Maxwell type) formulations of pair-wise interaction forces F^{cp}_{ik} and F^{tp}_{ik} [30, 31] for simulation of granular media and porous structures with non-linear/ductile rheological properties of the material of the skeleton walls or granules. However, such potentials make possible adequate description of the mechanical behavior of porous systems only at a "low" scale (the scale of discontinuities or granules).

The key problems that strongly limit the range of application of the traditional implementation of DEM with pair-wise interaction forces are well known. They are (i) dependence of the macroscopic properties of an ensemble of discrete elements on the type of packaging and size distribution of elements, (ii) incorrect description of the plastic strain of an ensemble of elements (for example, plastic deformation of a sample may be accompanied by an uncontrolled change in its volume), and other related problems.

Various approaches to solving these fundamental difficulties within the framework of the concept of non-deformable elements have been proposed in last decades. In particular, stochastic dense packing of non-uniform-sized circular (2D) or spherical (3D) elements [22, 28, 29] is used to solve the problems of packing-induced anisotropy of the elastic response and packing-dependent ratio of elastic modules of an ensemble of elements. An alternative approach is to use the formalism of spring network model (lattice model [32, 33]) to build relationships for the forces of interaction of regularly packed uniform-sized elements [34–36]. The lattice model is based on the postulation of the form of interaction potential (harmonic potential is usually used for both central and angular interactions) and equalization of elastic strain energy stored in a unit cell of volume to the associated elastic strain energy of the modelled continuum. The material parameters derived from this equality are included in the relationships for the forces of element-element interaction. The above approaches made it possible to adequately describe the mechanical (and thermomechanical [36]) behavior of brittle materials under complex loading conditions. At the same time, they do not allow solving the key problem of incorrect modeling of nonlinear (and/or inelastic) mechanical behavior of materials with complex rheological properties (including rubber-like viscoelastic materials as well as metallic and polymer materials, whose macroscopic plasticity is not related to discontinuities).

The problem of correct modeling of nonlinear mechanical behavior of consolidated materials by the method of discrete elements can be generally solved only by using the approximation of the deformable element. In turn, the deformability of an element can be realized only within the framework of a many-body interaction of elements. This means that the potential interaction force must depend not only on the relative motion of the elements in the pair, but also on the interaction of each of them with other neighbors. The formulation of the general structural form of the potential interaction force and its specific realizations for materials with various rheological properties has been among the critical challenges for the DEM until recently.

2 Distinct Element Method with Deformable Elements

A meaningful contribution to the development of the formalism of DEM was done by Professor Sergey G. Psakhie and his team. Professor Psakhie was a founder of the new particle-based method, namely, the method of movable cellular automata (MCA) [37, 38].The basic principles of this method were developed in collaboration with Professor Yuki Horie (North Carolina State University, Los Alamos National

Lab). Originally, the MCA method was designed as a hybrid technique to model mechanically activated chemical reactions in powder mixtures [39, 40]. This original implementation combined the formalisms of discrete elements and cellular automata, in which the mechanical response of the particle was described using the DEM formalism, while the non-mechanical thermodynamic aspects of particle–particle interaction (including melting and mechanically activated chemical reaction) were modelled on the basis of the concept of cellular automata.

The most important achievement of S. G. Psakhie in the development of numerical particle-based modelling techniques is the proposed general formalism of the method of homogeneously (simply) deformable discrete elements.

The keystones of this formalism were laid in the framework of collaboration with Professor Valentin L. Popov (Technische Universität Berlin) and his scientific group. In a joint work of Professors Psakhie and Popov [41], the basic principles for describing the mechanical behavior of a discrete element (movable cellular automaton) as a deformable area of the medium were formulated. For a special case of an ensemble of close packed elements of the same equivalent radius, which models an isotropic two-dimensional continuum, a relation was proposed for the potential force of the central interaction of elements in the many-body approximation:

$$F_{ik}^{cp} = E^* \delta L_{ik} = E^* \left(\delta r_{ik} + D \sum_{j=1}^{N} \delta r_{ij} + D \sum_{m=1}^{N} \delta r_{km} \right). \tag{2}$$

Here, the symbol δ denotes the difference between the current and initial values of the corresponding parameter, $r_{ik} = |\vec{r}_i - \vec{r}_k|$ is the distance between the centers of mass of the elements i and k, L_{ik} is the effective distance between the elements, N is the number of neighbors in the first coordination sphere. The coefficients E^* and D are expressed in terms of the elastic constants of the material and the element packing parameters. Derivation of these coefficients is based on the condition for ensuring the required values of Young's modulus and Poisson's ratio of the material [41]. This model is based on the same principles as the classic spring network models, but it has a fundamental difference. The force of the central interaction of the two elements is represented in the form of a superposition of the pair-wise component $E^* \delta r_{ik}$ and "hydrostatic" components. The latter are proportional to the change in the volumes of the interacting elements (here, we use the particular form of expression for the element's volume change in regular packing). The reasonableness of this formulation is confirmed by the linear relationship of the diagonal components of the stress tensor and the volume strain in the vast majority of macroscopic rheological models of materials (linear and non-linear elasticity, viscoelasticity, plasticity).

The proposed formalism actually uses the approximation of homogeneously deformable elements. It was further developed to describe plastic flow of the elements based on constitutive equations of the macroscopic continuum theory of defects [41]. Despite the clear advantages of the proposed formalism, it has the same key limitations as traditional lattice-based models. Among them, are the absence of the

tangential interaction of elements (shear resistance force), packing-dependent artificial anisotropy of the integral response of the ensemble of elements at a significant distortion of the initial symmetry of the lattice, and the lack of a general and simple algorithm of implementation of complex rheological material models.

A generalized formulation of the method of homogeneously deformable discrete elements was proposed later in the works of Professor S. G. Psakhie with co-authors. It applies the concept of many-body interaction for the ensemble of arbitrarily packed different-size elements and is based on the following principles:

1. Approximation of equivalent disks/spheres (Fig. 2). Within the framework of this approximation, the dynamics of elements is described by Eq. (1), and the forces of central and tangential interaction are assumed to be formally unrelated to each other. Elements interact with each other through flat contact areas. The geometry and squares of these areas are determined by the local packing and sizes of the elements [22, 42, 43]. The potential interaction of the two elements is conveniently described in terms of specific forces of interaction (normal σ_{ik} and tangential τ_{ik} contact stresses):

$$\begin{cases} F_{ik}^{cp} = \sigma_{ik} S_{ik} \\ F_{ik}^{tp} = \tau_{ik} S_{ik} \end{cases}. \tag{3}$$

2. A discrete element is assumed to be homogeneously deformable, i.e., its stress–strain state is characterized by tensors of stress and strain (hereinafter called

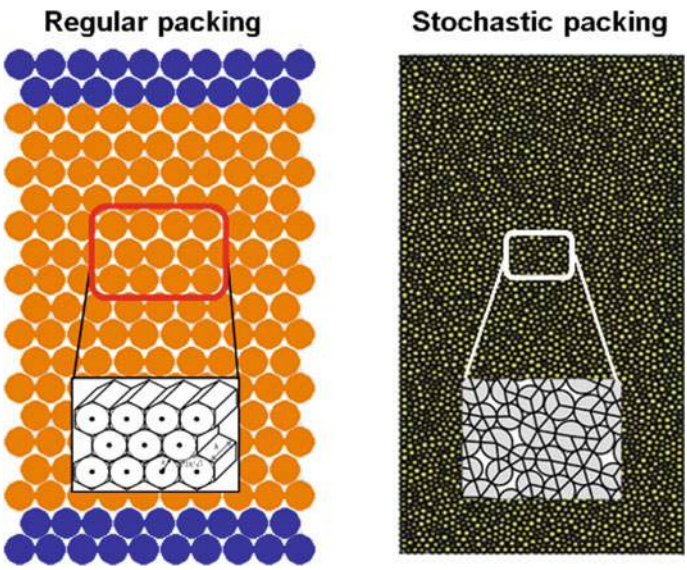


Fig. 2 Typical examples of 2D samples modelled by the ensembles of regularly and stochastically (dense) packed discrete elements. Figures show equivalent disks

average stresses $\overline{\sigma}_{\alpha\beta}$ and strains $\overline{\varepsilon}_{\alpha\beta}$, where $\alpha,\ \beta = x,\ y,\ z$). To determine the components of average stress tensor, we use well-known classical relations for average stresses and "local" values of traction vectors at the contact areas of the element with neighbors (stress homogenization) [22, 42, 44]:

$$\overline{\sigma}^{i}_{\alpha\beta} = \frac{R_i}{\Omega^0_i}\sum_{k=1}^{N_i} S^0_{ik}(\vec{n}_{ik})_\alpha\left(\vec{\Sigma}_{ik}\right)_\beta = \frac{R_i}{\Omega^0_i}\sum_{k=1}^{N_i} S^0_{ik}(\vec{n}_{ik})_\alpha\left(\sigma_{ik}(\vec{n}_{ik})_\beta + \tau_{ik}\left(\vec{t}_{ik}\right)_\beta\right), \tag{4}$$

where R_i is the radius of equivalent sphere approximating the element i, Ω^0_i is the volume of unstrained element, S^0_{ik} is the contact square in unstrained pair $i-k$, $\vec{\Sigma}_{ik}$ is the traction vector at the area of contact of elements i and k (normal and tangential contact stresses are the components of this vector), $\vec{n}_{ik}$ is the unit normal vector directed along the line connecting the mass centers of the elements, $\vec{t}_{ik}$ is the unit tangent vector directed in the tangential plane, $(\vec{W})_\beta$ is the projection of some vector $\vec{W}$ onto the β-axis. We emphasize the generality of definition (4), which is applicable for arbitrary local packing of elements of various sizes (packing and sizes determine the values of the direction cosines and contact areas).

3. A consequence of the deformability of an element is the need to divide the spatial parameters of its interaction with a neighbor (pair overlap and relative tangential displacement) into two components, namely, the contributions of both elements:

$$\begin{cases} \Delta r_{ik} = \Delta q_{ik} + \Delta q_{ki} = R_i\Delta\varepsilon_{ik} + R_k\Delta\varepsilon_{ki} \\ \Delta l^{sh}_{ik} = R_i\Delta\gamma_{ik} + R_k\Delta\gamma_{ki} \end{cases}, \tag{5}$$

where q_{ik} and q_{ki} are the distances from the mass centers of the interacting elements i and k to the central point of the contact area (they are equal to equivalent radii R_i and R_k respectively for the case of unstrained elements), ε_{ik} and ε_{ki} are central pair strains of discrete elements i and k, l^{shear}_{ik} is the value of relative tangential displacement of the elements (it is calculated with taking into account the element rotations [22, 39, 42]), γ_{ik} and γ_{ki} are the shear angles of discrete elements i and k (contributions to the total shear angle). In the general case, $\varepsilon_{ik} \neq \varepsilon_{ki}$ and $\gamma_{ik} \neq \gamma_{ki}$. Relations (5) are given in the incremental form [hereinafter the symbol Δ denotes an increment of some parameter over the time step of numerical integration of the motion equations (1)] because this form is convenient for the numerical implementation of complex rheological models. Note that the strains ε_{ik} and γ_{ik} are the components of the "local" strain vector, which is used in the definition of $\overline{\varepsilon}^{i}_{\alpha\beta}$ by the analogy to (4).

4. In the framework of approximation of deformable element, the specific normal and tangential forces (contact stresses) σ_{ik} and τ_{ik} are interpreted as the components of the specific force of mechanical response of the element i to the mechanical loading by the neighboring element k. These stresses are the functions of the

i-th element strains ε_{ik} and γ_{ik} in the pair $i-k$. We proposed the general form of these functions, which assumes homogeneous deformability of the element and linear relation between volume strain and mean stress (or pressure) [42–44]:

$$\begin{cases} \sigma_{ik} = \sigma_{ik}^{pair}(\varepsilon_{ik}, \dot{\varepsilon}_{ik}) + B_i \overline{\sigma}^i_{mean} = \sigma_{ik}^{pair}(\varepsilon_{ik}, \dot{\varepsilon}_{ik}) - B_i \overline{P}_i \\ \tau_{ik} = \tau_{ik}^{pair}(\gamma_{ik}, \dot{\gamma}_{ik}) \end{cases}. \tag{6}$$

Here, the upper index "*pair*" denotes pair-wise function, $\overline{\sigma}^i_{mean} = -\overline{P}_i = \left(\overline{\sigma}^i_{xx} + \overline{\sigma}^i_{yy} + \overline{\sigma}^i_{zz}\right)/3$, B_i is the material parameter. The first relation in (6) suggests that the normal (compressive/tensile) resistance of an element is determined by both the strain of this element ε_{ik} along the loading axis and the hydrostatic (liquid-like) component. The second relation is written in the pair-wise approximation, which ideologically corresponds to the relations connecting the off-diagonal components of the stress and strain (or strain rate) tensors in most rheological models of solids. The specific form of the pair-wise components σ_{ik}^{pair} and τ_{ik}^{pair} as well as the values of the material coefficients are determined by the applied rheological model of the material modeled by a discrete element. The necessity to satisfy Newton's third law ($\sigma_{ik} = \sigma_{ki}$ and $\tau_{ik} = \tau_{ki}$) leads to the following systems of equations, which are used to calculate the current value of interaction forces $\vec{F}_{ik}^{cp}$ and $\vec{F}_{ik}^{tp}$ for the motion equation (1):

$$\begin{cases} \sigma_{ik}^{pair}(\varepsilon_{ik}, \dot{\varepsilon}_{ik}) + B_i \overline{\sigma}^i_{mean} = \sigma_{ki}^{pair}(\varepsilon_{ki}, \dot{\varepsilon}_{ki}) + B_i \overline{\sigma}^i_{mean} \\ \Delta r_{ik} = \Delta q_{ik} + \Delta q_{ki} = R_i \Delta \varepsilon_{ik} + R_k \Delta \varepsilon_{ki} \end{cases}, \tag{7}$$

$$\begin{cases} \tau_{ik}^{pair}(\gamma_{ik}, \dot{\gamma}_{ik}) = \tau_{ki}^{pair}(\gamma_{ki}, \dot{\gamma}_{ki}) \\ \Delta l_{ik}^{sh} = R_i \Delta \gamma_{ik} + R_k \Delta \gamma_{ki} \end{cases}. \tag{8}$$

The solutions to each pair of equations are the values of pair strains. These strains are then used to calculate the current values of the forces of interaction of elements according to (6).

5. A pair of elements modeling a part of a consolidated material is assumed to be chemically bonded (linked). The central interaction of linked elements includes resistance to compression and tension, and the tangential interaction typically takes into account shear and bending resistance [42]. In the framework of the discrete element method, the elementary act of fracture at the considered spatial scale is the breaking of the chemical bond between the two elements. The condition of linked-to-unlinked transition is governed by a specified fracture criterion for a pair. This criterion is determined as a fracture condition at the contact area. Most fracture criteria in the mechanics of a deformable solid are formulated in force-like form in terms of the invariants of the stress tensor (Mises, Mohr–Coulomb, Drucker-Prager and other failure criteria). We proposed an approach to implementation of such kind of criteria within the framework of the formalism of deformable elements [42, 44]. It is based on determining the local stress tensor on the contact area of the linked pair of elements and

calculating its invariants. The local stress tensor $\sigma^{ik}_{\alpha'\beta'}$ is determined in the local coordinate system of the pair $i - k$. The specific forces σ_{ik} and τ_{ik} are used as the diagonal and off-diagonal components of this tensor. The missing 4 components are determined on the contact surface by linear interpolation of the corresponding components of average stress tensors in the interacting elements: $\sigma^{ik}_{\alpha'\beta'} = \left(\overline{\sigma}^{i}_{\alpha'\beta'} q_{ki} + \overline{\sigma}^{k}_{\alpha'\beta'} q_{ik}\right) / r_{ik}$. Here the accent means that average stresses are considered in the local coordinate system.

A bond break leads to a change in the interaction in a pair of discrete elements: the central interaction includes only compression resistance, and the dry [22] or viscous friction force is usually used as the tangential force. A pair becomes non-interacting if the value of the central force becomes equal to zero. We also note that the deformability of elements leads to a generalized formulation of contact detection condition, which takes into account a change in the linear dimensions of elements [42, 44].

6. Contact interaction of unlinked discrete elements is traditionally treated as non-adhesive. However, in many real systems, the adhesion of surfaces is an important factor determining the laws of friction and wear [45]. To adequately model the adhesive contact of surfaces, we assume attractive normal force acting between elements even after they are debonded [46, 47]. This force varies with separation of the surfaces of interacting elements according to prescribed model of adhesion (Dugdale's, Van-der-Vaals or other interaction potential). The value of separation is determined with taking into account deformation of elements along the normal $\vec{n}_{ik}$. The tangential force of interaction of unlinked and noncontact elements is assumed to be zero.

   At large values of surface energy, the mechanical contact of chemically clean and smooth surfaces can be accompanied by the formation of a chemical bond (this effect is often called cold welding). This effect is taken into account in the method of deformable discrete elements based on the use of the criterion of unlinked-to-linked transition for contacting pairs of elements. The pair of elements becomes linked if the criterion is satisfied. We proposed some formulations of such kind of criteria including critical values of the contact normal stress and plastic work of deformation [42].

3 Principles of Implementation of Rheological Models

The most important advantage of the proposed general formulation of element-element interaction (6)–(8) is the possibility of simple implementation of various rheological models of solids. One can see from relation (4) that the components of average stress tensor are linearly related to the forces of interaction of the element with its neighbors. In turn, the interaction forces are linearly related to the components of average stress tensor by relation (6). A fully similar interconnection takes place between average strains and pair strains of the element. We first showed that the similarity of the relationship between average and local (contact) stress and strain

parameters of the element inevitably leads to the fact that the specific formulation of relations (6) should be similar to the formulation of the constitutive relations for the material of a discrete element. In other words, the relation for the specific central force σ_{ik} of the ith element response to the mechanical loading by the neighbor k has to be formulated by means of direct rewriting of the constitutive equation for the diagonal components of the stress tensor. The specific force of tangential response τ_{ik} is formulated by direct rewriting the corresponding constitutive relation for off-diagonal stress components. Using these principles, we implemented macroscopic mechanical models of elasticity, viscoelasticity, and plasticity and applied them to study the behavior of various materials, including metals, ceramics, composite materials, rocks, rubbers, and even bone tissues [38, 42–44].

Below there is an example of the relations (6) for the quite general case of locally isotropic viscoelastic (described by the Prony series) material of the element. These relations are written in an incremental form:

$$\begin{cases} \Delta\sigma_{ik} = 2G_{i,\Sigma}\Delta\varepsilon_{ik} + \left(1 - \dfrac{2G_i}{3K_i}\right)\Delta\overline{\sigma}^{i}_{mean} \\ \qquad\qquad - 2\sum\limits_{p}\dfrac{G_{i,Mp}}{\eta_{i,Mp}}\Delta\sigma_{ik,Mp} + 2\dfrac{\Delta\overline{\sigma}^{i}_{mean}}{3K_{i,\Sigma}}\sum\limits_{p}\dfrac{G_{i,Mp}K_{i,Mp}}{\eta_{i,Mp}}. \\ \Delta\tau_{ik} = 2G_{i,\Sigma}\Delta\gamma_{ik} - 2\sum\limits_{p}\dfrac{G_{i,Mp}}{\eta_{i,Mp}}\Delta\tau_{ik,Mp} \end{cases} \tag{9}$$

Here $G_{i,\Sigma} = G_{i,K} + \sum_p G_{i,M}$ is the instant shear modulus of viscoelastic material ($G_{i,K}$ is the shear modulus of the Kelvin element, $G_{i,Mp}$ is the shear modulus of the pth Maxwell element in a series), $K_{i,\Sigma}$ is the total bulk modulus determined in a similar way, $\eta_{i,Mp}$ is the dynamic viscosity of the p-th Maxwell element, $\sigma_{ik,Mp}$ and $\tau_{ik,Mp}$ are the contributions of the pth Maxwell element to the total specific force. It is easy to show that substitution of the relations (9) to the definition of average stress tensor (4) leads to rigorous fulfilment of the constitutive equation for the viscoelastic material in terms of average stresses and strains. The particular case of (9) is the linear-elastic model (the only Kelvin element), which is typically used to numerically study the elastic behavior of brittle and ductile materials.

It is well known that although the mechanisms of plasticity can qualitatively differ for the materials of various natures (defects of the crystal lattice in metals and alloys and discontinuities in microscopically brittle materials), their macroscopic inelastic behavior is adequately described on the basis of similar models based on the principles of the classical theory of plastic flow. In the framework of this theory, plasticity is described as instantaneous relaxation of "excess" stresses when the plasticity criterion (which is traditionally expressed in terms of the stress tensor invariants) exceeds a specified threshold value (yield shear stress).

Macroscopic inelastic (ductile) behavior of materials is conventionally modelled using associated (for metals and polymers) and non-associated (for ceramic materials, rocks and bone tissues under mechanical confinement) plastic flow models.

The most popular way to implement these models within explicit numerical methods of continuum mechanics is the use of radial return algorithm of Wilkins [48]. We were the first to show that this algorithm can be easily adapted to the formalism of deformable discrete elements. By the analogy with elasticity, the prescribed law of mapping the average stresses $\overline{\sigma}^{i}_{\alpha\beta}$ is rigorously satisfied when applying the stress correcting expressions to the specific response forces σ_{ik} and τ_{ik}. Using this way, we numerically implemented the widely used macroscopic model of plasticity of metals with von Mises yield criterion [42, 43] and non-associated plastic flow model of Nikolaevsky [44] (the macroscopic rock plasticity model with von Mises-Schleicher yield criterion [49]). Other models of elasticity and plasticity can be implemented within the formalism of deformable discrete elements using this direct way. In particular, recently we developed a numerical model of inelastic deformation and fracture of brittle materials, which takes into account the finite incubation time of structural defects and applies principles of the structural-kinetic theory of strength [50].

The generality of the developed formalism allows one to implement not only mechanical but also coupled (thermomechanical, poromechanical and so on) material models. One of the most significant recent achievements of Professor S. G. Psakhie and his scientific team is the development of a hybrid (coupled) DEM-based technique to model the mechanical behavior of "contrast" materials, namely the porous materials with solid skeleton and interstitial liquid [38, 51]. Well-known examples of materials with locally contrasting mechanical properties are watered porous rocks and rubbers, bone and soft tissues. The importance of adequate consideration of the liquid phase in the contrast materials is determined by the fact that such kind of materials possesses strongly nonlinear behavior and non-stationary mechanical characteristics (even in the case of elastic-brittle skeleton) due to redistribution of mobile interstitial fluid in the pore space. Note that the formalism of DEM-based hybrid method was developed in collaboration with Dr. S. Zavsek (Velenje Coal Mine) and Professor J. Pezdic (University of Ljubljana).

Within the framework of this hybrid technique, the discrete element is considered as porous and permeable solid. The mechanical behavior of both solid skeleton and interstitial fluid and their mutual influence are taken into account. In particular, using Biot's linear model of poroelasticity [52] and Terzaghi's concept of effective stresses, we developed the coupled macroscopic model of permeable brittle materials. The main constitutive equations of this model are:

1. Dependence of the pore volume of an element on the hydrostatic component of the external load and the pore pressure

$$\frac{\Omega_p - \Omega_{p0}}{\Omega} = \phi - \phi_0 = \frac{a}{K}\sigma_{mean} + \left(\frac{1}{K} - \frac{1+\phi}{K_s}\right)P_{pore}, \tag{10}$$

   where Ω_p and Ω_{p0} are the current and initial (in an undeformed element) values of the pore volume, φ and φ_0 are the corresponding values of porosity, Ω is the volume of the element, P_{pore} is the pore pressure in the volume of the element.
2. Constitutive equation of linear compressible liquid in the pore space

$$P_{pore} = P^0_{pore} + K_{fl}\left(\frac{\rho_{fl}}{\rho^o_{fl}} - 1\right) = P^0_{pore} + K_{fl}\left(\frac{m_{fl}}{\rho^o_{fl} V_p} - 1\right), \qquad (11)$$

where ρ^0_{fl} and P^0_{pore} are the equilibrium values of the density and pressure of the fluid under atmospheric conditions (in the absence of a mechanical confinement), ρ_{fl} is the current value of the density of the liquid in the pore space of the element, K_{fl} is the bulk modulus of the liquid.

3. Hooke's law for poroelastic material of the element

$$\Delta\sigma_{\alpha\beta} = 2G\left(\Delta\varepsilon_{\alpha\beta} - \delta_{\alpha\beta}\frac{a\Delta P_{pore}}{K}\right) + \delta_{\alpha\beta}\left(1 - \frac{2G}{K}\right)\Delta\sigma_{mean}, \qquad (12)$$

where $a = 1 - K/K_s$ is a coefficient of poroelasticity.

4. Modified formulations of yield and fracture criteria in terms of Terzaghi's effective stresses $\overline{\sigma}^{eff}_{\alpha\beta} = \overline{\sigma}_{mean} + P_{pore}$ [51].
5. The equations of motion of discrete elements (1) are supplemented by the classical equation of transport of interstitial fluid in the pore space of the material [53]. The transport equation is solved on an ensemble of discrete elements by the finite volume method [51].

Shown below examples demonstrate the capability of the developed formalism to implement various complex material models. This allows qualitative expansion of the range of simulated materials and the spatial scales under consideration.

4 Recent Applications of the Formalism of Deformable Elements

DEM is particularly efficient technique to study various aspects of the contact problems (including mechanisms of wear) in technical and natural friction pairs. One of the main principles of Professor Psakhie was a diversification of research activity and the use of the developed mathematical tools in a variety of scientific fields. The section is devoted to a brief outline of some recent results of computer study of contact problems in technical and biological systems. These studies were initiated and supervised by S. G. Psakhie in close collaboration with Professor V. L. Popov.

4.1 Surface Adhesion as a Factor Controlling Regimes of Adhesive Wear

Over the past two decades, the authors have carried out numerical studies of the laws of friction of rough surfaces of ductile and brittle materials. The key results of the

studies are the features of the formation of the third body (quasi-liquid nano-layer) [54], the obtained generalized functional dependence of the friction coefficient on dimensionless combinations of material parameters and loading parameters [54, 55], the formulated principles of nanotribospectroscopy as a promising nondestructive-testing technique for assessment of nanostructured coating and surface layer damage [56].

Recent joint research in this field has focused on studying the "elementary" wear mechanisms (that is, the modes of fracture of individual asperities) and analyzing the effect of attractive (adhesive) force between spatially separated surfaces on the involved mechanism of asperity wear under the condition of low-angle collision. Note that adhesive wear of the surface layers of contacting bodies is a widely studied but still poorly predicted phenomenon [45]. A key factor determining the regime and the rate of wear is the adhesive interaction of surfaces in the contact spots of asperities. Here, the term "adhesive interaction" includes (a) attractive interaction between detached surfaces and (b) effect of "cold welding" (chemical bonding of contacting surfaces) for chemically clear surfaces with high surface energy. The first well-known attempt to generalize the patterns of adhesive wear is the analytical model of Rabinowicz [57, 58]. He examined two qualitatively different mechanisms of asperity wear (plastic smoothing and breakaway) and showed that the involving of a specific mechanism is determined by the size parameter, which is a combination of specific material parameters, including shear strength, elastic constants, and specific surface energy. In recent years, a number of scientific groups carried out extensive numerical studies of the laws of interaction of single asperities [59, 60]. These studies have shown that the dependence of the power of asperity wear on material parameters generally has a significantly more complex nonlinear form, and the spectrum of realized mechanisms is not limited to those considered by Rabinowicz. However, the vast majority of theoretical results were obtained for nanoscale asperities using atomistic simulation.

We were the first to make a systematic numerical analysis of typical modes of asperity wear on higher (micro- and mesoscopic) scales. Due to the capabilities of the developed formalism of deformable discrete elements, such a study was carried out for ductile and brittle materials.

The main result of the study is that we revealed and substantiated two dimensionless material parameters that control the regime of asperity wear [46].

The first one is the ratio of attractive stress σ_0 between the detached surfaces (the adhesion stress) to the shear strength of the material σ_j.We showed that the range of values of this dimensionless parameter could be divided into two intervals with its own wear regime in each of them. The border value of the ratio $\left(\sigma_0^*/\sigma_j\right)$ and the specific involved mechanisms (specific modes) of asperity wear are determined by the value of the second dimensionless parameter.

The second one is the dimensionless parameter a, which characterizes shear strength sensitivity to the applied normal stress (or mean stress). It can be expressed in terms of the ratio of material strength values under different loading conditions. Note that shear strength sensitivity parameter effectively characterizes fracture toughness and is closely related to material brittleness: $a \approx 1$ for highly ductile materials,

$1 < a < 1.5$ for moderately ductile materials, a reaches 5–10 for elastic-brittle solids. We showed that increase in the ratio of adhesion stress to shear strength is accompanied by the transitions:

1. from slipping (wear at the atomic scale) to grinding-based wear regime for the case of highly and moderately ductile materials;
2. from breakaway (separation of the asperity from the foundation) to grinding-based wear regime for the case of materials with a limited ductility or brittle.

We have built the qualitative map of asperity wear regimes in terms of dimensionless material parameters σ_0/σ_j and a [46]. Figure 3 shows a rough schematic representation of this classification of wear regimes, which agrees with the results of atomistic studies by other researchers, supplements and generalizes them to higher scales of surface roughness.

An important result of the DEM-based study is the determined dependence of the position of the boundary between the wear regimes (the border value σ_0^*/σ_j) separating "low adhesion" and "high adhesion" wear modes) on asperity size L. We showed that the value is determined not by the absolute value of L, but its relation to the length d of action of attractive potential between spatially separated surfaces [47]. This dependence has a nonlinear increasing profile with reaching a saturation level at the scale $\sim 10^4$ of the ratio L/d. For larger asperities $\left(L/d > 10^4\right)$, the results of the analysis of the wear regimes are scale invariant under the condition of scale invariance of the mechanical characteristics of the material.

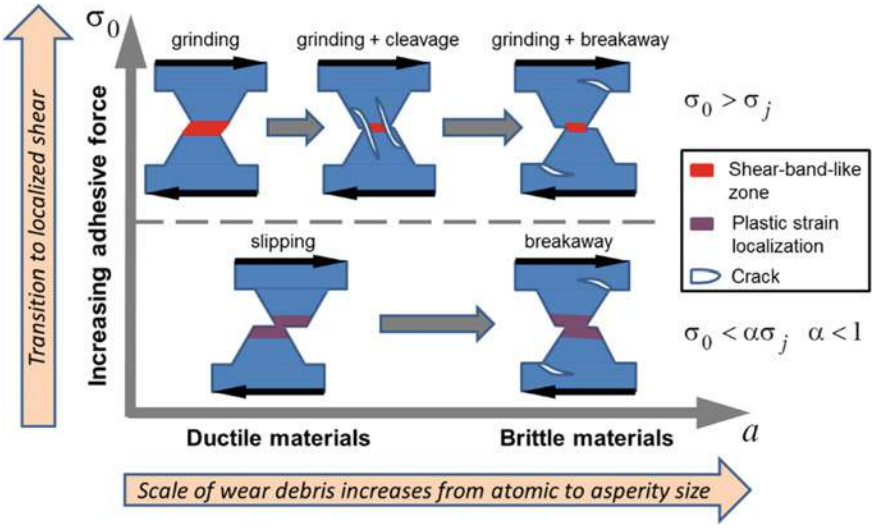


Fig. 3 Schematic classification of wear regimes of asperities for the case of tangential dry contact. The coefficient α is close to 1 for highly ductile materials and about 0.5–0.6 for brittle materials $\left(\alpha = \sigma_0^*/\sigma j\right)$ [46].

So, the surface adhesion stress is a criterion that determines the wear regime of asperities under the condition of tangential contact. The obtained map of asperity wear (Fig. 3) together with the revealed asperity size (scale) effect have both fundamental and practical significance as they allow forecasting the dominating mode of asperity wear for quite different materials from brittle to highly ductile.

4.2 Influence of Interstitial Fluid on the Sensitivity of the Femur to the Rate of Contact Loading

Contact loading is capable of determining not only surface wear and structural modification of surface layers, but also structural changes in the volume of contacting bodies. This is particularly relevant for biological (bone and cartilage) tissues. Functioning of these tissues is largely determined by the redistribution of interstitial fluid. Pore fluid has a complex nonlinear effect on the state and behavior of these biological materials. There are two key aspects of this influence. The first one is fluid flow in the pore space. Fluid flow provides the transfer of nutrients and oxygen and serves as a prerequisite for cell proliferation and tissue regeneration [61, 62]. The second aspect is the mechanical effect of pore pressure. It causes local tensile stresses in the skeleton and contributes to local fracture and gradual degradation of bone tissue. At extreme values of pore pressure it makes a significant contribution to the formation of cracks. The aforesaid argues the existence of optimal distributions of pore fluid (that is, optimal maximum local values and their gradient), which on the one hand provide sufficient fluid flows to ensure normal (regenerative) tissue activity, and on the other hand do not cause local fracture. Such distribution is formed under certain ("optimal") condition of the mechanical (contact) loading of the analyzed organ.

This problem is especially important in application to elements of the human musculoskeletal system (joints) because functioning of bone tissues in these regions strongly depends on the normal contact load and mode of tangential contact interaction. One of the topical problems in this field is the analysis of stress evolution and redistribution of interstitial fluid in the femur near the hip joint under dynamic loading. The developed formalism of permeable fluid-saturated discrete elements is an efficient tool for such a dynamic analysis up to the stage of macroscopic crack formation. Note that despite the topicality of the problem, there are practically no works devoted to the numerical study of the dynamic mechanical response of femur within the consideration of bone tissue as a multiphase fluid-saturated solid. The preliminary results of the numerical DEM-based study presented below are, therefore, pioneering in some way.

The aim of the study is to determine the loading rate sensitivity of healthy bone tissue of the proximal femur and bone tissue at different stages of osteoporosis. The main attention was paid to the regions of the femur in which the volume stresses are positive. This is due to the fact that bone is a brittle material and is characterized by

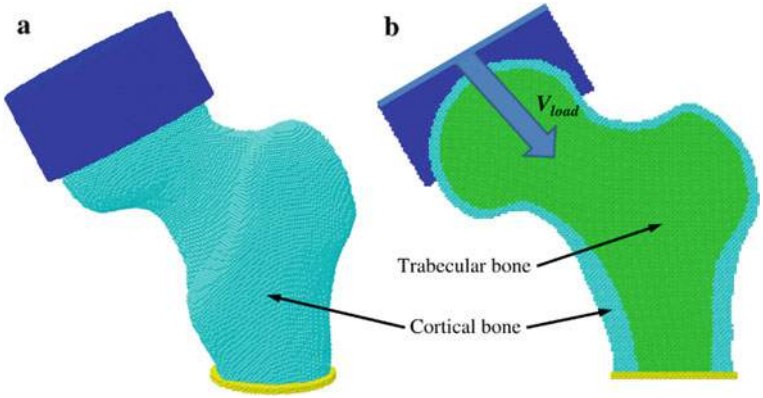


Fig. 4 General view of the model of the proximal femur (**a**) and its cross section (**b**) with differently colored cortical and spongy (trabecular) bone tissues

significantly lower tensile strength than compressive one. First of all, we analyzed the femoral neck (the typical region of the origin of the cracks).

The femur was modeled as a heterogeneous (shell) structure containing two fluid-saturated porous layers (Fig. 4): an inner spongy bone with low stiffness and an outer layer of cortical bone tissue. The latter plays the role of a hard shell and determines constrained conditions for the deformation of the soft spongy bone. The standard CAD model of the femur was used as the geometric basis of the DEM model [63]. We modelled compression of the proximal part of the femur along its main physiological direction (at an angle of about 45° relative to the orientation of the stem, Fig. 4b). The load was applied by setting the constant velocity V_{load} to the outermost elements of a special "cap" on the bone head. The base of the model was fixed. The model was loaded until the resistance force $F_{\max} = 10\,\text{kN}$ was reached (10 kN, is about 70% of the critical load [64]). A velocity range from 1 to 8 m/s was considered. This range covers various modes of motor activity up to the "extreme" one.

Various authors previously showed that the strain rate dependence of the mechanical characteristics of heterogeneous (block-structured) brittle materials under confined loading conditions has not just a non-linear, but non-monotonic profile with a local minimum [65]. The reason for the non-monotonic nature of the dependence is the competition of two factors: pore fluid flow, which affects the redistribution of local stresses in the skeleton, and change in the equilibrium linear dimensions of the material due to change in pore pressure. The non-monotonic nature of the influence of interstitial fluid is especially pronounced under conditions of constrained deformation, when a change in linear dimensions causes a corresponding change in the degree of constraint. The biomechanical system under consideration (femur) also has a heterogeneous structure and is constrained by a hard cortical bone "shell". The shell causes confined deformation of a much softer inner region. This gives reason to suggest that under dynamic loading of such a system, the dependence of local stress values (including stresses in the femoral neck) on the strain rate can also

be nonmonotonic with a local minimum. The magnitude of corresponding "optimal" loading velocity should be related to the characteristics of the porosity and permeability of the bone.

The simulation results confirmed these assumptions and, moreover, allowed us to estimate the characteristic values of the "optimal" loading velocity for healthy bone and bone at various stages of osteoporosis. The key result of the numerical study is the revealed non-linear and nonmonotonic dependence of the stress concentration in the femoral neck on the loading rate. Figure 5 shows examples of the distribution of mean and equivalent stresses in a healthy femur sample at different loading velocities. The simulation results show that an increase in V_{load} is accompanied by an increase in the values of the parameters characterizing the stress concentration in the upper part of the femoral neck. In particular, the volume of the region of maximum stresses and the magnitude of the maximum stress in the neck increase. However, when approaching the "optimal" value of the loading velocity (~3 m/s for the healthy bone), stress concentration decreases, then reaches minimum value at the "optimal" velocity, and then (with further increase in V_{load}) increases monotonously again. This effect takes place both for mean and equivalent stresses. The relative magnitude of the reduction in peak stresses in the neck reaches 10–20%, and the maximum decrease in the volume of the region of stress concentration amounts to 10%. Note that the loading velocities 3–4 m/s correspond to the regime of training motion of a human.

The described effect of reducing the heterogeneity of the stress state of bone tissue (particularly in the femoral neck) in the vicinity of the "optimal" loading velocity is directly related to the influence of pore fluid pressure and redistribution. Special simulations for the "dry" femur showed a monotonic dependence of the stress

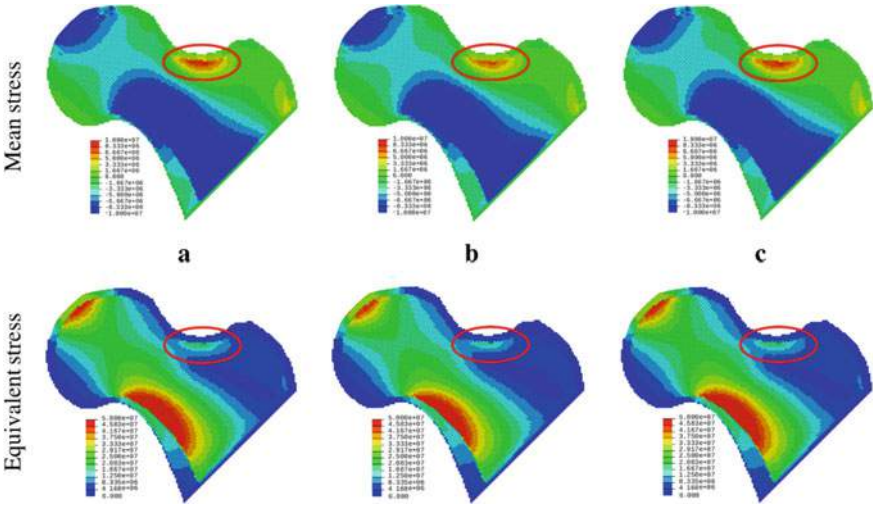


Fig. 5 Distributions of mean stress (upper row) and equivalent stress (lower row) in the proximal femur at different values of loading velocity: **a** 1 m/s; **b** 3 m/s; **c** 5 m/s. All pictures correspond to the same value of applied load $F_{\max} = 10$ kN

concentration in the sample (including the femoral neck) on the loading rate in the entire considered velocity range.

Preliminary numerical studies have also shown that a change in bone tissue parameters corresponding to the successive stages of osteoporosis leads to a systematic increase in the values of "optimal" loading velocity. Bone tissue suffering from osteoporosis is characterized by the reduced strength and higher brittleness. Therefore, our results indicate that the selection of the optimal regime of motor activity for such a femur is questionable.

Finally, we note the importance of the revealed effect of the "optimal" loading rate. Despite the relatively small decrease in the maximum tensile stresses in the critical region of the bone (upper part of the neck), this effect can have a significant positive value in the long term (with multiple repetitions of the load). In addition to reducing the risk of microdamage nucleation/accumulation in the upper part of the neck, a decrease in the magnitude of positive volume stresses contributes to a more intensive and uniform circulation of interstitial fluid in this region of the bone. The consequence is a more complete and rational flow of substances necessary for the normal functioning of bone tissue at optimal loading intensity. So, the results obtained are critically important for determining the optimal modes of motor activity of people, as well as for developing strategies for treating osteoarthritis and preventing the negative consequences of osteoporosis.

5 Conclusion

A formalism of homogeneously deformable elements, which was developed by Professor Sergey G. Psakhie and his colleagues, made possible qualitative enhancement of the capabilities of the particle-based discrete element method. The main advantage of DEM is the ability to correctly describe the mechanical behavior of various materials with taking into account the accompanying thermal, hydromechanical and other effects. Deformability of elements is of particular importance when studying various aspects of dynamic contact interaction, for example, stick-to-slip transition in technical and geological contact zones, mechanisms of friction and wear, redistribution of pore fluid in the surface layers of geological and biological joints, etc.

Two key factors that identified the advantages of the formalism of deformable elements should be especially noted. The first one is the postulation of the many-particle form of relations for the element-element interaction forces. In this regard, DEM has much in common with molecular dynamics method. Pair-wise interatomic potentials are able to catch basic properties of crystal lattice, but do not describe many fundamentally important effects in the bulk and on the surface of solids, including those determining plasticity and phase transformations. Note that the approximation of a homogeneously deformable element is an efficient alternative to more computationally expensive combined DEM-FEM technique. The second factor is the proposed universal method for determining the specific form (and the values of

constants) of the interaction forces. This method is based on the replication of the corresponding rheological relationships of the applied mechanical model. It "circumvents" the fundamental problem of traditional DEM, namely, the need to find a vector analogue of constitutive equations written in tensor form. The particular cases of rheological models mentioned in this chapter are related to locally isotropic materials, however, models of elasticity, ductility and local fracture of anisotropic materials can be implemented in a similar way.

At present time, the formalism of deformable DEM is actively developed and applied in various fields and in the wide range of spatial scales. Moreover, international scientific teams leaded by well-recognized scholars adopt these ideas and develop them in their own way. DEM with deformable elements becomes not just an addition to the classical continuum numerical methods, but to some extent their competitive even in the areas of mechanics, where they traditionally dominate.

Acknowledgements The work was performed according to the Government research assignment for ISPMS SB RAS.

References

1. Cauchy AL (1823) Recherches sur l'equilibre et le mouvement interieur des corps solides ou fluides, elastiques ou non lastiques. Bulletin des sciences par la Societe Philomatique de Paris 9–13
2. Navier CL (1823) Sur les lois de l'equilibre et du mouvement des corps solides elastiques. Bulletin des sciences par la Societe Philomatique de Paris 177–181
3. Kocsis A, Challamel N (2018) Generalized models and non-classic approaches in complex materials 1. Springer, Berlin, pp 451–486 (On the foundation of a generalized nonlocal extensible shear beam model from discrete interactions)
4. Zienkiewicz OC, Taylor RL, Fox DD (2014) The finite element method for solid and structural mechanics, 7th edn. Butterworth-Heinemann, Oxford
5. Balokhonov RR, Romanova VA, Kulkov AS (2020) Microstructure-based analysis of deformation and fracture in metal-matrix composite materials. Eng Fail Anal 110:104412. https://doi.org/10.1016/j.engfailanal.2020.104412
6. Moczo P, Kristek J, Galis M (2014) The finite-difference modelling of earthquake motions: waves and ruptures. Cambridge University Press, Cambridge (UK)
7. Garavand A, Stefanov YP, Rebetsky YL, Bakeev RA, Myasnikov AV (2020) Numerical modeling of plastic deformation and failure around a wellbore in compaction and dilation modes. Int J Numer Anal Meth Geomech 44(6):1–28. https://doi.org/10.1002/nag.3041
8. Cheng AHD, Syngellakis S (eds) (2019) Boundary elements and other mesh reduction methods XLI. WIT Press, Southampton (UK)
9. Popov VL, Pohrt R, Li Q (2017) Strength of adhesive contacts: influence of contact geometry and material gradients. Friction 5(3):308–325. https://doi.org/10.1007/s40544-017-0177-3
10. Yazid A, Abdelkader N, Abdelmajid H (2009) A state-of-the-art review of the X-FEM for computationalfracture mechanics. Appl Math Model 33(12):4269–4282. https://doi.org/10.1016/j.apm.2009.02.010
11. Pohrt R, Popov VL (2015) Adhesive contact simulation of elastic solids using local mesh-dependent detachment criterion in boundary elements method. Facta Univ Mech Eng 13(1):3–10

12. Liu MB, Liu GR (2010) Smoothed particle hydrodynamics (SPH): an overview and recent developments. Arch Comput Methods Eng 17:25–76. https://doi.org/10.1007/s11831-010-9040-7
13. Li X, Zhao J (2019) An overview of particle-based numerical manifold method and its application to dynamic rock fracturing. J Rock Mech Geotech Eng 11(3):684–700. https://doi.org/10.1016/j.jrmge.2019.02.003
14. Nabian MA, Farhadi L (2016) Multiphase mesh-free particle method for simulating granular flows and sediment transport. J Hydraul Eng 143(4):04016102. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001275
15. Munjiza A, Smoljanović H, Živaljić N, Mihanovic A, Divić V, Uzelac I, Nikolić Ž, Balić I, Trogrlić B (2019) Structural applications of the combined finite–discrete element method. Comput Part Mech. https://doi.org/10.1007/s40571-019-00286-5
16. Rodriguez JM, Carbonell JM, Cante JC, Oliver J (2016) The particle finite element method (PFEM) in thermo-mechanical problems. Int J Numer Meth Eng 107(9):733–785. https://doi.org/10.1002/nme.5186
17. Cerquaglia ML, Deliege G, Boman R, Papeleux L, Ponthot JP (2017) The particle finite element method for the numerical simulation of bird strike. Int J Impact Eng 109:1–13. https://doi.org/10.1016/j.ijimpeng.2017.05.014
18. Jing L, Stephansson O (2007) Fundamentals of discrete element method for rock engineering: theory and applications. Elsevier, Amsterdam (NL)
19. Bicanic N (2017) Encyclopaedia of computational mechanics, 2nd edn. Wiley, Hoboken, pp 1–38 (Discrete element methods)
20. Cundall PA, Strack ODL (1979) A discrete numerical model for granular assemblies. Geotechnique 29(1):47–65
21. Cundall PA, Hart RD (1992) Numerical modelling of discontinua. Eng Comput 9(2):101–113. https://doi.org/10.1108/eb023851
22. Potyondy DO, Cundall PA (2004) A bonded-particle model for rock. Int J Rock Mech Min Sci 41:1329–1364. https://doi.org/10.1016/j.ijrmms.2004.09.011
23. Ivars DM, Pierce ME, Darcel C, Reyes-Montes J, Potyondy DO, Young RP, Cundall PA (2011) The synthetic rock mass approach for jointed rock mass modelling. Int J Rock Mech Min Sci 48(2):219–244. https://doi.org/10.1016/j.ijrmms.2010.11.014
24. Mustoe GGW (1992) A generalized formulation of the discrete element method. Eng Comput 9(2):181–190. https://doi.org/10.1108/eb023857
25. Hatzor YH, Ma G, Shi G-H (2017) Discontinuous deformation analysis in rock mechanical practice. CRC Press, Boca Raton (Florida, US)
26. Schneider B, Ramm E (2019) Conceptual experiments and discrete element simulations with polygonal particles. Granular Matter 21(91). https://doi.org/10.1007/s10035-019-0930-6
27. Podlozhnyuk A, Pirker S, Kloss C (2017) Efficient implementation of superquadric particles in discrete element method within an open-source framework. Comput Part Mech 4:101–118. https://doi.org/10.1007/s40571-016-0131-6
28. Sinaie S (2017) Application of the discrete element method for the simulation of size effects in concrete samples. Int J Solids Struct 108:244–253. https://doi.org/10.1016/j.ijsolstr.2016.12.022
29. Potyondy DO (2015) The bonded-particle model as a tool for rock mechanics research and application: current trends and future directions. Geosyst Eng 18(1):1–28. https://doi.org/10.1080/12269328.2014.998346
30. Nosewicz S, Rojek J, Pietrzak K, Chmielewski M (2013) Viscoelastic discrete element model of powder sintering. Powder Technol 246:157–168. https://doi.org/10.1016/j.powtec.2013.05.020
31. Rojek J, Lumelskyj D, Nosewicz S, Romelczyk-Baishya B (2019) Numerical and experimental investigation of an elastoplastic contact model for spherical discrete elements. Comput Part Mech 6:383–392. https://doi.org/10.1007/s40571-018-00219-8
32. Wang G, Al-Ostaz A, Cheng AH-D, Mantena PR (2009) Hybrid lattice particle modeling: theoretical considerations for a 2D elastic spring network for dynamic fracture simulations. Comput Mater Sci 44(4):1126–1134. https://doi.org/10.1016/j.commatsci.2008.07.032

33. Puglia VB, Kosteski LE, Riera JD, Iturrioz I (2019) Random field generation of the material properties in the lattice discrete element method. J Strain Anal Eng Des 54(4):236–246. https://doi.org/10.1177/0309324719858849
34. Pieczywek PM, Zdunek A (2017) Compression simulations of plant tissue in 3D using a mass-spring system approach and discrete element method. Soft Matter 13:7318–7331. https://doi.org/10.1039/C7SM01137G
35. Zabulionis D, Rimsa V (2018) A lattice model for elastic particulate composites. Materials 11(9):1584. https://doi.org/10.3390/ma11091584
36. Rizvi ZH (2019) Lattice element method and its application to multiphysics. Dissertation in fulfilment of the requirements for the degree "Dr.-Ing." of the Faculty of Mathematics and Natural Sciences at Kiel University, Christian-Albrechts-Universität, Kiel
37. Psakhie SG, Horie Y, Ostermeyer GP, Korostelev SYu, Smolin AYu, Shilko EV, Dmitriev AI, Blatnik S, Špegel M, Zavšek S (2001) Movable cellular automata method for simulating materials with mesostructured. Theoret Appl Fract Mech 37(1–3):311–334
38. Psakhie SG, Shilko EV, Smolin AY, Dimaki AV, Dmitriev AI, Konovalenko IS, Astafurov SV, Zavshek S (2011) Approach to simulation of deformation and fracture hierarchically organized heterogeneous media, including contrast media. Phys Mesomech 14(1–5):224–248. https://doi.org/10.1016/j.physme.2011.12.003
39. Psakhie SG, Horie Y, Korostelev SYu, Smolin AYu, Dmitriev AI, Shilko EV, Alekseev SV (1995) Method of movable cellular automata as a tool for simulation within the framework of mesomechanics. Russ Phys J 38(11):1157–1168. https://doi.org/10.1007/BF00559396
40. Psakhie SG, Shilko EV, Smolin AY, Dmitriev AI, Korostelev SY (1996) Computer aided study of reaction-assisted powder mixture shock compaction at meso-scale. New computational technique. In: Proceedings of US-Russian workshop "Shock induced chemical processing", Saint-Petersburg, 23–24 June, 1996
41. Popov VL, Psakhie SG (2001) Theoretical principles of modeling elastoplastic media by movable cellular automata method. I. Homogeneous media. Phys Mesomech 4(1):15–25
42. Psakhie S, Shilko E, Smolin A, Astafurov S, Ovcharenko V (2013) Development of a formalism of movable cellular automaton method for numerical modeling of fracture of heterogeneous elastic-plastic materials. Frattura Ed Integrità Strutturale 24(7):26–59. https://doi.org/10.3221/IGF-ESIS.24.04
43. Shilko EV, Psakhie SG, Schmauder S, Popov VL, Astafurov SV, Smolin AYu (2015) Overcoming the limitations of distinct element method for multiscale modeling of materials with multimodal internal structure. Comput Mater Sci 102:267–285. https://doi.org/10.1016/j.commatsci.2015.02.026
44. Psakhie SG, Shilko EV, Grigoriev AS, Astafurov SV, Dimaki AV, Smolin AYu (2014) A mathematical model of particle–particle interaction for discrete element based modeling of deformation and fracture of heterogeneous elastic–plastic materials. Eng Fract Mech 130:96–115. https://doi.org/10.1016/j.engfracmech.2014.04.034
45. Wen S, Huang P (2017) Principles of tribology, 2nd edn. Wiley, London (UK)
46. Dimaki AV, Shilko EV, Dudkin IV, Psakhie SG, Popov VL (2020) Role of adhesion stress in controlling transition between plastic, grinding and breakaway regimes of adhesive wear. Sci Rep 10:1585. https://doi.org/10.1038/s41598-020-57429-5
47. Dimaki AV, Dudkin IV, Popov VL, Shilko EV (2019) Influence of adhesion force and strain hardening coefficient of the material on the rate of adhesive wear in a dry tangential frictional contact. Russ Phys J 62(8):1398–1408. https://doi.org/10.1007/s11182-019-01857-y
48. Wilkins ML (1999) Computer simulation of dynamic phenomena. Springer, Berlin
49. Nikolaevsky VN (1996) Geomechanics and fluidodynamics with application to reservoir engineering. Kluwer Academic Publishers (Springer), Berlin
50. Grigoriev AS, Shilko EV, Skripnyak VA, Psakhie SG (2019) Kinetic approach to the development of computational dynamic models for brittle solids. Int J Impact Eng 123:14–25. https://doi.org/10.1016/j.ijimpeng.2018.09.018
51. Psakhie SG, Dimaki AV, Shilko EV, Astafurov SV (2016) A coupled discrete element-finite difference approach for modeling mechanical response of fluid-saturated porous materials. Int J Numer Meth Eng 106(8):623–643. https://doi.org/10.1002/nme.5134

52. Detournay E, Cheng AH-D (1993) Comprehensive rock engineering: principles, practice and projects. Pergamon Press, Oxford (UK), pp 113–171 (Fundamentals of poroelasticity)
53. Basniev KS, Dmitriev NM, Dmitriev NM, Chilingar GV (2012) Mechanics of fluid flow. Wiley, London (UK)
54. Popov VL, Psakhie SG, Dmitriev AI, Shilko E (2003) Quasi-fluid nano-layers at the interface between rubbing bodies: simulations by movable cellular automata. Wear 254(9):901–906. https://doi.org/10.1016/S0043-1648(03)00244-8
55. Bucher F, Dmitriev AI, Ertz M, Knothe K, Popov VL, Psakhie SG, Shilko EV (2006) Multiscale simulation of dry friction in wheel/rail contact. Wear 261(7–8):874–884. https://doi.org/10.1016/j.wear.2006.01.046
56. Psakhie SG, Shilko EV, Popov VL, Starcevic J, Thaten J, Astafurov SV, Dimaki AV (2009) Assessment of nanostructured ceramic coating damage Nanotribospectroscopy. Russ Phys J 52(4):380–385. https://doi.org/10.1007/s11182-009-9242-3
57. Rabinowicz E (1958) The effect of size on the looseness of wear fragments. Wear 2(1):4–8
58. Popova E, Popov VL, Kim DE (2018) 60 years of Rabinowicz' criterion for adhesive wear. Friction 6(3):341–348. https://doi.org/10.1007/s40544-018-0240-8
59. von Lautz J, Pastewka L, Gumbsch P, Moseler M (2016) Molecular dynamic simulation of collision-induced third-body formation in hydrogen-free diamond-like carbon asperities. Tribol Lett 63:26. https://doi.org/10.1007/s11249-016-0712-9
60. Molinari J-F, Aghababaei R, Brink T, Frérot L, Milanese E (2018) Adhesive wear mechanisms uncovered by atomistic simulations. Friction 6:245–259. https://doi.org/10.1007/s40544-018-0234-6
61. Christen P, Ito K, Ellouz R, Boutroy S, Sornay-Rendu E, Chapurlat RD, van Rietbergen B (2014) Bone remodelling in humans is load-driven but not lazy. Nat Commun 5:4855. https://doi.org/10.1038/ncomms5855
62. Wittkowske C, Reilly GC, Lacroix D, Perrault CM (2016) In vitro bone cell models: impact of fluid shear stress on bone formation. Front Bioeng Biotechnol 4:87. https://doi.org/10.3389/fbioe.2016.00087
63. Cheung G, Zalzal P, Bhandari M, Spelt JK, Papini M (2004) Finite element analysis of a femoral retrograde intramedullary nail subject to gait loading. Med Eng Phys 26(2):93–108. https://doi.org/10.1016/j.medengphy.2003.10.006
64. Todo M (2018) Biomechanical analysis of hip joint arthroplasties using CT-image based finite element method. J Surg Res 1:34–41
65. Shilko EV, Dimaki AV, Psakhie SG (2018) Strength of shear bands in fluid-saturated rocks: a nonlinear effect of competition between dilation and fluid flow. Sci Rep 8:1428. https://doi.org/10.1038/s41598-018-19843-8

A Tool for Studying the Mechanical Behavior of the Bone–Endoprosthesis System Based on Multi-scale Simulation

Alexey Yu. Smolin, Galina M. Eremina, and Evgeny V. Shilko

Abstract The chapter presents recent advances in developing numerical models for multiscale simulation of the femur–endoprosthesis system for the case of hip resurfacing arthroplasty. The models are based on the movable cellular automaton method, which is a representative of the discrete element approach in solid mechanics and allows correctly simulating mechanical behavior of a variety of elastoplastic materials including fracture and mass mixing. At the lowest scale, the model describes sliding friction between two rough surfaces of TiN coatings, which correspond to different parts of the friction pair of hip resurfacing endoprosthesis. At this scale, such parameters of the contacting surfaces as the thickness, roughness, and mechanical properties are considered explicitly. The next scale of the model corresponds to a resurfacing cap for the femur head rotating in the artificial acetabulum insert. Here, sliding friction is explicitly computed based on the effective coefficient of friction obtained at the previous scale. At the macroscale, the proximal part of the femur with a resurfacing cap is simulated at different loads. The bone is considered as a composite consisting of outer cortical and inner cancellous tissues, which are simulated within two approaches: the first implies their linear elastic behavior, the second considers these tissues as Boit's poroelastic bodies. The later allows revealing the role of the interstitial biological fluid in the mechanical behavior of the bone. Based on the analysis of the obtained results, the plan for future works is proposed.

Keywords Bone · Endoprosthesis · Mechanical behavior · Friction · Computer simulations · Particle-based method

A. Yu. Smolin (✉) · G. M. Eremina · E. V. Shilko
Institute of Strength Physics and Materials Science of Siberian Branch of Russian Academy of Sciences, Tomsk, Russian Federation
e-mail: asmolin@ispms.ru

Tomsk State University, Tomsk, Russian Federation

G. M. Eremina
e-mail: anikeeva@ispms.ru

E. V. Shilko
e-mail: shilko@ispms.tsc.ru

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_5

1 Introduction

Endoprosthetics is an effective way to treat such common degenerative diseases of large human joints as osteoporosis and arthritis. Wear in a pair of friction of the constituent elements of the joint endoprosthesis design has a significant impact on its operational resource. Primarily, this refers to the prostheses of the hip and knee joints. The key role in the wear process belongs to the structure of the surface layers of the contacting elements. To improve the tribological characteristics of metal endoprostheses, which are currently the most widely used, reinforcing coatings are used. In practice, titanium is usually used as a metal and titanium nitride (TiN) as a coating.

Since endoprostheses are intended for a long stay in the human body, they should not lead to any discomfort. Therefore, an important stage in the development of endoprostheses is their testing. Endoprosthesis tests are divided into several stages; these are preclinical and clinical trials. Clinical trials are conducted through the installation of an endoprosthesis in a living human body. Using this test, the final result is determined. However, when conducting clinical trials, there is a danger that a poorly or improperly selected endoprosthesis may adversely affect the patient's health. Therefore, much more attention is paid to preclinical trials when developing endoprostheses. Talking about the mechanical behavior of the endoprosthesis, its preclinical studies can be divided into experimental and theoretical. Experimental studies are tests using special equipment that simulates real dynamic loading experienced by an endoprosthesis in the body. In particular, standard methods, such as instrumented indentation, scratching, and three-point bending, are used to determine the tribological properties of the surface layers of materials. It is worth noting that conventional experimental studies are carried out under standard atmospheric conditions and their results can differ greatly from the behavior of materials in real conditions inside a living body. Therefore, part of the experimental studies is carried out in a medium simulating a living body (in vitro), or on test samples placed into the animal (in vivo). However, for obvious reasons, the number and possibilities of such studies are very limited.

Theoretical studies of the mechanical behavior of the endoprosthesis are most often carried out using computer simulation. It should be noted that modern numerical methods of solid mechanics allow a detailed study of the mechanical behavior of endoprostheses, which takes into account the influence of a variety of factors. However, to fully utilize all the features of this approach, it is necessary to use special software.

In the last decade, the development of computer-aided tools for studying the mechanical behavior of endoprostheses has been the subject of extensive work by leading world scientific groups in such well-known centers as McGill University (Canada), Liverpool John Moores University (Great Britain), Iowa State University (USA) and many others. The high relevance of this problem is also confirmed by numerous publications on this topic in leading scientific journals.

However, tools taking into account the influence of structural features of surface layers on the wear of contacting surfaces, as well as the effect of wear on the mechanical behavior of the endoprosthesis as a whole, have not yet been developed. The second important factor that has to be considered in modeling the mechanical behavior of living bone is the presence of interstitial fluid in it. The availability of such tools would be very relevant in the study of the durability of a new class of endoprostheses made from new composite materials, including titanium alloys with a nanostructured TiN coating.

2 State-of-the-Art

There are two methods of hip arthroplasty: total hip replacement and hip resurfacing. Total hip replacement suggests cutting the femur head and inserting a prosthetic implant into the bone. In the case of hip resurfacing only a very thin layer of the femur head is replaced by a metal cap, which is hollow and shaped like a mushroom. Despite the more complicated procedure of the operation, the latter method is preferable for young people who lead an active life. Previously, only metals were used for the friction pair of the resurfacing endoprosthesis, in particular, the CoCrN alloy. This led to the fact that during life after the surgical operation, severe wear was observed in the friction pair with a large release of metal particles, which led to necrosis of cells and small phagocytes, and severe allergic vasculitis was also observed [1]. Therefore, in recent years there has been active scientific research on the use of titanium alloys with superthin ceramic coatings in resurfacing endoprosthetics. Moreover, titanium is usually used as a metal and titanium nitride (TiN) as a coating [2, 3]. The structure of the surface layer of the coating plays a key role in the wear process and is determined by the methods and regimes of its application. There are several ways to apply this coating: vacuum spraying (PVD), chemical vapor deposition (CVD), powder nitriding (PIRAC-powder immersion reaction assisted coating nitriding). The coatings obtained by PVD and CVD have insufficient adhesion to the substrate and are prone to delaminating under dynamic loads, while the adhesion of the coating obtained by the PIRAC method is significantly higher. Therefore, the use of metal implants with a PIRAC coating is very promising. Compounds of carbides, borides, nitrides are used as coating materials applied by the PIRAC method [4, 5].

Computer simulation allows one to study the mechanical behavior of endoprostheses, taking into account the influence of a variety of factors on it. Therefore, in preclinical studies, computer simulation is widely used to predict the mechanical behavior of prostheses. For example, modeling based on numerical methods of continuum mechanics allows a detailed study of the behavior of the coating and surface layer under contact loading. Thus, the influence of coating thickness on the load–displacement curve of instrumented indentation of ceramic coatings on a metal substrate was studied in [6–8]. The role of the surface roughness of the ceramic coating in the features of the mechanical behavior of the coating-substrate system during instrumented indentation was numerically studied in [9–12]. The authors of

[13–17] simulated the mechanical behavior of the coating–substrate system in instrumented indentation and scratching and analyzed stress fields in contact areas, as well as features of the load–displacement curves.

There is practically no theoretical work on the study of wear in a friction pair of a resurfacing endoprosthesis of a hip joint. Most of the work have been devoted to wearing in a friction pair of endoprostheses for total hip replacement. In theoretical studies of the wear process in a friction pair of an endoprosthesis, two approaches are used: with explicit consideration of wear debris and its implicit consideration. In the numerical study of wear in a friction pair by the finite element method, an implicit consideration of wear particles is used. First of all, this circumstance is associated with large computational costs and the complexity of explicit modeling of material failure. Most of the research [18, 19] on the theoretical study of wear in the friction pair of the endoprosthesis are based on the technique proposed in [20], where the ratio of Archard/Lancaster is used to describe the wear via calculating the wear coefficient for adhesive or abrasive wearing. This ratio defines linear wear as a function of contact pressure, loading rate, and wear coefficient, which in turn depends on the tribological characteristics of two contacting surfaces, such as roughness and various physical and mechanical characteristics of the materials of the contacting pair. The surface roughness is taken into account implicitly [21–23].

Most of the works on numerical modeling of the mechanical behavior of the bone–endoprosthesis system is performed using the methods of continuum mechanics. Thus, research of Kuhl and Balle [24], the influence of the type of prosthetics on the stress state in the system was investigated; establishing that the stress field of the system during resurfacing endoprosthetics is close to the stresses in a healthy joint. The results of a numerical study of the mechanical behavior of the bone–endoprosthesis system depending on the geometric characteristics of the implant and its design features are presented in [25], and on the implant material in [26].

Work on studying the mechanical behavior of bone tissue with biological fluid started long ago [27]. At the same time, the Biot model of poroelasticity was most widely used to describe the mechanical behavior of bone tissues. In papers [28–30] the Biot model of the isotropic poroelastic medium was used to describe bone tissues in the framework of continuum mechanics methods.

Taken together, this analysis of the literature shows that for the correct prediction of the mechanical behavior of the bone–endoprosthesis system, the development of numerical models that can describe the bone within the framework of a poroelastic body taking into account possible fracture of bone tissues, as well as wear in the friction pair of the endoprosthesis, is in demand.

3 The Problem Statement

The aim of this chapter is to present recent advances in development of numerical models for multi-scale simulation of the femur–endoprosthesis system for the hip resurfacing arthroplasty. The models are based on the movable cellular

automaton method, which is a representative of the discrete element approach in solid mechanics. According to the chosen multi-scale approach, the model of the lowest scale (mesoscale) describes sliding friction between two rough surfaces of TiN coatings, which correspond to different parts of the friction pair of hip resurfacing endoprosthesis. At this scale, we consider explicitly such parameters of the endoprosthesis friction pair as the thickness, roughness, and mechanical properties of the corresponding TiN coatings. The next scale of the model is a resurfacing cap rotating in the artificial acetabulum insert with an explicit account of sliding friction based on the effective coefficient of friction obtained at the previous scale. At the macroscale, we consider compression of the proximal part of the femur with a resurfacing cap. This macro-model considers the bone as a composite consisting of outer cortical and inner cancellous tissues. Here we use two approaches: the first implies simple linear elastic behavior of both tissues, the second considers these tissues as Boit's poroelastic bodies with accounting for the role of the interstitial biological fluid in the mechanical behavior of the bone. For comparison, we also consider the macroscopic model for healthy bone in compression.

4 Description of the Modeling Method

For simulating the mechanical behavior of the materials for the bone and prosthesis, we use the particle-based method of movable cellular automata (MCA), which was proposed by professors Sergey Psakhie and Yasuyuki Horie in 1994 and firstly published in 1995 [31]. Since that time, the method has been being actively developed in Prof. Psakhie's lab and its latest description can be found, for example, elsewhere in papers [32, 33] as well as in book chapters [34, 35].

MCA is a representative of so-called discrete element methods (DEM). The main principles of the method are as follows. A simulated body is represented by an ensemble of bonded equiaxial discrete elements of the same size (called movable cellular automata), which spatial position and orientation, as well as state, can change due to local interaction with nearest neighbors. Automata interact with each other through their contacts. The initial value of the contact area, as well as the automaton volume, is determined by the size of automata and their packing. The main advantage of the MCA-method in comparison with DEM is the generalized many-body formulas for central interaction forces acting between the pair of elements, similar to the embedded atom force filed used in molecular dynamics. It is based on computing components of the average stress and strain tensors in the bulk of automaton according to the homogenization procedure described in [32]. Use of many-body interaction forces allows correct simulation within discrete element approach of such important features of the mechanical behavior of solids like Poisson effect and plastic flow.

When describing the kinematics and dynamics of an automaton motion, its shape is approximated by an equivalent sphere. This approximation is the most widely used in the discrete element method and allows one to consider the forces of central and tangential interaction of automata as formally independent. This makes also possible

to use the simplified Newton–Euler equations of motion to govern translational motion and rotation of the movable automata.

$$\begin{cases} m_i \dfrac{d^2\mathbf{R}_i}{dt^2} = \displaystyle\sum_{j=1}^{N_i} \mathbf{F}_{ij}^{\text{pair}} + \mathbf{F}_i^{\Omega}, \\ \hat{J}_i \dfrac{d\boldsymbol{\omega}_i}{dt^2} = \displaystyle\sum_{j=1}^{N_i} \mathbf{M}_{ij} \end{cases}, \tag{1}$$

where $\mathbf{R_i}$, $\boldsymbol{\omega}_i$, m_i and $\widehat{J_i}$ are the location vector, rotation velocity vector, mass and moment of inertia of ith automaton respectively, $\mathbf{F}_{ij}^{\text{pair}}$ is the interaction force of the pair of ith and jth automata, $\mathbf{F}_i^{\Omega}$ is the volume-dependent force acting on ith automaton and depending on the interaction of its neighbors with the remaining automata. In the latter equation, $\mathbf{M}_{ij} = q_{ij}(\mathbf{n}_{ij} \times \mathbf{F}_{ij}^{\text{pair}}) + \mathbf{K}_{ij}$, here q_{ij} is the distance from the center of ith automaton to the point of its interaction (contact) with jth automaton, $\mathbf{n}_{ij} = (\mathbf{R}_j - \mathbf{R}_i)/r_{ij}$ is the unit vector directed from the center of ith automaton to the jth one and r_{ij} is the distance between automata centers, $\mathbf{K}_{ij}$ is the torque caused by relative rotation of automata in the pair.

Movable automata are treated as deformable. Strains and stresses are assumed to be uniformly distributed in the volume of each automaton. Within the framework of this approximation, the values of averaged stresses in the automaton volume may be calculated as the superposition of forces applied to different parts of the automaton surface. In other words, averaged stress tensor components are expressed in terms of the interaction forces with neighbors [32]:

$$\overline{\sigma}_{\alpha\beta}^{i} = \frac{R_i S_{ij}^0}{\Omega_i^0} \sum_{j=1}^{N_i} \left[f_{ij} \left(\vec{n}_{ij}\right)_{\alpha} + \tau_{ij} \left(\vec{t}_{ij}\right)_{\beta} \right] \tag{2}$$

where i is the automaton number, $\overline{\sigma}_{\alpha\beta}^{i}$ is the component $\alpha\beta$ of the averaged stress tensor, $\alpha,\ \beta = x, y,\ z$ (XYZ is the global coordinate system), Ω_i^0 is the initial volume of the automaton i, S_{ij}^0 is the initial value of the contact area between the automata i and j, R_i is the radius of the equivalent sphere (semi-size of automaton i), f_{ij} and τ_{ij} are specific values of central and tangential forces of interaction between the automata i and j, $\left(\vec{n}_{ij}\right)_{\alpha}$ and $\left(\vec{t}_{ij}\right)_{\alpha}$ are the projections of the unit normal and unit tangent vectors onto the α-axis , N_i is the number of interacting neighbors of automaton i.

Invariants of the averaged stress tensor $\overline{\sigma}_{\alpha\beta}^{i}$ are used to calculate the central interaction forces $\left(f_{ij},\ \tau_{ij}\right)$ and the criterion of an inter-element bond breaking (local fracture). The components of the averaged strain tensor $\overline{\varepsilon}_{\alpha\beta}^{i}$ are calculated in increments using the specified constitutive equation of the simulated material and the calculated increments of mean stress.

In [32] it is shown that the relation for the force of central interaction of automata can be formulated based on the constitutive equation of the material for the diagonal components of the stress tensor, while the force of tangential interaction can be formulated on the basis of similar equations for non-diagonal stress components. When implementing the linear elastic model, the expressions for specific values of the central and tangential forces of the mechanical response of the automaton i to mechanical action from the neighboring automaton j are written as follows:

$$\begin{cases} \Delta f_{ij} = 2G_i \Delta\varepsilon_{ij} + D_i \Delta\sigma_i^{\text{mean}} \\ \Delta\tau_{ij} = 2G_i \Delta\gamma_{ij} \end{cases}, \tag{3}$$

where the symbol Δ means increment of the corresponding variable during time step Δt of the numerical scheme of integration of the motion equations, $\Delta\varepsilon_{ij}$ and $\Delta\gamma_{ij}$ are the increments of normal and shear strains of the automaton i in pair $i-j$, G_i is the shear modulus of the material of the automaton i, K_i is the bulk modulus, $D_i = 1 - 2G_i \big/ K_i$.

Due to the necessity of the third Newton's law $\left(\sigma_{ij} = \sigma_{ji} \text{ and } \tau_{ij} = \tau_{ji}\right)$, the increments of the reaction forces of the automata i and j are calculated based on the solution of the following system of equations.

$$\begin{cases} \Delta f_{ij} = \Delta f_{ji} \\ R_i \Delta\varepsilon_{ij} + R_j \Delta\varepsilon_{ji} = \Delta r_{ij} \\ \Delta\tau_{ij} = \Delta\tau_{ji} \\ R_i \Delta\gamma_{ij} + R_j \Delta\gamma_{ji} = \Delta l_{ij}^{\text{sh}} \end{cases}, \tag{4}$$

where Δr_{ij} is the change in the distance between the centers of the automata for a time step Δt, $\Delta l_{ij}^{\text{sh}}$ is the value of the relative shear displacement of the interacting automata i and j. The system of equations (4) is solved for finding the increments of strains. This allows calculation of the increments of the specific interaction forces. When solving the system (4), the increments of mean stress and the values of specific forces in the right-hand sides of relations (3) are taken from the previous time step or are evaluated and further refined within the predictor–corrector scheme.

A pair of automata can be in one of two states: bound and unbound. Thus, in MCA fracture and coupling of fragments (crack healing, microwelding etc.) is simulated by the corresponding switching of the pair state. Switching criteria depend on physical mechanisms of material behavior [32, 33]. Note, that knowing stress and strain tensor in the bulk of an automaton, makes possible direct application of conventional fracture criteria written in the tensor form, herein we used von Mises criterion based on a threshold value of the equivalent stress.

5 Results and Discussion

5.1 Modeling Friction Pair of the Hip Resurfacing

We start with the simulation of the friction pair of the hip resurfacing endoprosthesis, which consists of two contacting bodies made of titanium alloy with TiN coating. First, we choose the materials parameters and validate the materials models to compare the simulation results for instrumented indentation with available experimental data. After that, we can simulate friction of the contacting bodies at the mesoscale to get an estimation of the coefficients of friction, which was used in the macroscopic model for rotation a resurfacing cap in the artificial acetabulum insert.

5.1.1 Materials Characterization

The values for the main physico-mechanical properties of the titanium alloy Ti6Al4V that we chose from the literature [36] are shown in Table 1 (corresponding Young's modulus $E = 110$ GPa). Geometric features of the TiN coating and its physico-mechanical properties are determined by the deposition modes at PIRAC forming [37]. So, at a deposition temperature of 700 °C and a treatment time of 48 h, the coating on the titanium substrate has a thickness of 1.3 μm with an average roughness height of 0.15 μm (mode 1) and elastic modulus, depending on the deposition regime, was equal to $E_1 = 258$ GPa; at a deposition temperature of 800 °C and a treatment time of 4 h the coating on the titanium substrate has a thickness of 1.4 μm with an average roughness height of 0.132 μm (mode 2) and an elastic modulus of $E_2 = 258$ GPa; at a deposition temperature of 900 °C and a treatment time of 2 h the coating on the titanium substrate has a thickness of 1.5 μm with an average roughness height of 0.265 μm (mode 3) and an elastic modulus of $E_3 = 321$ GPa. According to this information, we chose the values for the material properties of the TiN coating that are presented in Table 1. Data for yield stress σ_y and ultimate strength σ_b and strain ε_b were obtained using reverse analysis of the load–displacement curve for instrumented indentation [38].

Table 1 Properties of the model materials for the friction pair

Material	Bulk modulus, K, GPa	Shear modulus, G, GPa	Density, ρ, kg/m^3	Yield stress, σ_y, GPa	Ultimate strength, σ_b, GPa	Ultimate strain, ε_b
Ti6Al4V	92	41	4420	0.99	1.07	0.100
TiN mode 1	173	104	5220	4.50	5.50	0.075
TiN mode 2	173	104	5220	4.50	5.50	0.075
TiN mode 3	205	129	5220	4.50	5.50	0.075

5.1.2 Validation of the Models for Materials

The first numerical test, which we performed for validation of the models used for further simulations, was instrumented indentation [39]. General view of the model geometry for this test is shown in Fig. 1a. The model specimen was a parallelepiped consisting of titanium substrate, interface and TiN coating. The loading was simulated by moving all the automata of the Berkovich indenter with constant velocity $V_z = -1$ m/s until the required penetration depth is reached, then we apply them the velocity $V_z = -1$ m/s for unloading (Fig. 1b).

Using the procedure proposed by Oliver and Pharr [40] applied to processing the simulation results we plot the dependence of the material hardness on the penetration depth (Fig. 2) for three kinds of coatings obtained in different regimes. It can be seen from Fig. 2 that the hardest is the coating obtained by mode 3. All the curves correspond to experimental data from [37]. This allows us to conclude that the models of the materials behavior are validated for the normal contact loading conditions and can be used in further steps of our work.

The model specimen for scratch testing was also a parallelepiped consisting of titanium substrate, interface and TiN coating but elongated along the axis Y (Fig. 3). To simulate the force acting on the indenter along axis Z in the experiment, in our calculations we set the velocity of indenter automata to $V_Z = -0.5$ m/s (Fig. 3b) until the indenter was immersed into a predetermined penetration depth and after that the vertical velocity was set to zero. Here we considered penetration of the indenter only up to the interface layer; the possibility of the coating to detach from the substrate was not allowed. To move the indenter along the sample surface, a constant velocity of the indenter automata along the axis Y was set to $V_Y = 1$ m/s.

Based on the results of our simulations, images of the deformed sample were created, and the values of the critical force characteristics at certain stages of fracture were obtained. It was found that the value of the critical force for the onset of fracture

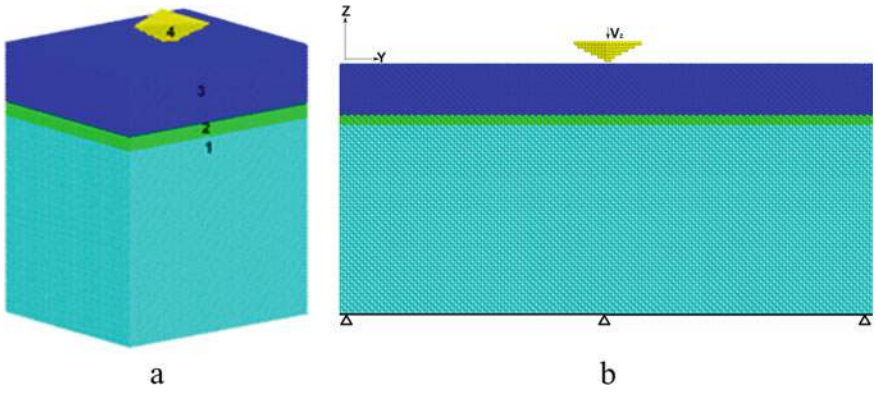


Fig. 1 The model specimen for indentation (**a**) and its cross-section with loading parameters (**b**) represented by automata packing (the numbers indicate the model materials: 1—titanium, 2—interface, 3—coating, 4—diamond)

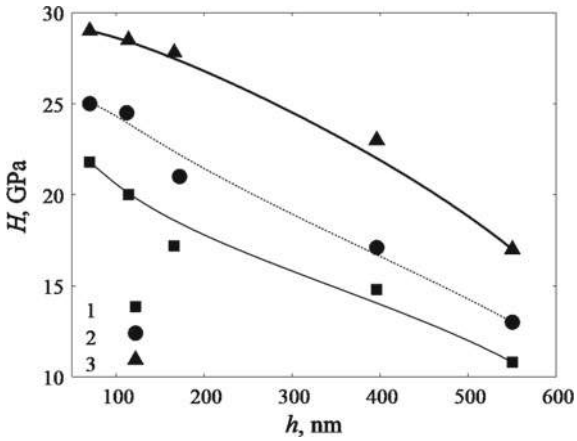


Fig. 2 Dependence of the hardness on the depth of penetration (the numbers indicate the modes of coating deposition)

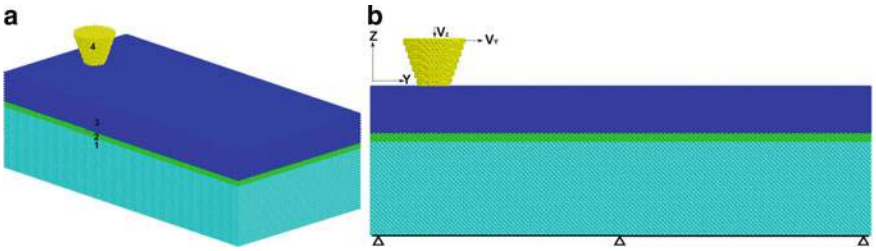


Fig. 3 The model specimen for scratch testing (**a**) and its cross-section with the loading parameters (**b**) represented by automata packing (the numbers indicate the model materials: 1—titanium, 2—interface, 3—coating, 4—diamond)

and cracking is largest for the coating thickness of 1.4 μm and roughness of 0.132 μm, and smallest for the thickness of 1.5 μm and roughness of 0.265 μm. At delamination of the coating, the maximum of critical force is typical for the specimen with a coating thickness of 1.5 μm and roughness of 0.265 μm, and the minimum for the specimen with a coating thickness of 1.3 μm and roughness of 0.15 μm (Fig. 4).

Critical loads obtained by modeling scratch testing are in very good agreement with the experimental data [41].

5.1.3 Modeling of Sliding Friction at the Meso-scale

At the next step, we developed a meso-model of sliding friction of the contacting surfaces of the endoprosthesis that explicitly accounts for the surface roughness, mean height of which is determined by the regime of the coating deposition [42].

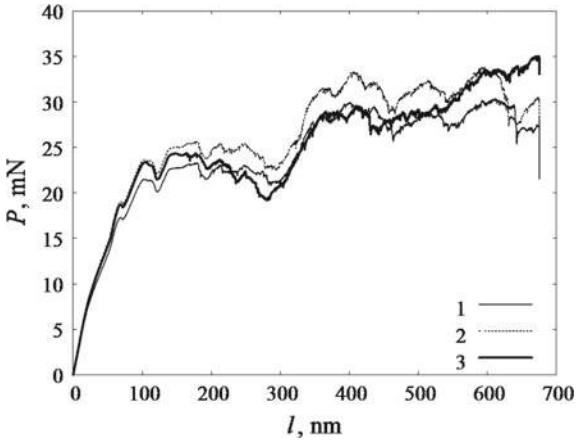


Fig. 4 The force acting on the indenter from the coating P versus the scratching path l of the indenter (the numbers indicate the modes of coating deposition)

Because at the macroscale we need just effective value of the friction coefficient, whereas the substrate and interface were not considered in this model.

Geometrically, the model for studying friction consists of two bodies in the form of parallelepipeds (Fig. 5). Each parallelepiped is usually divided into two parts: one is the contact surface, and the other is its base. The different roughness relief of the two contacting surfaces was set in the preprocessor of the MCA_3D software package. To simulate the environment of the developed meso-scale simulation box during sliding friction, periodicity conditions along the X and Y axes were set on its side faces.

In the initial state, the interacted bodies are separated in spaced so that the asperities of the rough surfaces do not touch each other. For their approaching and the beginning of contact interaction, the velocity V_Z along the Z-axis was applied to the automata of the top and bottom layers of the sample. After touching the upper and lower parts of the sample, traction forces (pressure $P = 0.75\sigma_y$) directed along the Z-axis, as well as horizontal velocity $V_y = 1$ m/s along the Y-axis, were applied to the automata of the top and bottom layers of the sample (Fig. 5b). The specified loading conditions at the top and bottom of the model had symmetrical character and the same magnitude, but the directions of the action were opposite. To avoid too hard loading in the friction zone, a gradual increase in the applied loads was implemented. To imitate the length of the real bodies along the vertical direction, it is necessary to damp the elastic waves arising in the friction zone and propagating along this direction. To this end, a special viscous damping force along the Z-axis was introduced for the automata of both loading regions.

According to the simulation results obtained, the dependence of the friction coefficient on the calculation time was plotted for different combinations of the roughnesses and reliefs (Fig. 6). Figure 6a shows that at the moment of contact of two

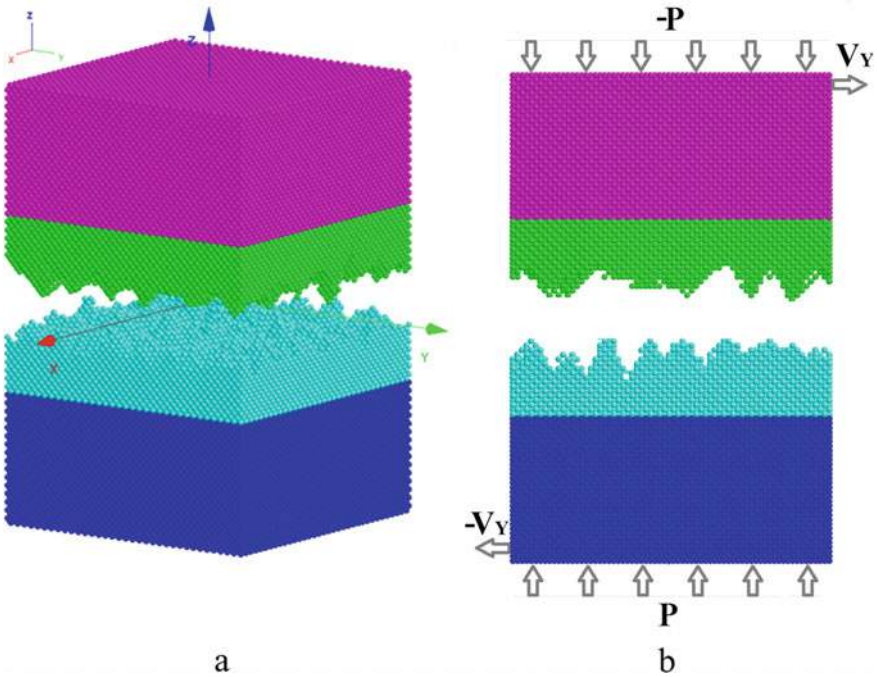


Fig. 5 The model sample for sliding friction at mesoscale (**a**) and its cross-section with loading parameters (**b**), the colors indicate different geometry blocks of the model

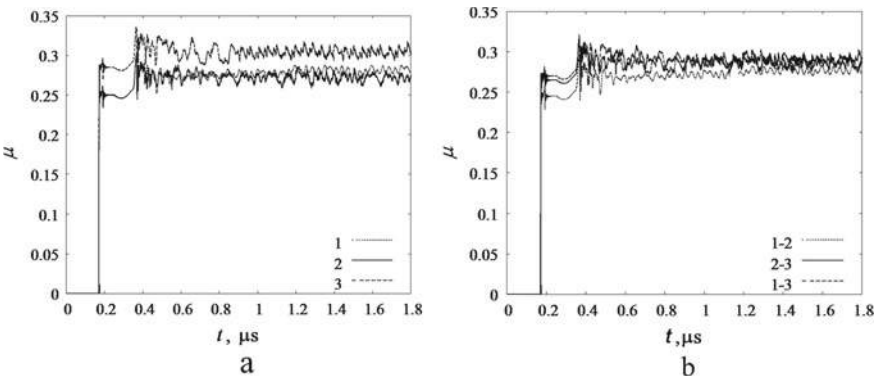


Fig. 6 Plots of the friction coefficient in the model at meso-scale versus the calculation time with the same (**a**) and different (**b**) modes of coating deposition (the numbers indicate these modes)

rough surfaces of different relief, but of the same roughness, the peak of the friction coefficient occurs, then the process goes to the stationary mode and the friction coefficient becomes almost constant. The plot shows that the greatest coefficient of friction $\mu_3 = 0.3$ is observed on the surfaces of the TiN coating in the first and third deposition modes, which corresponds to the experimental data [43], and the minimal $\mu_2 = 0.25$ in the second deposition mode. The thickness of the quasi-liquid mixing layer for the first and second regime is 0.08 μm, for the third regime—0.12 m.

It is known that an increase in the hardness of materials in contact with friction leads to a decrease in the friction coefficient, and an increase in the surface roughness leads to an increase in the friction coefficient [44, 45]. However, the materials used for the manufacture of endoprostheses should have increased hardness. Therefore, we have aimed to identify the most effective materials used in a friction pair. For this purpose, the calculations were carried out for the samples with different roughness and mechanical characteristics. According to the results of these numerical calculations, it was found that in the friction pair of the second and third modes $\mu_{2-3} = 0.27$; for the first and third modes $\mu_{1-3} = 0.28$ (Fig. 6b). The results indicate the promising use of materials obtained under different application conditions for different parts of the friction pair. At the modes 1–3 and 2–3, the thickness of the mixing layer is 0.1 μm, and for the pair of the modes 1–2—0.8 μm.

5.1.4 Modeling Friction in the Rotating Friction Pair

The next step in the numerical study of the friction in the hip joint with resurfacing endoprosthesis was the development of the macro-model of friction between the endoprosthesis casing cap rotating in the artificial insert of acetabular cup [46].

In the case of a one-component material, the geometric model consisted of a hemisphere simulating a casing cap for the femur head with an external diameter D_ext_cap = 36 mm and an interior diameter D_int_cap = 33 mm; a hemisphere simulating an acetabular cup of the hip prosthesis with an outer diameter of D_ext_insert = 41 mm and an interior diameter of D_int_insert = 38 mm; and also a conical "shell" for the cup imitating the surrounding bone tissue (Fig. 7b). In the case of a two-component (coated) material, the hollow hemispheres were additionally specified for a casing cap with an outer diameter of 35.9 mm and an inner diameter of 33.1 mm (Fig. 7a).

The load was applied by specifying the translational and rotational velocities for the automata of the resurfacing cap. These velocities corresponded to rotation of the cap as perfectly rigid body around the axis of symmetry of the corresponding sphere, which in our case was parallel to the axis *OX*. The value of the corresponding rotational velocity gradually increases from 0 to 10 s^{-1}. The bottom layer of the automata of the conical shell of the bone tissue was rigidly fixed (Fig. 7b).

When simulating a single rotational cycle, the maximum reaction force was not greater than 3 kN, which corresponds to the load of a walking man, and the angle of rotation of the resurfacing cap was 120°, which is typical for standard daily physical activities for a healthy person.

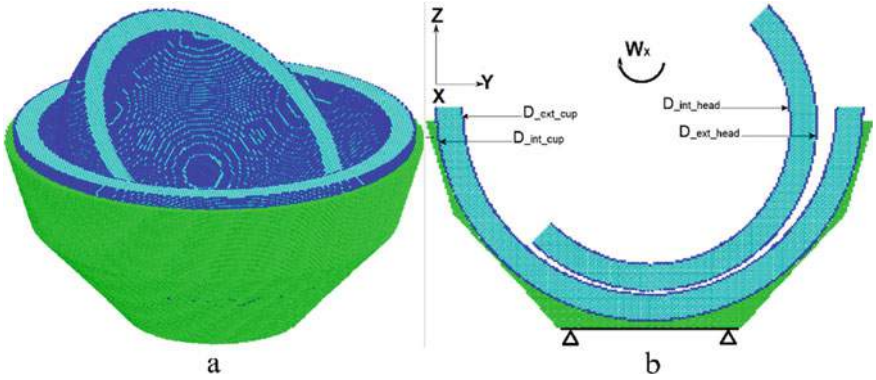


Fig. 7 General view of the model for friction pair of hip resurfacing endoprosthesis (**a**) and its cross section with loading parameters (**b**), represented by automata packing

Table 2 Properties of the model cortical bone

Material	Bulk modulus, K, GPa	Shear modulus, G, GPa	Density, ρ, kg/m^3	Ultimate strength, σ_b, GPa
Cortical bone	14	3.3	1850	0.12

Friction of Bone-Bone Pair

For comparison, based on the developed model we also studied friction between two cortical bone hemispheres (assuming them as a healthy joint) [47]. Cortical bone was considered as a linear elastic brittle material with the properties taken from [48] and shown in Table 2.

The simulation results for healthy friction pair of bone tissues showed that in the contact interaction zone of the acetabulum and femur head and behind this zone at extreme positions (the edge of the acetabulum), large compressive stresses aroused in the head and in the acetabulum with a maximum value not reaching 10 MPa. Such a load does not exceed the strength and, therefore, reduces the likelihood of premature wear of the femoral joint (Fig. 8).

Friction of Ti–Ti Pair

Then we considered stress fields in the simulated specimens for the titanium endoprosthesis joints without any coating, which are depicted in Figs. 9 and 10. The simulation results showed that in the case of a friction pair of a homogeneous metal material a large tensile stress with a maximum value not reaching 990 MPa appeared in the cap. Namely, it was observed at extreme positions in the zone of contact interaction of the acetabulum insert and resurfacing cap and behind it (edge of the

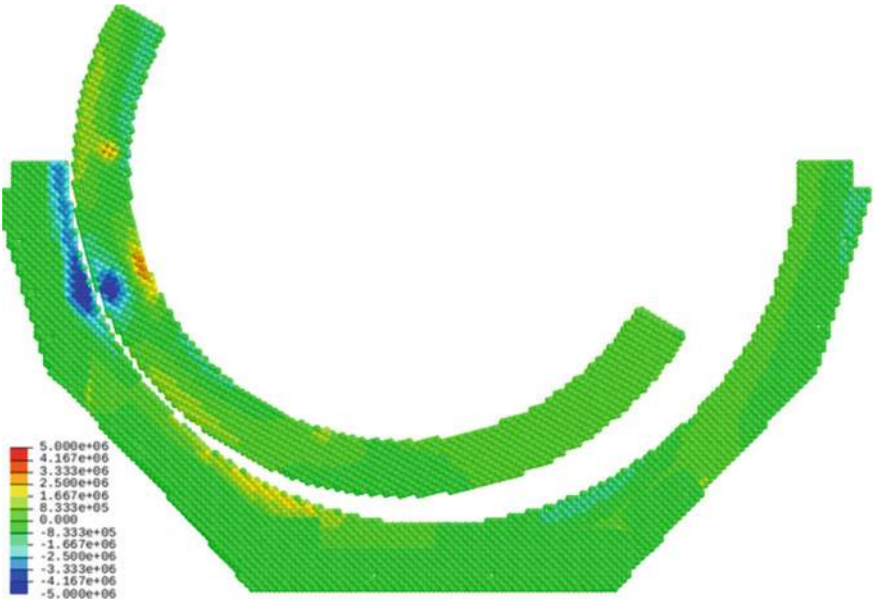


Fig. 8 Field of mean stress in the friction pair of the healthy hip joint (bone)

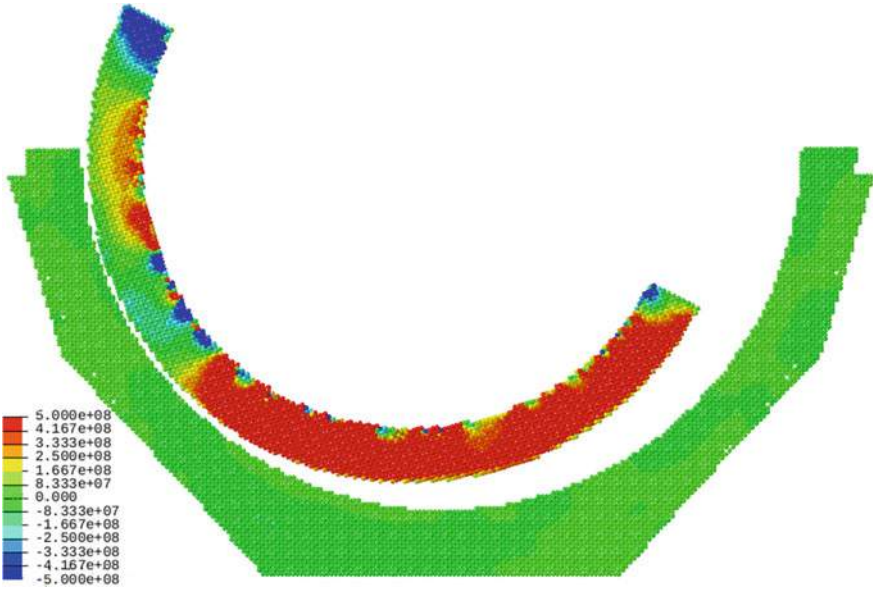


Fig. 9 Field of mean stress in the friction pair of the hip resurfacing endoprosthesis made of titanium alloy

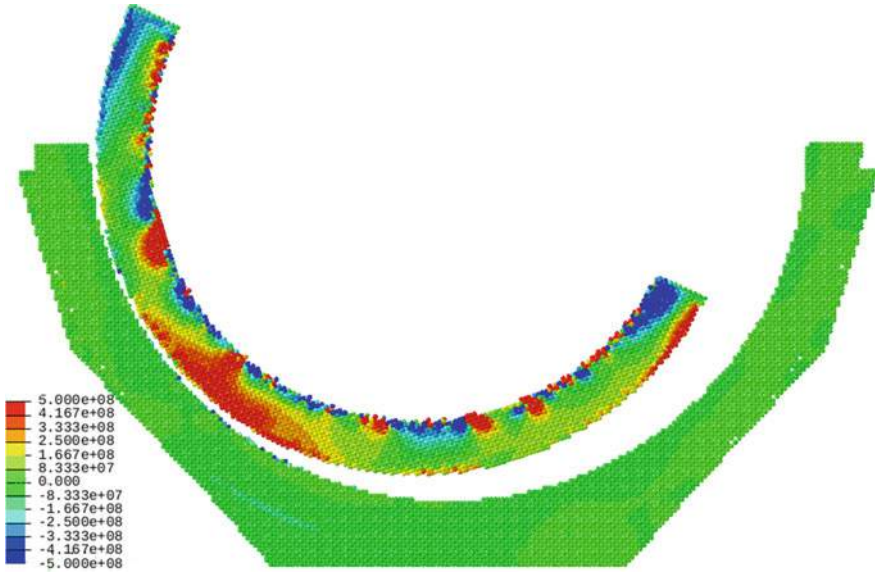


Fig. 10 Field of mean stress in the friction pair of the hip resurfacing endoprosthesis made of titanium alloy with TiN coating

acetabulum). Such a load exceeds the yield strength and, therefore, can lead to rapid wear of the surface of the resurfacing cap of the femur head of the joint (Fig. 9). These results are consistent with the data on the stress distribution in the metal head obtained in [49]. At the same time, the stress in the acetabulum insert did not exceed 100 MPa.

Friction of Ti–NiTi Pair

In the case of a coated endoprosthesis, a zone of tensile stresses with a maximum value of 1.1 GPa was observed in the contact zone, but this area was significantly smaller and concentrated mainly in the coating (Fig. 10). In addition, when using two-component materials in the friction pair, there was no noticeable increase in stress values in the cap when it was in the extreme positions. Consequently, the use of titanium alloys with a ceramic coating allows avoiding premature wear at the extreme positions of the femur head in the acetabulum. In the insert consisting of titanium alloy and coating, the value of compressive stresses reached 300 MPa. It should be noted that the magnitude of such stresses is not critical for the coating.

Thus, the presented results of the numerical simulations and their analysis suggest that the use of titanium alloy coated by TiN in the friction pair of hip resurfacing endoprosthesis can help avoiding premature wear of the endoprosthesis.

5.2 Modeling Bone–Endoprosthesis System

Finally, we developed a macromodel of hip resurfacing endoprosthesis with proximal part of the femur bone shown in Fig. 11. The geometry of the model was based on the so-called 3rd generation composite femur [50], which provides geometries of the cortical and cancellous bones as different solid bodies. A numerical model of the bone-endoprosthesis system was constructed for the resurfacing endoprosthesis with real geometric parameters. A CAD model from [50] was taken as a tubular femur, according to the parameters of which a personalized solid model for the endoprosthesis was created using FreeCAD software. For our purpose, we cut the top part of the bone geometry, added the resurfacing endoprosthesis (colored in cyan, and its coating colored in blue) and special loading part (colored in red) (Fig. 11b). Based on these solid models of the femur and endoprosthesis, mesh models were constructed in STL format, which then were imported into the MCA preprocessor as it shown [39].

Based on the developed model we simulated compression of the proximal part of the femur with resurfacing endoprosthesis, which can have a hardening coating or have not. The loading was applied by setting constant velocity in both horizontal and vertical directions as shown in Fig. 11b up to reaching the critical values of the resisting force of 3 and 10 kN [51]. The simulation results are presented in Fig. 12 as fields of mean stress.

It is reported in [52] that strength limit for the cancellous bone of healthy people reaches about 10 MPa in compression and 5 MPa in tension; and may be lower than 5 MPa and even 3 MPa for some diseases, respectively. Analysis of the stress field in the model system showed that maximum tensile stress was observed near the endoprosthesis, and did not reach critical values. Therefore, herein we tried to analyze the compression stress in the scale up to 5 MPa. From Fig. 12 one can see

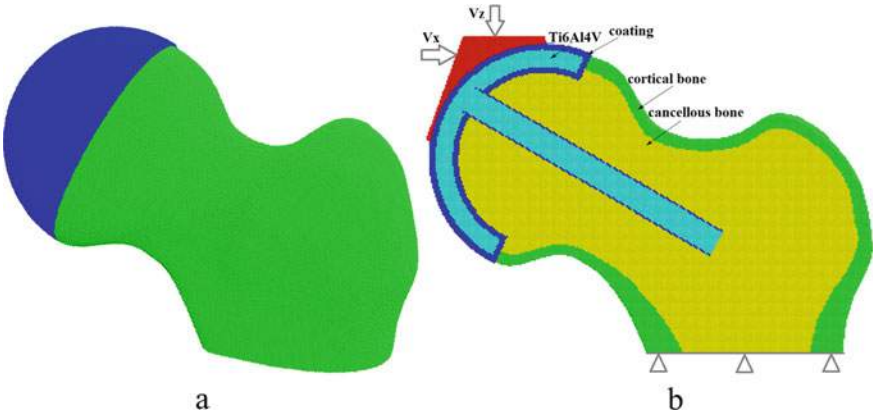


Fig. 11 The model of hip resurfacing endoprosthesis (**a**) and its cross-section with the scheme of loading (**b**)

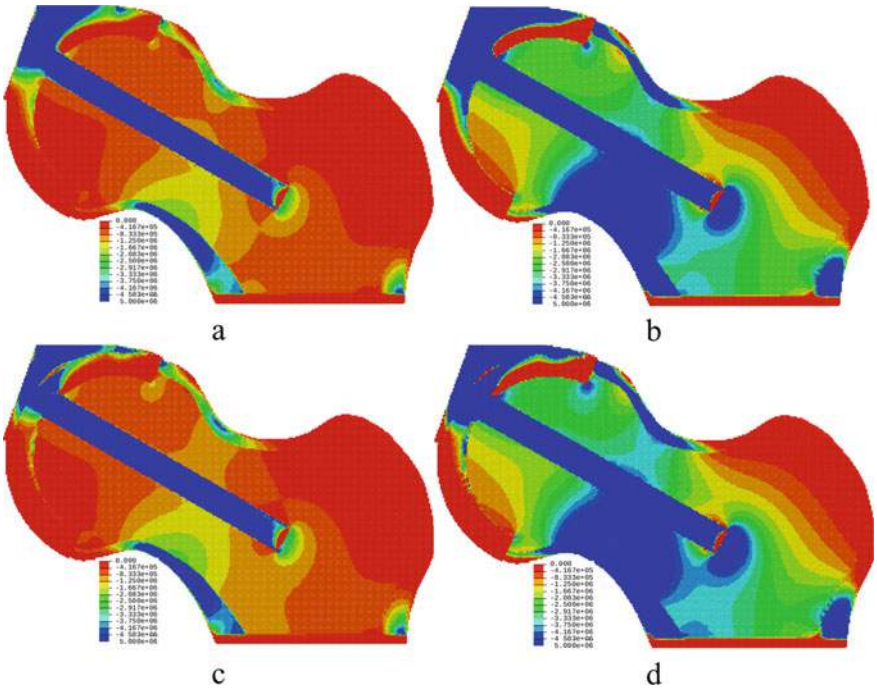


Fig. 12 Fields of mean stress in the cross-section of the system "bone-endoprosthesis" loaded by the force of 3 kN (**a**, **c**) and 10 kN (**b**, **c**) for titanium endoprosthesis (**a**, **b**) and titanium with TiN coating (**c**, **d**)

that maximum compression stress, which that can lead to fracture, is observed in the femoral neck.

5.3 Modeling of Biomaterials Based on Poroelastic Approach

It is well known, that the main difference of the living bone from one stored for a long time after removal from the body is the presence of the biological fluid. This is one of the reasons for the difference in the mechanical properties of living and "dead" bones. Moreover, this biological fluid may transfer across the whole bone and dramatically change its mechanical behavior under dynamic loading. That is why the next step in the further development of the numerical model described above is taking into account the biological fluid and its influence on the mechanical response of the femur.

5.3.1 Modification of the MCA Method to Enable Simulating Fluid-Saturated Materials

Automata that model fluid-saturated material are considered as porous and permeable. Pore space of such an automaton can be saturated with liquid. The characteristics of the pore space are taken into account implicitly through the specified integral parameters, namely, porosity ϕ, permeability k, and the ratio $a = 1 - K/K_S$ of the macroscopic value of bulk modulus K to the bulk modulus of the solid skeleton K_S. The mechanical influence of the pore fluid on the stresses and strains in the solid skeleton of an automaton is taken into account on the basis of the linear Biot's model of poroelasticity [53, 54]. Within this model, the mechanical response of a "dry" automaton is assumed linearly elastic and is described based on the above-shown relations. The mechanical effect of the pore fluid on the automaton behavior is described in terms of the local pore pressure P^{pore} (fluid pore pressure in the volume of the automaton). In the Biot model, the pore pressure affects only the diagonal components of the stress tensor. Therefore, it is necessary to modify only the relations for the central interaction forces in Eq. (3):

$$\Delta f_{ij} = 2G_i \left(\Delta \varepsilon_{ij} - \frac{a_i \Delta P_i^{\text{pore}}}{K_i} \right) + D_i \Delta \sigma_i^{\text{mean}} \tag{5}$$

Interstitial fluid is assumed to be linearly compressible. The value of fluid pore pressure in the volume of an automaton is calculated based on the relationships of Biot's poroelasticity model with the use of the current value of the pore volume. The pore space of the automata is assumed to be interconnected and provides the possibility of redistribution (filtration) of the interstitial fluid between the interacting elements. A pore pressure gradient is considered as the "driving force" of filtration. The fluid redistribution between automata is carried out by numerical solution of the classical equation of the fluid density transfer [55]. This equation is numerically solved using the finite volume method adopted for the ensemble of automata.

5.3.2 Choosing Poroelastic Parameters for Bone Tissues

The aim of this section is to choose the correct values of the model for both cortical and cancellous bone tissues. The first who considered bones like poroelastic bodies was Cowin [27]. In paper [56] he with co-authors provided the values of the main parameters of poroelastic body for cortical and cancellous tissues of the human bone. These parameters are as follows: Young's modulus, Poisson's ratio, the permeability, Biot's coefficient (or bulk modulus of the solid phase), the porosity, densities of the solid grain and the fluid. Later, several authors have made experimental and theoretical studies aimed to get the values of poroelastic parameters for some specific bones of humans and animals [48, 57–60]. Based on data published in the literature, one may conclude that there is a large scatter of the main poroelastic properties of the bone tissues. For example, geometric permeability estimates span across several

Table 3 Poroelastic properties of the model bone tissues

Bone tissue	Bulk modulus of the solid phase, K_S, GPa	Bulk modulus of the matrix, K, GPa	Shear modulus of the matrix, G, GPa	Density of the matrix, ρ, kg/m^3	Porosity ϕ	Geometric permeability, k, m^2
Cortical	17.0	14.0	5.55	1850	0.04	1.0×10^{-16}
Cancellous	17.0	3.3	1.32	600	0.80	3.5×10^{-13}

orders of magnitude $\left(10^{-25}-10^{-10}\ \mathrm{m}^2\right)$, and values for Young's modulus vary from 1 up to 25 GPa [48]. That is why it is of special interest to study the peculiarities of the mechanical behavior of small model specimens in compression depending on the variation of these properties.

Basic values of the poroelastic properties chosen for our numerical models for both cortical and cancellous bone tissues are shown in Table 3. The fluid in both bone tissues is assumed to be the same and equivalent to salt water, with a bulk modulus $K_\mathrm{f} = 2.4$ GPa, and density $\rho_\mathrm{f} = 1000$ kg/m^3.

The developed model was applied to study the dynamic mechanical behavior of the fluid-saturated porous materials under uniaxial compression at a constant speed. We studied and analysed the dependences of the effective Young's modulus of fluid-saturated materials on the strain rate and the characteristic time of fluid redistribution in the pore space. In our calculations, the material parameters and the strain rate varied within wide limits: the permeability of the material varied within 4 orders of magnitude, the viscosity of the fluid varied within 2 orders of magnitude, the sample size changed within the order of magnitude, and the strain rate varied within 3 orders of magnitude [61].

We simulated uniaxial compression of 3D cubic specimens along the vertical axis (Z). The size of the automata for all models in this study was equal to 2 mm. The base size of the cubic specimens was chosen to be 5 cm. The initial pore pressure of interstitial fluid was assumed to be equal to atmospheric. Fluid could freely flow out from the compressed specimen through the side surface.

Analysis of the simulation results showed that under compression, the values of the mechanical characteristics of the fluid-saturated material are determined by the balance of two competing processes [62, 63]:

- deformation of the solid skeleton, providing compression of the pore space and a corresponding increase in the pore pressure of the interstitial fluid;
- outflow of the interstitial fluid through the side surface, which leads to the inverse effect of lowering pore pressure.

We revealed a key control parameter that determines the specific dynamic value of the mechanical characteristics of fluid-saturated materials, namely the dimensionless Darcy number:

$$D_a = \frac{T_{\text{Darcy}}}{T_{\text{load}}} \sim \frac{\eta_{\text{fl}} L^2}{k \Delta P} \dot{\varepsilon}_{\text{def}}, \tag{6}$$

where T_{Darcy} is the characteristic time of fluid filtration (Darcy time scale), $T_{\text{load}} \sim 1 \big/ \dot{\varepsilon}_{\text{def}}$ is the time scale of the specimen deformation, $\dot{\varepsilon}_{\text{def}}$ is the specimen strain rate, η_{fl} is the dynamic viscosity of the pore fluid, L is the characteristic length of the pore pressure difference ΔP (a half of length of the cubic side in the considered case). The parameter D_a characterizes the ratio of the timescales of deformation of the fluid-saturated porous specimen and filtration of the pore fluid.

The simulation results showed that the effective Young's modulus of fluid-saturated bone tissues non-linearly depends on the strain rate $\dot{\varepsilon}_{\text{def}}$, the specimen size L, the dynamic viscosity of pore fluid η_{fl}, and the permeability of solid skeleton k. In particular, Young's modulus of a fluid-saturated sample is minimal (equal to E_{min}) at infinitely small strain rates and tends to the maximum value (Young's modulus of undrained sample E_{max} [54]) at large ones. The key result is the established ability to build single "gauge" dependence applicable to specimens of porous materials of various sizes, characterized by different permeability of solid skeleton, different fluid viscosities, and deformed at different strain rates. An argument of such a "master curve" is the dimensionless Darcy number D_a (Fig. 13):

$$E = E_{\text{max}} + \frac{E_{\text{min}} - E_{\text{max}}}{1 + \left(D_a \big/ D_{a_0} \right)^p}, \tag{7}$$

where E_{min} corresponds to $D_a \to 0$ ("dry" specimen), E_{max} corresponds to $D_a \to \infty$ (undrained specimen), D_{a_0} and p are the fitting constants. This master curve has a logistic (sigmoidal) form but the fitting values of D_{a_0} and p (as well as E_{min} and E_{max}) are different for cortical and cancellous bones.

Based on the presented results we chose the values of the poroelastic parameters for both bone tissues applied to the model proximal femur corresponding to Darcy

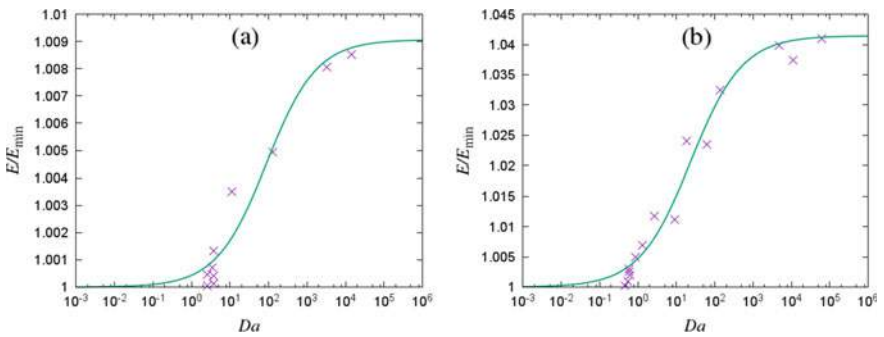


Fig. 13 Dependences of the normalized Young's moduli of the modeled specimens of cortical (**a**) and cancellous (**b**) bone tissues with different poroelastic properties on the Darcy number. Crosses show numerically obtained values, lines are approximating curves

numbers about 50, i.e. the middle of the plots range shown in Fig. 13. This means that at the loading rate of the model femur the effect of interstitial fluid flow is expected to be well pronounced.

5.3.3 Validation of the Materials Model

During life, the properties of the human bones change considerably; usually they become fragile. In [64], the concept of bone fragility was formulated as follows: bone tissue can adapt its shape and size in response to mechanical stress through a remodeling mechanism, during which the bones are formed or rearranged under the independent action of osteoclasts and osteoblasts. Remodeling is a process that supports the mechanical resistance of the skeleton, allowing you to selectively restore and replace damaged bone tissue. During the period of growth, these processes form a structure capable of adapting to the loads and maintaining strength. With age, the natural remodeling processes slow down under the influence of such factors as decreasing muscle mass and physical activity, malnutrition. Consequently, bone embrittlement occurs. Osteoporosis is a systemic skeletal disease characterized by low bone mass and impaired microarchitecture of bone tissue, which leads to increased bone fragility (reduced bone density) and a tendency to fractures. Researchers and clinicians continue to persist in finding ways to prevent the 8 million osteoporotic fractures occurring annually around the world [65]. An early and accurate assessment of the risk of fracture, with timely initiated treatment—this approach seems to be the most appropriate for reducing this number of fractures and associated personal and social losses.

There are many methods for the direct assessment of the structural strength and the material composition of the bone, such as testing of whole bone, bone mass, strength assessment using microindentation [66]. However, all of these methods are used for in vitro or in vivo research and are therefore not suitable for clinical practice. Recently, however, indentation has become positioned as a procedure for clinical use with minimal invasiveness [67]. In the indentation method, the bones are penetrated using an indenter tip with a depth sensor. Among the advantages of this method, there is an ability to measure material properties and microstructural features and identify local changes in bone material caused, for example, by disease or medication [68]. However, the method remains invasive and extremely local, therefore, at present, the development of methods for non-invasive diagnostics of the strength properties of local bone tissue is actively underway.

One of the new directions in this area is the combination of methods for visualizing the structure of bone tissue (CT and MRI) with computer modeling [69]. Therefore, the development of numerical models of the mechanical behavior of bone tissue during micro and macroindentation is relevant.

Here we describe a numerical model of the mechanical behavior of fluid-saturated cancellous tissue during indentation used for validation of the developed model of fluid-saturated bone tissues [70].

Geometrically, the bone tissue indentation model was a parallelepiped with a Berkovich pyramid (Fig. 1). During indentation, the counter-body was set as the rigid non-deformable indenter, which motion was set through the velocity in the vertical direction. This velocity was set to $V_z = -1$ m/s until the indenter was penetrated at a given depth, after which the velocity $V_z = 1$ m/s was set to simulate unloading. The simulation results were processed using the Oliver-Farr method [40].

The mechanical properties for the cancellous bone were chosen from Table 3, and the fluid in bone tissue is assumed to be equivalent to salt water as before.

It is known that bone tissue, when indented, behaves like a soft material, and gives an error in determining the elastic modulus using a standard experimental procedure; thus, indentation with a time delay of indentation depth (holding) looks to be better for this material [71]. Therefore, one of the first experiments to test the developed model was aimed to establish the fact of the influence of holding time on the mechanical properties of bone tissue. Three variants of loading were considered: (1) without holding $t_{\text{hold}} = 0$; (2) holding time corresponded to loading time $t_{\text{hold}} = t_{\text{loading}}$; (3) holding time was two times longer than loading time $t_{\text{hold}} = 2 \cdot t_{\text{loading}}$ (Fig. 14a).

When analyzing the load–displacement curves (Fig. 14b) using the Oliver-Farr method, the following values of recoverable characteristics were obtained: the hardness of 47 MPa for all loading regimes, the elastic modulus 3.7 GPa for no holding, $E = 3.5$ GPa for $t_{\text{hold}} = t_{\text{loading}}$, $E = 3.33$ GPa for $t_{\text{hold}} = 2 \cdot t_{\text{loading}}$. The obtained results showed that the elastic modulus is determined correctly at $t_{\text{hold}} = t_{\text{loading}}$ and corresponds to the specified input value. In addition, these calculations suggest that at constant deformation, stress relaxation occurs, which indicates that the model specimen of cancellous bone tissue possesses viscoelastic properties.

At the next stage, the effect of interstitial fluid flow in the porous skeleton of the material was investigated. According to the simulation results (Fig. 15a), it was established that the hardness of the undrained bone tissue specimen was 55 MPa, while the hardness of the fluid-saturated bone was 47 MPa, and the elastic moduli

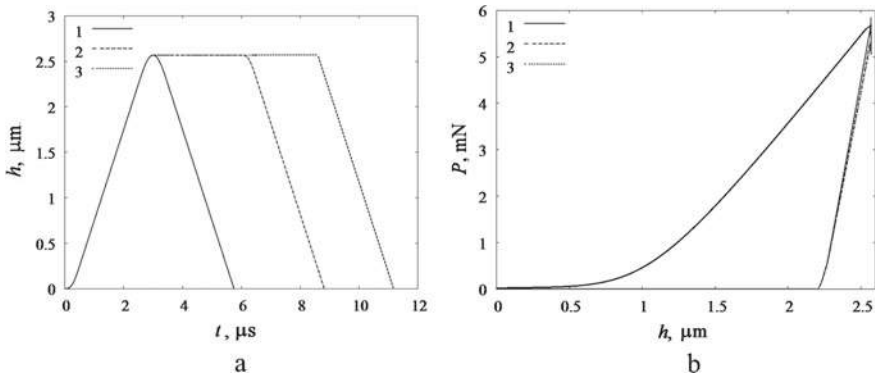


Fig. 14 **a** The indenter penetration depth versus time; **b** load–displacement curve for different holding times; numbers indicate loading regimes: (1) $t_{\text{hold}} = 0$, (2) $t_{\text{hold}} = t_{\text{loading}}$, (3)$t_{\text{hold}} = 2 \cdot t_{\text{loading}}$

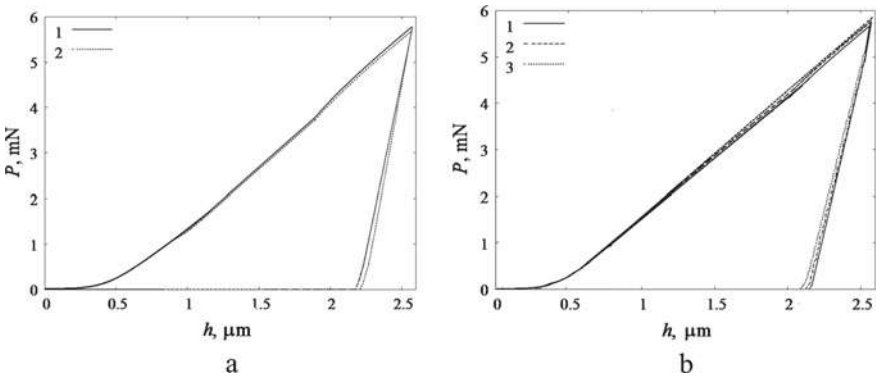


Fig. 15 Load–displacement curves for undrained (**a**), numbers indicate loading rates: (1) undrained, (2) fluid-saturated and fluid-saturated model specimens (**b**), numbers indicate loading rates: (1) $v = 1$ m/s, (2) $v = 5$ m/s, (3) $v = 10$ m/s

were 3.9 and 3.6 GPa, respectively. The results obtained correspond qualitatively to the data on the influence of fluid on the mechanical response of bone tissue by other authors [72–74].

Further calculations were carried out to study the effect of porosity value of the bone tissue and its permeability on the mechanical response of the bone in indentation. However, the results obtained indicate that with the chosen values of porosity (80%), the above factors do not have a significant effect on the mechanical response of the fluid-saturated bone tissue.

The viscoelastic behavior of the bone tissue is also characterized by the dependence of the mechanical characteristics on the loading rate; therefore, at the next stage of testing the numerical model development, experiments were performed on the indentation of a model specimen of cancellous bone tissue with different loading rates. The study of the effect of loading rate on the mechanical response of fluid-saturated bone tissue showed that an increase in loading rate leads to an increase in the values of recoverable characteristics (Fig. 15b), thus at the speed of $V = 1$ m/s we got $H = 47$ MPa, $E = 3.5$ GPa, at the speed of $v = 5$ m/s we got $H = 53$ MPa, $E = 3.7$ GPa, and at the speed of $v = 10$ m/s we got $H = 56$ MPa, $E = 3.95$ GPa. The obtained values, again, corresponds qualitatively to the results of other authors [75–77].

5.3.4 Modeling the Bone Compression

Similar as in the previous section, the geometry of the bone is based on the 3rd generation composite femur [50], which consists of the cortical and cancellous parts as different solid bodies. General view of the model represented as fcc packing of automata and its cross-section are shown in Fig. 16.

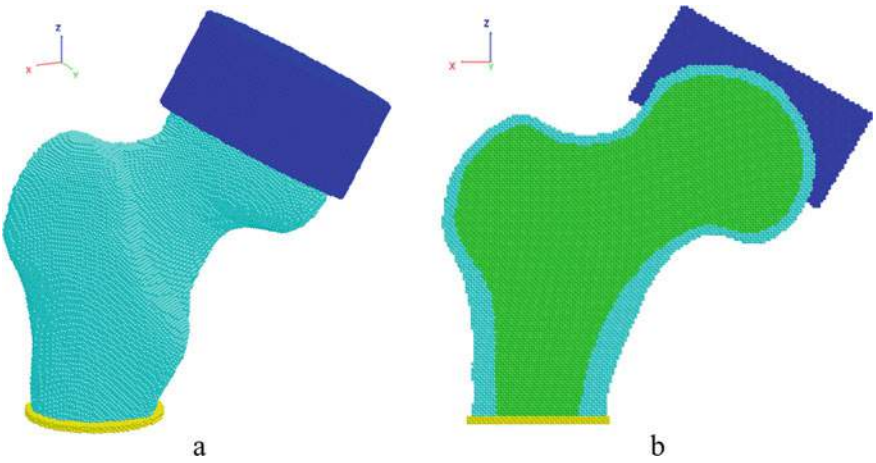


Fig. 16 General view of the model for proximal femur (**a**) and its cross-section (**b**)

At the bottom of the model, we place a disk with properties of the cortical bone; the automata of this disk are fixed. For compressing the femur, we place a special cylindrical "cap" on the femur head. This "cap" is shown in blue color in Fig. 16. Automata of the "cap" have the properties corresponding to cartilage. Loading is applied by setting the constant velocity $V = 1$ m/s to the automata of the upper face of the "cap". The velocity vector is directed along the face normal. Note, that this loading results in both compression and small bending of the bone.

To reveal the role of the interstitial fluid flow in the model femur under compression, we studied three different cases. In the first case, all bone tissues contained no fluid, i.e. were "dry" (so-called drained test). In the second case, we used the chosen poroelastic parameters from Table 3. In the third case, all pores containing fluid were assumed to be closed, which means no permeability of the materials (undrained test). Then we analyzed the distributions of mean stress, equivalent stress and pore fluid pressure at the final point at the total strain about 2%.

Fields of mean stress in the cross section of the model femur for all three cases are shown in Fig. 17. One can see that fluid filtration cause the increase in mean stress in the cortical part of the femur head, especially in the area of its contact with the loading "cap". At the same time, the drained and undrained tests do not differ considerably from each other. However, fields of equivalent stress in the cross section of the model for these cases shown in Fig. 18, clearly demonstrate that shear stress in the same area is much smaller for the undrained test, while for the two other cases are practically the same.

It can be clearly seen from Figs. 16 and 18 that the maximum stresses occur in the cortical part of the femur. Figure 19 shows the 3D view of both equivalent and mean stresses distribution in the model (particularly, the outer part of the cortical bone). It is obvious that the main stresses are localized in the femoral neck. The shear stresses propagate along the loading direction into the main part of the bone up to the

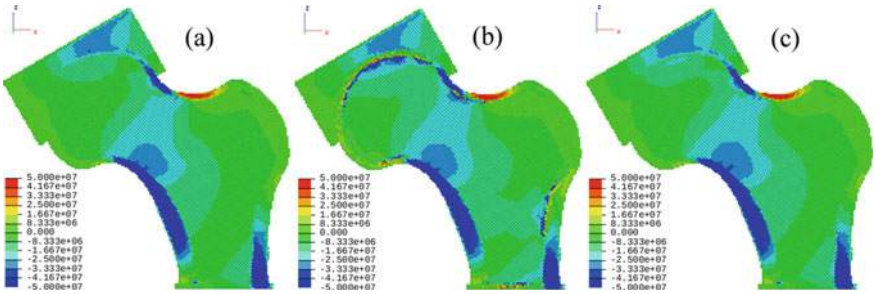


Fig. 17 Distribution of mean stress in cross-section of the model femur at drain test (**a**), test with chosen poroelastic parameters (**b**), and undrained test

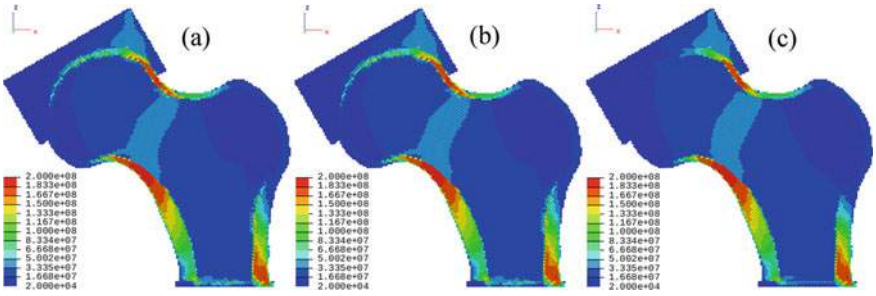


Fig. 18 Distribution of equivalent stress in cross-section of the model femur at drained test (**a**), test with chosen poroelastic parameters (**b**), and undrained test

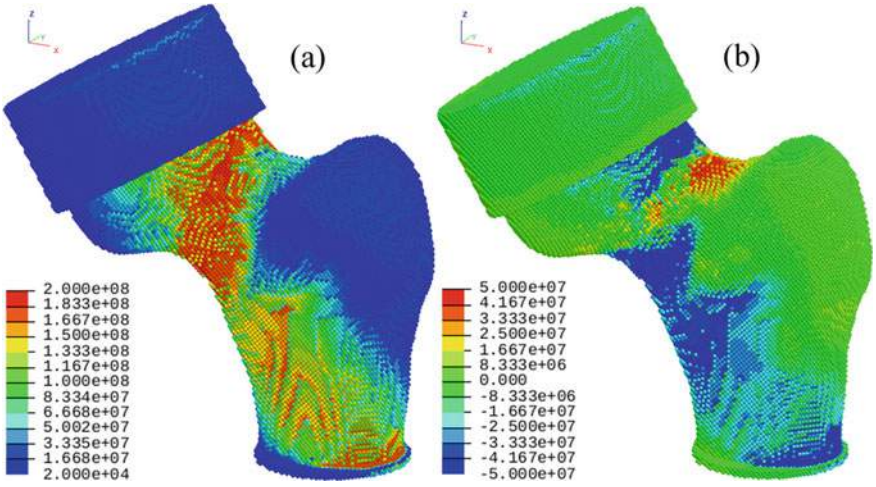


Fig. 19 Distribution of equivalent stress (**a**) and mean stress (**b**) in the model femur with chosen poroelastic parameters

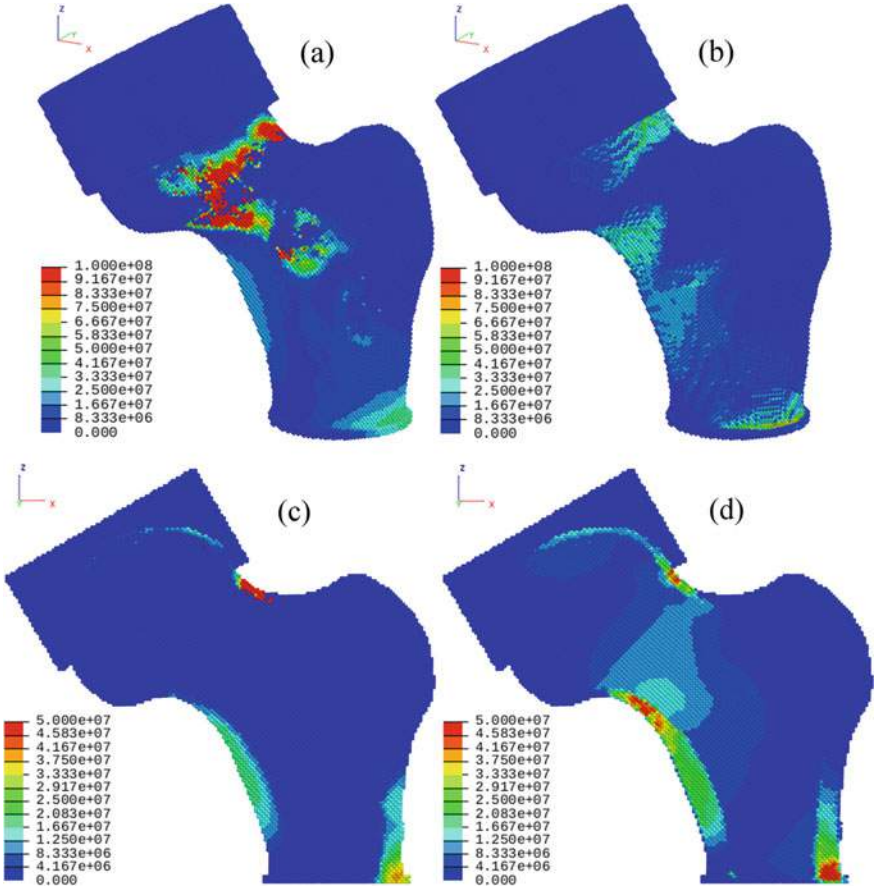


Fig. 20 Distribution of pore pressure in the model for the test with chosen poroelastic parameters (**a**, **c**) and undrained test (**b**, **d**), **c** and **d** are corresponding cross-sections

supporting plate. But the most dangerous seems the tensile mean stress in the upper part of the femoral neck (Fig. 19b).

Figure 20 shows fields of the pore pressure in the cases with interstitial fluid. It can be seen that ability to flow results in filtration of the fluid to the regions of large tensile and shear stresses of the cortical bone, but not the maximum tensile mean stress (upper part of the femoral neck). At the same time, one can see negligible pore pressure in the cancellous bone in the case of fluid filtration (Fig. 20c).

5.3.5 Modeling the Bone-Endoprosthesis System

In the last sub-section, we consider the femur–endoprosthesis system, where the bone is described using poroelastic approach and the implant is made of the TiN-coated

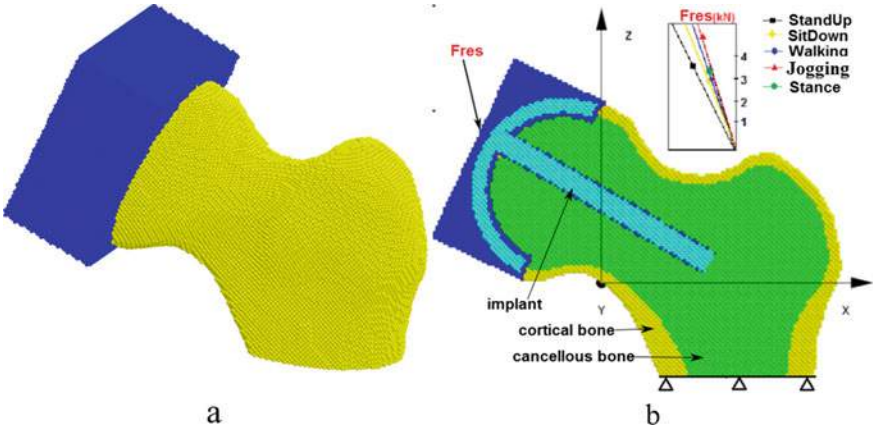


Fig. 21 Model system "bone-endoprosthesis". **a** 3D view; **b** cross-section view with scheme of loading

titanium alloy Ti6Al4V. The values for the main physico-mechanical properties of the titanium alloy and TiN coating produced by the PIRAC deposition mode 2 are taken from Table 1. The properties for the bone tissues are taken from Table 3. A model of the bone-endoprosthesis system was constructed by analogy with Sect. 5.2 and is shown in Fig. 21a.

The main feature of this section is that here we vary the loading similar to the different types of human activity. Thus, dynamic loading F_{res}, which is equivalent to the physiological one for a person weighing 75 kg, was applied to the upper part of the implant (Fig. 21b). According to [78], this force lies in the medial plane ZX and is inclined under different angles relative to the bone axis Z for different kinds of activity. Standing up load is characterized by the total force of 3.6 kN and applied at the angle of 24°; sitting down load of 3 kN is applied at the angle of 20°; a load of walking is 3 kN and applied at the angle of 17°; jogging is characterized by 4.5 kN and applied under angles of 15°; stance position is characterized by 3.2 kN and applied under angle of 16°. Here, the loading is simulated through the setting constant velocity to the automata of the external surface of the loading block marked by the blue color in Fig. 21 up to the moment when the required loading value of the resistance force is reached, similarly as it was done in [78, 79]. The value of the loading velocity is 1 m/s for walking, sitting, standing up and position while standing, for jogging the loading velocity is 2 m/s [79].

Typical patterns of the mean stress fields obtained by simulations are shown in Fig. 22. According to the presented results, the highest tensile and compressive stresses are concentrated in the upper and lower parts of the femur neck, respectively.

Analysis of the images also shows that the angle of the load application for the same value of the resulting force significantly affects the distribution of compressive stresses in the proximal femur (Fig. 22a–c). Thus, an increase in the angle of the load application by 4° with a simultaneous increase in the force by 20% leads to an increase

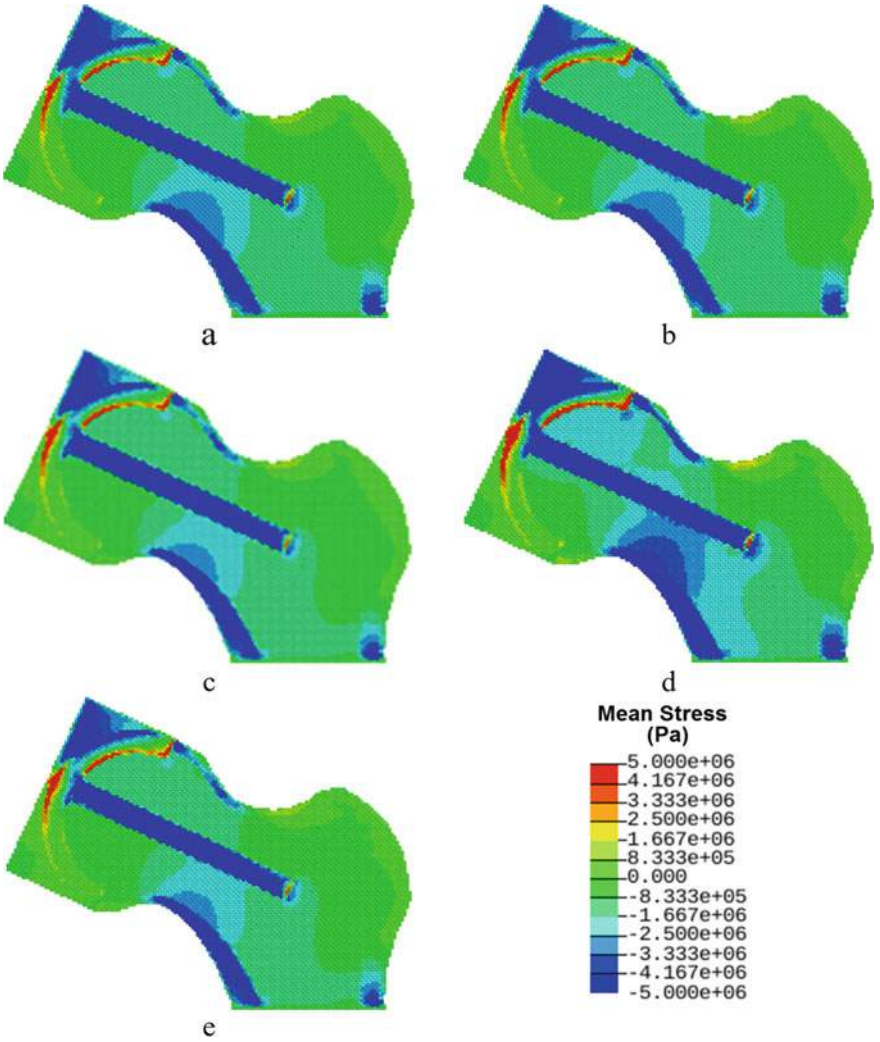


Fig. 22 Fields of mean stress in the system "bone-endoprosthesis" under different physiological loading: **a** standing up, **b** sitting down, **c** walking, **d** jogging and **e** stance position

in the area of compressive stresses by 20% (Fig. 22b, a, respectively). A decrease in the angle of the load application by 3° at the same value of the force (the cases of sitting down and walking) causes an increase in the localization area of maximum compressive stresses by about 15% (Fig. 22b, c, respectively). At the same time, a decrease in the angle of the load application by 2° relative to the nominal direction with an increase in the loading force by 7% (the cases of walking and standing position) leads to a decrease in the localization area of maximum compressive stresses by 5% (Fig. 22c, e, respectively). The highest stress concentration is observed for

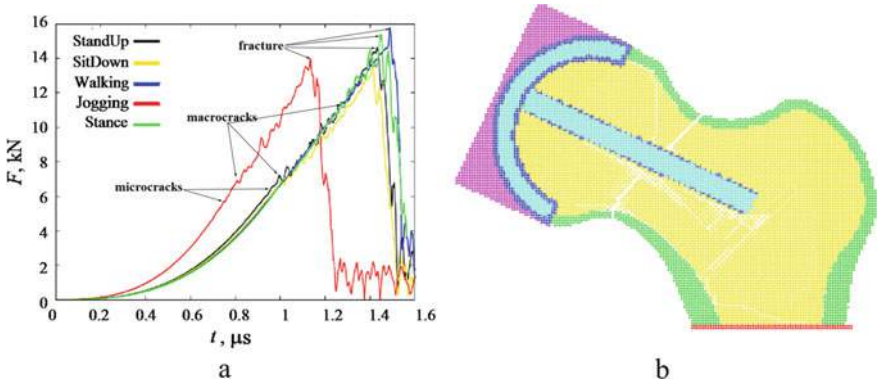


Fig. 23 **a** Plots of the loading force versus time for the considered types of physiological loading; **b** fracture pattern of the bone–endoprosthesis system (the lines connect bonded automata, the cracks correspond to the white zigzag "curves" in the network of the bonded automata)

jogging: compressive stresses are observed in the lower part of the neck under the implant pin, as well as in the femoral head area under the casing cap (Fig. 22d).

Analysis of the loading plots shown in Fig. 23a and the corresponding fracture patterns (one example is depicted in Fig. 23b) allows us to conclude that for all types of physiological loads microcracks arise at the loading of 5–6 kN, while formation of a macrocrack is observed at the force above 6 kN. When the force value reaches 14–16 kN the complete failure of the femoral neck takes place; according to [80] the resulted pattern of crack depicted in (Fig. 23b) may be defined as the subcapital fracture. It worth noting that the least probability of fractures occurs at walking, while the greatest when sitting down.

One more interesting peculiarity that can be seen from the loading curves depicted in Fig. 23a, is a lower failure force at jogging (red curve). It can be explained by the minimum angle of the load inclination for this case, which means the maximum arm of the force for breaking the femoral neck. Failure, in this case, occurs earlier because of the higher loading velocity.

Thus, we finally considered the mechanical behavior of the femur bone with hip resurfacing under conditions of real physiological loads. The simulation data obtained, indicate that with a decrease in the angle of the force application relative to the femur axis, an increase in the area of compressive stresses is observed, as well as the appearance of such stresses in the femur head in the region closed to the prosthesis cap. With a further increase in loading force, corresponding to an increase in the body weight or the performance of physical exercises with weighting, a tendency to fracture is observed. Furthermore, the accumulation of micro and macro-cracks in the bone tissue under the loading force values, slightly exceeding the natural physiological levels, can also under certain circumstances lead to failure. A critical physiological load, in which the destruction of bone tissue starts to appear in the case of resurfacing endoprosthesis, corresponds to sitting down in excess weight conditions or using additional weights.

6 Conclusions and Future Work

This chapter presents the multiscale numerical models that can be used as a simulation tool for the virtual analysis of the femur–endoprosthesis system for the case of hip resurfacing arthroplasty. The movable cellular automaton method, which is used for simulation of the mechanical behavior of the materials, allows explicit accounting for contact loading of rough surfaces of the TiN coatings used in the friction pair of the endoprosthesis as well as fracture of the bone tissues at different scales of the model. The implementation of the Boit theory of poroelasticity within the movable cellular automaton method allows revealing the role of the interstitial biological fluid in the mechanical behavior of the bone tissues.

Based on the analysis of the obtained simulation results, the following plan for future works is proposed. First, it is necessary to use real geometry of the bone and endoprosthesis, which are more complicated than was considered here. This means that we need to consider not only the proximal part of the femur, but at least a whole bone, and maybe with the knee joint. Real resurfacing caps have many geometrical features that also have to be included in future models. All these requirements lead to the use of the small size of the automata and hence the huge number of them and large computational costs. Second, the loads have to be also refined in order to be more realistic. The third important problem that has to be considered in the future is the detained description of the interface between an implant and bone, including osseointegration.

Acknowledgements The research presented in this chapter was supported by the Russian Foundation for Basic Research (Grant No. 20-08-00818, simulation results) and the Government research assignment for ISPMS SB RAS (in-house software development).

References

1. Bitar D, Parvizi J (2015) Biological response to prosthetic debris. World J Orthop 6(2):172–189. https://doi.org/10.5312/wjo.v6.i2.172
2. Sovak G, Weiss A, Gotman I (2000) Osseointegration of Ti6A14V alloy implants coated with titanium nitride by a new method. J Bone Joint Surg 82(2):290–296. https://doi.org/10.1302/0301-620X.82B2.0820290
3. Gotman I, Gutmanas EY, Hunter G (2017) In: Ducheyne P (ed) Comprehensive biomaterials II. Elsevier, Amsterdam, pp 165–203. https://doi.org/10.1016/B978-0-12-803581-8.09795-2
4. Wu SJ, Li H, Wu SY, Guo Q, Guo B (2014) Preparation of titanium carbide–titanium boride coatings on Ti6Al4V by PIRAC. Surf Eng 30(9):693–696. https://doi.org/10.1179/1743294414Y.0000000304
5. Wu Si, Ma S, Wu Sh, Zhang G, Dong N (2018) Composition, microstructure, and friction behavior of PIRAC chromium carbide coatings prepared on Q235 and T/P 24. Int J Appl Ceram Technol 15:501–507. https://doi.org/10.1111/ijac.12803
6. Kot M, Rakowski W, Lackner JM, Major T (2013) Analysis of spherical indentations of coating-substrate systems: experiments and finite element modeling. Mater Des 43:99–111. https://doi.org/10.1016/j.matdes.2012.06.040

7. Lofaj F, Németh D (2017) The effects of tip sharpness and coating thickness on nanoindentation measurements in hard coatings on softer substrates by FEM. Thin Solid Films 644:173–181. https://doi.org/10.1016/j.tsf.2017.09.051
8. Marchiori G, Lopomo N, Boi M, Berni M, Bianchi M, Gambardella A, Visani A, Russo A (2016) Optimizing thickness of ceramic coatings on plastic components for orthopedic applications: a finite element analysis. Mater Sci Eng C 58:381–388. https://doi.org/10.1016/j.msec.2015.08.067
9. Bouzakis KD, Michailidis N, Hadjiyiannis S, Skordaris G, Erkens G (2002) The effect of specimen roughness and indenter tip geometry on the determination accuracy of thin hard coatings stress–strain laws by nanoindentation. Mater Charact 49(2):149–156. https://doi.org/10.1016/S1044-5803(02)00361-3
10. Jiang WG, Su JJ, Feng XQ (2008) Effect of surface roughness on nanoindentation test of thin films. Eng Fract Mech 75(17):4965–4972. https://doi.org/10.1016/j.engfracmech.2008.06.016
11. Sliwa A, Mikuła J, Gołombek K, Tanski T, Kwasny W, Bonek M, Brytan Z (2016) Prediction of the properties of PVD/CVD coatings with the use of FEM analysis. Appl Surf Sci 388(PartA):281–287. https://doi.org/10.1016/j.apsusc.2016.01.090
12. Skordaris G, Bouzakis KD, Kotsanis T, Charalampous P, Bouzakis E, Breidenstein B, Bergmann B, Denkena B (2017) Effect of PVD film's residual stresses on their mechanical properties, brittleness, adhesion and cutting performance of coated tools. CIRP J Manuf Sci Technol 18:145–150. https://doi.org/10.1016/j.cirpj.2016.11.003
13. Zlotnikov I, Dorogoy A, Shilo D, Gotman I, Gutmanas E (2010) Nanoindentation, modeling, and toughening effects of zirconia/organic nanolaminates. Adv Eng Mater 12(9):935–941. https://doi.org/10.1002/adem.201000143
14. Li J, Beres W (2006) Three-dimensional finite element modelling of the scratch test for a TiN coated titanium alloy substrate. Wear 260:1232–1242. https://doi.org/10.1016/j.wear.2005.08.008
15. Holmberg K, Laukkanen A, Ronkainen H, Wallin K, Varjus S (2003) A model for stresses, crack generation and fracture toughness calculation in scratched TiN-coated steel surfaces. Wear 254(3–4):278–291. https://doi.org/10.1016/S0043-1648(02)00297-1
16. Pandure PS, Jatti V, Singh TP (2014) Three dimensional FE modeling and simulation of nanoindentation and scratch test for TiN coated high speed steel substrate. Int J Appl Eng Res 9(15):2771–2777
17. Toparlj M, Sasaki S (2002) Evaluation of the adhesion of TiN films using nanoindentation and scratch testing. Philos Mag A 82(10):2191–2197. https://doi.org/10.1080/01418610208235729
18. Zhang T, Harrison NM, McDonnell PF, McHugh PE, Leen SB (2013) A finite element methodology for wear–fatigue analysis for modular hip implants. Tribol Int 65:113–127. https://doi.org/10.1016/j.triboint.2013.02.016
19. Kruger KM, Tikekar NM, Heiner AD, Baer TE, Lannutti JJ, Callaghan JJ, Brown TD (2014) A novel formulation for scratch-based wear modelling in total hip arthroplasty. Comput Methods Biomech Biomed Eng 17(11):1227–1236. https://doi.org/10.1080/10255842.2012.739168
20. Donati D, Colangeli M, Colangeli S, Di Bella C, Mercuri M (2008) Allograft-prosthetic composite in the proximal tibia after bone tumor resection. Clin Orthop Relat Res 466(2):459–465. https://dx.doi.org/10.1007%2Fs11999-007-0055-9
21. Koukal M, Fuis V, Florian Z, Janíček P (2011) A numerical study of effects of the manufacture perturbations to contacts of the total hip replacement. Eng Mech 18(1):33–42
22. Ashkanfar A, Langton DJ, Joyce TJ (2017) A large taper mismatch is one of the key factors behind high wear rates and failure at the taper junction of total hip replacements: a finite element wear analysis. J Mech Behav Biomed Mater 69:257–266. https://doi.org/10.1016/j.jmbbm.2017.01.018
23. Askari E, Flores P, Dabirrahmani D, Appleyard R (2016) A review of squeaking in ceramic total hip prostheses. Tribol Int 93(A):239–256. https://doi.org/10.1016/j.triboint.2015.09.019
24. Kuhl E, Balle F (2005) Computational modeling of hip replacement surgery: total hip replacement vs. hip resurfacing. Technische Mechanik 25(2):107–114

25. Dickinson A, Taylor A, Browne M (2012) Implant–bone interface healing and adaptation in resurfacing hip replacement. Comput Methods Biomech Biomed Eng 15(9):935–947. https://doi.org/10.1080/10255842.2011.567269
26. Dickinson AS, Brown M, Roques AC, Taylor AC (2014) A fatigue assessment technique for modular and pre-stressed orthopaedic implants. Med Eng Phys 36(1):72–80. https://doi.org/10.1016/j.medengphy.2013.09.009
27. Cowin SC (1999) Bone poroelasticity. J Biomech 32(3):217–238. https://doi.org/10.1016/S0021-9290(98)00161-4
28. Zhang D, Cowin SC (1996) Mechanics of poroelastic media. Springer, Dordrecht (NL), pp 273–298. https://doi.org/10.1007/978-94-015-8698-6_16
29. Manfredini P, Cocchetti G, Maier G, Redaelli A, Montevecchi FM (1999) Poroelastic finite element analysis of a bone specimen under cyclic loading. J Biomech 32(2):135–144. https://doi.org/10.1016/S0021-9290(98)00162-6
30. Abousleiman Y, Cui L (1998) Poroelastic solutions in transversely isotropic media for wellbore and cylinder. Int J Solids Struct 35(34–35):4905–4929. https://doi.org/10.1016/S0020-7683(98)00101-2
31. Psakhie SG, Horie Y, Korostelev SY, Smolin AY, Dmitriev AI, Shilko EV, Alekseev SV (1995) Method of movable cellular automata as a tool for simulation within the framework of physical mesomechanics. Russ Phys J 38(11):1157–1168. https://doi.org/10.1007/BF00559396
32. Shilko EV, Psakhie SG, Schmauder S, Popov VL, Astafurov SV, Smolin AYu (2015) Overcoming the limitations of distinct element method for multiscale modeling of materials with multimodal internal structure. Comput Mater Sci 102:267–285. https://doi.org/10.1016/j.commatsci.2015.02.026
33. Smolin AYu, Shilko EV, Astafurov SV, Kolubaev EA, Eremina GM, Psakhie SG (2018) Understanding the mechanisms of friction stir welding based on computer simulation using particles. Defence Technol 14(6):643–656. https://doi.org/10.1016/j.dt.2018.09.003
34. Smolin AY, Smolin IY, Shilko EV, Stefanov YP, Psakhie SG (2019) Handbook of mechanics of materials. Springer, Singapore, pp 1675–1714. https://doi.org/10.1007/978-981-10-6884-3_35
35. Psakhie SG, Smolin AY, Shilko EV, Dimaki AV (2019) Handbook of mechanics of materials. Springer, Singapore, pp 1311–1345. https://doi.org/10.1007/978-981-10-6884-3_79
36. Datasheet M (2000) Properties and processing of TIMETAL 6–4s. Titanium Metals Corporation, Dallas
37. Bonello T, Avelar-Batista Wilson JC, Housden J, Gutmanas EY, Gotman I, Matthews A, Leyland A, Cassar G (2014) Evaluating the effects of PIRAC nitrogen-diffusion treatments on the mechanical performance of Ti–6Al–4V alloy. Mater Sci Eng A 619:300–311. https://doi.org/10.1016/j.msea.2014.09.055
38. Giannakpoulos AE, Suresh S (1999) Determination of elastoplasic properties by instrumented sharp indentation. Scripta Mater 40(10):1191–1198. https://doi.org/10.1016/S1359-6462(99)00011-1
39. Eremina GM, Smolin AY (2019) Multilevel numerical model of hip joint accounting for friction in the hip resurfacing endoprosthesis. Facta Universitas Ser Mech Eng 17(1):29–38. https://doi.org/10.22190/FUME190122014E
40. Oliver W, Pharr GM (1992) An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. J Mater Res 7(6):1564–1583. https://doi.org/10.1557/JMR.1992.1564
41. Avelar-Batista Wilson JC, Wu S, Gotman I, Housden J, Gutmanas EY (2015) Duplex coatings with enhanced adhesion to Ti alloy substrate prepared by powder immersion nitriding and TiN/Ti multilayer deposition. Mater Lett 157:45–49. https://doi.org/10.1016/j.matlet.2015.05.054
42. Eremina GM, Smolin AYu (2019) Numerical modeling of wearing two rough surfaces of a biocompatible ceramic coating. AIP Conf Proc 2167:020089. https://doi.org/10.1063/1.5131956
43. Attard B, Leyland A, Matthews A, Gutmanas EY, Gotman I, Cassar G (2018) Improving the surface characteristics of Ti-6Al-4V and Ti metal 834 using PIRAC nitriding treatments. Surf Coat Technol 339:208–223. https://doi.org/10.1016/j.surfcoat.2018.01.051

44. Kang J, Wang M, Yue W, Fu Z, Zhu L, She D, Wang C (2019) Tribological behavior of titanium alloy treated by nitriding and surface texturing composite technology. Materials 12(2):301. https://doi.org/10.3390/ma12020301
45. Kao WH, Su YL, Horng JH, Hsieh YT (2017) Improved tribological properties, electrochemical resistance and biocompatibility of AISI 316L stainless steel through duplex plasma nitriding and TiN coating treatment. J Biomater Appl 32(1):12–27. https://doi.org/10.1177/0885328217712109
46. Eremina GM, Smolin AY (2019) Numerical model of the mechanical behavior of coated materials in the friction pair of hip resurfacing endoprosthesis. In: Oñate E, Wriggers P, Zohdi T, Bischoff M, Owen DRJ (eds) VI international conference on particle-based methods. Fundamentals and applications. PARTICLES 2019, 28–30 Oct 2019, CIMNE, Barcelona, pp 197–203
47. Eremina GM, Smolin AYu (2019) Numerical modeling of the mechanical behavior of hip resurfacing endoprosthesis and healthy bone. AIP Conf Proc 2167:020087. https://doi.org/10.1063/1.5131954
48. Le Pense S, Chen Y (2017) Contribution of fluid in bone extravascular matrix to strain-rate dependent stiffening of bone tissue—a poroelastic study. J Mech Behav Biomed Mater 65:90–101. https://doi.org/10.1016/j.jmbbm.2016.08.016
49. Xia Z, Kwon YM, Mehmood S, Downing C, Jurkschat K, Murray DW (2011) Characterization of metal-wear nanoparticles in pseudotumor following metal-on-metal hip resurfacing. Nanomed Nanotechnol Biol Med 7(6):674–681. https://doi.org/10.1016/j.nano.2011.08.002
50. Cheung G, Zalzal P, Bhandari M, Spelt JK, Papini M (2004) Finite element analysis of a femoral retrograde intramedullary nail subject to gait loading. Med Eng Phys 26(2):93–108. https://doi.org/10.1016/j.medengphy.2003.10.006
51. Todo M (2018) Biomechanical analysis of hip joint arthroplasties using CT-image based finite element method. J Surg Res 1:34–41. https://doi.org/10.26502/jsr.1002005
52. Gerhardt LC, Boccaccini AR (2010) Bioactive glass and glass-ceramic scaffolds for bone tissue engineering. Materials 3(7):3867–3910. https://doi.org/10.3390/ma3073867
53. Biot MA (1957) The elastic coefficients of the theory of consolidation. J Appl Mech 24:594–601
54. Detournay E, Cheng AH-D (1993) Comprehensive rock engineering: principles, practice & projects, vol 2. Elsevier, Amsterdam, pp 113–171. https://doi.org/10.1016/B978-0-08-040615-2.50011-3
55. Basniev KS, Dmitriev NM, Chilingar GV, Gorfunkle M, Nejad AGM (2012) Mechanics of fluid flow. Wiley, Hoboken. https://doi.org/10.1002/9781118533628.ch3
56. Smita TH, Huygheb JM, Cowin SC (2002) Estimation of the poroelastic parameters of bone. J Biomech 35(6):829–836. https://doi.org/10.1016/S0021-9290(02)00021-0
57. Lim TH, Hong JH (2000) Poroelastic properties of bovine vertebral trabecular bone. J Orthop Res 18(4):671–677. https://doi.org/10.1002/jor.1100180421
58. Kohles SS, Roberts JB (2002) Linear poroelastic cancellous bone anisotropy: trabecular solid elastic and fluid transport properties. J Biomech Eng 124(5):521–526. https://doi.org/10.1115/1.1503374
59. Cardoso L, Schaffler MB (2015) Changes of elastic constants and anisotropy patterns in trabecular bone during disuse-induced bone loss assessed by poroelastic ultrasound. J Biomech Eng 137(1):011008. https://doi.org/10.1115/1.4029179
60. Sandino C, McErlain DD, Schipilow J, Boyd SK (2015) The poro-viscoelastic properties of trabecular bone: a micro computed tomography-based finite element study. J Mech Behav Biomed Mater 44:1–9. https://doi.org/10.1016/j.jmbbm.2014.12.018
61. Smolin AYu, Eremina GM, Dimaki AV, Shilko EV (2019) Simulation of mechanical behaviour of the proximal femur as a poroelastic solid using particles. J Phys Conf Ser 1391:012005. https://doi.org/10.1088/1742-6596/1391/1/012005
62. Shilko EV, Dimaki AV, Smolin AYu, Psakhie SG (2018) The determining influence of the competition between pore volume change and fluid filtration on the strength of permeable brittle solids. Procedia Struct Integrity 13:1508–1513. https://doi.org/10.1016/j.prostr.2018.12.309

63. Shilko EV, Dimaki AV, Psakhie SG (2018) Strength of shear bands in fluid-saturated rocks: a nonlinear effect of competition between dilation and fluid flow. Sci Rep 8:1428. https://doi.org/10.1038/s41598-018-19843-8
64. Seeman E, Delmas PD (2006) Bone quality-the material and structural basis of bone strength and fragility. New Engl J Med 354(21):2250–2261. https://doi.org/10.1056/NEJMra053077
65. Johnell O, Kanis JA (2006) An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporos Int 17(12):1726–1733. https://doi.org/10.1007/s00198-006-0172-4
66. Judex S, Boyd S, Qin YX, Miller L, Müller R, Rubin C (2003) Combining high-resolution micro-computed tomography with material composition to define the quality of bone tissue. Curr Osteoporos Rep 1(1):11–19. https://doi.org/10.1007/s11914-003-0003-x
67. Diez-Perez A, Güerri R, Nogues X, Cáceres E, Peña MJ, Mellibovsky L, Randall C, Bridges D, Weaver JC, Proctor A, Brimer D, Koester KJ, Ritchie RO, Hansma PK (2010) Microindentation for in vivo measurement of bone tissue mechanical properties in humans. J Bone Miner Res 25(8):1877–1885. https://doi.org/10.1002/jbmr.73
68. Tomanik M, Nikodem A, Filipiak J (2016) Microhardness of human cancellous bone tissue in progressive hip osteoarthritis. J Mech Behav Biomed Mater 64:86–93. https://doi.org/10.1016/j.jmbbm.2016.07.022
69. Link TM (2012) Osteoporosis imaging: state of the art and advanced imaging. Radiology 263(1):3–17. https://doi.org/10.1148/radiol.12110462
70. Eremina GM, Smolin AYu, Shilko EV (2019) Numerical modeling of the indentation of cancellous. AIP Conf Proc 2167(1):020090. https://doi.org/10.1063/1.5131957
71. Wahlquist JA, DelRio FW, Randolph MA, Aziz AH, Heveran CM, Bryant SJ, Neu CP, Ferguson VL (2017) Indentation mapping revealed poroelastic, but not viscoelastic, properties spanning native zonal articular cartilage. Acta Biomater 64:41–49. https://doi.org/10.1016/j.actbio.2017.10.003
72. Makuch AM, Skalski KR (2018) Human cancellous bone mechanical properties and penetrator geometry in nanoindentation tests. Acta Bioeng Biomech 20(3):153–164. https://doi.org/10.5277/ABB-01176-2018-02
73. Bembey AK, Oyen ML, Bushby AJ, Boyde A (2006) Viscoelastic properties of bone as a function of hydration state determined by nanoindentation. Phil Mag 86(33–35):5691–5703. https://doi.org/10.1080/14786430600660864
74. Wang B, Chen R, Chen F, Dong J, Wu Z, Wang H, Yang Z, Wang F, Wang J, Yang X, Feng Y, Huang Z, Lei W, Liu H (2018) Effects of moisture content and loading profile on changing properties of bone micro-biomechanical characteristics. Med Sci Monit 24:2252–2258. https://dx.doi.org/10.12659%2FMSM.906910
75. Marcián P, Florian Z, Horáčková L, Kaiser J, Borák L (2017) Microstructural finite-element analysis of influence of bone density and histomorphometric parameters on mechanical behavior of mandibular cancellous bone structure. Solid State Phenom 258:362–365. https://doi.org/10.4028/www.scientific.net/SSP.258.362
76. Fan Z, Rho JY (2003) Effects of viscoelasticity and time-dependent plasticity on nanoindenttion measurements of human cortical bone. J Biomed Mater Res Part A 67:208–214. https://doi.org/10.1002/jbm.a.10027
77. Chittibabu V, Rao KS, Rao PG (2016) Factors affecting the mechanical properties of compact bone and miniature specimen test techniques: a review. Adv Sci Technol Res J 10(32):169–183. https://doi.org/10.12913/22998624/65117
78. Stansfield BW, Nicol AC, Paul JP, Kelly IG, Graichen F, Bergmann G (2003) Comparison of calculated hip joint contact forces with those measured using instrumented implants. An evaluation of a three-dimensional mathematical model of the lower limb. J Biomech 36(7):929–936. https://doi.org/10.1016/s0021-9290(03)00072-1
79. Stansfield BW, Nicol AC (2002) Hip joint contact forces in normal subjects and subjects with total hip prostheses: walking and stair and ramp negotiation. Clin Biomech 17(2):130–139. https://doi.org/10.1016/s0268-0033(01)00119-x

80. Fabbri D, Orsini R, Moroni A (2018) Stress fracture of proximal femur after hip resurfacing treated with cannulated screw. Joints 6(2):128–130. https://dx.doi.org/10.1055%2Fs-0038-1660815

Abstract Methods on Mesoscopic Scales of Friction

Georg-Peter Ostermeyer and Andreas Krumm

Abstract In recent years, research has increasingly focused on the complex processes involved in friction contacts. Especially in tribological high-loaded contacts, characterized by the presence of contact modifying wear particles, macroscopic friction shows a surprisingly high dynamic complexity on many temporal and local scales. There are dominant effects on mesoscopic scales such as the geometric self-organization structures of the wear dust in the contact, which can significantly change the local contact surfaces. For the description and simulation of these phenomena, abstract methods have shown their effectiveness. One class of methods are cellular automata, both volume- and particle-based. The latter are in particular the Movable Cellular Automata developed by Sergey Psakhie. The scales of these discrete methods are freely selectable in wide ranges between the macro world and the atomic scale. Nevertheless, they provide reliable information on mesoscopic balances in the boundary layer and thus also on the macroscopic behavior of the tribocontact. The success of these methods is shown by the example of an automotive brake. The question of the relative insensitivity of the scales of these mesoscopic methods is examined in detail.

Keywords Dynamic friction · Dynamic equilibrium · Mesoscopic scales · Cellular automata · Particle methods · Dissipation of information

1 Introduction

Normally, the introduction to works on friction provides a quick overview or excerpt from 5000 years of history on friction. Starting with the fire generated by friction in the Stone Age to modern theories of friction. Although friction is omnipresent, it is still far from being understood well.

G.-P. Ostermeyer (✉) · A. Krumm
Institut für Dynamik und Schwingungen, Technische Universität Braunschweig, Schleinitzstraße 20, 38106 Braunschweig, Germany
e-mail: gp.ostermeyer@tu-braunschweig.de

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_6

This paper provides not a historical but an "engineering" view to friction as a "friction machine". Imagine you have to design a machine that generates the following:

- forces,
- heat,
- vibration,
- noise,
- particles and dust

and in addition includes:

- surface roughening,
- surface smoothing,
- welding processes,
- modifications of the crystalline surface structure and
- chemical oxidation processes.

It is very simple to realize this kind of machine: Two solid bodies in contact, which are shifted tangentially against each other, can show all the phenomena mentioned above (see Fig. 1).

The paradox between the geometric simplicity of the machine and the complexity of the physical phenomena observed is the reason for the difficulty in describing friction comprehensively. Only in very few cases is it possible to adjust some of the processes mentioned above a priori by designing this "friction machine" or "boundary layer machine" [1].

It is interesting that people have frequently tried to describe these machines with a single scalar quantity μ [2]. This quantity is defined as the quotient of frictional and normal force.

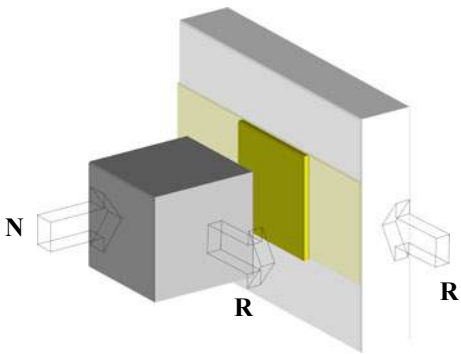


Fig. 1 Design of a friction machine

$$\mu = \frac{\left\| \vec{R} \right\|}{\left\| \vec{N} \right\|} \tag{1}$$

Measurements repeatedly confirmed that the coefficient of sliding friction depends on the speed, surface roughness, the materials involved, the ambient medium and many other parameters. It may even depend on quantities that can be explained by the measuring principle of the friction force rather than by the physics of friction. Recent investigations suggest that this coefficient is dynamic, i.e. it is only indistinctly represented by stationary measurements [3].

The following interpretation may explain why the friction coefficient is nevertheless useful. If the numerator and denominator in (1) are multiplied by the tangential velocity vector,

$$\mu = \frac{\vec{R} \cdot \vec{v}}{\left\| \vec{N} \right\| \cdot \|\vec{v}\|} \tag{2}$$

the friction coefficient can be interpreted as friction power related to a characteristic system performance [1]. Power is certainly a good measure to describe the effect of a machine. With this interpretation energetic connections to the above mentioned phenomena of a friction machine can be presented and explained. This applies in particular to heat regeneration or the averaged wear volume.

If you want to take a closer look at the physics of the machine, you have to move from the symbolic point contact (1) to the consideration of the area of the frictional contact. This is connected with a change of scale of the considered lengths and times. The scales are very different for the processes of the boundary layer machine mentioned above. They often lie between the macroscopic world of experience and the atomic scale. Consideration of mesoscopic scales have proven to be very useful for establishing connections both with macro and micro scales. To approach the meso scale, one can start from the atomic scale (“Bottom Up”) or the macroscopic scale (“Top Down”).

2 Bottom-Up View

On the atomic scale, with a characteristic length of 10^{-10} m, energy distribution processes can be traced by phonons and atomic movements. Today’s non-equilibrium molecular dynamics programs are able to describe atomic or molecular assemblies in a thermodynamically and quantum-mechanically correct way. In principle, this method yields approximately 10^{25} equations per cm^3 of matter. The time constants of molecular dynamics are on the order of femtoseconds. With today’s computers, therefore, no general statements about friction on the macroscopic scale are possible.

The transition from the atomic scale to larger length scales is combined with a drastic reduction in the number of descriptive equations of motion. Classical mechanics offers a perspective on this matter. There are different ways to transfer atomic information into the macroscopic world.

On the one hand, atomic or molecular assemblies can be made ever larger by reducing the number of particles but increasing their mass. This results in an ensemble of particles which are the core of the so-called Discrete Element Methods (DEM). If these atomic substitutes are further coarsened, only a few rigid bodies may remain. These are objects of the Multibody Systems Methods (MBS).

Another very successful way of mechanics to transform information from the atomic level to the macroscopic world is the smoothing out of the atomic structure to a massy continuum. Like the other two methods, this can be treated mathematically with very small effort. Numerical methods like finite element methods or boundary element methods (FEM, BEM) are being used.

In this way, mass-equivalent models can be created that are easy to handle mathematically. Interestingly, the force influences that are essential for the equations of motion are determined by measurement from the macroscopic world of experience alone and integrated into the models. Any phenomena for the material or motion behaviour can be approximated with arbitrary accuracy using a few generic force elements. This works quite perfectly in may situations, but fails completely in the case of friction forces.

One method for a better description of friction could be to transfer force information from the atomic scale to the macroscopic experiential using the mesoscopic particle method [4, 5].

In fact, as early as 1890, Boltzmann made the first proposals to connect the microscopic particle world with the macroscopic mechanical world [6]. Pioneering papers on this subject have been published by Greenspan [7, 8]. DEM provides excellent results for macroscopic motion and deformation dynamics, but it shows its weakness in thermodynamics. Although the thermodynamic processes can be calculated well with the MD simulation by statistical and stochastic methods, these methods cannot describe dynamic effects in the macroscopic world due to their time constants.

The approach of mesoscopic particles is to apply an atomic substitute world on a mesoscopic scale, which is below but close to the macroscopic scale (see Fig. 2). This mesoscopic world makes it possible to couple the particles to the mechanical bodies by “hidden” degrees of freedom. These additional degrees of freedom allow a correct thermodynamic description of mechanical systems even with particle worlds.

The idea is to describe the frictional system on the macroscopic time scale, but the friction boundary layer on the microscopic scale. The dynamics on the microscopic scale are hidden for the macroscopic scale, just as the atomic dynamics are hidden for the macroscopic world. It only serves to detect macroscopically observable phenomena, for example heat or wear. Thus, it is possible to describe the friction process and the associated phenomena (vibrations, heat, wear, etc.) with the mesoscopic particles with arbitrarily adjustable accuracy (see Fig. 3).

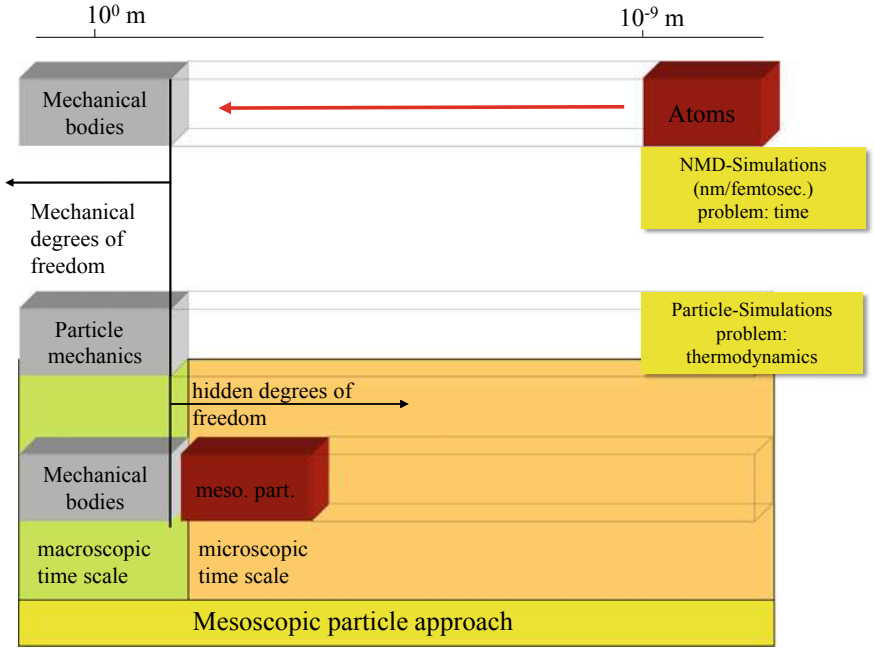


Fig. 2 Classification of the mesoscopic scale

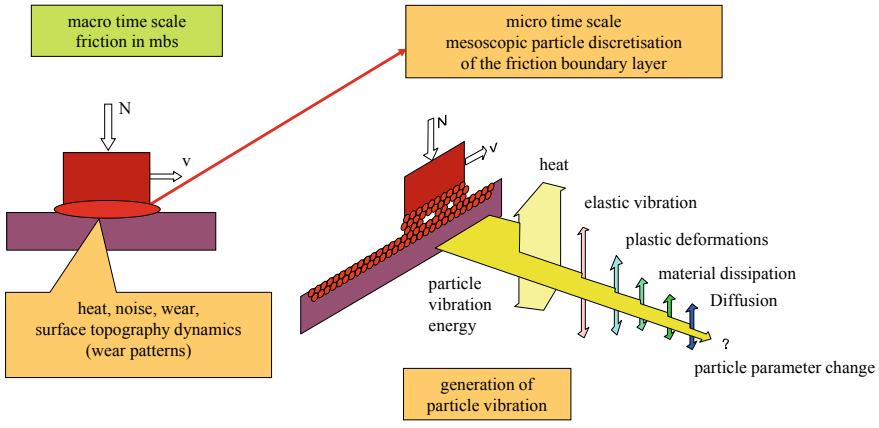


Fig. 3 Combination of macroscopic and microscopic scales in a friction system

To illustrate how this method works, a simple rod as in Fig. 4a is considered. If a hammer beats on the rod, the rod heats up. In this thought experiment the mechanical degrees of freedom of the rod are not considered, the elastic waves in the rod are not important. Therefore, the rod is only modelled with a sub-mechanical degree of

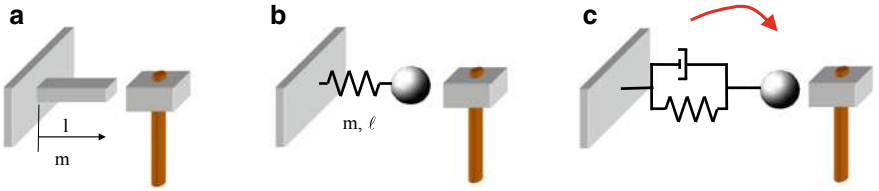


Fig. 4 **a** Heating a rod; **b** particle discretisation; **c** mesoscopic particle discretization [9]

freedom to detect heat in the rod. This sub-mechanical degree of freedom can be represented by a simple spring-mass system (see Fig. 4b).

Beaten with a hammer, the mass starts to vibrate. The system has absorbed energy from this impact. This energy can be interpreted as heat energy, the vibration itself is macroscopically irrelevant. If a second beat is given to this system, it usually leads to a violation of the thermodynamic laws, because with the second beat, the system depends essentially on the point in time at which the beat occurs, that is in which phase relation the beat and the already existing vibration stands. Thus, the vibration can be completely eliminated or the energy of the vibration can be doubled (see Fig. 5).

However, this means that the energy introduced by the second beat can no longer be interpreted as a heat gain. This energy should be added to the existing heat energy. This is exactly where the thermodynamic laws are violated.

If this mass point is, in addition to its mass property, given an internal dynamic variable, the temperature, then ultimately the sub-mechanical vibration dynamics induced by the beat can be transformed into heat. For this purpose, the vibration is dissipated by the damper, but the dissipated energy is stored in the inner variable T. In Fig. 6 this procedure is shown by means of the potential curve (blue) of the system. By the first beat a certain energy ΔE is entered into the system (red). Through the damper integrated in the system, the energy introduced as an vibration is now stored in the system as heat and the potential curve is raised to this energy level. Then, a second beat can be introduced and the sub-mechanical vibration of the system induced by this

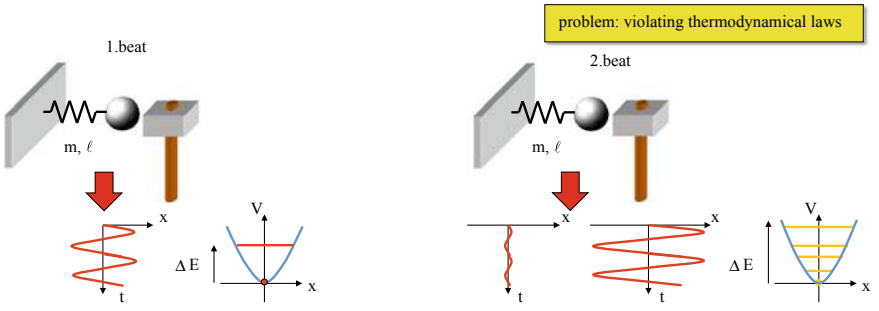


Fig. 5 Violation of the thermodynamic laws in conventional particle methods

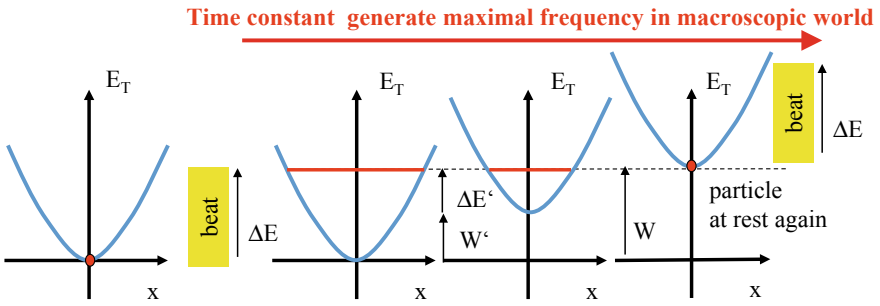


Fig. 6 Storage of thermal energy in mesoscopic particles

can be added to the existing heat energy. It should be noted that this method defines a minimum constant in terms of time, which defines the minimum distance between macroscopic impacts in order to correctly represent the thermodynamic behaviour of the system via the sub-mechanical vibrations. Such time grids can be found in all discrete systems, the grid width is a function of the typical frequencies in the grid. They limit the maximum possible frequency in the model. Every discretization method induces such grid constants. In numerical mathematics, this property is taken for granted as an essential element for the interpretation of results. However, these grid sizes can be made as small as desired by refining the sub-mechanical degrees of freedom.

The quantity of the time constant is determined by the time needed to adjust the potential curve to the correct energy level.

Mathematically this can be described by the following Lagrangian equations. The Lagrange function and the dissipation function of the system have the form

$$L = E_{kin} - V = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}cx^2 \text{ and } D = \frac{1}{2}b\dot{x}^2, \tag{3}$$

where m is the mass, c the stiffness of the spring, and b the damping coefficient.

The equations of motion are as follows:

$$\frac{d}{dt}\left(\frac{\partial}{\partial\dot{x}}L\right) - \frac{\partial}{\partial x}L = -\frac{\partial}{\partial\dot{x}}D \tag{4}$$

and the temperature change is calculated as

$$\dot{T} = \frac{1}{m \cdot c_v}\dot{x}\frac{\partial}{\partial\dot{x}}D, \tag{5}$$

where c_v is the specific heat capacity of the material (per unit mass) and T is the temperature.

The inner variable T gives an extended energy integral:

$$\frac{1}{2}m\dot{x}^2 - \frac{1}{2}cx^2 + mc_vT = const. \tag{6}$$

This approach can be transferred to all possible mechanical systems. In addition to temperature, other internal variables can also be used. For example, different energy fields with different time constants are superimposed with each other. The different levels provide their macroscopically observable phenomena, and the sub-mechanical texture of the model allows the thermodynamically correct transition from one form of energy to another.

Macroscopic forces either have a zero range, for example contact or impact forces, or are described by some interaction law with a specific coordinate dependence. Forces of finite range are characteristic of atomic forces. They can be described by Lennard Jones-like potentials. In many cases, the macroscopic forces can be approximated very accurately [7].

Such Lennard–Jones potentials can also be designed temperature-dependent in such a way that transitions of the mesoscopic particles from solid to fluid or gaseous phases can be described dynamically. Figure 7 shows a metal block that is melted on a heat plate.

At the beginning of the simulation the heating plate is switched off and the particles have a defined starting temperature. After the plate is switched on, it can be observed that the temperature in the block slowly increases. As expected, the temperature of the particles in contact with the heating plate increases first. Through heat transmission, the temperature of the particles in contact with the heating plate increases. After some time it can be observed that the melting temperature of the material is reached and particles are released from the block.

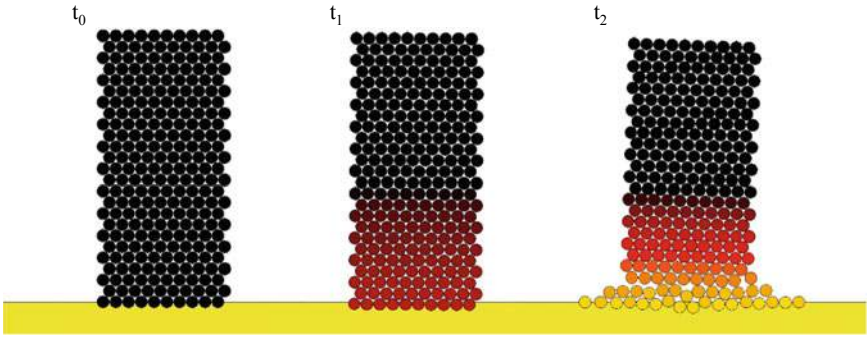


Fig. 7 A melting metal block

3 Top-Down View

Friction in the macroscopic world can also be investigated more intensively using a top-down method. This has been used intensively especially for tribological high-load contacts such as a vehicle brake. In Fig. 8, such a technical brake is shown.

Brakes show a highly dynamic friction coefficient. This is correlated with an extremely rich topographical and chemical dynamic in the boundary layer during the braking process.

Due to the complex material composition in brake pads (often more than 25 different materials), the friction process is supplied through a true a warehouse of materials in the boundary layer. Depending on the load and its duration, this process uses various chemical and physical processes which are technically designed to guarantee the quality and comfort of the brake pad. These processes are only understood to a small extend, and often appeared in a trial and error process to become the know-how of a company. In addition to chemical processes, the boundary layer shows a characteristic topography. The creation of this topography can be explained with Fig. 9.

The main chemical structure of the brake material is a relatively soft polymer matrix with embedded small and very hard particles (e.g. SiO_2 particles). All other components seem to only modify the process, which is described below and they will be ignored in this study [10]. A small section of the simplified coating with only one hard inhomogeneity is shown below.

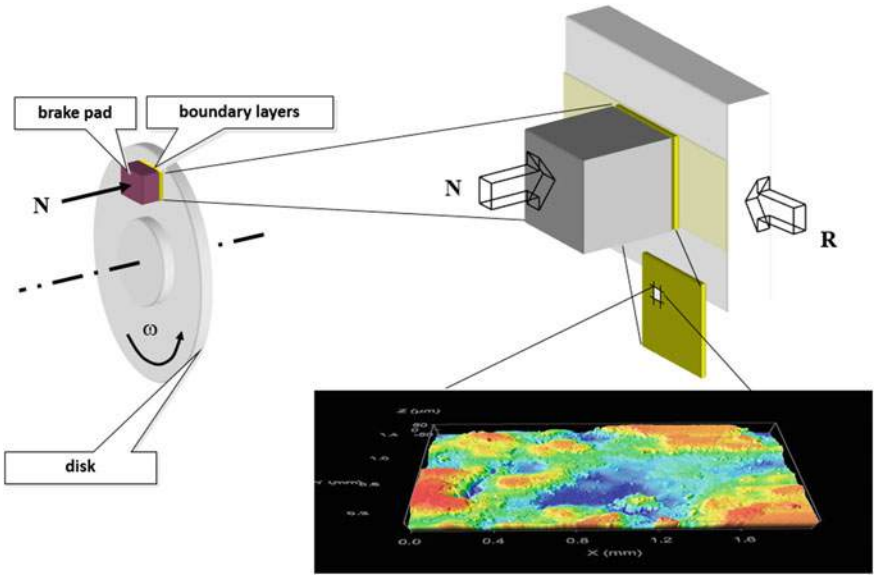


Fig. 8 Technical brake and presentation of the topography of the friction boundary layer

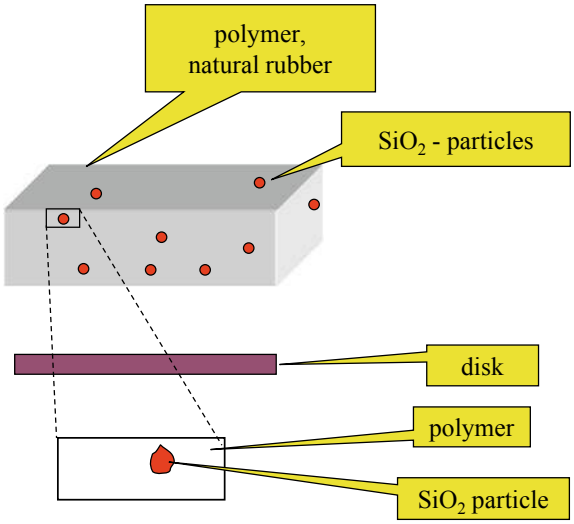


Fig. 9 Composition of a brake pad [3]

When the brake disc is pressed against the brake pad, abrasion occurs on the soft polymer matrix. The wear particles are transported along the contact zone. Some of the wear particles stay on the brake disc and after the disc has rotated, they return to the contact zone, while other particles are released into the environment [10].

This wear causes the hard particles to come to the surface of the pad. In the following, two modes may be observed. On the one hand, the flow of wear particles of the boundary layer is disturbed. On the other hand, the hard particles are pressed into the polymer matrix, as the soft matrix around the particle wears much more than the particle itself. As a result, an increase in the normal and tangential stress can be observed in the area of the hard particle. This results in an increase of the local temperature (see Fig. 10). The increase in stress and heat results in a process similar

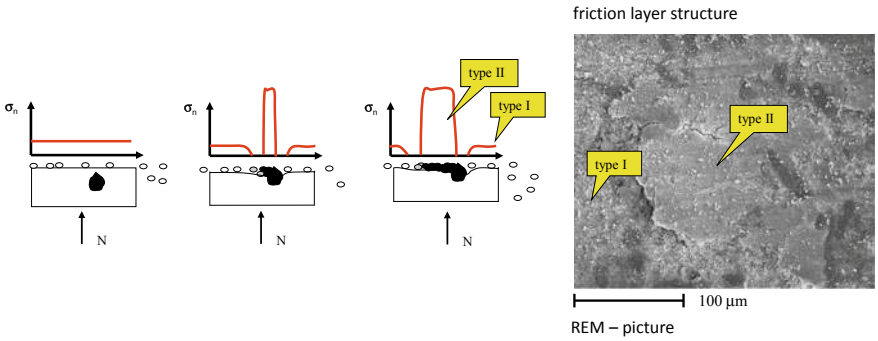


Fig. 10 Contact areas of the pad surface with normal stress distribution [10]

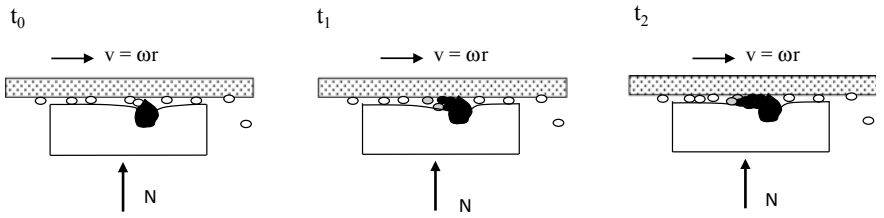


Fig. 11 Growth of contact patches [10]

to melting or sintering, whereby the wear particles together with the hard particle form an agglomeration modified by alloying processes and thin hard contact patches are formed (see Fig. 11) [10].

This process will divide the boundary layer surface into two contact areas. Type 1 is a contact area that consists of a rather soft polymer matrix that is relatively rough, heavily worn and has little frictional power. The second contact area is represented by the very hard and more wear-resistant patches (Type 2), which have a smooth surface and are mainly responsible for the main part of the friction power. They are ultimately responsible for the grey cast iron disc of the brake to wear and for iron entering the boundary layer chemistry.

These contact areas are dynamic. The patches grow due to friction power. However, their size is limited. If the total frictional power of a contact patch becomes too large, the patch cracks (see Fig. 12). The worn patch fragments represent an essential part of the wear particles of the braking process.

In the friction boundary layer, heat and wear are significant for the described dynamic process of topography change. Wear produces particles, and together with the heat the friction-intensive contact patches are born. Wear itself also ensures that the service life of these patches is limited. The total area of the patches is correlated with the current frictional power.

Friction is therefore a dynamic equilibrium. Not the friction coefficient itself, but its time derivative is defined by this process (see Fig. 13).

Dynamic friction laws can be derived from this idea [10]. Coulomb's friction law can be represented as the simplest dynamic equilibrium possible

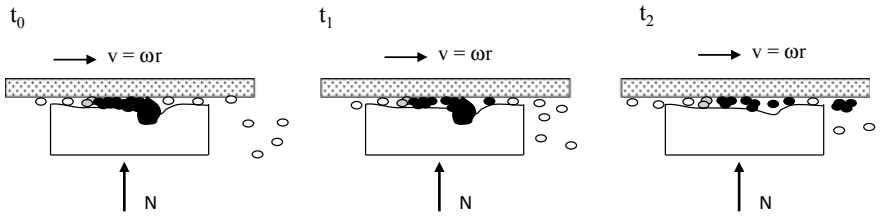


Fig. 12 Wear of contact patches [10]

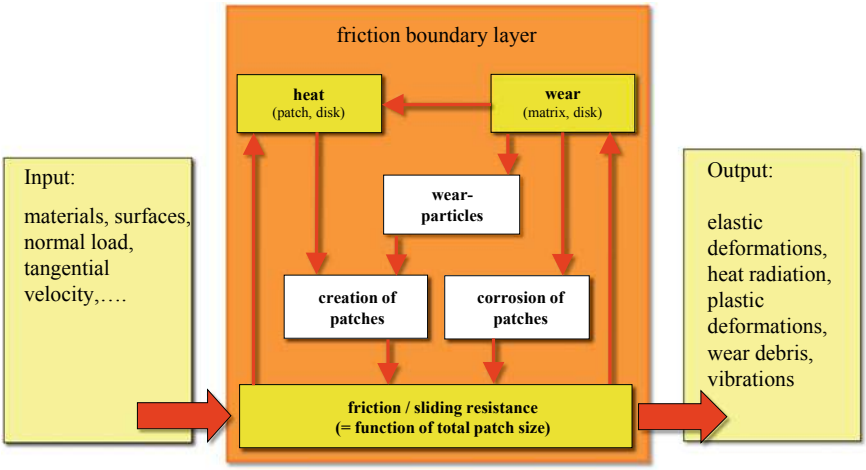


Fig. 13 Interactions in the boundary layer during the friction process [10]

$$\dot{\mu} = 0.$$

Volume-based Cellular Automata (CA) are a useful tool to simulate patch dynamics in the boundary layer (see Fig. 14). For the implementation, two areas are separated, as shown in Fig. 13. The relatively soft polymer matrix and the hard patches are discretized by cells with different normal elasticity.

In the first step, the normal stresses in the cells are calculated with a specified normal force. The contact forces in the patch areas are significantly higher. The

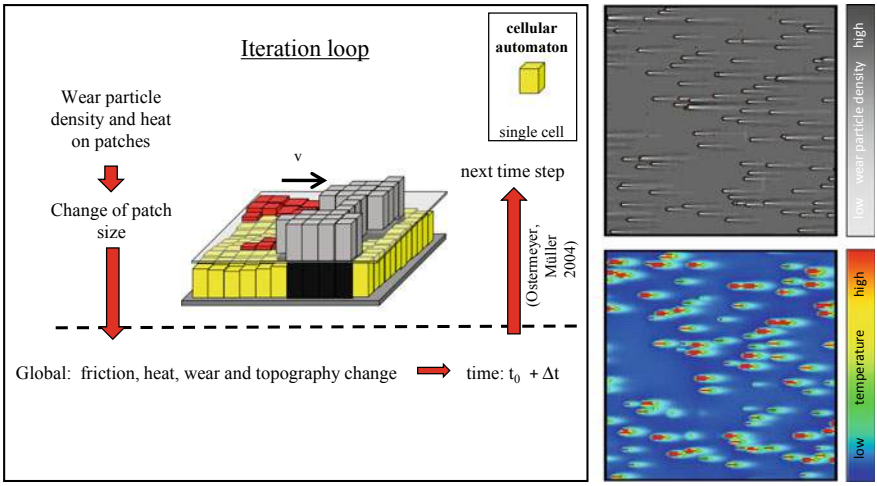


Fig. 14 Cellular automaton for the description of boundary layer dynamics

tangential speed of the brake disc above the cells is also given as well as different friction coefficients on the polymer cells and contact patch cells.

This allows a wear volume or a wear particle density to be calculated in the contact area of the polymer matrix (see Fig. 14, right). Calculating the frictional power on the patch cells the temperature can be obtained on the respective patches. This temperature uses the density of the wear particles from the matrix to determine the growth of the patches. The total dissipation over time, the temperature level, and the size control the time of patch destruction.

This simulation model creates an iterative process that maps the patch dynamics and allows a statement to be made about global friction, heat, wear and topography changes. This allows the CA to discretize the local geometry of the brake pad, to introduce local material data, to map the local chemical status of the boundary layer and to formulate local balance equations. The balance equations serve to formulate the thermodynamic, physical and chemical rules of the CA [11, 12].

The simulation model confirmed the dynamics of the surface structure and the heterogeneous heat distribution of the pad. However, the simulation also shows that this boundary layer dynamics is a very general effect of a stable and robust self-organisation process on highly loaded friction surfaces. These simulations confirmed the structure of the transient friction coefficient of Ostermeyer in [13, 10], that the temperature T and the wear as a function of the friction power according to (2) are considered in the flow balance.

$$\begin{aligned} \dot{\mu} &= -\alpha\big((\beta + |v_r \cdot N|) \cdot \mu - \gamma \cdot T_p\big) \\ \dot{T} &= -\delta\big(T_p - T_0 - \varepsilon \cdot |v_r \cdot N|\big) \end{aligned} \tag{7}$$

4 Natural Principles of Dissipation of Information

Mesoscopic scales between the macroscopic and atomic world seem to be particularly suitable for observing locally resolved dynamics in the friction boundary layer. The core of these considerations, however, is the reduction of information that would theoretically make an exact calculation of friction on the atomic scale possible.

The problem with friction is that its characteristic dissipation on the macroscopic level does not exist on the atomic scale. It is neither the dissipation of energy nor the dissipation of material that is the core of friction, but the dissipation of information that should explain the macroscopic manifestations.

In search for general mechanisms of information reduction, one quickly finds that the mechanisms that are being used in physics are a natural way.

One of them is heat. Heat is a scalar field that collects all high-frequency effects on a molecular basis in a single parameter. Another mechanism that can be interpreted as dissipation of information is wear. Wear constantly changes the topography on an atomic scale. Therefore, it does not make sense to describe the geometric details at

any time. Here, dynamic topography change rates as a scalar field of wear quantities make more sense.

Heat and wear are very good candidates to dissipate dynamic and geometric micro information, see Fig. 13, which is the basis for the dynamic friction models (7).

Certainly there are other elementary dissipation mechanisms of information which can be applied here in a useful way. Another mechanism is probably the uncertainty with which our experiences and measurements in the macroscopic world are recorded. A mathematically promising approach is offered by the Polymorphic Uncertainty Analysis.

Frictional systems can perform self-excited vibrations. During braking, they occur as an unpleasant noise, the so-called squealing. These self-excited states can be analyzed considering an eigenvalue problem. The uncertainty in the friction prediction induces an uncertainty in the stability limits of the model. Minimizing these uncertainties is necessary to enable more reliable brake designs. In Fig. 15 an uncertainty calculation has been carried out in the complex eigenvalue analysis of a simple brake model [14].

Here, Coulomb's friction law is compared to the dynamic friction law with respect to the confidence interval for stability limits in such an eigenvalue analysis. With the confidence interval, it is possible to quantify uncertainties that are present in the friction laws. If the parameters of Coulomb's friction law and the parameters of Ostermeyer's friction law are determined from concrete measured values, and the uncertainty of the friction coefficient is determined from these values, the model of the brake and the technique of complex eigenvalue analysis result in some uncertainty of the stability limits.

It is shown that the uncertainty interval for Coulomb's description is 15 times greater than for the dynamic friction law. This large difference is a consequence of the dynamic a-priori knowledge, which is represented in the flow balance conditions.

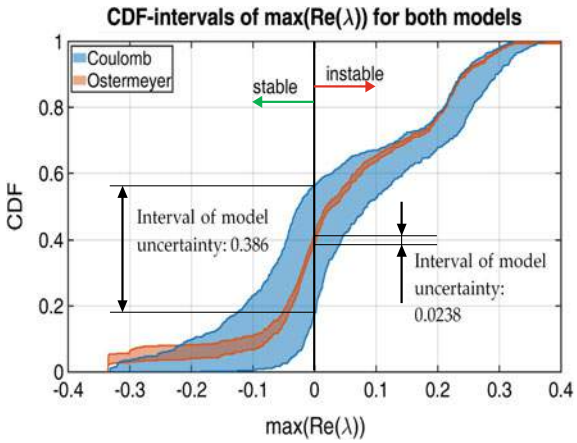


Fig. 15 Comparison of the Coulomb friction model and the Ostermeyer friction model with respect to polymorphic uncertainty and stability analysis [14]

5 Conclusion and Discussion

Friction is a highly complex and dynamic phenomenon that is still not fully understood. One of the challenges in describing friction are the manifold interactions of dynamic effects on different time and length scales. Mesoscopic methods are an excellent tool to cope with the complexity of friction.

In this paper, two different approaches are discussed, which can describe the friction process in technical brakes with higher precision. One is specific particle methods, the other is volume-based cellular automata. Both methods use the mobility on the temporal and local scales between the atomic and macroscopic scale.

More general is the method of Movable Cellular Automata, which combines the advantages of the particle world with the advantages of the world of Cellular Automata. The creator of this method, Sergey Psakhie was a visionary. His methods have found the way not only into friction, but into many other areas [15].

References

1. Ostermeyer G-P (2003) Thesen zur Modellierung von Reibung und Verschleiß. Tribol Schmierungstech 4:18–22
2. Coulomb CA (1785) Die Theorie einfacher Schwingungen. Memoires de mathematique et de physique de l'Academie des Sciences 10:161–331
3. Ostermeyer G-P (2010) Dynamic friction laws and their impact on friction induced vibrations. SAE technical paper 2010-01-1717, pp 1–27
4. Ostermeyer G-P (1999) A mesoscopic particle method for description of thermomechanical and friction processes. Phys Mesomech 6:25–32
5. Ostermeyer G-P (2007) The mesoscopic particle approach. Tribol Int 40(6):953–959
6. Boltzmann L (1897) Vorlesungen über die Prinzipe der Mechanik. Verlag von Johann Ambrosius Barth, Leipzig
7. Greenspan D (1972) A discrete numerical approach to fluid dynamics. Inf Process 71:1297–1304
8. Greenspan D (1988) Particle modeling in science and technology. In: Numerical methods, Proceedings 4th conference, Miskolc/Hung, 1986, Colloquia Mathematica Societatis Janos Bolyai 50, pp 51–66
9. Ostermeyer G-P (1996) Many particle systems. Dynamical problems in mechanical systems. In: Proceedings of the 4th Polish-German workshop, July 30–Aug 05, 1995, Polska Akademia Nauk, IPPT Warszawa, Warsaw (PL), pp 249–259
10. Ostermeyer G-P (2003) On the dynamics of the friction coefficient. Wear 254(9):852–858
11. Ostermeyer G-P, Müller M (2005) Dynamic Interaction of friction and surface topography in brake systems. Tribol Int 39(5):370–380
12. Ostermeyer G-P, Müller M (2008) New insights into the tribology of brake systems. Proc Inst Mech Eng Part D J Autom Eng 222(7):1167–1200
13. Ostermeyer G-P (2001) Friction and wear of brake systems. Forsch Ingenieurwes 66:267–272
14. Ostermeyer G-P, Müller M, Brumme S, Srisupattarawanit T (2019) Stability analysis with an NVH minimal model for brakes under consideration of polymorphic uncertainty of friction. Vibration 2:135–156
15. Psakhie SG, Horie Y, Ostermeyer G-P, Smolin AYu, Korostelev SYu, Shilko EV, Dmitriev AI, Blatnik S, Špegel M, Zavšek S (2001) Movable cellular automata method for simulating materials with mesostructured. Theoret Appl Fract Mech 37(1–3):311–334

Study of Dynamics of Block-Media in the Framework of Minimalistic Numerical Models

Alexander E. Filippov and Valentin L. Popov

Abstract One of the principal methods of preventing large earthquakes is stimulation of a large series of small events. The result is a transfer of the rapid tectonic dynamics in a creep mode. In this chapter, we discuss possibilities for such a transfer in the framework of simplified models of a subduction zone. The proposed model describes well the basic characteristic features of geo-medium behavior, in particular, statistics of earthquakes (Gutenberg Richter and Omori laws). Its analysis shows that local relatively low-energy impacts can switch block dynamics from stick–slip to creep mode. Thus, it is possible to change the statistics of seismic energy release by means of a series of local, periodic, and relatively low energy impacts. This means a principal possibility of "suppressing" strong earthquakes. Additionally, a modified version of the Burridge-Knopoff model including a simple model for state dependent friction force is derived and studied. The friction model describes a velocity weakening of friction between moving blocks and an increase of static friction during stick periods. It provides a simplified but qualitatively correct stability diagram for the transition from smooth sliding to a stick–slip behavior as observed in various tribological systems. Attractor properties of the model dynamic equations were studied under a broad range of parameters for one- and two-dimensional systems.

Keywords Earthquakes · Block-media motion · Numerical simulation · Burridge-knopoff model · Stick–slip · Phase diagram · Seismic shocks · Phase transition · Time dependent friction

A. E. Filippov (✉)
Donetsk Institute for Physics and Engineering, NASU, Donetsk 83114, Ukraine
e-mail: filippov_ae@yahoo.com

V. L. Popov
Technische Universität Berlin, 10623 Berlin, Germany
e-mail: v.popov@tu-berlin.de

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_7

1 Introduction

An important and interesting application of studies of spatial–temporal pattern formation of mechanical systems is the formation of geological faults and their dynamics. Of particular interest is the study of the possibility of changing the mode of the fault dynamics into a slow one, thus preventing strong seismic shocks. The fact that the statistics of earthquake magnitude and their time correlations meet the laws of Gutenberg Richter [1, 2] and Omori [1, 3], typical for self-organized critical systems [4, 5], is often used for the conclusion that it principally occurs on all spatial scales ranging from microscopic to continental plate scale. Therefore, it is impossible to exert a targeted influence on the dynamics of earthquakes by local effects of limited energy. However, the works [6–10], based both on modeling by movable cellular automaton and full-scale experiments (on one of the active faults of the Baikal rift zone) suggest the principal possibility of releasing the accumulated elastic energy due to controlled low energy actions (vibration load and watering).

Based on this idea, we develop and study a model of the behavior of contact zones of block media and analyze the possibility of controlling the mode of displacement as was found experimentally.

2 Mechanical Model

Minimalistic mechanical model demonstrating the time correlations typical for systems showing self-organized criticality has been suggested by the authors. In the conceptual form, it is shown in Fig. 1. The plate is moved by the external force of F_{ext}. The plate is inclined by an angle that determines the ratio between the vertical component of the force $F_{ext}\cos(\alpha)$ acting against the force of Archimedes, which supports the plate "in magma", and the horizontal component, which results in the

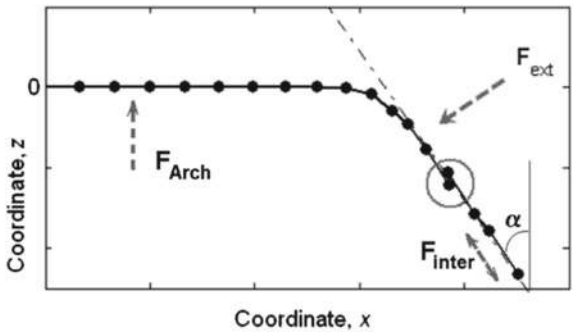


Fig. 1 Schematic diagram of the mechanical model. The force exerted by the mainland, F_{ext}, the force of Archimedes F_{Arch} and the force of elastic interaction F_{inter} are indicated. The circle shows the event of "bark fracture"

displacement of the entire system along the x-axis and deformation of the elastic plate.

In a numerical simulation, the plate is transformed into a set of discrete elements connected by a (nonlinear) elastic force that tries to maintain a fixed distance between them. Let us first consider the simplest, two-dimensional version of the problem. In this case, the plate is transformed into an elastic chain, and the model equations are reduced to the following form:

$$\begin{aligned} \partial x/\partial t &= F_{\text{inter},\,x} + F_x + \xi_x; \\ \partial z/\partial t &= F_{\text{inter},\,z} + F_{\text{Arh},\,z} + F_z + \varepsilon_z; \\ \partial X/\partial t &= \sum_j F_x(j) + F_{\text{ext},\,x} + \varepsilon. \end{aligned} \tag{1}$$

Here, $F_x = F_{\text{ext}} \sin(\alpha)$ and $F_z = F_{\text{ext}} \cos(\alpha)$ are the projections of the force of the chain's interaction with the "mainland plate" (we neglect the vertical movement of the heavy mainland plate here), and the summation is made over all elements of the chain. The external force is assumed to be constant (acting on the drifting continent from the magma side). The levitation force of Archimedes $F_{\text{Arch},\,z}$ is given by a condition:

$$F_{\text{Arch},\,z} = const \cdot U_0 > 0 \text{ at } z < 0 \text{ and } F_{\text{Arch},\,z} = 0 \text{ at } z > 0 \tag{2}$$

and $F_{\text{inter},\,x}$ and $F_{\text{inter},\,z}$, components of nonlinear elasticity between the segments of the chain, are equal:

$$F_{\text{inter},\,x} = -\partial U_{\text{inter}}/\partial x \text{ and } F_{\text{inter},\,z} = -\partial U_{\text{inter}}/\partial z, \tag{3}$$

where the distance-dependent effective interaction potential U_{inter} looks like:

$$U_{\text{inter}}(r) = Kr^2(1 - r^2/r_0^2). \tag{4}$$

Here, K is the elastic constant. To fix the distance between the elements, in the simplest, most widespread approach [11–14], the potential of the fourth order (4) is used, for which the components of the forces between the elements contain cubic nonlinearity, providing the required rigidity. At the same time, in 2D (or 3D) space, the chain (surface) can bend under the influence of the force F_{ext} and its elements move in the vertical (and/or horizontal) direction.

In numerical modeling, this leads to the following fracture condition. If the x- or z-projection of the vector connecting two consecutive segments of the chain is negative and its absolute value exceeds some threshold value, then the "fracture" occurs in this place. The fragment of the chain from its beginning to this point is subsequently removed. In the absence of resistance on the part of the removed fragment, the speed of the continental plate sharply increases, up to its deceleration by further segments of

the chain. Random influences from the surrounding subsystems are included through the δ-correlated noise source:

$$\left\langle \xi_{xz}(t,x,z)\xi_{xz}(t',x',z')\right\rangle = D_1\delta_{xz}\delta(t-t')\delta(x-x')\delta(z-z');$$
$$\left\langle \xi(t,x,z)\xi(t',x',z')\right\rangle = D_2\delta(t-t')\delta(x-x')\delta(z-z'). \quad (5)$$

Here $\delta(\ldots)$ is the impulse function of Dirac, δ_{xz} is the symbol of Kronnecker, and in each case some effective temperature can be assigned to the "diffusion coefficient" $D_{1,2} = 2k_B T_{1,2}$. The dissipative constant can be selected arbitrarily. It sets a characteristic time scale, and should be fitted a posteriori by experimental data. Random influences on the heavy mainland plate can be neglected, which leads to the assumption $D_1 \gg D_2$.

3 Statistical Properties of the Model

At a constant external force F_{ext}, the movement occurs with a constant—on the large time scales—average velocity $\langle V\rangle = dX/dt$. If we subtract Vt from the $X(t)$ curve, the fine structure of the derivative dX/dt becomes clearly visible when the chain breaks (see Fig. 2). The right side of Fig. 2 shows the distribution of the lengths of "jumps" and intervals between them.

Let us compare these distributions at different noise intensities. With exception of the expected reduction in the average interval between the jumps, the results do weakly dependend on the noise intensity up to the values comparable to the fracture threshold. Moderate noise (which does not exceed the dynamic chaos intensity of this

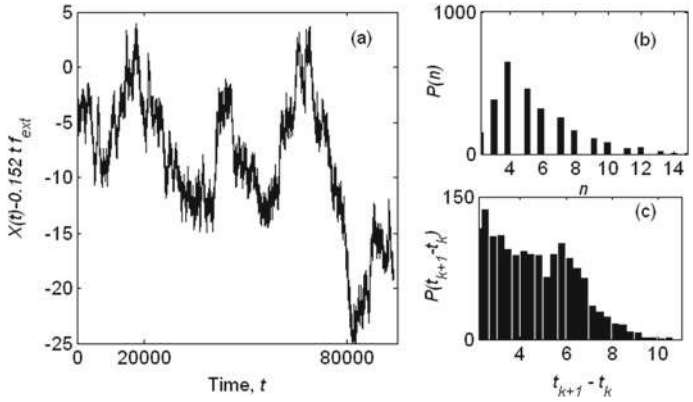


Fig. 2 The fine structure of the jumps **a** obtained by subtracting Vt and typical distributions of lengths of the chain fragments removed after fracture events (**b**) intervals between the jumps (**c**) obtained for parameter values $\alpha = \pi/6$, $F_{\text{ext}} = 10$, $U_0 = 0.1$, $K = 0.1 + 0.9\xi$

nonlinear system) only increases standard deviations. Therefore, we present below the results obtained at negligibly low noise.

Of interest is an analysis of the influence of regular (periodic) spatially localized actions on the system. We simulated the influence of sinusoidal and impulse actions of different intensity. Note that exactly the latter type of action is used in field experiments.

In the presence of noise, such actions reduce the average time between jumps. However, being spatially localized, they fix quite precisely both the time and length of each jump. This can be achieved by selecting resonant frequencies, amplitudes and force application points. Weak impact plays here the role of the trigger mechanism, provoking its own, more powerful processes in the system.

Formally, the model is able to achieve an accurate resonance optimum such that the random components are practically suppressed. In field conditions, the parameters are not as controlled as in the numerical experiment, and the histogram of jump distribution acquires additional lines near the resonant one.

4 Three-Dimensional System and Reduced Frontal Motion Model

Real tectonic systems are three-dimensional. Therefore, even for a minimalistic model, in addition to coordinates (x, z) also the y-coordinate along the edge of the fault has to be considered. In a self-consistent approximation, we can assume that the $3D$ system is composed of many equivalent $2D$ systems, which interact only through a common front of contact with the "continent". All system Eq. (1) acquire an additional index [layer number (x, z)] along the y-axis

$$\begin{aligned} \partial x(k)/\partial t &= F_{\text{inter},\,x}(k) + F_x(k) + \xi_x(k), \\ \partial z(k)/\partial t &= F_{\text{inter},\,z}(k) + F_{\text{Arh},\,z}(k) + F_z(k) + \xi_z(k). \end{aligned} \tag{6}$$

The last equation of the system (1) is modified into equation:

$$\partial X(k)/\partial t = \sum_j F_x(k)(j) + F_{\text{ext},\,x}(k) + K[X(k+1) + X(k-1) - 2X(k)]. \tag{7}$$

At small deviations between the neighboring layers, they behave quasi-independently according to the $2D$ model described above. If the deviations between $X(k+1) + X(k-1)$ and $2X(k)$ increase, they are suppressed by the elastic bond $K[X(k+1) + X(k-1) - 2X(k)]$ determined by the constant K.

In the following, we will shift each following action along the front by some value so that at the present moment the small jump will be (with certain probability) provoked in another place of the front. Figure 3 illustrates the resulting propagation

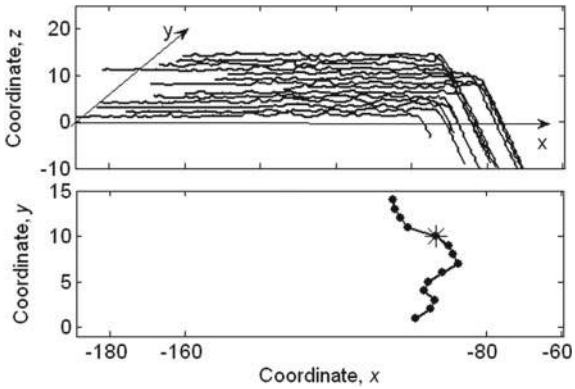


Fig. 3 The movement of the frontal area of $X(k)$ caused by periodic impacts along it

of the front on example fraction of the front. Moving the impact point along the front (asterisk in Fig. 3) generates multiple front jumps (instead of one), thus facilitating small jumps. The corresponding histograms contain some contribution of jumps, large in amplitude and time intervals between them, which, however, is much smaller than for an unperturbed system.

Taking into account the practical importance of such a problem, as well as the general scientific interest, it is useful to construct a simplified minimalist model of frontal motion, in which the connection between the layers would be taken into account in the rules of advancement of its fragments. For this purpose, let us consider a $2D$ front line in the plane (x, y), each segment of which moves forward under the action of a constant external force.

Such a minimalistic model successfully reproduces the basic properties of a more general $3D$-model, and is compact enough for large parameter studies. It is stable against varying model parameters in a very wide range. First, we checked that the change of the elastic constant by three orders of magnitude did not led to any substantial change in the distribution function $P\ [y(k)]$.

The method of inducing local surges described above works with the reduced model in the same manner as described above in a 3D model. In other words, it is possible to select such periodicity and distance between the local impacts that they provoke a wave of small jumps, which leads to an almost regular movement of the entire front line. The instantaneous state of this process is shown in Fig. 4. For the sake of clarity, the planar front has been chosen here as the initial condition. The area near the artificial influence is marked with a grey circle. The instantaneous position of the impact zone is shown by a dark solid circle; the spontaneous jumps caused by it are marked with bold dots inside the circle.

The numerical model above, in its minimalistic variant, was published by the authors [12] in 2006 and was used to study the dynamics of subduction zone dynamics. Later on, in 2008, the detected correlations were used by us [13] to select the optimal scenarios of other "weak and cheap" local energy effects. As a result, the

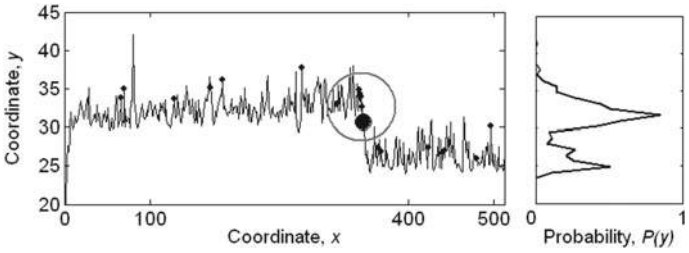


Fig. 4 The initial stage of the formation of an equilibrium distribution of probability $P[y(k)]$ in the minimalist model at the 2D front

efficiency of such impacts was significantly increased making it possible to switch the movement of blocks between "stick–slip" and creep modes. The average energy of single seismic shocks was significantly reduced and it became possible to "suppress" strong earthquakes. The model adequately described the laws of deformation for a block system and the temporal correlations typical of systems with self-organized criticality. The proposed model differs in principle from those studied previously (see, e.g., [2–11]) by taking into account the real topology of the creep of a continental platform on a thinner oceanic platform.

5 Correlation Functions

The simplicity and numerical efficiency of the model described in the previous Sections can be used to accumulate the statistics of correlations of the motion of separate blocks at the front. These correlations, in turn, can be used for selecting an optimum scheme of external action leading to the transformation of the system dynamics into a creep regime. Consider a $2D$ front in the (x, y) plane, each segment of which is moved forward by an over damped external force. Let us assume that every subsequent "tectonic" jump of each segment takes place upon its displacement by a certain distance, which is generated by a random number generator $\left\langle \xi(t,x,z)\xi(t',x',z')\right\rangle = D_2\delta(t-t')\delta(x-x')\delta(z-z')$ with zero mean value $\langle \xi(t,x,z)\rangle = 0$.

The jump magnitude ψ is also assumed to be a random quantity with zero mean value $\langle \psi(k,t)\rangle = 0$ and is set by the average weak noise intensity $D_\psi \ll D_\xi$ as follows:

$$\left\langle \psi(k,t)\psi(k',t')\right\rangle = D_\psi \delta_{kk'}\delta(t-t'). \tag{8}$$

The particular noise intensity D_ψ has to be reconstructed a posteriori from experimental data. Following article [12], we assumed that the coordinate $X(k)$ depends on the number of layers k and that different layers are coupled by the elastic force $F_{elastic} = K[X(k+1) + X(k-1) - 2X(k)]$ already used in Eq. (7). Other details

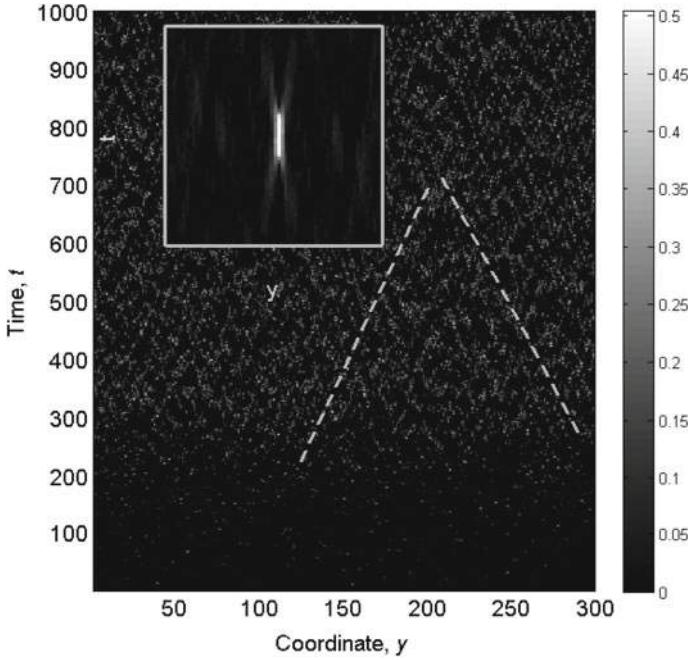


Fig. 5 Spatial–temporal map of jump distribution represented by the $M(t, x)$ matrix, with the intensities indicated by grayscale on the right. The initial stage corresponds to the establishment of a stationary process. Dashed lines correspond to the characteristic velocities of wave propagation (pre- and aftershocks). The inset shows the corresponding $G(t, x)$ correlation function

of the model are the same as described in [12] and above in this Chapter. The initial condition for the further study was selected in the form of a planar front. The resulting spatio-temporal map of jump length distribution along the front is depicted in Fig. 5.

At every step of the numerical procedure, system (1)–(3) is solved and a set of tectonic displacements $\delta X(k)$ (including zero shifts) distributed along the y-coordinate is obtained, which represents a row of the $M(t, y)$ matrix at the given time. This procedure is repeated and the entire geological history of the system is recorded in the form of the $M(t, y)$ matrix, which is represented in Fig. 5 by a gray scale map. This spatial–temporal map exhibits a clear initial transition period. A particular scenario is determined by the initial configuration (here, a planar front). As can be seen from Fig. 5, the stationary regime reveals a well-pronounced correlated character. The neighboring regions at the front interact by means of the elastic force $F_{elastic} = K[X(k+1) + X(k-1) - 2X(k)]$, so that large jumps of one segment induce several jumps in the neighboring segment that propagate as decaying waves in both directions from a strong local "earthquake." Arriving at a certain "weak" segment, i.e., a segment potentially close to a spontaneous break, such waves can initiate this break, inducing a new "tectonic shear" with accompanying waves and

so on. In other words, each significant event in the system is surrounded by a set of pre- and aftershocks that lead to correlations in the $M(t, y)$ matrix.

The dashed lines in Fig. 5 correspond to the characteristic velocities of wave propagation. Although the velocity of these waves is not universal and varies depending on the constants in Eqs. (1)–(3) their existence is a consequence of the structure of the system under consideration. Physically, the velocity of correlation spreading depends on the composition and strength of rocks, the level of friction forces between blocks, etc.; therefore, each geo-logical region has a certain characteristic velocity, which can be determined from experimental statistics of local secondary shocks accompanying earthquakes. The correlation can be quantitatively described by a spatial–temporal correlation function $G(t - t', y - y') = \langle M(t, y)M(t', y')\rangle$ similar to that depicted in the inset to Fig. 5.

This correlation function was calculated for a particular realization of the $M(t, y)$ matrix over the $t - t_0$ time interval (beginning with the time t_0 found from the termination of the transient process). The gray scale reflects the absolute values of correlations between jumps at the front. The slopped ridges of density $G(t-t', y-y')$ on both sides of the central maximum correspond to the averaged (typical) velocities of propagation of the interacting events in the given system.

The aim of our investigations is the practical usage of a theoretically justified effect of weak controlled spatially localized impacts on a given system. Previously, we modeled the effect of periodic pulses of variable intensity and preset on/off ratios. If the action is localized in a single (x, z)-layer, this layer gradually proceeds forward and pulls the neighboring layers behind to form a protrusion on the $X(k)$ front. Then, increasing deviations $X(k+1)+X(k-1)-2X(k)$ are sup-pressed by the elastic coupling with neighboring layers. Nevertheless, we succeeded in suggesting a strategy [12] that retained the applicability of the proposed method in a distributed system. For this purpose, each subsequent point of action was shifted over several (x, z) layers along the front so that a small jump would be initiated at a different site of the front, stimulating new neighboring regions. This shift was selected in both the *3D*-model and its reduced variant.

Figure 6a illustrates such an artificially stimulated process in the same system (and same notation) as in Fig. 5. Here, the clearly distinguished straight lines correspond to periodic impacts regularly shifted along the front, which virtually completely suppress the spontaneous jumps in the systems. Unfortunately, this scenario requires large-scale preliminary works irrespective of whether the probable earthquakes will actually take place. At the same time, the correlations of spontaneous events suggest a constructive idea; it is possible to apply the artificial impacts at the sites of statistically anticipated aftershocks rather than over the entire front, thus only producing a controlled initiation of small jumps at the sites where these jumps are stimulated by intrinsic correlations.

Figure 6b shows the distribution of events caused by such a self-consistent action. This pattern appears as more densely filled with jumps as compared to that in Fig. 6a. However, the scales of jump lengths in Fig. 6a, b are also substantially different. The main consequence of this procedure is a sharp drop in the fraction of spontaneous events taking place when the system reaches the level of critical stresses.

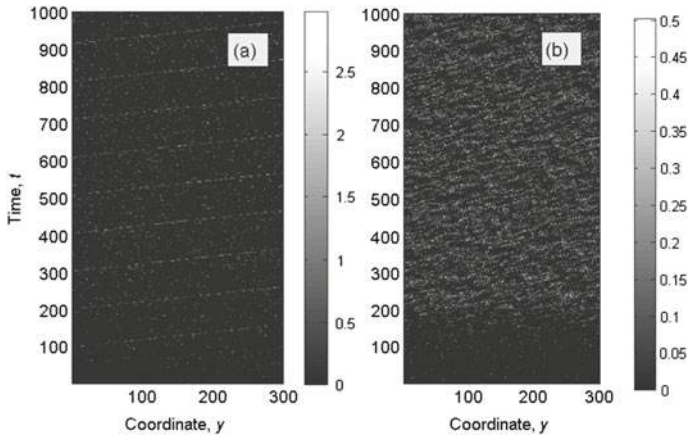


Fig. 6 Spatial–temporal map as in Fig. 5, but in the presence of artificial impacts initiating local jumps of the front segments in the case of **a** regular shift of the impact site along the entire front and **b** adaptive reaction to events selected using the $G(t, x)$ correlation function

Figure 7 shows the temporal variation of the number of such events that were not prevented by the economic adaptive scenario mentioned previously. As can be seen, the relative fraction of critical events is formally large (even reaching unity) only in the initial transient stage, where the events are not yet correlated. However, this stage is an evident artifact of the numerical procedure with a planar initial front (since the initial configuration was a priori not known). In a stationary stage, where the system attains a self-consistent regime, the fraction of spontaneous jumps not prevented by the economic adaptive action falls within 0.1–0.2. In other words, the

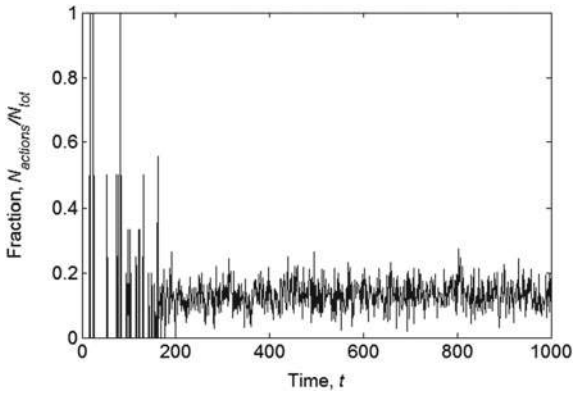


Fig. 7 Temporal variation of the fraction of spontaneous jumps not prevented by the proposed adaptive scenario (mapped in Fig. 6b). In the initial transient stage, the events are not yet correlated and the fraction of spontaneous jumps can be large (reaching unity), while in a stationary stage this fraction falls within 0.1–0.2

economic scenario allows 80–90% of spontaneous earthquakes in the system to be prevented.

6 Burridge-Knopoff (BK) Model

In addition to the above model, we also studied in [14] the well known Burridge-Knopoff (BK) model [15] initially proposed to investigate statistical properties of earthquakes. Numerical studies by Carlson et al. [16, 17] have demonstrated that the BK model can reproduce characteristic empirical features of tectonic processes such as the Gutenberg-Richter law for the magnitude distribution of earthquakes, or the Omori law for statistics of aftershocks [16–19], both properties stemming from the so called "self-organized criticality" of this system. It has been intensively used to simulate different aspects of the problem [18–38] and to discuss general properties of earthquakes statistics as well as predictability of earthquakes.

Numerical simulations give evidences that the self-organized criticality and the corresponding fractal attractor of the system is closely related to dynamic structures with "traveling waves" [21], their ordering and specific "phase transitions" [22] controlled by a number of parameters (external driving velocity, springs stiffness, number of blocks, their mutual interaction and so on). It is in particular the dependence of the dynamic properties of the BK model on the spring stiffness, which makes it necessary to introduce the generalization of the friction law proposed in this paper.

The physical reason for the stick–slip instability in the Burridge-Knopoff model is the assumed decrease of the friction force with the sliding velocity [16, 17]. Motion of a single block with this friction law is always unstable which does not correspond to properties of real tribological systems. The real law of rock friction is more complicated [26–31]. In this article, we proceed from more realistic friction laws described in [31, 32]. The main qualitative picture of a realistic law is: (a) approximately logarithmic increase of the static friction force as a function of contact time—the property, found already by Coulomb [39], and (b) a logarithmic dependence of the sliding friction on the sliding velocity. Both properties can be described in the framework of "state dependent" friction laws by introducing additional internal variables describing the state of the contact. Up to now, there were no attempts to study the dynamic and statistic properties of the Burridge-Knopoff model with a state dependent friction law.

In the paper [14], we proposed a modified version of the BK model with a state dependent friction force, reproducing in the simplest way both the velocity weakening friction and the increase of static friction with time when the block is not moving. To validate the model, we studied the stability diagram. It qualitatively reproduces typical diagrams found for almost all tribological systems. The state dependent friction law was also used in [40, 41] where an extensive numerical simulation of the one-dimensional spring-block model with such a friction law has been performed and the magnitude distribution and the recurrence-time distribution were studied.

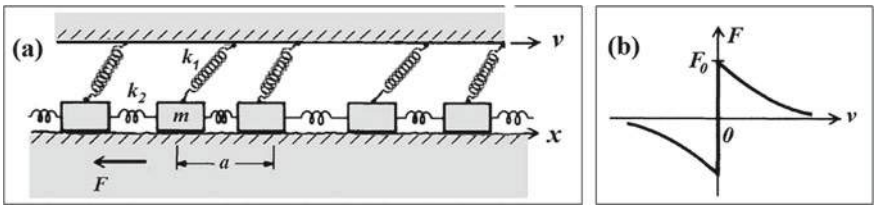


Fig. 8 Burridge-Knopoff (BK) model. Subplot **a** presents a conceptual scheme of the model. In the simulations we use $m = 1$ for the mass and $a = 1$ for the distance between the blocks. Subplot **b** reproduces the original velocity weakening stick–slip friction law used in the Carlson, Langer version of the BK model

Let us start from the original BK model, the conceptual structure of which is depicted in Fig. 8. Blocks of mass m are attached to a moving surface by springs with stiffness k_1 and are coupled to each other by springs with stiffness k_2. The moving surface has a velocity v, and the blocks are in contact with a rough substrate. The friction force F between the blocks and the rough surface is assumed to depend only on the velocity. Subplot (b) of Fig. 8 reproduces the original velocity dependence of the friction force used in Carlson and Langer version of the BK model. The equations of the BK model can be written in the following form:

$$m\frac{\partial^2 u_j}{\partial t^2} = k_2\left(u_{j+1} + u_{j-1} - 2u_j\right) + k_1\left(vt - u_j\right) - F(v_j) \tag{9}$$

where $v = const$ is the external driving velocity and $v_j \equiv \frac{\partial u_j}{\partial t}$ is an array of individual block velocities ($j = 1, \ldots, N$).

In the BK model, the sliding friction force is supposed to decreases monotonously from a constant initial value F_0. It is further supposed that the static friction $F(v_j \to 0)$ can possess any necessary negative value to prevent back sliding:

$$F(v_j) = \begin{cases} \frac{F_0}{1+2\alpha v_j/(1-\sigma)}, \; F_0 = 1-\sigma; & \partial u_j/\partial t > 0 \\ (-\infty, 0] & \partial u_j/\partial t = 0 \end{cases} \tag{10}$$

Here, according to the original works [16, 17], the parameter α defines a rate of friction decrease when block starts to slide, and σ is the acceleration of a block at the instant when slipping begins.

7 Modified BK Model

The friction Eq. (10) is a drastic oversimplification of real properties of static and kinetic friction. This equation does not reproduce the correct stability diagram for sliding; with this friction equation the system is always unstable. Experiments show

that, in most cases of dry friction, sliding stabilizes for either sufficiently large velocities or a sufficiently large stiffness of the system. These friction properties are now well understood and explained in details in the books [31, 32]. Based on friction experiments with rocks, Dieterich [26, 27] has proposed friction equations with internal variables. In his approach, the friction force depends on an additional variable that describes the state of the contact zone. This variable is, in a sense, a "melting parameter." The friction force at non-zero velocity drops down from its initial value due to a "shear melting" effect which may have various physical origins [14–17]. When the motion stops, the surfaces start to form new bonds and the static friction increases with time. These observations become especially important if the model is to be used for describing phenomena with geological characteristic times like earthquakes. Below we follow the ideology of "shear melting effect" and use additional kinetic equations for the friction force.

The dynamics of systems with state dependent friction has been investigated in a number of papers [26–34]. All these studies have been devoted to the simple one-particle version of the model. In the present paper, we investigate dynamics of the many-body BK model with a state dependent friction. The basic dynamics equations of the model are the same.

$$m\frac{\partial^2 u_j}{\partial t^2} = k_2\left(u_{j+1} + u_{j-1} - 2u_j\right) - \eta\frac{\partial u_j}{\partial t} + k_1\left(vt - u_j\right) - F_j[v_j(t)]. \qquad (11)$$

However, the friction force is not a function of velocity, but is defined by the additional kinetic equation:

$$\frac{\partial F_j[v_j(t)]}{\partial t} = \beta_1\left(F_0 - F_j\right) + \beta_2 v_j, \quad \text{with } \beta_2 < 0\, v_j > 0, \qquad (12)$$

$$F(v_j) = -\infty \quad v_j \le 0. \qquad (13)$$

The parameters in first of these equations β_1 and β_2 have the following physical (and geophysical) meaning. When a block starts to slide with $v_j > 0$, its friction force monotonously decreases from an initial value F_0. The general time scale of this process [in relation to other time-scales of the problem, defined by the terms of Eq. (11)] is determined by the first parameter β_1 and an effectiveness of the melting is given by a relation between the absolute value of the negative parameter $\beta_2 < 0$ and β_1.

Static friction $F(v_j \to 0)$ in second line of Eq. (12) can possess any necessary negative value to prevent back sliding, as in Eq. (10). It is the only nonlinear part of the system, which is found to be enough to create all nontrivial properties of the model. For general applicability to tribological problems Eq. (11) contains also a viscous term $\eta\partial u_j/\partial t$. Our calculations show that attractor properties are weakly influenced by the viscous term. Moreover, this influence exists only at sufficiently

high velocities v_j. For generality, in the results presented below, we keep small nonzero value $\eta = 0.05$.

It should be noted that other forms of the kinetic equations for $\partial F_j/\partial t = \beta_1(F_0 - F_j) + \beta_2 v_j$ have been proposed [14, 15] leading to qualitatively similar decreasing behavior of the dynamic friction at $v_j > 0$. Equation (12) is the simplest form. It is also linear. Another reason for this choice is the following. A realistic model must reproduce the correct stick–slip and sliding behavior in appropriate parameter regions. In particular, it must reproduce the correct stability diagram on the plane $\{v,\ k_1\}$, which typically has an unstable region at small system velocities and small system stiffness [14, 15]. Note that the original BK model in the form of Eqs. (9) and (10) does not reproduce a stable sliding regime at small sliding velocities and large stiffness. The modified model qualitatively reproduces a typical known stability diagram for sliding friction under various conditions. We have studied the stability diagram by direct dynamic simulation of the BK-system. For this sake, we control the behavior of the complete velocity arrays $\{v_j(t)\}$ for all blocks $j = 1, \ldots, N$. To visualize the dynamics of the system, it is convenient to calculate the mean velocity:

$$\langle v \rangle \equiv \langle v_j(t) \rangle = \frac{\sum_{j=1}^{N} v_j(t)}{N}. \tag{14}$$

Unless otherwise specified, we keep the number of the blocks equal to $N = 512$. The main control parameters of the system are the driving velocity v and the stiffness k_1 of the "springs" connecting the block with the external subsystem. In order to study general properties of the model, let us vary parameters v and k_1 around the values of order of unity. Typical time dependencies of the mean velocity $\langle v \rangle$ found at free boundary conditions for three representative regions of driving velocity and stiffness constant $v = 0.7$ and $k_1 = 0.2$; $v = 0.7$ and $k_1 = 2$; $v = 2$ and $k_1 = 0.2$ are shown in Fig. 9 subplots (a–c), respectively.

In regions (b) and (c), the fluctuations of the velocity disappear after a transient period, whereas in the stick–slip region (a) the velocity fluctuates continuously. The dynamics of the friction force

$$\langle F_{fric} \rangle = \langle F_j(t) + \eta v_j(t) \rangle \tag{15}$$

is shown in the same plot.

Performing the simulation for various combinations of parameters, we have found the desired stability diagram, shown in Fig. 10, for the modified BK model in the plane $\{v,\ k_1\}$. The line separates the regions of sliding and stick–slip motion. Higher intensity of gray color corresponds to bigger oscillations of friction force. As discussed above, this characterizes the amplitude of the force variation at different spring stiffness k_1 and driving velocity v. If the interaction between the blocks is relatively weak, these strong variations of friction can lead, in principle, to a state in which different blocks of the chain can simultaneously be found in moving and stacked states. This observation is important for studying of the model further. Below we

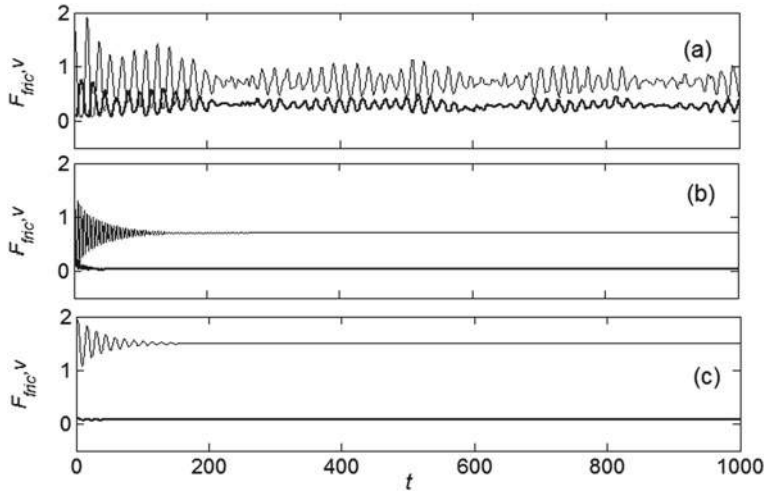


Fig. 9 Time dependencies of mean velocity $\langle v \rangle$ and friction force $\langle F_{fric} \rangle = \langle F + \eta v \rangle$ found for three representative regions of the driving velocity and stiffness constant $v = 0.7$ and $k_1 = 0.2$; $v = 0.7$ and $k_1 = 2$; $v = 2$ and $k_1 = 0.2$ shown in (**a**)–(**c**) subplots, respectively. In the sliding regions **b** and **c** the fluctuations of the velocity disappear after a transient period of initial oscillations, whereas in stick–slip region (**a**) both mean velocity and a correlated friction force remain perpetually variant

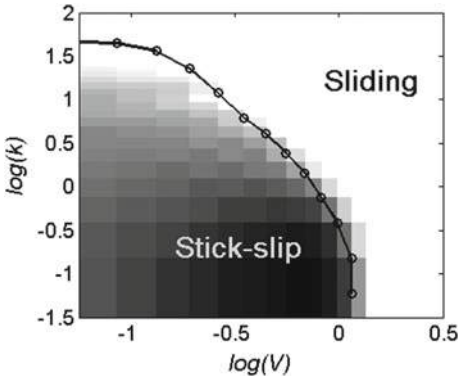


Fig. 10 Stability diagram for the modified BK model (MBK). The line separates the region of stable sliding from that of variable stick–slip motion. The grayscale map represents the standard deviation value of the complete friction force $\langle F_{fric} \rangle = \langle F_j[v_j(t)] + \eta v_j(t) \rangle$ in stationary stick–slip regime. It characterizes the amplitude of the force variation at different spring constants k and driving velocities v. Higher gray color saturation corresponds to bigger friction force oscillations (with black color corresponding to maximum of standard deviation equal to 0.72)

study the dynamic and statistical properties of the new model with the previously described realistic state dependent friction law. We will refer to the set of Eqs. (12), (13) as the modified BK model (MBK).

8 Attractor Properties, Wave State and Phase Transition in a 1-Dimensional Model

The equations of motion in (12) are a discrete representation of a nonlinear wave equation. During the last few decades various nonlinear wave equations have been widely studied, starting from the very early implementation of the numerical simulations [35]. In particular, in the context of dynamic "thermalization" it has been shown that N interacting segments of the nonlinear chain form a collective attractor with energy transfer performed by nonlinear excitations [36, 37, 42, 43]. Depending on the total energy and/or on the strength of the interaction between the blocks k_2, the chain can form (nearly) uniform or strongly non-uniform structures and phase patterns.

Analogous behavior should be expected in the MBK model as well. It is obvious that instantaneously moving blocks must be involved with stationary ones in the overall motion. This causes a "detachment" wave, propagating along the chain. Such solitary waves were found and studied in the original BK model [15], and the analogous process exists in the MBK model. There are two possible types of traveling waves: (a) after a transition time all the blocks are provoked to move simultaneously, (b) in steady-state, some of the blocks can be found instantaneously motionless while others move. One would expect that the response of the chain will depend on the strength of the interaction between the blocks. This has been studied for the original BK model in [17], and a specific "phase transition" between correlated and uncorrelated behavior has been found.

Let us study this problem for the MBK model. Figure 11 presents typical waves of local block velocities in a "stationary" regime for parameter values $k_2 = 4$, $k_1 = 1$ and $v = 0.2$. Non-zero velocities are shown here by the spatiotemporal mesh surface. Traveling and mutually scattering waves are clearly visible. Let us draw the attention to the areas of intensive local spikes appearing as result of mutually scattering of waves. These spikes have displacement amplitudes $|u_j| \gg \langle u_j \rangle$ and velocities $|v_j| \gg \langle v_j \rangle$ much higher than the mean values. They are relatively rare (the area occupied by these events is much smaller than the total space–time area), and in applications for geodynamics they should be treated as "earthquake events".

To characterize the difference of the correlated and uncorrelated behavior, an order parameter has been introduced in Ref. [22]. If we define a value h_j in block number j by the condition:

$$h_j = \begin{cases} 1 & \partial \mathrm{u}_j/\partial t \neq 0 \text{ (the block is moving);} \\ 0 & \partial \mathrm{u}_j/\partial t = 0 \end{cases} \tag{16}$$

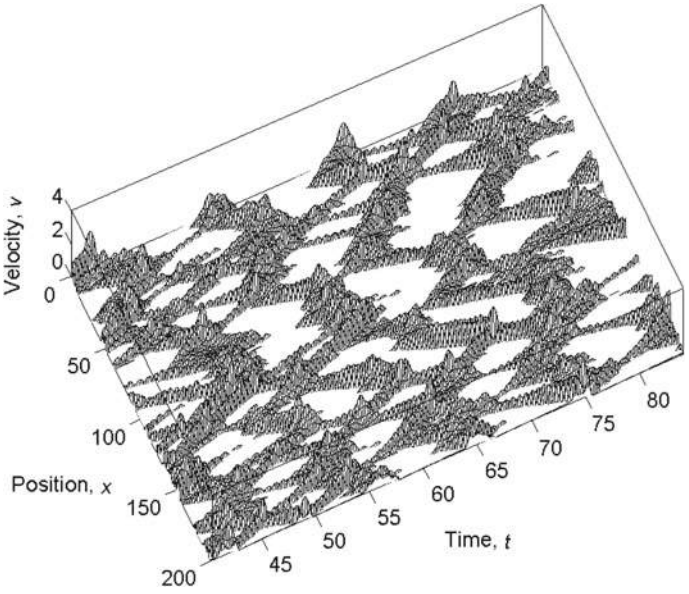


Fig. 11 Typical waves of the local velocity (of each discrete block of the system) for the developed regime in a one-dimensional system. The non-zero velocities are shown by the spatiotemporal mesh surface. The areas of intensive local events due to mutual scattering of waves are clearly visible. Parameter values: $v = 0.7,\ k_1 = 0.2,\ k_2 = 4,\ N = 512,\ \beta_1 = 1,\ \beta_2 = 25,\ \eta = 0.05,\ m = 1$.

then the local density of the order parameter H_j^* can be written as $H_j^* = h_j(h_{j+1} + h_{j-1})$. This function takes the unit value $H_j^* = h_j(h_{j+1} + h_{j-1}) = 1$ if block j is moving and exactly one of its nearest neighbors is also moving. Further, $H_j^* = h_j(h_{j+1}+h_{j-1}) = 2$ when both nearest neighbors of the moving block are in motion. All other cases yield $H_j^* = 0$.

Our observations with the MBK model show that even for blocks at vanishing inter-block interaction $k_2 \to 0$ (when the motion of neighboring blocks is almost uncorrelated), the fraction of configurations with moving sets of neighboring blocks is still relatively high. The combination $H_j^* = h_j(h_{j+1} + h_{j-1})$ does not vanish in such an uncorrelated system. It is therefore convenient to construct another simple combination:

$$H_j = h_j h_{j+1} h_{j-1} = \begin{cases} 1 & \text{all 3 blocks } j,\ j+1,\ j-1 \text{ in contact are moving} \\ 0 & \text{otherwise} \end{cases}. \quad (17)$$

This yields $H_j = h_j h_{j+1} h_{j-1} = 0$ zero in all cases except when both neighboring blocks of a central sliding block are also in motion. This combination can be used as an "order parameter".

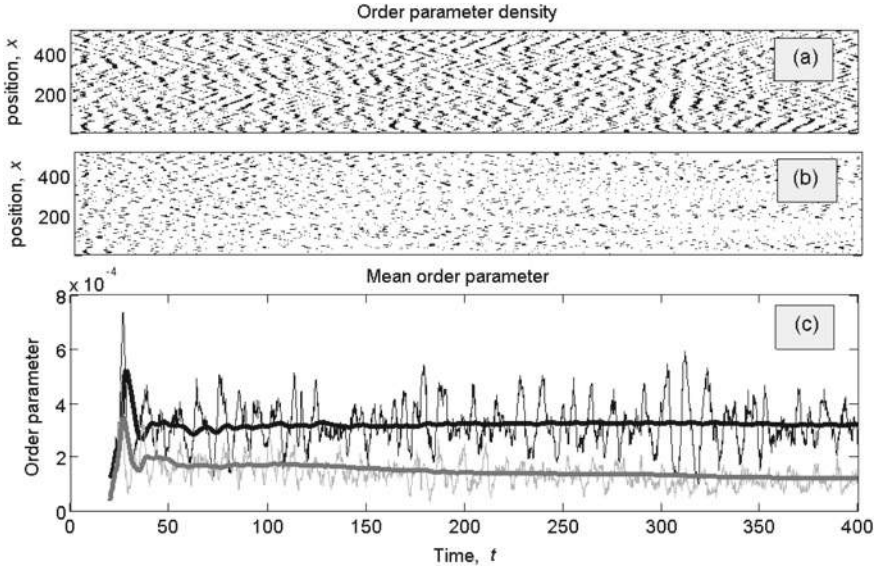


Fig. 12 Order parameter density and its time dependence in two different regimes. Subplots **a** and **b** show the grayscale maps of the order parameter density depending on time and space for strong and weak coupling between the neighboring blocks ($k_2 = 4$ and $k_2 = 1.2$ respectively). Time evolution of the ensemble-averaged order parameter for (**a**) and (**b**) cases are presented in the subplot (**c**) by black and gray tick lines respectively. The time averages for these values are shown by the bold lines of corresponding color. Parameter values: $v = 0.7,\ k_1 = 0.2,\ k_2 = 4,\ N = 512, \beta_1 = 1,\ \beta_2 = 25,\ \eta = 0.05,\ m = 1$.

Figure 12 shows the order parameter density and time dependence of its ensemble average $\langle H_j(t)\rangle$ in two different regimes. Subplots (a) and (b) show grayscale maps of the order parameter density depending on time and space for strong and weak coupling between neighboring blocks ($k_2 = 4$ and $k_2 = 1.2$ respectively). Time dependence of the mean value $\langle H_j(t)\rangle$ is presented in the subplot (c) with black and gray thick lines for the (a) and (b) cases, respectively.

To extract integral quantitative information about the steady ordered and disordered states let us calculate the time evolution of the ensemble-averaged order parameter:

$$H(t) \equiv \langle H_j(t)\rangle = \frac{1}{t}\int_0^t \langle H_j(t)\rangle dt. \tag{18}$$

These time-averaged values for the cases (a) and (b) are shown in the subplot (c) by the bold lines overlapping respective thick curves $\langle H_j(t)\rangle$. The long-time stationary asymptote $\langle H_j(t)\rangle \to const$ can characterize the behavior of the system in an integral manner.

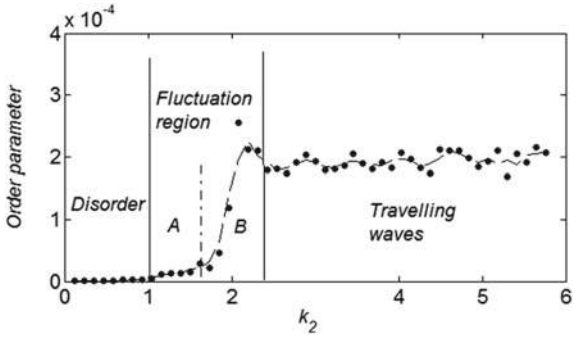


Fig. 13 Phase transition from correlated to uncorrelated motion of blocks in a one-dimensional system. The order parameter tends to a constant asymptote at high mutual interaction $k_2 \gg 1$ and vanishes below transition point $k_2 = k_2^{critical} \simeq 1$. Other parameters are the same as in Fig. 5. The ordered state corresponds to the nearly regular waves (seen clearly in Fig. 11). Two intermediate fluctuation regions A and B correspond to states with short-range and long-range correlated nonlinear excitations, respectively

The dependence of the integrated order parameter on the stiffness k_2, showing a transition from correlated to uncorrelated block motion in a one-dimensional system, is shown in Fig. 13. Two limiting cases can be identified: (a) the order parameter tends to a constant non-zero asymptote $\langle H_j(t) \rangle \to const \neq 0$ at strong interaction $k_2 \gg 1$, and (b) it vanishes $\langle H_j(t) \rangle \to 0$ below the transition point $k_2 \approx 1$.

The ordered state corresponds to the nearly regular waves presented in Fig. 12. We distinguish two fluctuation regions A and B at intermediate interaction. These regions can be characterized by two clearly different order parameter mean values in Fig. 13. One can verify further that they also differ dynamically and correspond to states with short range- and long-range correlated nonlinear excitations, respectively. This intermediate behavior may characterize the physically important features of the model under consideration.

9 Study of the 2-Dimensional Model

The real contact of two surfaces is two-dimensional. Let us generalize the MBK model for the 2D case. The generalized model is very similar to the 1-dimensional model, but incorporates a 2D array of blocks connected by elastic springs in both directions. All other components of the MBK remain unchanged.

The system of equations of motion takes the form:

$$m\frac{\partial^2 u_{j,n}}{\partial t^2} = k_2\left(u_{j+1,n} + u_{j-1,n} + u_{j,n+1} + u_{j,n-1} - 4u_{j,n}\right) - \eta\frac{\partial u_{j,n}}{\partial t} + k_1\left(vt - u_{j,n}\right) + F(v_{j,n}); \tag{19}$$

$$\frac{\partial F_{j,n}(v_{j,n}(t))}{\partial t} = \beta_1\left(F_0 - F_{j,n}\right) + \beta_2 v_{j,n};\ \text{with}\ \beta_2 < 0,\ v_{j,n} > 0, \qquad (20)$$

$$F(v_{j,n}) = -\infty \quad v_{j,n} \le 0$$

where: $j = 1, \ldots, N_x$ and $n = 1, \ldots, N_y$. N_x and N_y are the numbers of elements in the x- and y-directions. It is possible to repeat all the simulations of the previous section, reproduce all the results presented in Figs. 9, 10, 11, 12 and 13, and show that these properties are quite common between the 1D and 2D MBK models.

In particular, one can obtain a wave state in two dimensions. The only difficulty appears in a visualization of the results, depending on 3 coordinates $\{x, y; t\}$. As an example, Fig. 14 presents the mentioned wave state in the two-dimensional model. In contrast to the 2-dimensional $\{x; t\}$ space–time maps with a complete history of events in Figs. 11 and 12 the subplots (a) and (b) now represent instantaneous

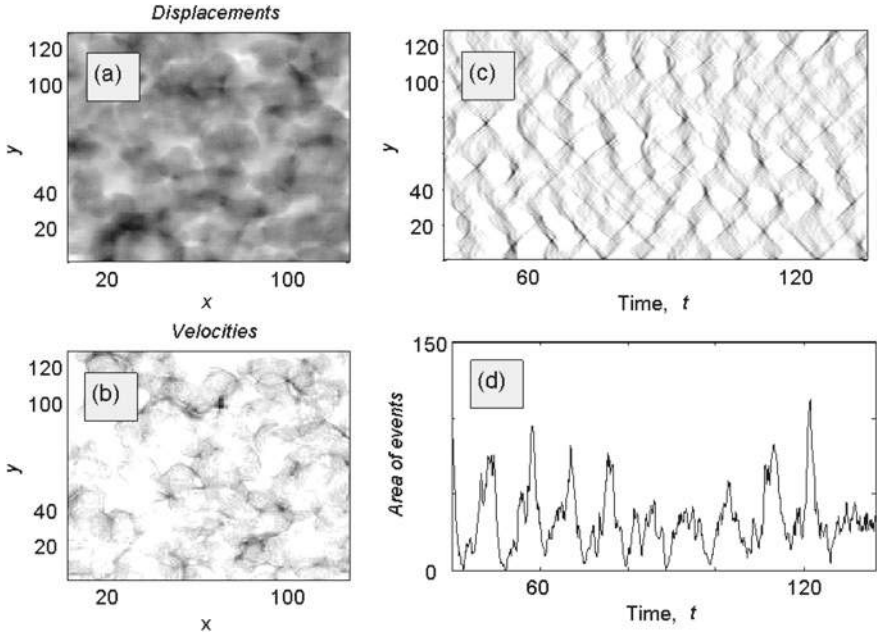


Fig. 14 Wave state in the two-dimensional model. Subplots **a** and **b** represent snapshots of the instantaneous densities of the local displacements $u = u(x, y; t)$ and velocities $v = v(x, y; t) \equiv \partial u/\partial t$, respectively. The darkest color corresponds to the value 3 indimensionless units and white corresponds to zero. Mutual scattering manifests itself in high sharp peaks of the "events" reproduced here by the dark gray spots of the corresponding densities. A time–space representation of this process is shown in (**c**) by the cross-section of the $\{x, y; t\}$-space along one of the $x = const$ planes. Ensemble averaged area of the events corresponding to the same process is plotted in subplot (**d**). The number of blocks is equal to $N_x \times N_y = 128 \times 128$ and other parameters are the same as in Fig. 12

snapshots of the density distribution (for the local displacements $u = u(x, y; t)$ and velocities $v = v(x, y; t) \equiv \partial u/\partial t$ respectively). However, direct observation of the time-dependent numerical simulations shows that the waves (clearly visible in the Fig. 14) are moving $2D$ fronts of the excitations. These fronts conserve their shape for relatively long periods of time. Their mutual scattering manifests itself in high and sharp peaks of the "events". Corresponding spikes are well reproduced by dark gray spots of the density distributions in subplots (a) and (b) of Fig. 14.

Some record of the process which has led to the presented instantaneous distributions is shown in subplot (c) by means of a cross-section of the $\{x, y; t\}$-space along one of the planes where $x = const$. One can also calculate an ensemble averaged area of the events corresponding to the same process. This is plotted in Fig. 14 subplot (d). There is obvious correlation between the subplots (c) and (d); however, the correlation is not complete. In reality, the total area of events includes a summation over all the planes $j = 1, \ldots, N_x$ and involves plenty of impacts from other $x = const$ planes which are invisible in the subplot (c).

Nevertheless, the correlation is easily seen and looks much stronger than one would expect in such case. The traveling waves influence the motion of blocks neighboring in both directions $\{x, y\}$. Therefore, there is a certain correlation between the densities along all 3 time–space coordinates $\{x, y; t\}$. In order to reproduce this in a static picture, we present a 3-dimensional density distribution for the exact same process in Fig. 15. This figure combines the density of "events" (black volumes) with the grayscale maps discreetly depicted for certain sub-planes: $t = const$, $x = const$ and $y = const$.

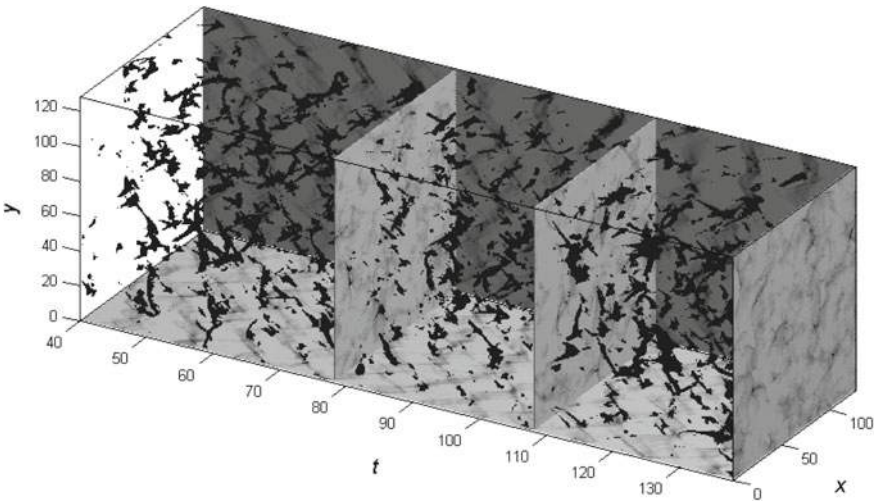


Fig. 15 Density of events (black volumes) in a two-dimensional system $N_x \times N_y = 128 \times 128$ at the same parameters as in Figs. 12 and 13. Grayscale maps (with the same gradations as in previous figure) for some representative planes $t = const$, $x = const$ and $y = const$ are added to compare with Fig. 13

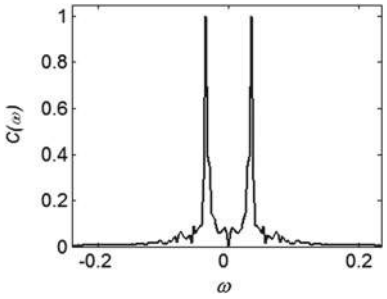


Fig. 16 Fourier transform $C(\omega)$ of the two-time correlation function $G(t_2 - t_1)$

There is a noticeable periodicity in the total area of events in Figs. 14 and 15. To illustrate this, we have calculated a two-time correlation function $G(t_2 - t_1)$ for the total area and taken its Fourier transform G_ω. The resulting Fourier transform is presented in Fig. 16. It smoothes random impacts from the time-fluctuations and possesses obvious maxima corresponding to a characteristic frequency of the total area oscillation. The frequency of these large-scale collective oscillations is determined by the parameters of the problem. According to our numerical experiments, the characteristic frequency can be varied mainly by changing the driving velocity, external springs and constants $\beta_{1,2}$ in the equation $\partial F_{j,n}[v_{j,n}(t)]/\partial t = \beta_1\left(F_0 - F_{j,n}\right) + \beta_2 v_{j,n}$. It is important to stress here that the observed behavior corresponds to a global attractor of the dynamic system (19) and (20). This means that it corresponds to the stationary asymptotic behavior of the system, independent of initial conditions.

In all the cases presented in previously mentioned figures, we omitted initial time intervals corresponding to the transient period. This part of the evolution can be different and depends on the initial conditions. We have checked this by starting the simulations from almost uniform distribution of the low velocities, from small displacements, from intensive random noise, or by changing open boundary conditions (normally used here) to the periodic ones, and so on. In all cases, the system quickly suppresses unfavorable fluctuations, vents to an attracting "large river" common for all the transient scenarios, and slowly attracts along the "river" to the stationary scenario. This kind of evolution is mathematically typical for many nonlinear systems [38] and the MBK model is no exception.

The attractor manifests itself in a stationary distribution $\rho = \rho\{u, F_{fric}, v, \ldots\}$ of the dynamic variables in a phase space. Figure 17a, b present its projections in two different sub-planes of the phase-space: planes $\{u, v\}$ and $\{F_{fric}, v\}$, respectively. By accumulating the density $\rho = \rho\{u, \ldots, v, \ldots\}$ onto a grayscale map in subspace $\{u, v\}$, one can see the correlation between the time-depending fluctuations of displacements $u(x, y; t)$ and velocities $v(x, y; t)$.

It can be shown that the dense central part of the distribution is mainly due to multiple but weak oscillations of small amplitude ("phonons") and basal areas of the traveling waves. The widely extended depopulated gentle slopes with low density

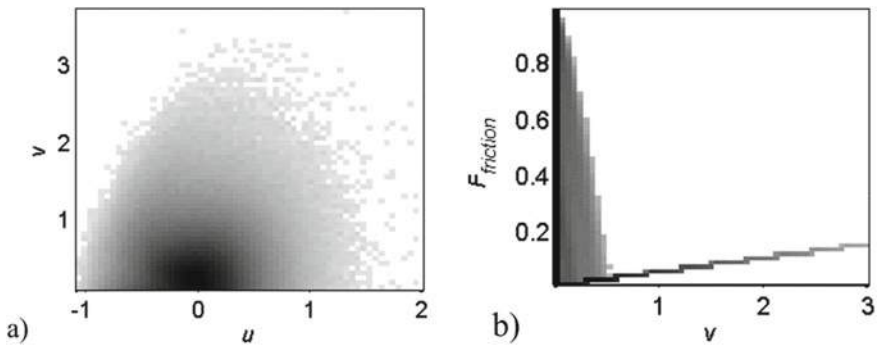


Fig. 17 **a** Grayscale map of the density of dissipative attractor states projected onto the phase sub-space $\{u, v\}$. **b** The same attractor as in Fig. 17a projected onto the sub-space $\{v, F_{friction}\}$, where $\langle F_{friction}\rangle = \langle F_j[v_j(t)] + \eta v_j(t)\rangle$

$\rho = \rho\{u, \ldots, v, \ldots\}$ in peripheral regions of the $\{u, v\}$ surface are produced by rare intensive "events" [which cause high spikes of displacements $u(x, y; t)$ and velocities $v(x, y; t)$]. In other words, statistical study of the rare "earthquake events" in the frame of the MBK model is equivalent to the study of the outer periphery of its dissipative attractor.

Another projection of the attractor onto the $\{F_{fric}, v\}$ subspace shown in the Fig. 17b can be used to control correct correspondence between statistically preferable behavior of the dynamically complete friction force $F_{fric} = F[v_{n,j}(t)] + \eta v_{n,j}(t)$ with the "naive", physically expected dependence $F = F_{fric}(v)$. Finally, let us return again to the discussion of Fig. 17a. The inherent structure of the attractor with extended gentle slopes of the density $\rho = \rho\{u, \ldots, v, \ldots\}$ corresponding to rare intensive "events" gives a simple and clear image for the origination of scaling asymptotic distributions. To obtain these, one must cut off the outer areas along both the displacement and velocity coordinates. Corresponding asymptotic distributions obtained after such a cut-off are reproduced in subplots (a) and (b) of Fig. 18, respectively. The inserts to the figures illustrate the power-law nature of both distributions.

Comparing the models one can conclude that the standard BK model utilizes a velocity weakening friction force to reproduce the correct statistical behavior of "events". In contrast, the MBK model includes an additional phenomenological equation, subsequently providing a self-consistent dynamic description of the velocity depending friction force. This modification has at least two advantages: it realistically generates the velocity weakening friction force of the moving blocks and provides growth of static friction for the locked blocks. The model was studied for different driving velocities and driving springs elastic constants. It was possible to build a stability diagram for the transition between smooth sliding and stick–slip behavior, which was in good qualitative agreement with what is expected experimentally. Further numerical study under a broad range of parameters proved that the MBK model reproduces all important features of the standard BK model (traveling

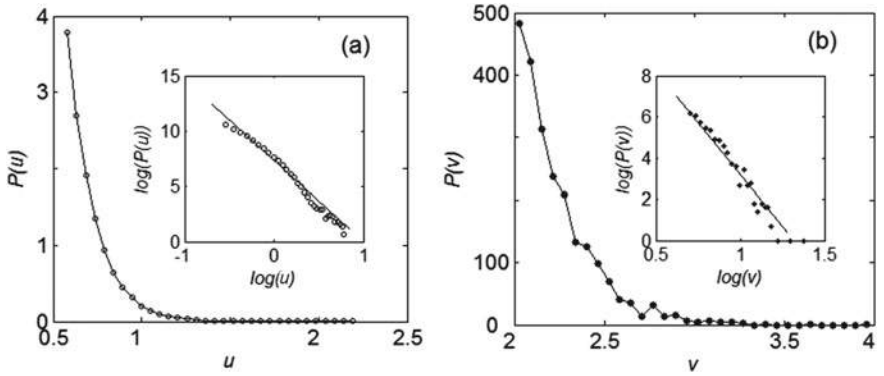


Fig. 18 Scaling relations in the model. To get scaling behavior one must cut-off the external light-gray regions with power-law density of states from the attractor in Fig. 17. Physically it corresponds to a selection of the rare but intensive "events", which is compatible with the ideology of the empirical Gutenberg-Richter law

waves, attractor properties of dynamic equations in one- and two-dimensional cases, and so on).

References

1. Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. Bull Seismol Soc Am 34(4):185–188
2. Bak P, Christensen K, Danon L, Scanlon T (2002) Unified scaling law for earthquakes. Phys Rev Lett 88:178501
3. Omori F, Coll J (1895) On the aftershocks of earthquakes. J Coll Sci Imperial Univ Tokyo 7:111–120
4. Bak P, Tang Ch, Wiesenfeld K (1987) Self-organized criticality: an explanation of the 1/f noise. Phys Rev Lett 59(4):381–384
5. Carlson JM, Langer JS (1989) Properties of earthquakes generated by fault dynamics. Phys Rev Lett 62:2632
6. Ruzhich VV, Smekalin OP, Shilko EV, Psakhie SG (2002) About nature of "slow waves" and initiation of displacements at fault regions. In: Proceedings of international conference "new challenges in mesomechanics", Aalborg University, Denmark, Aug 26–30, 2002, vol 1, pp 311–318
7. Ruzhich VV, Truskov VA, Chernykh EN, Smekalin OP (1999) Russ Geol Geophys 40:356
8. Psakhie SG, Shilko EV, Astafurov SV (2004) Peculiarities of the mechanical response of heterogeneous materials with highly deformable interfaces. Tech Phys Lett 30:237–239
9. Psakhie SG, Ruzhich VV, Shilko EV, Popov VL, Dimaki AV, Astafurov SV, Lopatin VV (2005). Influence of the state of interfaces on the character of local displacements in fault-block and interfacial media. Tech Phys Lett 31(8):712–715. https://link.springer.com/article/10.1134/1.2035374
10. Ruzhich VV, Psakhie SG, Bornyakov SA, Smekalin OP, Shilko EV, Chernykh EN, Chechelnitsky VV, Astafurov SV (2002) Investigation of influence of vibroimpulse excitations on regime of displacements in seismically active fault regions. Phys Mesomech 5(5–6):85

11. Fermi E, Pasta J, Ulam S, Tsingou M (1955) Studies of non linear problems. Los Alamos Report Laboratory of the University of California, LA-1940
12. Filippov AÉ, Popov VL, Psakhie SG, Ruzhich VV, Shilko EV (2006) Converting displacement dynamics into creep in block media. Tech Phys Lett 32:545–549
13. Filippov AÉ, Popov VL, Psakhie SG (2008) Correlated impacts optimizing the transformation of block medium dynamics into creep regime. Tech Phys Lett 34(8):689–692
14. Filippov AÉ, Popov VL (2010) Modified Burridge-Knopoff model with state dependent friction. Tribol Int 43:1392–1399
15. Burridge R, Knopoff L (1967) Model and theoretical seismicity. Bull Seismol Soc Am 57(3):341–371
16. Bruce SE, Carlson JM, Langer JS (1992) Patterns of seismic activity preceding large earthquakes. J Geophys Res Solid Earth 97(B1):479
17. Carlson JM, Langer JS, Shaw BE (1994) Dynamics of earthquake faults. Rev Mod Phys 66(2):657–670
18. Saito T, Matsukawa H (2007) Size dependence of the Burridge-Knopoff model. In: Journal of physics: conference series, International conference on science of friction, 9–13 Sept 2007, Irago, Aichi, Japan, vol 89, p 012016
19. Mandelbrot BB (1982) The fractal geometry of nature. Freeman and Co, San Francisco
20. Gutenberg B, Richter CF (1954) Seismicity of the earth and associated phenomena. Princeton University Press, Princeton (New Jersey)
21. Muratov CB (1999) Traveling wave solutions in the Burridge-Knopoff model. Phys Rev E 59:3847
22. Huisman BAH, Fasolino A (2005) Transition to strictly solitary motion in the Burridge-Knopoff model of multicontact friction. Phys Rev E 72(1Pt2):016107
23. Clancy I, Corcoran D (2005) Criticality in the Burridge-Knopoff model. Phys Rev E 71(4Pt2):046124
24. Clancy I, Corcoran D (2006) Burridge-Knopoff model: exploration of dynamic phases. Phys Rev E 73(4Pt2):046115
25. Mori T, Kawamura H (2008) Simulation study of earthquakes based on the two-dimensional Burridge-Knopoff model with long-range interactions. Phys Rev E 77(5Pt1):051123
26. Dieterich JH (1974) Earthquake mechanisms and modeling. Annu Rev Earth Planet Sci 2:275–301
27. Dieterich JH (1978) Time-dependent friction and the mechanics of stick-slip. Pure Appl Geophys 116:790–806
28. Rice JR (1983) Earthquake aftereffects and triggered seismic phenomena. Pure Appl Geophys 121:187–219
29. Gu JC, Rice JR, Ruina AL, Tse ST (1984) Slip motion and stability of a single degree of freedom elastic system with rate and state dependent friction. J Mech Phys Solids 32(3):167–196
30. Marone C (1998) Laboratory-derived friction laws and their application to seismic faulting. Annu Rev Earth Planet Sci 26:643–696
31. Persson BNJ (2000) Sliding friction, physical properties and applications. Springer, Berlin
32. Popov VL (2010) Contact mechanics and friction. Foundations and applications. Springer, Berlin
33. Popov VL (2000) A theory of the transition from static to kinetic friction in boundary lubrication layers. Solid State Commun 115(7):369–373
34. Filippov AE, Klafter J, Urbakh M (2004) Friction through dynamical formation and rupture of molecular bonds. Phys Rev Lett 92(13):135503
35. Fermi E, Pasta J, Ulam S, Tsingou M (1993) In: Mattis DC (ed) The many-body problem: an encyclopedia of exactly solved models in one dimension. World Scientific Publishing Co Pte Ltd., Singapore
36. Braun OM, Hu B, Filippov A, Zeltser A (1998) Traffic jams and hysteresis in driven one-dimensional systems. Phys Rev E Stat Phys Plasmas Fluids Relat Interdisc Top 58(2):1311
37. Fillipov A, Hu B, Li B, Zeltser A (1998) Energy transport between two attractors connected by a Fermi-Pasta-Ulam chain. J Phys A Gen Phys 31(38):7719

38. Filippov AE (1994) Mimicry of phase transitions and the large-river effect. JETP Lett 60(2):141
39. Popova E, Popov VL (2015) The research works of Coulomb and Amontons and generalized laws of friction. Friction 3(2):183–190
40. Persson BNJ, Popov VL (2000) On the origin of the transition from slip to stick. Solid State Commun 114(5):261–266
41. Ohmura A, Kawamura H (2007) Rate-and state-dependent friction law and statistical properties of earthquakes. EPL (Europhys Lett) 77(6):69001
42. Ford J (1992) The Fermi-Pasta-Ulam problem: paradox turns discovery. Phys Rep 213(5):271–310
43. Lichtenberg AJ, Lieberman MA (1992) Regular and chaotic dynamics. Springer, New York

Material Transfer by Friction Stir Processing

Alexander A. Eliseev, Tatiana A. Kalashnikova, Andrey V. Filippov, and Evgeny A. Kolubaev

Abstract Mechanical surface hardening processes have long been of interest to science and technology. Today, surface modification technologies have reached a new level. One of them is friction stir processing that refines the grain structure of the material to a submicrocrystalline state. Previously, the severe plastic deformation occurring during processing was mainly described from the standpoint of temperature and deformation, because the process is primarily thermomechanical. Modeling of friction stir welding and processing predicted well the heat generation in a quasi-liquid medium. However, the friction stir process takes place in the solid phase, and therefore the mass transfer issues remained unresolved. The present work develops the concept of adhesive-cohesive mass transfer during which the rotating tool entrains the material due to adhesion, builds up a transfer layer due to cohesion, and then leaves it behind. Thus, the transfer layer thickness is a clear criterion for the mass transfer effectiveness. Here we investigate the effect of the load on the transfer layer and analyze it from the viewpoint of the friction coefficient and heat generation. It is shown that the transfer layer thickness increases with increasing load, reaches a maximum, and then decreases. In so doing, the average moment on the tool and the temperature constantly grow, while the friction coefficient decreases. This means that the mass transfer cannot be fully described in terms of temperature and strain. The given load dependence of the transfer layer thickness is explained by an increase in the cohesion forces with increasing load, and then by a decrease in cohesion due to material overheating. The maximum transfer layer thickness is equal to the feed to rotation rate ratio and is observed at the axial load that causes a stress close to the yield point of the material. Additional plasticization of the material resulting from the acoustoplastic effect induced by ultrasonic treatment slightly reduces the transfer layer thickness, but has almost no effect on the moment, friction coefficient, and temperature. The surface roughness of the processed material is found to have a similar load dependence.

A. A. Eliseev · T. A. Kalashnikova · A. V. Filippov · E. A. Kolubaev (✉)
Institute of Strength Physics and Materials Science SB RAS, 634055 Tomsk, Russia
e-mail: eak@ispms.tsc.ru

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_8

Keywords Friction stir welding · Ultrasound · Aluminum alloy · Transfer layer · Roughness · Friction coefficient

1 Introduction

Friction stir welding is a solid-state process of permanent joining, which is based on mass transfer. A nonconsumable tool, rotating during welding, heats the two pieces of a material to a plastic state ($\approx 0.8 \cdot T_m$) due to friction force and mixes them as it moves along the joint line (Fig. 1). Being a nonmelting process, friction stir welding allows joining unweldable and dissimilar materials. Weld material moves in a complex way, undergoing not only circular but also upward and downward flow. In this case, the material can be taken as a quasi-viscous fluid, while still remaining a solid phase. Insufficient or excessive heat generation causes macrodefects to form in the seam, such as voids, wormholes, lack, etc. About 80% of heat is produced due to friction of the tool shoulders against the workpiece surface [1], and mass transfer in the bulk material is mainly implemented by the tool pin. The combination of elevated temperatures and severe plastic deformation during welding induces a series of processes at various stages: recovery, annealing, static and dynamic recrystallization, etc. [2].

These processes can significantly alter a structure and properties of the material. For example, the grain structure of rolled 2024 aluminum alloy becomes equiaxial and refines by an order of magnitude (Fig. 2). Restructuring occurs at all hierarchical levels of the material: grains, subgrains, intermetallic compounds, coherent and semi-coherent particles of secondary phases, and lattice curvature. This feature makes friction stir welding applicable to surface modification. In particular, friction stir processing is capable of hardening the surface due to plastic deformation, forming composite materials by mixing particles of dissimilar materials, removing surface defects, for example, pores in products of powder additive manufacturing, etc., which offers promise for many industries. However, this comparatively recent technology (developed in 1991 at the Welding Institute of the United Kingdom) has its limitations and is still not clearly understood. In particular, deformation and heat are combined in a complex way during welding, which makes it difficult to predict

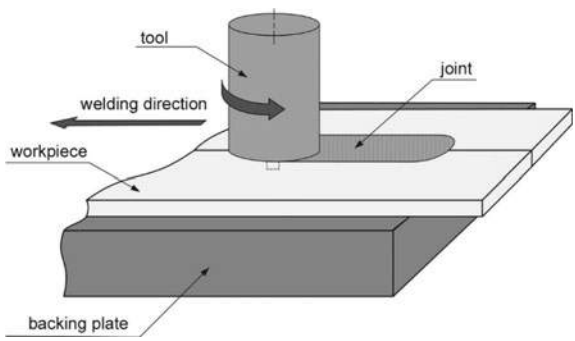


Fig. 1 Scheme of friction stir welding process

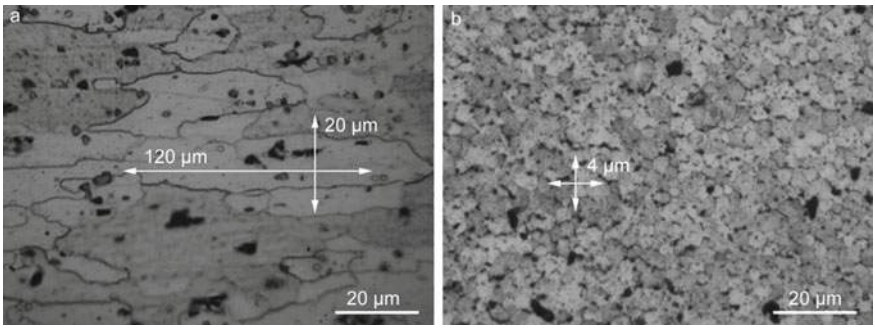


Fig. 2 Metallographic images of initial rolled sheets AA2024 (**a**) and after friction stir welding (**b**) with grain size

their effect on structure and properties. It is known from general materials science that heating coarsens grains of a solid solution and particles of secondary phases (an example is thermally hardened aluminum alloys), while deformation, due to dynamic recrystallization and strain-induced dissolution of phases, leads to grain refinement and dissolution of intermetallides in the solid solution. The influence of welding process parameters on heat generation and deformation is however nonlinear. Moreover, deformation and heat generation prove to be interrelated processes. A part of deformation energy goes into moving material macrovolumes, and a part is expended in increasing internal energy, i.e. temperature, due to internal friction. Heating, in its turn, changes viscosity of the material. Changed viscosity affects strength of the material and conditions of adhesion to the tool, i.e. deformation. To ensure the necessary temperature and deformation in practice, an optimal welding/processing mode is chosen using methods of parameter optimization, fuzzy logic, and neural networks. However, advances in the production of materials and compounds do not help to elucidate fundamental processes of friction and deformation. This work is aimed at summarizing the scientific results in this field and synthesizing them with the original experimental data.

2 Influence of Process Parameters

The main friction stir process parameters are tool rotation rate, tool feed rate, and axial load. Without adaptive adjustment during the process, the first two parameters are kept constant. As for load, the "soft" and "hard" modes are distinguished. The soft mode is characterized by constant axial load acting on the tool during its plunging and further processing/welding. In the hard mode, the tool plunge rate is set, and further processing occurs without significant axial load. The soft mode takes account of natural flow of the material around the tool. In the hard mode, the load varies greatly, making the process unstable as heat generation and deformation change at each

process stage. However, the hard mode is more convenient in mass production and therefore the most common. From considerations of intensity of the impact on a material and quality of products, processing under constant axial load is favored. It also provides another way to control the process. As tool pressure, the load enters one of the classical equations of heat generation due to friction stir welding [3]:

$$Q_S = \frac{4}{3}\pi^2 \mu P \omega R_S^3 \tag{1}$$

where Q_S is the friction heat generated by the tool shoulders, μ is the friction coefficient, P is the pressure under the shoulders, ω is the tool rotation rate, and R_S is the tool shoulder radius.

This equation can be conveniently used at the specified constant friction coefficient, which is often the case. However, the friction coefficient is impossible to directly measure during processing; tribological model experiments also fail to provide accurate values as the coefficient depends on the quality of surfaces, materials, rotation rate, and temperature. Thus, the friction coefficient will change not only at different process parameters but also during welding and even along the tool pin height. The coefficient can be determined analytically and verified indirectly. In one of the first equations proposed in [4], the friction coefficient for the shoulder was expressed via the moment:

$$\mu = \frac{3M}{2F_z R_S} \tag{2}$$

where M is the measured moment, and F_z is the axial force.

The calculated friction coefficients for aluminum alloys 5182 and F-357 are very much different at the same moment (0.35–0.55 and 0.6–1.3, respectively), which is explained by different axial load. The applied load, in turn, depends on the material properties. Kumar et al. [5] used the same approach to determine the friction coefficient but with moment in place of tangential force. He found that a higher load, under parameters being equal, increases the friction coefficient (in the range 500–1500 N and 200–1400 rev/min for alloy AA7020-T6). The friction coefficient rises to 1.4 at the maximum load and rotation rate. The measured temperature correlates with the friction coefficient and amounts to 450 °C at the maximum parameters. Thus, the temperature and friction coefficient exhibit synergy. Based on this, modeling of welding/processing should take account of the temperature dependence of the friction coefficient.

The last successful attempt was described in [6]:

$$\mu = \frac{\tau_0 - \tau_1}{\left(1 - \frac{\tau_1 - \tau_0 \sin\alpha}{(1-\sin\alpha)\tau_y}\right) P_0 (1 - \sin\alpha)} \tag{3}$$

where τ_0 are the shear stresses under the pin and shoulders, τ_1 are the shear stresses on the lateral surface of the pin, P_0 is the axial contact pressure, α is the tool pin

cone angle, τ_y are the shear stresses determined using the von Mises yield criterion and Johnson–Cook material model with regard to temperature and strain rate. The experimentally determined and modeled welding temperatures in various modes differed by no more than 3%, which confirms the adequacy of the approach.

Thus, a variation in load affects heat generation, and the related temperature variation can change the friction coefficient. These processes can influence the material adhesion to the tool and cohesion in the material. According to the adhesion-cohesion mechanism of friction stir processing, changes in mass transfer are possible, which, however, have been still disregarded. Analytical evaluation of strain is usually implemented via geometry, feed and rotation rates of the tool. For example, the classical equation of strain due to friction stir welding is [7]

$$\varepsilon = \ln\left(\frac{l}{APR}\right) + \left|\ln\left(\frac{APR}{l}\right)\right| \tag{4}$$

where *APR* is advance per revolution, and *l* is the maximum deformed length. In fact, by *APR* is meant a material volume deformed per tool revolution without consideration for its adhesive transfer. This evaluation method may be adequate if the tool does not slip in the material, i.e., at optimal adhesion. However, a deformed volume does not conform to *APR* in every instance and in full measure. For a better understanding of deformation during friction stir processing, it is necessary to study the features of mass transfer.

3 Adhesion-Cohesion Concept of Mass Transfer

Mass transfer is characterized by the thickness of a transfer layer. The transfer layer is used to mean a certain material volume that sticks to the tool surface due to adhesion forces and is transferred in the rotation direction. This layer, when moving, grows in mass by capturing the surrounding material due to cohesion forces. When the critical mass is reached, the transfer layer breaks away from the tool on the trailing edge where the driving shear force of the tool and shear stresses due to the surrounding material are differently directed. A continuous material is thus formed layer by layer behind the tool. In some materials, for example, aluminum alloys, the thickness of this layer can be precisely determined at the front surface where the tool passes and in an etched longitudinal section within the material (Fig. 3).

In most works (for example in [8]), the transfer layer thickness *h* is expressed as the ratio of the parameters:

$$h = \frac{V}{\omega} \tag{5}$$

where *V* is the feed rate. Similar results are derived when introducing a marking material into the joint between the workpieces or placing it at the front surface. In

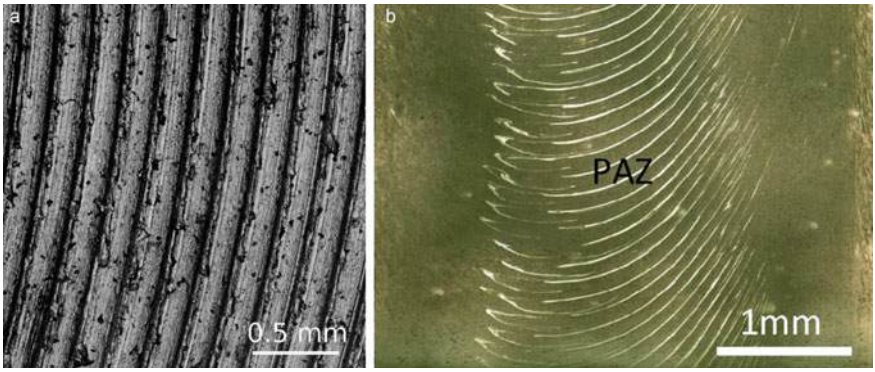


Fig. 3 Transfer layers on the front surface (**a**) and on the planar cross section in [9] (**b**)

this case, planar and cross sections demonstrate fragmented lamellae of the marking material at the boundaries of transfer layers [9]. Appearance of the marking material at the transfer layer boundary is evidence for the process of separation of materials with different densities, i.e. the vortical motion of a quasi-liquid material during processing.

However, the transfer layer thickness can be inconsistent with the given formula or can depend on conditions being neglected in it. For example, in [10] we revealed a variation in the transfer layer thickness when friction stir welding of 2024 aluminum alloy is accompanied by ultrasonic vibrations transmitted into the workpiece. Ultrasonic vibrations activate the acoustoplastic effect, which, in this context, means deformation intensification, making changes in adhesive-cohesive transfer possible. In [11] described an adhesion model based on the number of valence electrons and interatomic distance. Severe plastic deformation causes curvature of the crystal lattice and motion of electron gas [12]. Thus, any slight variation in temperature or load during friction stir welding can change transfer conditions. Under significant loads, strains, and strain rates, the material behaves differently. An example is cyclic phase transformation with intermediate amorphization [13], which can also affect the adhesive-cohesive transfer.

In [14] we performed a model experiment on ball-on-disk dry sliding friction at the ambient temperatures 25, 100, and 200 °C. Test specimens were balls made of bearing structural 100Cr6 steel 6 mm in diameter and a disk made of 5056 aluminum alloy 50 mm in diameter. The test conditions were the load acting on the ball 1 N, sliding velocity 0.5 m/min, and sliding distance 4 km. The friction force was observed to oscillate and vary in testing, which can be explained by the sticking and detaching cycles of the aluminum alloy as well as by local temperature increase. The higher was the test temperature, the greater was the friction coefficient variation. The friction coefficient decreased with sliding distance but increased on average with temperature (Fig. 4a).

The analysis of the wear surface of the steel balls showed the presence of a transfer layer in the form of separate islands of transferred material. In the longitudinal

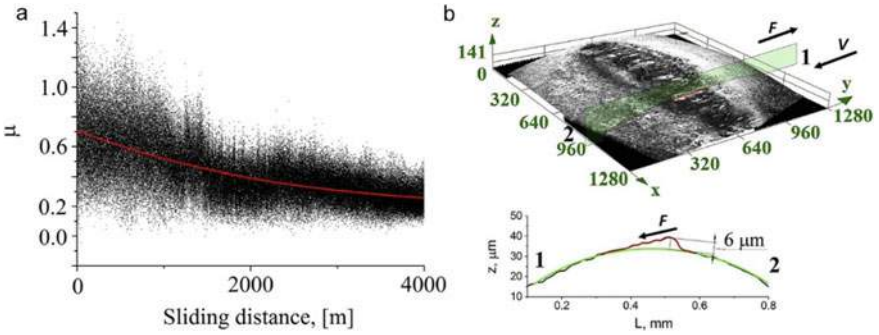


Fig. 4 Friction coefficient at sliding (**a**), and the wear scar and transfer layer on the ball bearing steel surface after sliding (**b**) [14]

section the islands had a wedge-shaped form, and the layer thickness was maximum at the trailing edge of the ball relative to its central axis (Fig. 4b). The transfer layer parameters were average thickness, maximum thickness, width of islands, and contact spot area. All the layer parameters increased with temperature. The most intensive transfer was on the specimen tested at 200 °C. The analysis of friction tracks shows the presence of aluminum particles transferred back and smoothed above the tribological layer. The mechanically stirred tribological layer had marks of plastic deformation in the form of sliding bands and curved grain boundaries. In so doing, the layer depth increased at higher test temperature. The presence of such a layer is typical for sliding friction, but bulges smoothed above it testify reverse transfer of the material from ball to disc. The analysis of the transfer layer thickness shows that the reverse process is most intensive at the ambient temperature 100 °C. At the temperature 25 °C, normal wearing with weak transfer of aluminum to the steel ball prevails, and at the temperature 200 °C direct transfer to the ball prevails. Reverse transfer is weakly pronounced in this case as in overheating the transfer layer serves as a lubricant and is uniformly distributed over the disc. In fact, these three modes take place between the tool and workpiece during friction stir processing, but underheating or overheating (with the adhesion-cohesion balance being broken) causes defects to form.

Technically, detachment of the transfer layer form the tool is as a rule incomplete during welding and should be so. After welding, there is always a certain layer of the workpiece material adherent to the tool. This occurs in welding of any materials: aluminum, titanium, copper alloys, etc. Therefore, the tool used to weld one alloy must not treat another material, unless required by the experimental condition. Moreover, a tool first time in use never demonstrates optimal processing results, even at the specified process parameters, until it is covered by a layer of the material being processed. This can be evidenced by the friction coefficient variation depending on the distance passed by the steel ball. Thus, a natural surface is formed on the tool, which provides an effective transfer of the material during friction stir welding. This layer is adherent due to diffusion, as was previously shown [15], and acts as a

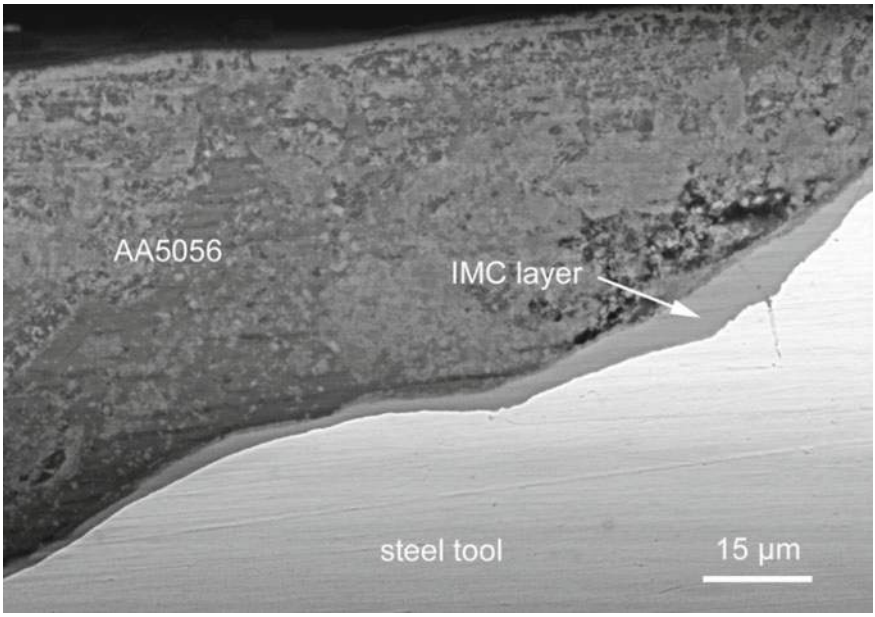


Fig. 5 Transfer layer and intermetallic compound on tool [15]

protector of the tool. In welding/processing of the aluminum alloy by the steel tool, a thin intermetallic Fe–Al layer up to 10 μm in thickness is formed at the interface due to diffusion, which is harder than the tool material (Fig. 5). The intermetallic layer is coated by the processed material, with the transfer layer sticking to it due to cohesion forces. In overheating and durable operation of the tool, diffusion of the material into the tool increases, grain-boundary diffusion occurs in the form of specific intermetallide protrusions and the tool surface fractures. Thus, the adhesive-cohesive transfer in friction stir processing is a two-step process: formation of a natural protective coating on the tool and cohesive transfer of the material by this coating.

4 Influence of Load on the Transfer Layer

In the present work, we study the dependence of the transfer layer on the axial load. Since the relation of the tool feed and rotation rates to the layer thickness is indisputable, of particular interest is load, which is rarely studied, although it can affect the heat generation and friction coefficient. The study is performed on 2024 aluminum alloy sheets 8 mm in thickness. The material is milled to remove 0.5 mm of its clad surface and then processed by the friction stir tool with the 5 mm long pin. The pin length is chosen so as not to touch the platform if the tool would sink at high loads. The processing scheme is shown in Fig. 6.

Fig. 6 Friction stir processing set up

Since the divergence of the measured transfer layer thickness from the previously calculated one is also observed at the acoustoplastic effect, friction stir processing is performed with ultrasonic assistance at the same parameters. Processing modes are given in Table 1. The ultrasonic power is 0.6 kW. Ultrasound is transmitted into the sheet at the rigid fixation at the free end. The ultrasound generator works in an adaptive mode and provides a resonant input of vibrations without attenuation throughout the workpiece length. This method is described and investigated elsewhere [16].

Polished and etched longitudinal sections cut from the center of the processed material show a typical pattern of transfer layers (Fig. 6). Due to variation of the material etchability across the thickness, the layers are of different contrast. From the front surface to the root, the layers rotate in the direction of the processing pass. Since the tool plunge depth is less than the material thickness, a layer of untreated material is observed at the root surface, with a structure corresponding to the thermomechanically affected zone and the base material. At high loads, the untreated layer becomes narrower due to a deeper penetration of the tool shoulders.

The longitudinal section clearly demonstrates that rings on the front surface correspond to the etched transfer layers, i.e. they are of the same nature, despite the fact that they are formed by the shoulders on the front surface and by the pin in the bulk. Since transfer layers on the front surface are more pronounced, they are viewed in an Olympus LEXT-OLS4100 laser scanning microscope. 3D images of friction surfaces are used to construct 2D profiles of transfer layers (Fig. 7). From the profiles it is seen that the layers are asymmetric, often with two—high and low—peaks. Asymmetry of the peaks is apparently associated with material extrusion from under the tool shoulders in the opposite direction. However, where the layer starts and how the second peak is formed are nontrivial questions that require further research. Although peak asymmetry was observed earlier, for example, by Zuo et al. [17], the morphology

Table 1 Friction stir processing modes

#	Feed rate, (mm/min)	Rotation rate (RPM)	Axial load (kg)	US	#	Feed rate (mm/min)	Rotation rate (RPM)	Axial load (kg)	US
1	90	450	2400	–	11	90	450	2400	+
2	90	450	2450	–	12	90	450	2450	+
3	90	450	2500	–	13	90	450	2500	+
4	90	450	2550	–	14	90	450	2550	+
5	90	450	2600	–	15	90	450	2600	+
6	90	450	2650	–	16	90	450	2650	+
7	90	450	2700	–	17	90	450	2700	+
8	90	450	2800	–	18	90	450	2800	+
9	90	450	2900	–	19	90	450	2900	+
10	90	450	3000	–	20	90	450	3000	+

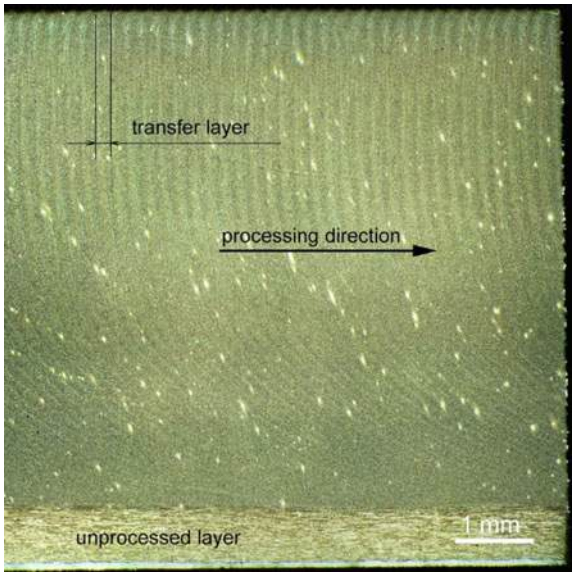


Fig. 7 Longitudinal sections of friction stir processed AA2024

remains unexplained. Evidently, the transfer layer formation is a gradual process. A layer possibly starts at the beginning of the first peak, and ends at the subsequent valley. Assuming that a layer is formed per one revolution of the tool, the layer thickness will vary from 0 to 2π (0°–360°). Thus, the shoulders gradually build up material layers with the peak formation, and the load causes them to rotate and extend. Each transfer layer actually consists of smaller layers of the material that is spread over the entire surface of the shoulders.

Pronounced peaks in the 2D profile allow an accurate construction of the load dependence of the transfer layer thickness. The measurement results are plotted in Fig. 8. The results show that the deviations of the transfer layer thickness are quite small and are within 6 μm. However, there is a clear dependence of these

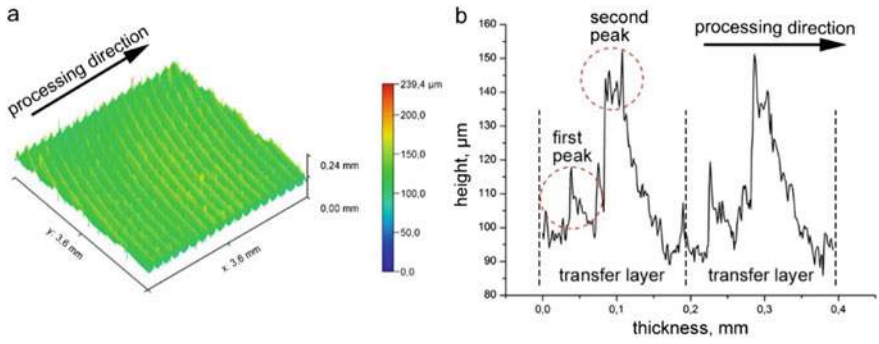


Fig. 8 Typical 3D images of the friction surface (**a**) and 2D profile of transfer layers (**b**)

deviations on the load, which was previously predicted [10]. The diagram resembles a Gaussian or parabola, and its maximum values correspond to the feed to rotation rate ratio during processing at the load 2650–2700 kg. The application of ultrasonic vibrations alters slightly the diagram, though the general pattern remains. The transfer layer thickness decreases slightly in this case (up to 3%). A decrease in the transfer layer thickness with lower or higher load relative to the maximum is associated with insufficient adhesion/cohesion force during mass transfer. At insufficient load, the adhesive-cohesive bond is weaker, and, if the load is too high, the material overheats, working as a lubricant, which reduces the reverse transfer from the tool. In terms of technology and material quality, such a small change in the transfer layer thickness will probably be of little importance, but this allows a better understanding of the fundamental processes that occur during processing/welding.

The maximum thickness of the transfer layer is observed at the load 2700 kg with and without ultrasonic treatment, which is obviously associated with the material properties. During processing, the longitudinal force is measured along the tool path. Since aluminum alloys have high thermal conductivity, this force decreases during processing due to heating of the material in front of the tool. The longitudinal force is however independent of the load and ultrasonic vibrations and amounts to 409 ± 35 kg. By relating this value to the pin projection area, we derive the longitudinal welding stress, which is equal to 99 MPa. This stress corresponds to the yield point of 2024 alloy in compression at a temperature of 450 °C [18]. A close axial stress can be obtained at the load 3000 kg related to the projected pin area (101 MPa). As mentioned above, at this load the tool shoulders become heated and sink into the workpiece. The axial stress at the load 2700 kg is 91 MPa, which is somewhat lower but still close. Thus, the largest mass transfer is observed in conditions close to the yield point of the material. A further increase in the load can lead to overheating of the material, a drop in its yield point and decrease in mass transfer.

During friction stir processing, the tool moment is also measured. The moment increases with load and ranges 15–21 Nm. Consequently, the friction force also increases. The experimentally measured moment is used to calculate the coefficient of friction of the pin by a formula similar to (2):

$$\mu = \frac{M}{F_z \overline{R_p} \sin(\alpha/2)} \tag{6}$$

where $\overline{R_p}$ is the average pin radius (4 mm), and α is the pin cone angle (30°). Geometrically, the sine of the half-cone angle is equal to the cosine of the angle between axial force and support force.

The dependence of the calculated friction coefficient on the processing time shows (Fig. 9) that an increase in the load leads to an insignificant but stable decrease in the friction coefficient, despite a rise in the moment and friction force. This means that the moment and load growth is disproportional due to changes in the friction behavior. Ultrasound impact exerts no effect on the moment and friction coefficient at any load.

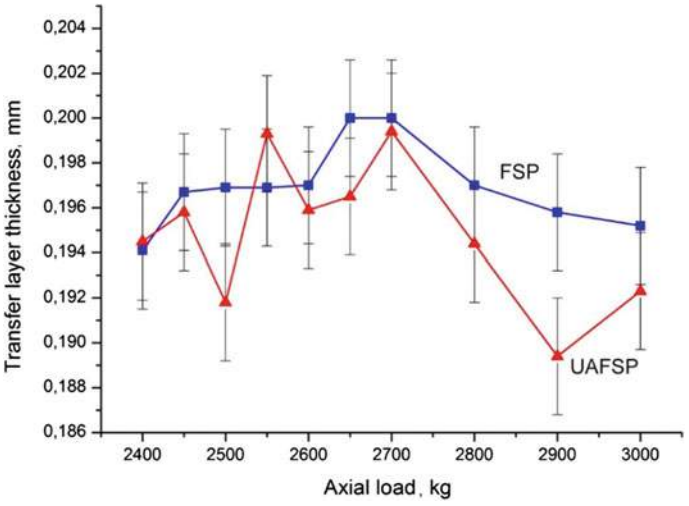


Fig. 9 The relationship between the transfer layer thickness and the axial load at friction stir processing and ultrasonic-assisted friction stir processing

This method of evaluating the friction coefficient or friction force is not quite adequate since the measured moment is a characteristic of the entire system. The evaluated value is an averaged one, but each tool region has its own moment, friction force, and friction coefficient on account of different sliding velocities at each radius. As is known, the friction coefficient depends on the speed not only in the case of viscous friction, but also during sliding. For example, in friction of a steel ball against lead or indium, the friction coefficient increases with speed, achieves a plateau, and falls again in the speed range 10^{-10} to 10 cm/s [19]. Thus, the real friction coefficient at the average pin radius can differ strongly from the evaluated one. However, the above evaluation can be used to compare technological modes at the same radius (Fig. 10).

The process temperature is measured using a FLIR 655 SC thermal imaging camera. Thermal images are taken only of the material near the tool contour, which quickly cools, so the measured temperature will be less than the real one. However, the measurement results allow an estimation of the influence of the technological mode on the heat generation. The average process temperature, from the tool penetration to its removal, is used as a factor (Table 2). An increase in the load is expected to rise the average process temperature, which is associated with a more intense heat generation. An increase in the temperature during deformation additionally plasticizes the material, resulting in a decrease in the friction coefficient.

The application of ultrasonic vibrations, on the contrary, leads to a slight decrease in temperature, which is less obvious since the ultrasonic treatment itself causes heating of the material. During friction stir welding at the slip contact of the waveguide, the material is usually heated by 10–15 °C [20]. At the rigid fixation of the

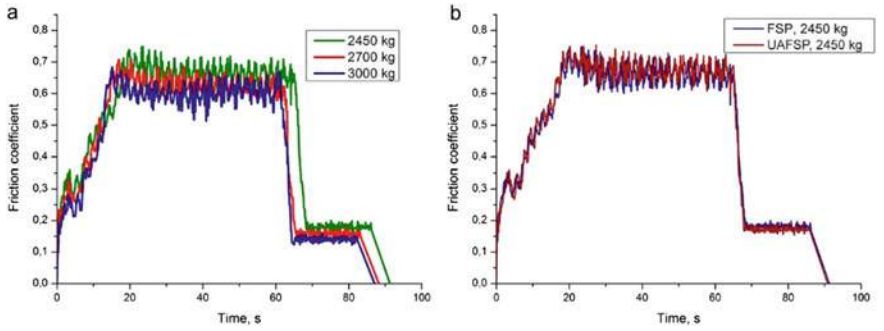


Fig. 10 The dependence of the calculated friction coefficient on the load (**a**); influence of ultrasonic impact on the calculated friction coefficient (**b**)

Table 2 The average friction stir processing temperature

#	Load (kg)	US	$\bar{t}$ (°C)
2	2450	–	216
12	2450	+	213
7	2700	–	221
17	2700	+	212
10	3000	–	227
20	3000	+	222

waveguide, the activation of the acoustoplastic effect apparently enhances the dissipation of thermal energy into the surrounding material. Nevertheless, this hardly affects the friction coefficient and the general moment.

Thus, the most effective mass transfer during friction stir processing occurs at loads that induce stresses close to the yield point of the material at the process temperature. Under these conditions, the transfer layer thickness is equal to the feed to rotation rate ratio. A reduction in the load decreases adhesion/cohesion and consequently mass transfer. An increase in the load also decreases mass transfer due to overheating of the material and reduces the friction coefficient. Intensification of deformation resulting from the acoustoplastic effect activated by ultrasound affects insignificantly the mass transfer characteristics.

5 Surface Topography and Roughness

The quality of structural components produced by friction stir welding/processing involves not only the strength characteristics of the material. An important criterion is also the surface quality of the components. The characteristic surface relief in the form of “onion rings”, which is formed behind the advancing tool and is associated

with the mass transfer process, can lead to undesirable consequences. The given part of the structure may be more prone to contamination, oxidation, corrosion, wear, etc. However, this issue has not yet been adequately addressed. The surface quality after welding/processing is evaluated only visually to identify macrodefects, such as shrinkage, tunnels, holes, oxidation, etc., because the fracture of an operating structure begins at these defects, if there are no other larger scale internal macrodefects [21]. Another defect is burr formation on both sides of the advancing tool, which reduces the cross-sectional area of the material in the stir zone. Burrs usually indicate that the welding/processing parameters were not properly selected, which leads, e.g., to overheating. However, the surface quality does not always imply a good quality of the joint in terms of strength. For example, as was shown in [22], an increase in the rotation rate during welding of AA5052 alloy led to a visually smoother surface, as well as to the formation of a tunnel defect. Thus, when selecting the optimal parameters, one should be guided by the quality criteria that are closest to the performance requirements.

The performance characteristics are affected not only by the presence of visible defects, but also by roughness. Visible defects on the surface are only a first approximation. With optimal parameters in terms of strength, these defects are usually absent, but the surface roughness is still pronounced. The roughness influences the fatigue characteristics and resistance to corrosion and wear [23]. In order to reduce the surface roughness, the process parameters can be further optimized within the range of previously selected optimal parameters or the surface can be post-processed. From a fundamental point of view, the surface topography and roughness can explain the mass transfer processes occurring in friction stir processing.

Within a certain range of process parameters, the roughness is significantly reduced with increasing rotation rate and decreasing feed rate, which is explained by a change in the transfer layer thickness according to Eq. (5). This was shown for 7075 aluminum alloy [17] and in dissimilar welding of A5052P-O aluminum and AZ31B-O magnesium alloys [24]. In [17] it was shown that the topography of the front surface is self-similar, and its fractal dimension linearly correlates with the roughness. However, the given regularity is not observed for all materials. For example, for friction stir processed 7075 aluminum alloy/CBA, WFA, CSA PKSA or CFA matrix composites, the dependence of the roughness on the rotation rate is unstable, up to directly proportional one, i.e., the larger the rotation rate, the higher the roughness [25]. With somewhat higher or lower parameters, the roughness dependence is also nonlinear.

The search for optimal processing parameters in terms of roughness was made for 2017 aluminum alloy [26]. The authors clearly showed that the dependence of the roughness on process parameters is not always linear, even for homogeneous materials. In particular, an increase in the rotation rate at a low feed rate can lead to material overheating and numerous overlaps, which increases the roughness. The axial load, which has not been previously investigated in the given context, also has a nonlinear effect on the roughness. For example, with a rotation rate of 900 rpm and a feed rate of 50 mm/min, an increase in the load from 500 to 1500 N led to an increase in the roughness, which obviously resulted from overheating due to large heat input.

However, with other rotation rate and feed rate values, the roughness decreased with increasing load. These results indicate that various mechanisms are involved in the surface relief formation, and a linear dependence is observed only in special cases. In this regard, an important factor in addition to the process parameters is the thermal conductivity of the material and the amount of heat generated by friction.

Note that not all materials exhibit a well-defined morphology in the form of the onion ring structure. For example, this kind of structure was not observed in friction stir processed commercially pure titanium [27]. Judging by the relief topography, mass transfer on the surface was extremely unstable. The roughness was reduced with increasing load and the surface was generally smoothed, but unlike more ductile materials the surface demonstrated overlaps and tear.

In the present work, the surface roughness of friction stir processed 2024 aluminum alloy was examined using a laser scanning microscope Olympus LEXT-OLS4100. The processing parameters are given in Table 1. Since the front surface is undulated and consists of rings (transfer layers), the roughness was analyzed for two cases: with and without subtracting the undulation (Fig. 11). In the case of subtracted undulation, the roughness is obviously lower, because it is measured with respect to a curved surface. In both cases, the load dependence of the roughness approximately resembles the load dependence of the transfer layer thickness with a maximum at 2700 kg. In general, this is consistent with the results obtained in [26]. An increase in the axial load enhances the extrusion of the material behind the tool and hence the roughness increases, but above a threshold load value it decreases. As noted earlier, the yield point of the material is reached at a 3000 kg load, which may explain surface smoothing. Without subtracting the undulation, the roughness values changed drastically within the range of 7–21 μm. The roughness may also increase due to increasing plunge depth of the tool.

The application of ultrasonic vibrations destabilized the load-roughness dependence, which points to a less uniform dependence of the transfer layer thickness. This behavior is observed only when the undulation is subtracted. Without subtracting the undulation, the roughness increases with increasing load. The above behavior of the

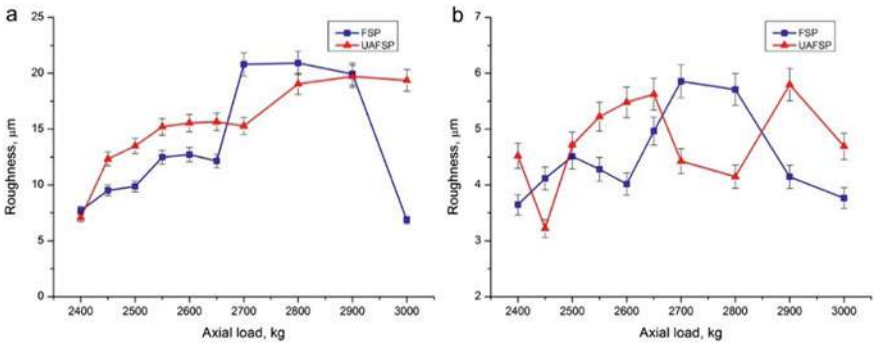


Fig. 11 Roughness of front surface without (**a**) and with subtracting the undulation (**b**)

curve generally indicates that it is inappropriate to control the roughness through the load, because its maximum is achieved at an optimal load from the viewpoint of adhesive-cohesive mass transfer in friction stir processing. But it is precisely at this load that the surface roughness decreases as a result of ultrasound application. It is currently unclear whether the result is occasional or regular, but ultrasonic assistance can be considered as a way to reduce roughness.

At present, if necessary, the surface of the component in the stir zone is smoothed by mechanical post-processing, which is considered technologically inefficient in modern production. The use of more advanced post-processing techniques does not always provide the desired result. For example, as shown in [23], laser peening and shot peening did not significantly reduce the surface roughness of the joint. The friction characteristics of various samples differed slightly, although laser peening made the weld surface less stepped. Obviously, it is not always possible to achieve a good surface quality by varying the process parameters within the admissible parameter range; the quality often depends on the material. That is why the friction stir processing technology is being improved. The surface quality can be improved by using a stationary shoulder that allows for a smoother surface [28]. Drawbacks of this approach are that when the surface is smoothed immediately behind the tool it is not always possible to visually inspect surface defects.

6 Conclusion

The friction stir welding and processing technology has greatly evolved over almost 30 years of its development. Nevertheless, the fundamental explanation of the processes can be found in the works of 15 years ago. The then proposed scientific foundations in terms of thermomechanical processes generally well explain changes in the mechanical properties and microstructure, but they do not fully predict the behavior of materials. This is due to the nonlinear dynamics of the friction stir processes. In particular, deformation and heat generation are interdependent, which is often not taken into account in analytical solutions. Moreover, the material behaves differently at such large strains and strain rates in comparison with ordinary deformation. The melting temperature rises under high loading conditions, and other material properties, e.g., fluidity, change. Thus, the process dynamics implicitly depends on process parameters. These problems are approximately solved by modeling methods, without complete understanding of the friction stir processes as they are based on adhesive-cohesive mass transfer.

The complex mass transfer pattern is due to a combination of temperature, load, friction coefficient, material properties and sliding velocity, which are also largely interdependent. Here we showed that the load increase at a constant feed rate and tool rotation rate first causes an increase in the transfer layer thickness and then a decrease. This cannot be fully explained by the change in the coefficient of friction and temperature, because the average moment and temperature increase while the friction coefficient decreases. With increasing load, the adhesive-cohesive force between the

tool and the surrounding material increases, resulting in larger material transfer. The maximum thickness of the transfer layer is observed at a stress close to the yield point of the material. A further increase in the load overheats the material and impairs mass transfer. The application of ultrasonic vibrations during friction stir processing, which causes the acoustoplastic effect and amplifies the deformation but does not increase the temperature and does not change the friction coefficient, leads to a decrease in the transfer layer thickness. This is evidently due to a change in the path length of the transferred material. Thus, the mass transfer in friction stir processing depends on the fundamental physical processes of bond and structure formation in the material at a fine level. The same effects influence the surface roughness of the processed material. In general, the load dependence of the roughness resembles that of the transfer layer thickness. An increase in the load enhances the extrusion of the transfer layer material from under the tool, but in overheating the shoulders smooth the surface. The roughness changes only slightly with ultrasound application, but it was found that at the optimal load in terms of the transfer layer thickness the ultrasonic treatment reduces the roughness.

These issues concerning friction stir processing/welding are rarely discussed as they are not of acute practical interest. However, their investigation can provide better fundamental understanding of mass transfer mechanisms. Presumably, the answers can be obtained from the studies of the transfer layer structure. Of particular interest is the asymmetry of the material asperities on the front surface as well as the different etchability of aluminum alloys over the layer thickness, which will be discussed in subsequent papers.

Acknowledgements The reported study was funded by RFBR and Tomsk Oblast according to Research Project No. 19-42-700002 (the results are reported in Sections 4–5) and performed within the frame of the Fundamental Research Program of the State Academies of Sciences for 2013–2020, line of research III.23.2.4 (the results in Sections 1–3).

References

1. Schmidt HNB, Hattel JH, Wert J (2004) An analytical model for the heat generation in friction stir welding. Modell Simul Mater Sci Eng 12(1):143–157
2. Mishra RS, De PS, Kumar N (2014) Friction stir welding and processing: science and engineering. Springer, Basel
3. Frigaard Ø, Grong Ø, Midling OT (2001) A process model for friction stir welding of age hardening aluminum alloys. Metall Mater Trans A 32(5):1189–1200
4. Colligan KJ, Mishra RS (2008) A conceptual model for the process variables related to heat generation in friction stir welding of aluminum. Scripta Mater 58:327–331
5. Kumar K, Kalyan C, Kailas SV, Srivatsan TS (2009) An investigation of friction during friction stir welding of metallic materials. Mater Manuf Process 24(4):438–445
6. Meyghani B, Awang MB, Poshteh RGM, Momeni M, Kakooei S, Hamdi Z (2019) The effect of friction coefficient in thermal analysis of friction stir welding (FSW). In: IOP Conference series: materials science and engineering, vol 495, Art. 012102. Science and engineering (CUTSE) international conference, 26–28 Nov 2018, Sarawak, Malaysia

7. Long T, Tang W, Reynolds AP (2007) Process response parameter relationships in aluminium alloy friction stir welds. J Sci Technol Weld Joining 12(4):311–317
8. Schneider JA, Nunes AC Jr (2004) Characterization of plastic flow and resulting microtextures in a friction stir weld. Metall Mater Trans B 35(4):777–783
9. Liu XC, Wu CS (2015) Material flow in ultrasonic vibration enhanced friction stir welding. J Mater Process Technol 225:32–44
10. Eliseev AA, Kalashnikova TA, Gurianov DA, Rubtsov VE, Ivanov AN, Kolubaev EA (2019) Ultrasonic assisted second phase transformations under severe plastic deformation in friction stir welding of AA2024. Mater Commun 21:100660. https://doi.org/10.1016/j.mtcomm.2019.100660
11. Wills JM, Harrison WA (1984) Further studies on interionic interactions in simple metals and transition metals. Phys Rev B 29:5486–5490
12. Panin VE, Surikova NS, Lider AM, Bordulev YuS, Ovechkin BB, Khayrullin RR, Vlasov IV (2018) Multiscale mechanism of fatigue fracture of Ti-6A1-4V titanium alloy within the mesomechanical space-time-energy approach. Phys Mesomech 21(5):452–463
13. Glezer M, Metlov LS (2010) Physics of megaplastic (severe) deformation in solids. Phys Solid State 52(6):1162–1169
14. Tarasov SY, Filippov AV, Kolubaev EA, Kalashnikova TA (2017) Adhesion transfer in sliding a steel ball against an aluminum alloy. Tribol Int 115:191–198. https://doi.org/10.1016/j.triboint.2017.05.039
15. Tarasov SY, Kalashnikova TA, Kalashnikov KN, Rubtsov VE, Eliseev AA, Kolubaev EA (2015) Diffusion-controlled wear of steel friction stir welding tools used on aluminum alloys. AIP Conf Proc 1683:020228. https://doi.org/10.1063/1.4932918
16. Tarasov SY, Rubtsov VE, Fortuna SV, Eliseev AA, Chumaevsky AV, Kalashnikova TA, Kolubaev EA (2017) Ultrasonic-assisted aging in friction stir welding on Al-Cu-Li-Mg aluminum alloy. Weld World 61(4):679–690
17. Zuo L, Zuo D, Zhu Y, Wang H (2018) Effect of process parameters on surface topography of friction stir welding. Int J Adv Manuf Technol 98:1807–1816
18. Seidt JD, Gilat A (2013) Plastic deformation of 2024–T351 aluminum plate over a wide range of loading conditions. Int J Solids Struct 50:1781–1790. https://doi.org/10.1016/j.ijsolstr.2013.02.006
19. Burwell JT, Rabinowicz E (1953) The nature of the coefficient of friction. J Appl Phys 24:136–139. https://doi.org/10.1063/1.1721227
20. Shi L, Wu CS, Sun Z (2018) An integrated model for analysing the effects of ultrasonic vibration on tool torque and thermal processes in friction stir welding. Sci Technol Weld Joining 23(5):365–379. https://doi.org/10.1080/13621718.2017.1399545
21. Derazkola HA, Aval HJ, Elyasi M (2015) Analysis of process parameters effects on dissimilar friction stir welding of AA1100 and A441 AISI steel. Sci Technol Weld Joining 20(7):553–562. https://doi.org/10.1179/1362171815Y.0000000038
22. Moshwan R, Yusof F, Hassan MA, Rahmat SM (2015) Effect of tool rotational speed on force generation, microstructure and mechanical properties of friction stir welded Al–Mg–Cr–Mn (AA 5052-O) alloy. Mater Des 66:118–128. https://doi.org/10.1016/j.matdes.2014.10.043
23. Hatamleh O, Smith J, Cohen D, Bradley R (2009) Surface roughness and friction coefficient in peened friction stir welded 2195 aluminum alloy. Appl Surf Sci 255(16):7414–7426. https://doi.org/10.1016/j.apsusc.2009.04.011
24. Shigematsu I, Kwon YJ, Saito N (2009) Dissimilar friction stir welding for tailor-welded blanks of aluminum and magnesium alloys. Mater Trans 50(1):197–203. https://doi.org/10.2320/matertrans.MER2008326
25. Ikumapayi OM, Akinlabi ET (2019) Experimental data on surface roughness and force feedback analysis in friction stir processed AA7075 – T651 aluminium metal composites. Data Brief 23:103710
26. Langlade C, Roman A, Schlegel D, Gete E, Noel P, Folea M (2017) Influence of friction stir process parameters on surface quality of aluminum alloy A2017. In: MATEC web of conferences, vol 94, Art. 02006. https://doi.org/10.1051/matecconf/20179402006

27. Eliseev AA, Amirov AI, Filippov AV (2019) Influence of axial force on the pure titanium surface relief during friction stir processing. AIP Conf Proc 2167(1):020077
28. Zhou Z, Yue Y, Ji S, Li Z, Zhang L (2017) Effect of rotating speed on joint morphology and lap shear properties of stationary shoulder friction stir lap welded 6061–T6 aluminum alloy. Int J Adv Manuf Technol 88:2135–2141

Nanomaterials Interaction with Cell Membranes: Computer Simulation Studies

Alexey A. Tsukanov and Olga Vasiljeva

Abstract This chapter provides a brief review of computer simulation studies on the interaction of nanomaterials with biomembranes. The interest in this area is governed by the variety of possible biomedical applications of nanoparticles and nanomaterials as well as by the importance of understanding their possible cytotoxicity. Molecular dynamics is a flexible and versatile computer simulation tool, which allows us to research the molecular level mechanisms of nanomaterials interaction with cell or bacterial membrane, predicting in silico their behavior and estimating physicochemical properties. In particular, based on the molecular dynamics simulations, a bio-action mechanism of two-dimensional aluminum hydroxide nanostructures, termed aloohene, was discovered by the research team led by Professor S. G. Psakhie, accounting for its anticancer and antimicrobial properties. Here we review three groups of nanomaterials (NMs) based on their structure: nanoparticles (globular, non-elongated), (quasi)one-dimensional NMs (nanotube, nanofiber, nanorod) and two-dimensional NMs (nanosheet, nanolayer, nanocoated substrate). Analysis of the available in silico studies, thus can enable us a better understanding of how the geometry and surface properties of NMs govern the mechanisms of their interaction with cell or bacterial membranes.

Keywords Nanoparticle · Nanotube · Nanosheet · Lipid membrane · Molecular dynamics

A. A. Tsukanov
Skolkovo Institute of Science and Technology, Moscow 121205, Russia

A. A. Tsukanov (✉) · O. Vasiljeva
Institute of Strength Physics and Materials Science SB RAS, Tomsk 634055, Russia
e-mail: a.a.tsukanov@yandex.ru

O. Vasiljeva
e-mail: olga.vasiljeva@ijs.si

O. Vasiljeva
Jozef Stefan Institute, 1000 Ljubljana, Slovenia

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_9

1 Introduction

Over the past decade, the variety of different nanoparticles (NPs) and nanomaterials (NMs) were considered an important contributors to multiple medical diagnostics and the therapy applications [1, 2]. Nanomaterials can be used as devices [3], contrast drug carriers, drug delivery systems [4, 5], adjuvant, therapeutic and theranostic agents [6–8]. All these applications need an understanding of how the nanomaterial interacts with the membranes of the cell and intracellular organelles. Before the consideration of a NM-biomembrane interaction, it is important to describe what the typical cell membrane is.

Amphiphilic organic molecules, lipids, are the basic building units of the typical cell membrane [9]. Lipid consists of two parts: long hydrophobic tails and a comparatively compact hydrophilic head. Due to their composition, lipids form stable structures in water solution: e.g. liposomes, spherical and cylindrical micelles, bilayer membrane, etc. Lipid bilayer is stabilized by hydrophobic interaction between lipid tails in the inner part of the membrane, as well as by the interaction of head groups with water and with each other at the surface regions [10]. Lipid bilayer is a basis of the cell membrane.

The rapid development of multiple techniques, as well as growth and availability of high-performance computers, contribute to an increasing number of computer simulation studies of complex molecular systems, which may include sophisticated biological objects and complex nanomaterials. There are several computational approaches to study the NM-biomembrane interaction: a self-consistent field and density functional theory approach [11, 12], stochastic-elastic modeling [13, 14], all-atom [15] and coarse-grained [16] molecular dynamics, including classical unbiased and constrained or steered molecular dynamics (SMD) [17], Monte Carlo methods [18], etc.

It is worth mentioning several previously published reviews concerning the interaction of molecules and NMs with biomembrane models. Small compounds, drug molecules, biomolecules, and fullerenes interaction with the cell membranes was considered in [19, 20]. In particular, cases of fullerenes, their aggregates and derivatives impact to the membranes were summarized as, pure fullerenes C_{60} and their clusters tend to penetrate inside the lipid bilayer accumulating in the hydrophobic interior of the membrane (membrane width is several times larger than the diameter of C_{60} fullerene) [21–23]. It was noted that larger fullerenes or fullerenes in high concentrations can cause significant disturbances in the membrane structure [24–26]. However, accordingly to [27] the presence of C_{60}–C_{180} fullerenes inside the membrane with fullerene-to-lipid ratio about ~1:1 results in a mechanical strengthening of the lipid bilayer. It was also reported, that the pure C_{60}, C_{70} fullerenes can bind with ion channels, embedded into the lipid membrane, thereby affecting their structure or function [28, 29]. Fullerenes with a functionalized surface exhibit tendency to anchor their polar or charged groups either in a lipid–water interface [21, 23] or in hydrophilic-hydrophobic interface between lipid heads and tails parts [26, 30]. The review of molecular dynamics studies of small compounds permeation

through the lipid bilayer as well as the interaction of proteins with the membranes can be found in [31]. Simulations of the impact of different carbon NMs as fullerenes, nanotubes, and combustion-generated carbon NPs, having an arbitrary structure, on the cell membrane are reviewed in [32]. In-depth review describing the computer modeling studies of nanomaterials interaction with the cell membranes and other different biological nano-objects can be also found in a work [33].

2 Nanoparticles

In this section, we consider papers focused on the numerical simulations of the interaction of lipid membranes with organic and inorganic nanoparticles such as, dendrimers, functionalized gold NPs, bimetallic NPs of immiscible metals, Janus NPs and abstractive nanoparticles with a surface charge.

2.1 Dendrimers and Dendritic Nanostructures

Dendrimers and dendritic nanostructures attract much attention in biochemistry, nanotechnology, and pharmaceutical sciences, due to the possibility to precise control of its size, shape, and location of functional groups [34, 35]. Dendrimers play an important role in biomedicine, as contrast agents, gene-transfection agents, and antibacterial substrates [36]. The number of branches and the size of dendrimer of a certain kind depends on its generation number G. A schematic of poly(amidoamine) dendrimers (PAMAM) G2 and a fragment of G11 dendrimers in the vicinity of the cell membrane is shown in Fig. 1. The 3D model of PAMAM G11 was built using the structural data from [37].

Lee and Larson using MD simulations showed that charged PAMAM, interacting with a cell, can induce pore formation in the lipid membrane. Herewith, this was observed at a temperature of 310 K, while at lower $T = 277$ K the described phenomenon did not occur, which is explained by transition of lipids into a condensed phase [38]. Moreover, it was also found that at a high ion concentration in water (about 0.5 M of NaCl) large charged dendrimers demonstrate no tendency to penetrate into the bilayer because of the screening of electrostatic interaction between dendrimer surface and lipid head groups. A simulation of PAMAM dendrimers and several copies of a peptide—poly-L-lysine (PLL) near the membrane showed that the ability to disturb the membrane is dependent on the generation number of the dendrimer, and hence on its size, as well as on the dendrimers concentrations near the membrane surface [39, 40]. Large dendrimers in a high concentration are capable to induce significant distortion of the bilayer structure, as well as the formation and stabilization of pores in the membrane. Kelly and co-workers performed MD simulations of neutral, positive, and negative dendrimers with the membrane in implicit water model [41]. It was shown that charged PAMAM dendrimers of third generation (G3)

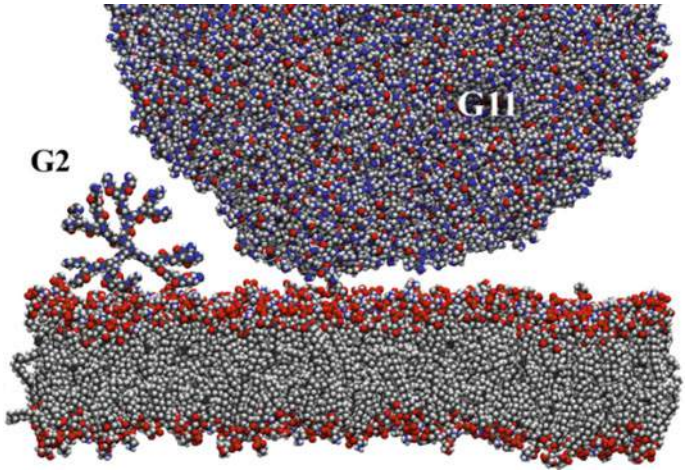


Fig. 1 Schematic of PAMAM dendrimers with generation numbers G2 and G11 near a lipid bilayer. Colors: carbon—grey, hydrogen—white, oxygen—red, nitrogen—blue, phosphorus—brown

more intensively interact with the membrane surface than neutral dendrimers. The interaction of charged dendrimers with the membrane in a liquid-phase state may also be accompanied by a hydrophobic interaction between the dendrimer interior and the lipid tail groups [42]. The permeation of the charged dendrimers through the lipid membrane under an elastic tension was studied using the coarse-grained MD simulations in [43]. The obtained results showed that the elastic tension enhances the permeability of the membrane for the charged PAMAM dendrimers of G3, G5 and G7 generation. Ting and Wang, using a self-consistent field theory, showed that in the absence of the lateral tension membrane prefer to wrap the charged dendrimer partially [44]. An increase in the dendrimer charge density increases the extent and stability of the wrapped state. But in a case of slightly tensioned membrane large charged nanoparticles as G5 (or higher) dendrimers can induce the formation of metastable pores in a lipid bilayer whereas for G3 dendrimers the pores are unstable.

2.2 Abstractive Nanoparticles

An abstractive NP means a model particle, which is not representing a certain material with a defined chemical composition, but mimics the NP with defined properties such as shape, size, charge, hydrophilic-hydrophobic balance etc.

In order to find out how the size and charge of nanoparticles affect their interaction with the cell membrane, Ginzburg and Balijepalli [45] used the self-consistent field/density functional theory of block copolymer/nanoparticle mixtures proposed in [11, 12]. The results revealed that a neutral nanoparticle of a diameter 16–32 Å penetrates into the bilayer, forming a single-layer hybrid micelle in which lipids orient

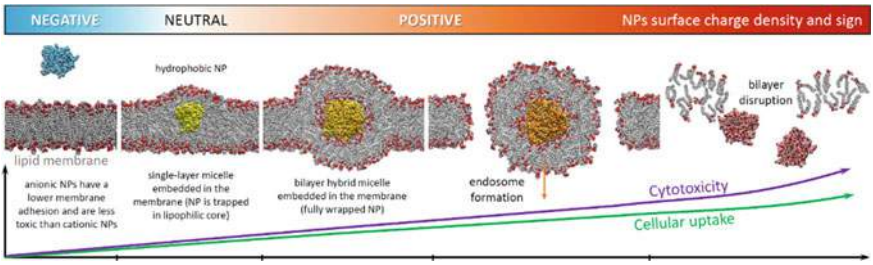


Fig. 2 A generalized scheme of the interaction of charged and neutral NPs with a cell membrane, combining the results from [45] and [49]. Anionic NP is less likely to translocate across the cell membrane than cationic NP. Hydrophobic NPs can integrate into the membrane, forming a single-layer embedded micelle. Cationic NPs having high enough surface charge density intensively participate in endosomes formation and cellular uptake. Strongly charged cationic NPs may generate pores in the bilayer, causing the overt membrane disruption

their hydrophobic tails toward the nanoparticle (see Fig. 2). If the NP charge density increases, a hybrid bilayer micelle is formed in the membrane. Further increase in the charge density or/and size of NP causes separation of the endosome, containing the nanoparticle, which may be followed by the loss of membrane integrity (Fig. 2).

The interaction of hydrophobic and semihydrophilic NPs with the membranes was studied in [46]. Using the coarse-grained MD simulations it was shown that a hydrophobic nanoparticle with a radius of about 5 nm can easily penetrate into a lipophilic bilayer interior, while a semihydrophilic NPs remain on the lipid membrane surface. An amphiphilic nanoparticle with Janus structure, consisting of both a hydrophilic and a hydrophobic part, can exhibit more complex behavior. A coarse-grained MD study of the interaction of a Janus nanoparticles with a cell membrane were conducted by Alexeev et al. [47]. The considered Janus NPs have a size comparable to the bilayer thickness. The capability of such nanoparticles to stabilize pores in the lipid membrane was found. Stabilized nanopores can be opened at relatively low elastic stress applied to the membrane, thereby providing permeability of membranes to water, ions and other compounds. Thus, using Janus NPs, it is possible to control membrane permeability altering its tension, which may be caused by changing local environmental conditions such as temperature or pH.

Besides the physicochemical properties of the NP surface, the shape and local curvature of the surface play an important role in NP interaction with a cell membrane. Yang and Ma investigated the translocation of NP with different shapes and sizes using coarse-grained model [48]. It was found that the shape anisotropy and the initial orientation of a particle are decisive for the character of its interaction with the lipid membrane. The penetrability of NP through a lipid bilayer is determined by the contact area between the particle and bilayer, and by the local surface curvature of the particle at the point of contact. The increase in curvature facilitates the translocation. It was also reported that NP volume indirectly affects penetration to a lesser extent.

The influence of the elastic lateral tension on the character of membrane interaction with abstractive dendrimer-like soft nanoparticles, having different generation

number is investigated in [50]. It was found that the way, in which the membrane interacts with soft particles, depends on both the value and the sign of membrane surface tension. The researchers defined three typical phases of interaction of a soft NP and the membrane: penetration at high positive tension, penetration and partial wrapping at low positive tension, and full wrapping at low negative tension.

In addition, the interaction of multi-molecular complex comprising a membrane-soluble outer shell and nanoporous core with the cell membrane can be considered in this subsection. Studying the membrane-NP interaction, Carr et al. proposed a possible route for forming a synthetic ion channel in the cell membrane by embedding a supramolecular complex of dimethyldioctadecylammonium (DODA) surfactant capsule with a porous polyoxomolybdate (POM) particle into the lipid bilayer [51]. The POM nanoparticle has a strong negative charge (−72 e) that makes their embedment in the membrane impossible (Fig. 2). Carr and colleagues using coarse-grained MD showed that the mixed capsule of amphiphilic cations DODA and POPC lipids, formed around the POM nanoparticle, facilitates the NP embedment into the membrane. The positive charge of the detergent liposomal structure partially screens the negative POM charge, and the further liposome fusion with the lipid bilayer leads to the embedding and stabilization of NP in the membrane center.

2.3 Metallic Nanoparticles

Gold nanoparticles (AuNP) have a wide range of possible biomedical applications [52] that explains the great attention paid to studies of the interaction of AuNP with the membrane, including in silico approaches. The interaction of functionalized gold nanoparticles with electroneutral and negatively charged lipid membranes was investigated in [49]. The considered AuNPs were functionalized with charged and/or hydrophobic ligands. Cationic ammonium groups and anionic carboxylate groups were used to provide AuNP with a positive or negative surface charge, respectively. Different ratios of charged/hydrophobic ligands coated gold core to make different surface charge densities of the NPs. The results of coarse-grained computer simulations showed that the AuNP may spontaneously adhere to the membrane surface or penetrate into the bilayer. The way they interact with membrane depends on both sign and density of NPs surface charge (Fig. 2). Using SMD simulations it was found that the approach of anionic AuNP to the negatively charged membrane is complicated by electrostatic repulsion. In both cases of the neutral and negative membranes, anionic NPs have free energy minima near bilayer surface in the adsorbed state. The free energy profile for gold NPs with hydrophobic ligands only reach minimal values inside a lipophilic bilayer region. In case of neutral lipid membrane, the energetically more preferable configuration for cationic AuNP (with 70% of charged ligands) as well as for anionic one is the adsorbed state without any significant bilayer distortions. Interaction of AuNP, having cationic ligands, with the negatively charged membrane is more intensive, strong binding, and immersion of NP were observed, which cause a large deformation of bilayer as well as a formation of hydrated region within the

membrane. The free energy profile has a minimum value corresponding to NP position inside the lipid head groups region. Increasing the surface charge density of positively charged AuNP enhances the ability of such particles for cellular uptake; however, beginning with a certain surface charge, severe deformation and disintegration of the cell membrane take place (Fig. 2). It was pointed out that there is a range of surface charges of cationic NPs in which a balance between cellular uptake and cytotoxicity may be attained [49].

The influence of the shape and surface functionalization of gold nanoparticles on the character of its interaction with a negatively charged membrane was investigated in [53]. Based on the coarse-grained simulations, the estimates of the free energy barriers and translocation rate constants were obtained depending on the nanoparticle shape and charge density. It was shown that anionic NPs were electrostatically repelled from the membrane surface and their translocation through the bilayer is less probable in comparison with cationic NPs. Furthermore, shape anisotropy may result in the reorientation rotations of the charged NPs in the contact region with membrane, thus distorting the lipids self-assembly and possibly causing cytotoxic effect. For the studied cases, it was also found that translocation rate constants may differ 60 orders of magnitude, in spite of equal sizes of NPs [53].

It is important that both the sign and the density of NP surface charge can be tuned by grafting of the ligands with certain charge, polarity and hydrophilic-hydrophobic balance. Recent theoretical research of Professor S. G. Psakhie's scientific group, devoted to bimetallic nanoparticles of immiscible or partially miscible metals and their interaction with bacterial and cell membranes, showed that the electrostatic properties of NP surface can be tuned without grafting of the charged ligands [54]. Using the embedded atom method (EAM) [55, 56] and MD simulations it was shown that the surface of bimetallic Ag–Cu NP is formed by silver atoms independent on nanoparticle composition ($Ag_{70}Cu_{30}$ or $Cu_{70}Ag_{30}$) due to lower surface energy of Ag (Fig. 3a). Furthermore, due to different electronegativity of Ag and Cu atoms, silver and copper, in average, will have different partial charges. Using the density functional theory (DFT) [57, 58], the mean values of partial atomic charges of Ag and Cu metals in small Ag–Cu clusters were estimated as a function of Ag-to-Cu ratio in the nanocluster. Based on this result it was concluded that magnitude of the surface charge density of considered bimetallic Ag–Cu NP is adjustable in the range from -2.9 to -7.3 e/nm^2, by the changing of Ag-to-Cu ratio from 7/3 to 3/7. Moreover, choosing metals by taking into account both the difference in their electronegativity and the difference in their surface energies, it is possible to synthesize bimetallic NP having a given permanent surface charge density of a certain sign [54]. In addition, using the force field based SMD simulations, in which metal atoms are uncharged, the free energy of interaction of $Ag_{70}Cu_{30}$, $Cu_{70}Ag_{30}$ and $Cu_{70}Ag_{30}O_4$ nanoparticles of a diameter about 40 Å with bacterial and cell membranes were estimated. It was found, that lipopolysaccharide (LPS-DPPE) membrane adsorbs the bimetallic NPs on the membrane-water interface (Fig. 3b), whereas lipid bilayer membrane demonstrates the tendency to wrap (or partially wrap) pure $Ag_{70}Cu_{30}$, $Cu_{70}Ag_{30}$ NPs (Fig. 3c). In the case of $Cu_{70}Ag_{30}O_4$ NP, the presence of surface oxide groups prevent the wrapping of the NP by the lipid membrane.

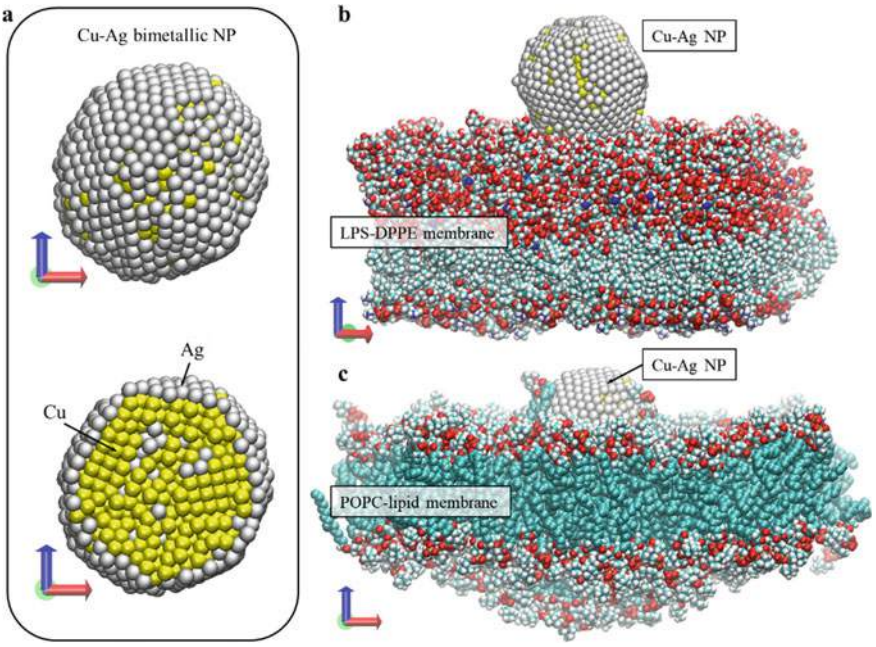


Fig. 3 Bimetallic NPs of immiscible or partially miscible metals: **a** structure of $Cu_{70}Ag_{30}$ NP, silver as a metal with lower surface energy forms a surface of NP; **b**, **c** interaction of Cu–Ag NP with lipopolysaccharide (LPS-DPPE) membrane and POPC-bilayer, respectively. The images were available using the 3D data from work [54]. Colors: Ag—grey, Cu—yellow, C—cyan, O—red, H—white, N—blue (water is not shown)

3 One-Dimensional Nanomaterials

3.1 Carbon Nanotubes

Nanomaterials with a structure as nanotube (NT), nanofiber, nanowire, nanorod, etc. belong to the class of (quasi)one-dimensional NMs (1D-NMs). The carbon nanotubes (CNT) occupies a special place among 1D-NMs and remains for a long time in the focus of researchers attention. CNTs are promising structures for the biotechnology and biomedicine, due to their biocompatibility, high stability, mechanical elasticity, thermal, electrical, and optic properties as well as the adaptability for chemical modification with bioactive compounds [59]. The single-walled carbon nanotubes (SWNTs) structure can be open or can be terminated with caps (hemispheres of a fullerene). In the former case, NT has a nanocapsule-like structure with the internal cavity isolated from the environment. Ligands in functionalized CNTs can be chemically bound with the cylindrical surface or ends. Functional groups can also be bound with CNT by non-covalent bonds. Nanotubes can be wrapped with surfactants, proteins, lipids, DNA, etc., forming the supramolecular complexes

due to the Van der Waals forces, hydrophobic interaction, and electrostatic polarization. Depending on chirality, SWNTs can exhibit metallic or semiconductor properties. Electronic properties of the SWNTs provide a formation of stable DNA-CNT nanohybrids [60, 61]. It is important to note, that typically in the force field-based MD simulation the polarizability of CNT is not taken into account, and carbon atoms are modeled as Lennard-Jones particles with zero partial charges.

One of the pioneering MD studies of the interaction of CNT with the membrane is reported in [62]. The results showed that open carbon SWNT of 13 Å in a diameter with hydrophilic groups at the ends, due to the formation of salt bridges with lipid head groups, is capable of the spontaneous embedment into the membrane, forming a transmembrane channel permeable for water molecules. Two phases of a transmembrane channel formation were distinguished: (1) location and embedment of the NT into the membrane surface, and (2) turn to a transmembrane position. As a hydrophilic NT end moved through the membrane, lipids attached from the nearest membrane monolayer were transferred to the opposite one. The example of a transmembrane position of a zigzag-type SWNT with chirality indices (25, 0) and with ends functionalized by OH-groups is shown schematically in Fig. 4 (left CNT).

The formation and transport properties of CNT-based transmembrane channel were investigated by Zimmerli and Koumoutsakos, using the all-atom MD simulations [63]. The results showed a possibility of electrophoretic transport of short RNA segments (20 adenosine nucleotides) through a synthetic channel based on transmembrane SWNT with a diameter of 18.7 Å. It was found that an electrostatic potential difference of about 1–2 V maintains the RNA fragment translocation with a velocity of about 1–30 nucleotides per nanosecond.

The study of interaction of non-functionalized NTs, having different length, with the membrane showed that short NTs, being inside the lipid bilayer, prefer to be oriented parallel to lipid molecules [64]. An increase in the length of NT change the preferred orientation to parallel to the membrane plane. The example of such

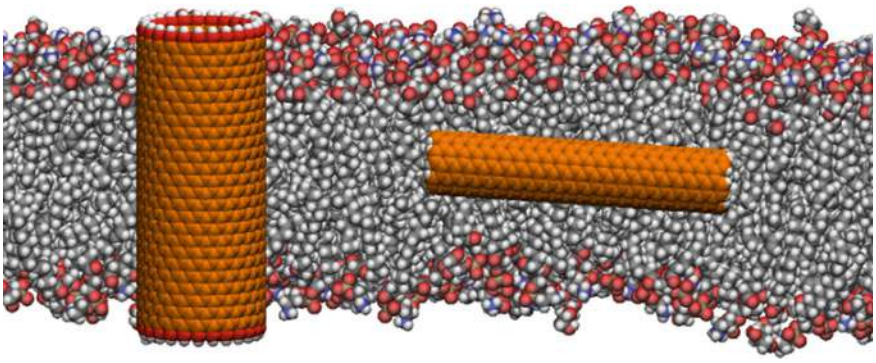


Fig. 4 Schematic of a transmembrane SWNT with functionalized faces (at the left) and hydrophobic thin CNT inside the bilayer. Color legend: carbon of CNT—orange, C (lipids)—grey, H—white, O—red, N—blue

an orientation of armchair-type SWNT with chirality (6, 6) in the bilayer center is shown schematically in Fig. 4 (right CNT).

Based on a coarse-grained MD model, two ways of CNT penetration through the membrane were revealed: by wrapping with lipids and by direct piercing [65]. It is reported that the key factor, which defines the way of penetration, is the diameter of NT. Nanotubes of a rather small diameter pierce membranes, whereas those of larger diameter penetrate through the bilayer via wrapping. The wrapping of a thinner NT is accompanied by considerable bending of the membrane, which is less energetically favorable than the separation of lipids to a comparatively small distance. For a NT of diameter comparable with the membrane thickness, the distance to which lipids need to be displaced for piercing is significant, however, the membrane curvature would be low if membrane wraps the NT. This explains why penetration via wrapping become more energetically preferable than piercing.

According to the computation results [66] based on a single chain mean field theory (SCMF) [67], the energy required for penetration of a perpendicularly oriented NT with a hydrophilic surface is about hundred of $k_B T$. It was noted, that an orthogonal penetration produces the least damages to the membrane and has the least energy barrier. A nanotube with a hydrophobic surface is "attracted" by the hydrophobic core of the bilayer, which prevents its movement and separation from the membrane due to thermal motion. The SCMF theory was also applied for estimating the penetration parameters of a NT decorated with alternate hydrophilic and hydrophobic bands on its surface [68]. It was shown that a specific pattern on the surface of a NT can facilitate its penetration through a lipid bilayer.

The penetration of NTs through the membrane is dependent on a composition of the membrane, e.g. on cholesterol content, as it was shown in a comparative study of a POPC-lipids bilayer and POPC/cholesterol membrane [69]. Based on a constant velocity SMD simulation of NT penetration, it was demonstrated that the presence of 30% of cholesterol molecules stabilizes the membrane and increases its rigidity. Moreover, MD modeling of lipid membranes reinforced with SWNTs showed that intercalated nanotubes limit the degrees of freedom of neighboring lipids, strengthen membranes, and make them more stable and resistant to temperature increase [70]. The decrease in lipid mobility, particularly, the diminution of self-diffusion coefficient of lipids in the presence of NTs intercalated in the bilayer was also reported [71].

Kraszewski and co-workers, investigated the interaction of pristine (non-functionalized) and amino derivative-functionalized SWNTs with the membrane [72]. The results of MD simulations showed that a closed pristine SWNT freely penetrates into the bilayer through three phases of passive diffusion: landing and floating on the bilayer surface (60 ns), fast penetration through the zone of head groups (20–40 ns), and finally, sliding through the lipid tails region. The penetration of functionalized SWNTs is similar, except that with a comparatively large number of functional groups where the phase of landing is in fact absent. During the functionalized SWNT translocation through the bilayer, amine groups were deprotonated in the simulation and then, when NT approached the opposite membrane side, the charges were recovered. It is pointed out that the presence of functional groups slows

down the penetration of NT through the bilayer. Open SWNT can cause sticking of lipids at their open ends, thereby violating the local structure of the membrane. This effect increases the free energy barrier for translocation of open SWNTs through the membrane in comparison with capped ones. The simulation results [73] lead to the same conclusions. The penetration of open and closed SWNTs of the armchair type through a cholesterol-containing lipid membrane was studied using the all-atom model. It was shown that the penetration of a closed NT causes smaller membrane disturbances than an open one does. A closed SWNT has a lower free energy barrier to penetration through the membrane, which makes it a successful choice for the drug delivery, serving as nano-carriers or nano-containers.

As an example of the use CNT as nano-carrier, the delivery of paclitaxelum (PTX) encapsulated into an open nanotube (PTX@SWNT) through a lipid membrane was considered in [74]. Paclitaxelum is a mitotic inhibitor used in cancer chemotherapy, which molecule has a polar core. Four SMD calculations were performed for the penetration of a PTX@SWNT complex through a lipid bilayer at different velocities. The PTX molecule was located in the far region of the SWNT. In the simulation, external forces were applied both to the PTX and to the SWNT. The highest resistance to penetration was observed when the complex passed through the region of lipid tails which was due to the formation of hydrophobic bonds between the lipophilic groups and the outer SWNT surface. Entry of both the water molecules and the lipids into the inner volume of the nanotube was observed during the simulation. Moreover, the simulations revealed the formation of hydrogen bonds between the PTX and water molecules penetrated into the SWNT. It was noted that a stabilization of the PTX molecule inside the SWNT occurs due to both the Van der Waals forces between the PTX and inner surface of SWNT wall, and the hydrogen bonding between the PTX and water molecules penetrated inside the nanotube cavity.

In order to investigate the features of cell membrane interaction with pristine and functionalized SWNTs differing in diameter, length, chemical modification, and location of functional groups, the simulations with seven types of SWNTs with closed ends and aggregates of SWNTs have been conducted in [75]. The nanotubes had different positions of hydrophilic groups: fully hydrophilic SWNTs, hydrophilic end groups, and fully hydrophobic SWNTs. It was shown that small nanotubes spontaneously penetrate into the membrane. The most stable position of closed hydrophobic SWNTs is the membrane center with an orientation parallel to the bilayer plane. However, in the case of NT with a length smaller than the membrane thickness, the energy minimum for perpendicular (or transmembrane) orientation in the bilayer is somewhat lower. For SWCNTs with functionalized ends and a length slightly greater than the membrane thickness, it is advantageous to have a transmembrane position with some angle to the bilayer at which the difference between the nanotube length and the membrane thickness is compensated. Fully functionalized hydrophilic SWNTs are preferably adsorbed by the membrane at the water-lipid interface in an orientation parallel to the membrane plane. The probability of crossing the membrane or embedding in the bilayer for such SWNTs is extremely low. It is also shown that bundles of several nanotubes self-aggregate in a water solution. These aggregates

entering the bilayer significantly disrupt the membrane structure and the perturbation increases with increasing cluster size.

3.2 Boron Nitride Nanotubes

Despite the similar to CNT tubular geometry, analogous crystal structure, and unique physicochemical properties, the nanoscale mechanisms of the boron nitride nanotubes (BNNT) interaction with the biomembranes and other bio-object are mostly unknown and very poorly covered by MD-based studies. To date, there are only two MD studies of BNNT-lipid bilayer interaction with the atomic description of the system [76, 77].

The pioneering MD study of the BNNT interaction with the lipid membrane was conducted by Hilder and co-workers [76]. The BNNTs with chirality (10, 0) and (10, 10), having a length of 2 nm, were considered. The partial atomic charges of B and N atoms were 0.4 e and –0.4 e, respectively. It was shown that the mechanism of BNNT insertion is similar to the mechanism of insertion of amino-functionalized carbon nanotubes, previously considered by Kraszewski et al. [72], into the lipid bilayer, except for the final stage of realignment. The following four stages were determined for the BNNT insertion into the membrane: (1) formation of BNNT-bilayer contact, (2) BNNT reorientation to become almost parallel to the membrane plane, (3) an interaction of one end of BNNT with the lipid head groups, and (4) an insertion of BNNT, allowing it to slide into the bilayer at an angle of about 45°. As a result the BNNT is partially inserted and is stable inside the bilayer. In general, the behavior of BNNTs is similar to the behavior of functionalized carbon nanotubes, which is less cytotoxic than pure CNTs. Although Hilder et al. performed both the unbiased and steered MD simulations, the energy characteristics of BNNT-membrane interaction were not obtained.

In order to estimate the free energy profile, including the energy barrier, and the depth of energy well, the SMD simulations with a potential of mean force (PMF) analysis of short BNNT and analogous CNT insertion into the lipid bilayer were conducted by the Professor S. G. Psakhie scientific group [77]. Previously, using DFT calculation, it has been found that the partial atomic charges of B and N are strong environmentally dependent [78]. In particular, in case of BNNT with chirality (5, 5) in a vacuum, boron atoms have partial charge of +0.4 e, nitrogen −0.4 e, but for the BNNT containing 5–7 water molecules inside, the absolute values of partial charges are significantly higher ±1.05 e. For this reason, two models of BNNT were examined in the study [77]: BNNT (±0.4) and BNNT (±1.05). A diameter of BNNT was 6.9 Å, length ~11.3 Å.

The result of SMD simulations with the PMF analysis showed that: (1) both BNNT have local minimum of energy about −10 kJ/mol at the lipid-water interface, (2) the BNNT (±0.4) has a global minimum of −72 kJ/mol inside the lipophilic core of the bilayer, the depth is about 30 kJ/mol smaller than that in case of analogous CNT, (3) the CNT was staying empty inside the lipid membrane whereas BNNT

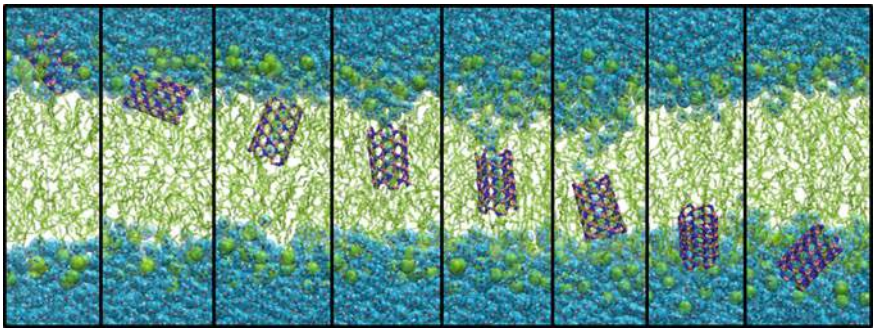


Fig. 5 Boron nitride nanotube (BNNT) translocation through the POPC membrane in steered MD simulation. The orientation of BNNT is transmembrane when it is crossing the tail region of the membrane. BNNT tilts orientation when overcoming the head group—water interface. Water is observed inside BNNT. Images use the XYZ-data from work [77]. Colors: B—pink, N—blue, lipids—light-green, water—light-blue

brings in several water molecules, (4) the penetration of the BNNT into the bilayer is accompanied by a formation of "water defect"—a several-water-molecule long tail behind the BNNT (Fig. 5), (5) the difference in orientation of short BNNT and CNT translocating into the bilayer is that the CNT tends to be in a perpendicular orientation to the membrane plane during whole time of insertion, whereas BNNT (± 0.4) has a tilt from 0 (parallel) to 45° being in the head-group region (Fig. 5), (6) the insertion of the BNNT (± 1.05) is not energetically favorable. The former means an absence of impact on the cell membrane, which indirectly indicates a less cytotoxicity of this nanomaterial.

Besides the two abovementioned researches there is a study of BNNT-membrane interaction, in which, using the mixed model—the Lennard-Jones potential together with the continuum approximation, the relationship between the location of energy minimum, the radius of a cylindrical hole in the membrane, and the perpendicular distance of the BNNT from the hole was determined [79].

4 Two-Dimensional Nanomaterials

4.1 Graphene

One of the pioneering MD studies on the interaction of graphene nanosheet (GNS) with a cell membrane was reported in [80]. Using a coarse-grained MD model, it was shown that GNS can form composite sandwich-like structures in the internal hydrophobic region of a lipid bilayer. Such hybrid graphene–membrane structure can be obtained by forming hydrated micelles of individual GNS coated with a phospholipid layer which can then be absorbed by the membrane. The GNS dimensions

in the simulation were about 60 × 60 Å^2. It was found that the absorbed graphene nanosheet residing at the bilayer center parallel to the membrane plane is a stable system. Slow diffusion motion of the frustrated graphene along the membrane was observed. It is pointed out that the nanosheet does not change the bilayer thickness. The absorption of a lipid micelle containing the GNS by the membrane was also investigated. It was shown that during the absorption of the micelle, the graphene sheet is fixed in the hydrophobic core of the bilayer and lipids from the shell transfer into the nearest membrane leaflet. Full absorption and stabilization of the nanosheet in the bilayer center required about 500 ns. The dependence of displacement of graphene balanced inside the bilayer on the force applied to its edge perpendicular to the system plane was also estimated. Further integration of the obtained data gave an estimate of the energy required for the extraction of graphene from the membrane. In the case of three parallel nanosheets are absorbed, the membrane increases in thickness by about 1 nm, the absorption of eight-GNS sandwich leads to an increase of thickness more than 1.5 times. If the size of a nanosheet on one of the sides coincides with the bilayer thickness and its opposite faces are functionalized by hydrophilic groups, the graphene fragment takes a stable transmembrane position inside the bilayer perpendicular the membrane plane.

The penetration of single square and circular fragments of pristine graphene GNSes of different sizes into the membrane was studied in [81]. It was shown that as the nanosheet size is increased, the bilayer tends to form semi-spherical vesicle, and thus the membrane experiences substantial deformation. During internalization, the angle between the membrane plane and GNS was estimated. The behavior of this angle revealed three stages: (1) spontaneous orientation before the nanosheet touches the membrane surface, (2) embedment of the nanosheet into the bilayer mainly at an angle of 47°, and (3) rotation of the nanosheet into membrane-parallel position with the formation a sandwich-like superstructure. The MD simulation was also performed to study how the thickness of single- and multilayered graphene nanosheets and the degree of their surface oxidation influence the interaction with a lipid membrane [82]. The penetration of a GNS coated with lipids was also considered. It was shown that for the pristine GNS, the position in the bilayer center parallel to the membrane plane is stable, whereas a nanosheet with oxide at the boundary (10% degree of oxidation) is attached at an angle to the membrane plane. The degree of boundary oxidation strongly affects the final GNS orientation in the bilayer. At oxidation values of less than 5%, the GNS behaves as a pristine one. It was noted that there are two appreciable energy drops in the system during penetration: (1) when the GNS is embedded in the bilayer, and (2) when it rotates inside the bilayer to take its final position. The second energy drop is lower than the first one. The simulation of coated GNS was performed for different amounts of coating lipids. At a low surface absorption density, the nanosheet edge zones were open to water. The penetration was no different from the internalization of pristine graphene sheets. At a relatively high lipid concentration on the GNS surface, no penetration within the rather limited time of MD simulation was observed. For penetration of a multilayered graphene into the bilayer, piercing of the membrane by its corner and parallel orientation inside the membrane, like for single GNS, was detected.

Using coarse-grained MD and all-atom SMD simulations, the mechanism of spontaneous penetration of multilayered (few-layer) GNS into a cell though piercing the membrane by their edge or corner was studied in [83]. The model took into account the roughness (asperities) of the GNS edge the texture of which was borrowed from an image of the real structure taken with a transmission electron microscope. The results of MD calculations showed that monolayer graphene in the form of a rhomb (with an edge length of 64 Å and most acute angle of 30° and 60°) that turns due to thermal motion resulting in one of its acute corner sites to be directed to the bilayer. When the region of head groups is pierced, complete GN absorption by the membrane can be initiated through hydrophobic interaction of graphene with lipid tail groups. Additionally, a series of calculations with triangular, square, and hexahedral fragments of both pristine GNS and GNS functionalized at the corners or along the perimeter were performed. The results of MD calculations showed that absorption occurs with the non-functionalized corner forward, and if no such corners are present, the GNS remains on the membrane surface. Using the SMD method, the free energy barrier for graphene penetration into the membrane was estimated. It was shown that the penetration of the GNS with an ideal boundary into the membrane is almost impossible at room temperature, whereas inhomogeneous topography of its boundary can greatly decrease the free energy barrier. It was noted that when absorbing the GNS, the membrane structure is hardly disturbed, except that the lipid tail groups are straitened along the nanosheet surface due to high adhesion.

A series of coarse-grained calculations were performed to study the interaction of pristine graphene and graphene oxide with a lipid membrane for different nanosheet sizes and degrees of oxidation [84]. The graphene sheets were square flakes with side lengths of 35, 70, and 105 Å; the density of oxide groups was varied from 0 to 40% of the total number of carbon atoms. Nanosheets oxidized only along the perimeter were also considered. In total, four variants were simulated: pristine GNS, graphene oxidized along the perimeter (epitaxial graphene oxide, EGO), graphene with oxide along the perimeter and 20% on the surface (GO20), and graphene with oxide along the perimeter and 40% on the surface (GO40). Each calculation took about 45 μs. It was found that the graphene interaction with the membrane provides one of the four configurations: a sandwich-like structure with a nanosheets located in the bilayer center and parallel to it (for GNS with sides $l = 35, 70$ Å); a graphene-containing semi-spherical vesicle immersed into the bilayer (for GNS and EGO with $l = 105$ Å); a nanosheet lying on the membrane surface and parallel to it (GO20 and GO40 with $l = 35$ Å); and crosswise oriented graphene inside the bilayer. It was found that the scale of observed membrane distortions depends on the degree of graphene oxidation: the more this degree, the higher the distortion of the bilayer structure. Increasing the side of square graphene fragments increases these disturbances up to the point of membrane continuity disruption.

The results of experimental and numerical studies of the interaction of pristine graphene and graphene oxide with a bacterial membrane were reported in [85]. According to them, GNS can embed and cut the external and internal cell membranes of *E. coli*. This involves degradation of the membrane with possible destruction of the cell wall and partial cytoplasm loss. Using the all-atom MD model, it was shown

that there exist three phases of graphene and graphene oxide interaction with lipid membranes. The first phase is diffusion motion of the nanosheet till it touches the membrane; the second phase is its comparatively fast insertion into the bilayer; the third one is an extraction of lipids from the bilayer due to Van der Waals forces and hydrophobic interaction on the GNS surface. This involves substantial depression and deformation of the membrane, facilitating disruption of its continuity. Such an action decreases the bacterial cell viability and allows consideration of graphene as an antibacterial nanomaterial. Schematic molecular model illustrating the extraction of the lipids on GNS surface is presented at Fig. 6.

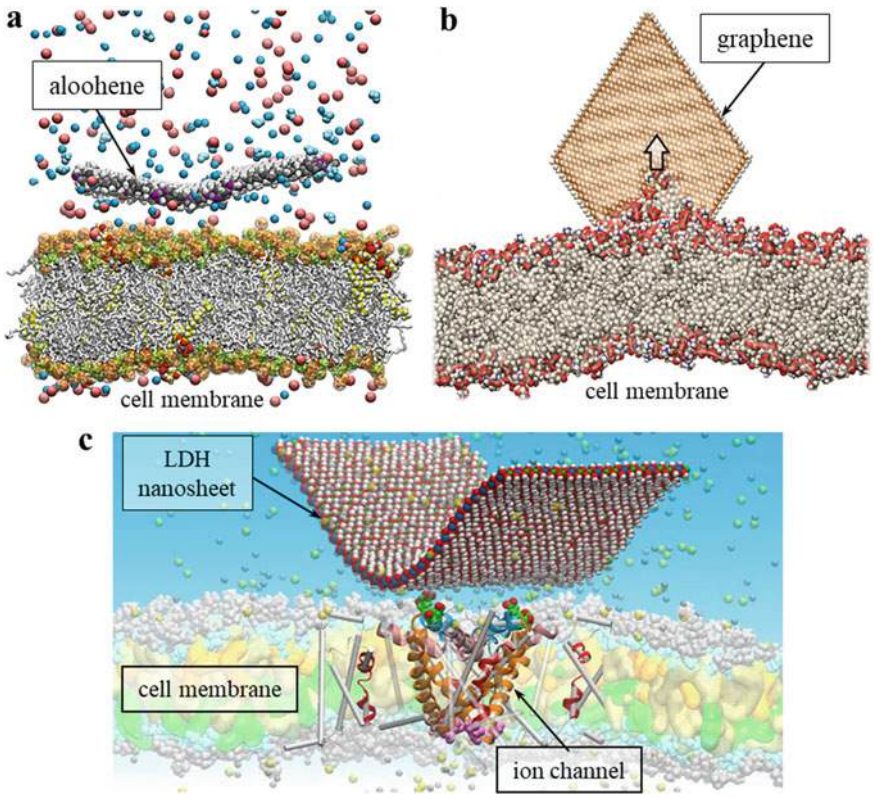


Fig. 6 Two-dimensional nanomaterials and cell membrane interaction: **a** aloohene doesn't penetrate to the bilayer and doesn't extract lipids, its impact to a cell is a dysregulation of ionic balance (the result of long unbiased molecular dynamics simulation). Reproduced with permission from ACS [6]. **b** Schematic molecular model illustrating the extraction of the lipids onto GNS surface (orange) from the cell membrane. **c** Nanosheet of layered double hydroxide interacting with outer loops of bacterial sodium channel Na_VAb, embedded into the membrane (the model was built using XYZ-data from [86])

4.2 Metal (Oxy)hydroxides

Unlike GNS, the layered metal hydroxides and oxyhydroxides are hydrophilic compounds with large amount of polar OH-groups on the surface. Layered metal hydroxides and oxyhydroxides as a rule are chemically inert, non-toxic, biocompatible, have a large specific surface area with non-zero surface charge, can form the hybrid complexes with organic molecules, and be used as selective ion exchangers with a high proton and/or hydroxyl buffer capacity. These unique physicochemical properties make the layered metal (oxy)hydroxides an important material for a wide range of applications in different areas, including biomedicine.

One of the first MD research of biomembrane interaction with metal (oxy)hydroxide was performed by Shroll and Straatsma [87]. They considered an adhesion of the gram-negative bacteria outer membrane on FeOOH (goethite). Although the goethite FeOOH is not a layered oxyhydroxide but the oxyhydroxide with bulk crystal structure, we will consider this case, since the modeled fragment has a thickness of several atomic layers. The modelled outer membrane of *Pseudomonas aeruginosa* bacteria has bilayer structure and consist of lipopolysaccharide (LPS) monolayer (outer leaflet) and the monolayer of typical phospholipids (inner leaflet). The LPS has several carboxyl and phosphate groups, which are the acceptors of proton. Each LPS molecule brings strong negative charge of −13 e. The strong adhesion within LPS-NM interface was observed. It was found that outermost LPS groups must compete with water molecules resided near the FeOOH surface and may replace them forming multiple hydrogen bonds with hydroxyl groups of the mineral surface. Due to the formation of these bonds the significant changes in the structure of outermost saccharide groups of the membrane were observed. No obvious tendency of FeOOH-induced membrane disruption is reported.

A similar result was obtained in the unbiased MD simulation of the layered double hydroxide (LDH) near the outer surface of the membrane with bacterial ion channel embedded [87]. The hydrogen bonding of outer carboxyl-contained groups of the Na_VAb voltage-gated sodium-selective ion channel with the hydroxide groups on the LDH surface was observed, without any tendency to disrupt the membrane (Fig. 6).

The layered metal oxyhydroxides nanomaterial based on the AlOOH (boehmite), termed “aloohene”, was developed by the scientific group of Professor S. G. Psakhie [6] and showed promise for multiple biomedical applications. The all-atom MD simulations of the aloohene fragment in the vicinity of the cell membrane demonstrate that aloohene doesn’t tend to internalize into the lipid bilayer. Moreover, due to the electrostatic and amphoteric properties of its surface, the aloohene dysregulates the composition and the ion balance of the tumor cell microenvironment.

It is supposed that the electrostatic action on the membrane and perturbation of the ion concentration near its surface can greatly affect the functioning of membrane proteins sensitive to changes in membrane polarization. This effect can be critical for some metabolic and intracellular processes, including those which support viability and proliferation of tumor cells.

5 Summary

Computer simulations of interactions on the molecular level provide a unique opportunity to explore the processes and mechanisms in such complex and heterogeneous systems as biomembrane-nanomaterial interfaces at picosecond-microsecond timescales. The understanding of the basic rules, which govern adsorption, insertion, accumulation, wrapping, uptake, and disruption in these systems allows one to determine the key factors to control the cell-nanomaterial or cell-nanodevice behavior. This knowledge is of fundamental importance for the further progress in nanomedicine, theranostics, oncology, and related fields.

Acknowledgements The part of work related to the carbon nanomaterials was performed in the framework of the research project supported from the Russian Foundation for Basic Research (RFBR), Grant No. 18-29-19198.

References

1. Shi J, Kantoff PW, Wooster R, Farokhzad OC (2017) Cancer nanomedicine: progress, challenges and opportunities. Nat Rev Cancer 17(1):20–37
2. Lee H, Lee Y, Song C, Cho HR, Ghaffari R, Choi TK, Kim KH, Lee YB, Ling D, Lee H, Yu SJ, Choi SH, Hyeon T, Kim D-H (2015) An endoscope with integrated transparent bioelectronics and theranostic nanoparticles for colon cancer treatment. Nat Commun 6:10059
3. Song S, Faleo G, Yeung R, Kant R, Posselt AM, Desai TA, Tang Q, Roy S (2016) Silicon nanopore membrane (SNM) for islet encapsulation and immunoisolation under convective transport. Sci Rep 6(1):1–9
4. Mikhaylov G, Mikac U, Magaeva AA, Itin VI, Naiden EP, Psakhye I, Babes L, Reinheckel T, Peters C, Zeiser R, Bogyo M, Turk V, Psakhye SG, Turk B, Vasiljeva O (2011) Ferri-liposomes as an MRI-visible drug-delivery system for targeting tumours and their microenvironment. Nat Nanotechnol 6(9):594–602
5. Li D, Zhang Y-T, Yu M, Guo J, Chaudhary D, Wang C-C (2013) Cancer therapy and fluorescence imaging using the active release of doxorubicin from MSPs/Ni-LDH folate targeting nanoparticles. Biomaterials 34(32):7913–7922
6. Lerner MI, Mikhaylov G, Tsukanov AA, Lozhkomoev AS, Gutmanas E, Gotman I, Bratovs A, Turk V, Turk B, Psakhye SG, Vasiljeva O (2018) Crumpled aluminum hydroxide nanostructures as a microenvironment dysregulation agent for cancer treatment. Nano Lett 18(9):5401–5410
7. Gupta A, Landis RF, Li C-H, Schnurr M, Das R, Lee Y-W, Yazdani M, Liu Y, Kozlova A, Rotello VM (2018) Engineered polymer nanoparticles with unprecedented antimicrobial efficacy and therapeutic indices against multidrug-resistant bacteria and biofilms. J Am Chem Soc 140(38):12137–12143
8. Zhou W, Pan T, Cui H, Zhao Z, Chu PK, Yu X-F (2019) Black phosphorus: bioactive nanomaterials with inherent and selective chemotherapeutic effects. Angew Chem (International ed. English) 131(3):779–784
9. Gennis RB (2013) Biomembranes: molecular structure and function. Springer Science & Business Media, New York
10. Mitaku S (1993) The role of hydrophobic interaction in phase transition and structure formation of lipid membranes and proteins. Phase Transit Multinational J 45(2–3):137–155
11. Thompson RB, Ginzburg VV, Matsen MW, Balazs AC (2001) Predicting the mesophases of copolymer-nanoparticle composites. Science 292(5526):2469–2472

12. Thompson RB, Ginzburg VV, Matsen MW, Balazs AC (2002) Block copolymer-directed assembly of nanoparticles: forming mesoscopically ordered hybrid materials. Macromolecules 35(3):1060–1071
13. Qian J, Gao H (2010) Soft matrices suppress cooperative behaviors among receptor-ligand bonds in cell adhesion. PLoS ONE 5(8):e12342
14. Gao H, Qian J, Chen B (2011) Probing mechanical principles of focal contacts in cell–matrix adhesion with a coupled stochastic–elastic modelling framework. J R Soc Interface 8(62):1217–1232. https://doi.org/10.1098/rsif.2011.0157
15. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
16. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink S-J (2008) The MARTINI coarse-grained force field: extension to proteins. J Chem Theory Comput 4(5):819–834
17. Izrailev S, Stepaniants S, Isralewitz B et al (1997) Computational molecular dynamics: challenges, methods, ideas. Springer, Berlin, pp 39–65 (Steered molecular dynamics)
18. Pogodin S, Werner M, Sommer JU, Baulin VA (2012) Nanoparticle-induced permeability of lipid membranes. ACS Nano 6(12):10555–10561
19. Orsi M, Essex JW (2010) Molecular simulations and biomembranes: from biophysics to function. Royal Society of Chemistry, London, pp 76–90 (Passive permeation across lipid bilayers: a literature review)
20. Tsukanov AA, Psakhie SG (2015) A review of computer simulation studies of cell membrane interaction with neutral and charged nano-objects. Quasi-zero-dimensional nanoparticles, drugs and fullerenes. Adv Biomater Dev Med 2(1):44–53
21. Qiao R, Roberts AP, Mount AS, Klaine SJ, Ke PC (2007) Translocation of C_{60} and its derivatives across a lipid bilayer. Nano Lett 7(3):614–619
22. Wong-Ekkabut J, Baoukina S, Triampo W, Tang I-M, Tieleman DP, Monticelli L (2008) Computer simulation study of fullerene translocation through lipid membrane. Nat Nanotechnol 3:363–368
23. D'Rozario RSG, Wee CL, Wallace EJ, Sansom MSP (2009) The interaction of C_{60} and its derivatives with a lipid bilayer via molecular dynamics simulations. Nanotechnology 20:115102
24. Jusufi A, DeVane RH, Shinoda W, Klein ML (2011) Nanoscale carbon particles and the stability of lipid bilayers. Soft Matter 7(3):1139–1146
25. Zhang S, Mu Y, Zhang JZH, Xu W (2013) Effect of self-assembly of fullerene nano-particles on lipid membrane. PLoS ONE 8(10):e77436
26. Bozdaganyan ME, Orekhov PS, Shaytan AK, Shaitan K (2014) Comparative computational study of interaction of C_{60}-fullerene and tris-malonyl-C_{60}-fullerene isomers with lipid bilayer: relation to their antioxidant effect. PLoS ONE 9(7):e102487
27. Lai K, Wang B, Zhang Y, Zheng Y (2013) Computer simulation study of nanoparticle interaction with a lipid membrane under mechanical stress. Phys Chem Chem Phys 15(1):270–278
28. Kraszewski S, Tarek M, Treptow W, Ramseyer C (2010) Affinity of C_{60} neat fullerenes with membrane proteins: a computational study on potassium channels. ACS Nano 4:4158–4164
29. Monticelli L, Barnoud J, Orlowski A, Vattulainen I (2012) Interaction of C_{70} fullerene with the Kv1.2 potassium channel. Phys Chem Chem Phys 14(36):12526–12533
30. Kraszewski S, Tarek M, Ramseyer C (2011) Uptake and translocation mechanisms of cationic amino derivatives functionalized on pristine C_{60} by lipid membranes: a molecular dynamics simulation study. ACS Nano 5:8571–8578
31. Tieleman DP (2006) Computer simulations of transport through membranes: passive diffusion, pores, channels and transporters. Proc Aust Physiol Soc 37:15–27

32. Monticelli L, Salonen E, Ke PC, Vattulainen I (2009) Effects of carbon nanoparticles on lipid membranes: a molecular simulation perspective. Soft Matter 5(22):4433–4445
33. Makarucha AJ, Todorova N, Yarovsky I (2011) Nanomaterials in biological environment: a review of computer modelling studies. Eur Biophys J 40:103–115
34. Esfand R, Tomalia DA (2001) Poly(amidoamine) (PAMAM) dendrimers: from biomimicry to drug delivery and biomedical applications. Drug Discov Today 6(8):427–436
35. Kolotylo M, Holovatiuk V, Bondareva J, Lukin O, Rozhkov V (2019) Synthesis of sulfonimide-based dendrimers and dendrons possessing mixed $1 \rightarrow 2$ and $1 \rightarrow 4$ branching motifs. Tetrahedron Lett 60(4):352–354
36. Rozhkov VV, Kolotylo MV, Onys'ko PP, Lukin O (2016) Synthesis of sulfonimide-based branched arylsulfonyl chlorides. Tetrahedron Lett 57(3):308–309
37. Maiti PK, Çağın T, Wang G, Goddard WA (2004) Structure of PAMAM dendrimers: generations 1 through 11. Macromolecules 37(16):6236–6254
38. Lee H, Larson RG (2006) Molecular dynamics simulations of PAMAM dendrimer-induced pore formation in DPPC bilayers with a coarse-grained model. J Phys Chem B 110:18204–18211
39. Lee H, Larson RG (2008) Coarse-grained molecular dynamics studies of the concentration and size dependence of fifth- and seventh-generation PAMAM dendrimers on pore formation in DMPC bilayer. J Phys Chem B 112:7778–7784
40. Lee H, Larson RG (2008) Lipid bilayer curvature and pore formation induced by charged linear polymers and dendrimers: the effect of molecular shape. J Phys Chem B 112:12279–12285
41. Kelly CV, Leroueil PR, Nett EK, Wereszczynski JM, Baker JR, Orr BG, Banaszak Holl MM, Andricioaei I (2008) Poly(amidoamine) dendrimers on lipid bilayers I: free energy and conformation of binding. J Phys Chem B 112(31):9337–9345
42. Kelly CV, Leroueil PR, Orr BG, Banaszak Holl MM, Andricioaei I (2008) Poly(amidoamine) dendrimers on lipid bilayers II: effects of bilayer phase and dendrimer termination. J Phys Chem B 112:9346–9353
43. Yan LT, Yu X (2009) Enhanced permeability of charged dendrimers across tense lipid bilayer membranes. ACS Nano 3(8):2171–2176
44. Ting CL, Wang ZG (2011) Interactions of a charged nanoparticle with a lipid membrane: implications for gene delivery. Biophys J 100(5):1288–1297
45. Ginzburg VV, Balijepalli S (2007) Modeling the thermodynamics of the interaction of nanoparticles with cell membranes. Nano Lett 7(12):3716–3722
46. Li Y, Chen X, Gu N (2008) Computational investigation of interaction between nanoparticles and membranes: hydrophobic/hydrophilic effect. J Phys Chem B 112(51):16647–16653
47. Alexeev A, Uspal WE, Balazs AC (2008) Harnessing Janus nanoparticles to create controllable pores in membranes. ACS Nano 2(6):1117–1122
48. Yang K, Ma YQ (2010) Computer simulation of the translocation of nanoparticles with different shapes across a lipid bilayer. Nat Nanotechnol 5(8):579–583
49. Lin J, Zhang H, Chen Z, Zheng Y (2010) Penetration of lipid membranes by gold nanoparticles: insights into cellular uptake, cytotoxicity, and their relationship. ACS Nano 4(9):5421–5429
50. Guo R, Mao J, Yan LT (2013) Unique dynamical approach of fully wrapping dendrimer-like soft nanoparticles by lipid bilayer membrane. ACS Nano 7(12):10646–10653
51. Carr R, Weinstock IA, Sivaprasadarao A, Müller A, Aksimentiev A (2008) Synthetic ion channels via self-assembly: a route for embedding porous polyoxometalate nanocapsules in lipid bilayer membranes. Nano Lett 8(11):3916–3921
52. Dykman L, Khlebtsov N (2012) Gold nanoparticles in biomedical applications: recent advances and perspectives. Chem Soc Rev 41:2256–2282
53. Nangia S, Sureshkumar R (2012) Effects of nanoparticle charge and shape anisotropy on translocation through cell membranes. Langmuir 28(51):17666–17671
54. Tsukanov AA, Pervikov AV, Lozhkomoev AS (2020) Bimetallic Ag–Cu nanoparticles interaction with lipid and lipopolysaccharide membranes. Comput Mater Sci 173:109396
55. Daw MS, Baskes MI (1984) Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals. Phys Rev B 29(12):6443

56. Williams PL, Mishin Y, Hamilton JC (2006) An embedded-atom potential for the Cu–Ag system. Modell Simul Mater Sci Eng 14(5):817
57. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev 136:B864
58. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140(4A):A1133
59. Bekyarova E, Ni Y, Malarkey EB, Montana V, McWilliams JL, Haddon RC, Parpura V (2005) Applications of carbon nanotubes in biotechnology and biomedicine. J Biomed Nanotechnol 1(1):3–17
60. Rotkin SV (2010) Electronic properties of nonideal nanotube materials: helical symmetry breaking in DNA hybrids. Annu Rev Phys Chem 61:241–261
61. Tsukanov AA, Grachev EA, Rotkin SV (2007) Modeling of the SWNT-DNA complexes in the water solution. APS March meeting, American Physical Society, 5–9 Mar 2007, abstract id. V28.002
62. Lopez CF, Nielsen SO, Moore PB, Klein ML (2004) Understanding nature's design for a nanosyringe. PNAS Proc Nat Acad Sci U S A 101:4431–4434
63. Zimmerli U, Koumoutsakos P (2008) Simulations of electrophoretic RNA transport through transmembrane carbon nanotubes. Biophys J 94(7):2546–2557
64. Höfinger S, Melle-Franco M, Gallo T, Cantelli A, Calvaresi M, Gomes JANF, Zerbetto F (2011) A computational analysis of the insertion of carbon nanotubes into cellular membranes. Biomaterials 32:7079–7085
65. Shi XH, Kong Y, Gao HJ (2008) Coarse grained molecular dynamics and theoretical studies of carbon nanotubes entering cell membrane. Acta Mech Sin 24:161–169
66. Pogodin S, Baulin VA (2010) Can a carbon nanotube pierce through a phospholipid bilayer? ACS Nano 4:5293–5300
67. Ben-Shaul A, Szleifer I, Gelbart WM (1985) Chain organization and thermodynamics in micelles and bilayers. AIP J Chem Phys 83(7):3597–3611
68. Pogodin S, Slater NKH, Baulin VA (2011) Surface patterning of carbon nanotubes can enhance their penetration through a phospholipid bilayer. ACS Nano 5:1141–1146
69. Gangupomu VK, Capaldi FM (2011) Interactions of carbon nanotube with lipid bilayer membranes. J Nanomater 2011:830436
70. Shityakov S, Dandekar T (2011) Molecular dynamics simulation of POPC and POPE lipid membrane bilayers enforced by an intercalated single-wall carbon nanotube. NANO 6(01):19–29
71. Parthasarathi R, Tummala NR, Striolo A (2012) Embedded single-walled carbon nanotubes locally perturb DOPC phospholipid bilayers. J Phys Chem B 116:12769–12782
72. Kraszewski S, Bianco A, Tarek M, Remaseyer C (2012) Insertion of short amino-functionalized single-walled carbon nanotubes into phospholipid bilayer occurs by passive diffusion. PLoS ONE 7(7):e40703
73. Raczyński P, Górny K, Pabiszczak M, Gburski Z (2013) Nanoindentation of biomembrane by carbon nanotubes—MD simulation. Comput Mater Sci 70:13–18
74. Mousavi SZ, Amjad-Iranagh S, Nademi Y, Modarress H (2013) Carbon nanotube-encapsulated drug penetration through the cell membrane: an investigation based on steered molecular dynamics simulation. J Membr Biol 246(9):697–704
75. Baoukina S, Monticelli L, Tieleman DP (2013) Interaction of pristine and functionalized carbon nanotubes with lipid membranes. J Phys Chem B 117:12113–12123
76. Thomas M, Enciso M, Hilder TA (2015) Insertion mechanism and stability of boron nitride nanotubes in lipid bilayers. J Phys Chem B 119(15):4929–4936
77. Tsukanov AA, Psakhie SG (2016) Potential of mean force analysis of short boron nitride and carbon nanotubes insertion into cell membranes. Adv Biomater Dev Med 3(1):1–9
78. Won CY, Aluru NR (2008) Structure and dynamics of water confined in a boron nitride nanotube. J Phys Chem C 112(6):1812–1818
79. Alshehri MH (2018) Interactions of boron nitride nanotubes with lipid bilayer membranes. J Comput Theor Nanosci 15(1):311–316

80. Titov AV, Kral P, Pearson R (2010) Sandwiched graphene–membrane superstructures. ACS Nano 4:229–234
81. Guo R, Mao J, Yan L-T (2013) Computer simulation of cell entry of graphene nanosheet. Biomaterials 34(17):4296–4301
82. Wang J, Wei Y, Shi X, Gao H (2013) Cellular entry of graphene nanosheets: the role of thickness, oxidation and surface adsorption. RSC Adv 3(36):15776–15782
83. Li Y, Yuan H, von dem Bussche A, Creighton M, Hurt RH, Kane AB, Gao H (2013) Graphene microsheets enter cells through spontaneous membrane penetration at edge asperities and corner sites. PNAS Proc Nat Acad Sci U S A 110(30):12295–12300
84. Mao J, Guo R, Yan L-T (2014) Simulation and analysis of cellular internalization pathways and membrane perturbation for graphene nanosheets. Biomaterials 35(23):6069–6077
85. Tu Y, Lv M, Xiu P, Huynh T, Zhang M, Castelli M, Liu Z, Huang Q, Fan C, Fang H, Zhou R (2013) Destructive extraction of phospholipids from *Escherichia coli* membranes by graphene nanosheets. Nat Nanotechnol 8(8):594–601
86. Tsukanov AA, Psakhie SG (2016) Adsorption of charged protein residues on an inorganic nanosheet: computer simulation of LDH interaction with ion channel. AIP Conf Proc 1760(1):020066
87. Shroll RM, Straatsma TP (2003) Molecular basis for microbial adhesion to geochemical surfaces: computer simulation of *Pseudomonas aeruginosa* adhesion to goethite. Biophys J 84(3):1765–1772

Application of Crumpled Aluminum Hydroxide Nanostructures for Cancer Treatment

Aleksandr S. Lozhkomoev, Georgy Mikhaylov, Vito Turk, Boris Turk, and Olga Vasiljeva

Abstract The tumor microenvironment regulates tumor progression and the spread of cancer in the body. Applications of nanomaterials that can dysregulate tumor-microenvironment are emerging as a promising anti-cancer approaches, which can improve the efficacy of existing cancer treatments. We have reported that agglomerates of radially assembled Al hydroxide crumpled nanosheets with the disordered defective surface structure have a large positive charge and therefore can lead to ion imbalance at the cell perimembranous space through the selective adsorption of extracellular anionic species. This effect was demonstrated in vitro by reduced viability and proliferation of tumor cells, and further validated in a murine melanoma cancer model. Furthermore, crumpled Al hydroxide nanostructures showed a much stronger suppressive effect on tumor growth in combination with a minimally effective dose of doxorubicin. Taken together, the described approach of tumor microenvironment dysregulation through selective adsorption properties of folded crumpled nanostructures opened a new avenue for development of innovative anticancer therapy strategies.

Keywords Aluminum hydroxide · Nanosheets · Folded crumpled nanostructures · Tumor microenvironment · Anti-cancer therapy · Aloohene · Ion imbalance

G. Mikhaylov · V. Turk · B. Turk · O. Vasiljeva (✉)
Jozef Stefan Institute, 1000 Ljubljana, Slovenia
e-mail: olga.vasiljeva@ijs.si

A. S. Lozhkomoev
Institute of Strength Physics and Materials Science, Tomsk 634055, Russia

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_10

1 Introduction to Low-Dimensional Aluminum (Hydro)oxides

Low-dimensional (2D) nanostructures have been extensively studied in the last 10–15 years [1–5]. The interest can be explained by the fact that two-dimensional nano-objects exhibit physicochemical properties significantly different from those of three-dimensional (bulk) materials. This is due to both the dominant role of the free surface and the morphological instability of the planar shape of 2D systems [6–9]. It should be noted that the current trend in the development of nanotechnology in the field of synthesis of low-dimensional nanostructures, possessing the necessary physicochemical and biomedical properties, is characterized by a broader use of the methods based on controlled self-organization (or self-assembling) of the structure directly in the synthesis process [10, 11]. In case of ultrathin 2D-nanostructures, the self-assembly of such ensembles is of high importance to fundamental and applied sciences, since it gives an opportunity not only to investigate a substance behavior in the two-dimensional state, but also to produce complex hybrid (organic-inorganic) nanosystems for a wide range of applications ranging from elements for sensor devices to nanostructures for biomedical applications [10, 12]. One of the promising directions in the formation of 2D nanostructures with a controlled structure is self-assembly during the oxidation of nanoparticles [10, 13].

The perspectives of new drugs development for the diagnostics and treatment of various diseases, including cancer, are associated with studies of low-dimensional nanostructures interaction with living cells [14–16]. The effects of colloidal gold particles [17], nanofibers or nanowires of zinc oxide [18, 19], titanium oxide [20], carbon nanotubes [21] and graphene oxide [22] have been the most studied as yet. Low-dimensional structures via self-assembling can form more complex ensembles [23]; however, the interaction of cells with such ensembles can be also studied at the level of individual low-dimensional components of the complex.

From the point of view of medical applications and targeted action on biological objects the low-dimensional aluminum (hydro)oxides should be considered as a promising basis for development of nanostructures with bio-activity. This is due to both their unique surface properties [12, 24], which determine their biological activity [25, 26], as well as their low toxicity [27] and the possibility to generate nanostructures with certain morphology during their synthesis [28].

2 Synthesis of Aluminum Oxyhydroxide Low-Dimensional Nanostructures

A simple and fast method for the synthesis of nanostructures based on the aluminum oxide and hydroxide phases is oxidation of aluminum nanopowder by water obtained via electrical dispersion of an aluminum wire [29, 30]. The synthesis proceeds in

a single stage at atmospheric pressure and a temperature of 50–70 °C. The reaction mixture contains only water and aluminum. Therefore, the reaction products do not contain organic and ionic contaminants. This is a major advancement from the previous studies, where oxidation of aluminum nanoparticles by water was studied only as a method for a hydrogen production, and much less attention was paid to the reaction products [31–36]. Nanoparticles of the aluminum nitride composition (Al/AlN) provide a better starting material for the reaction due to the presence of a thin oxide film on the surface of the particles and the release of ammonia during hydrolysis, which increases the pH of the reaction medium [37]. The oxidation process of Al/AlN nanoparticles proceeds without an induction period, due to a thinner oxide film, and is accompanied by complete oxidation of aluminum as a result of ammonia excretion during the reaction. In order to obtain the aluminum oxyhydroxide low-dimensional nanostructures, the AlN/Al nanopowder produced by electric explosion of an aluminum wire in a nitrogen atmosphere, was used in our research.

Nanopowder is represented by a mixture of spherical and faceted particles. According to the results of the elemental analysis, nitrogen is found in all particles, however it is distributed in spherical particles non-uniformly (Fig. 1a), indicating

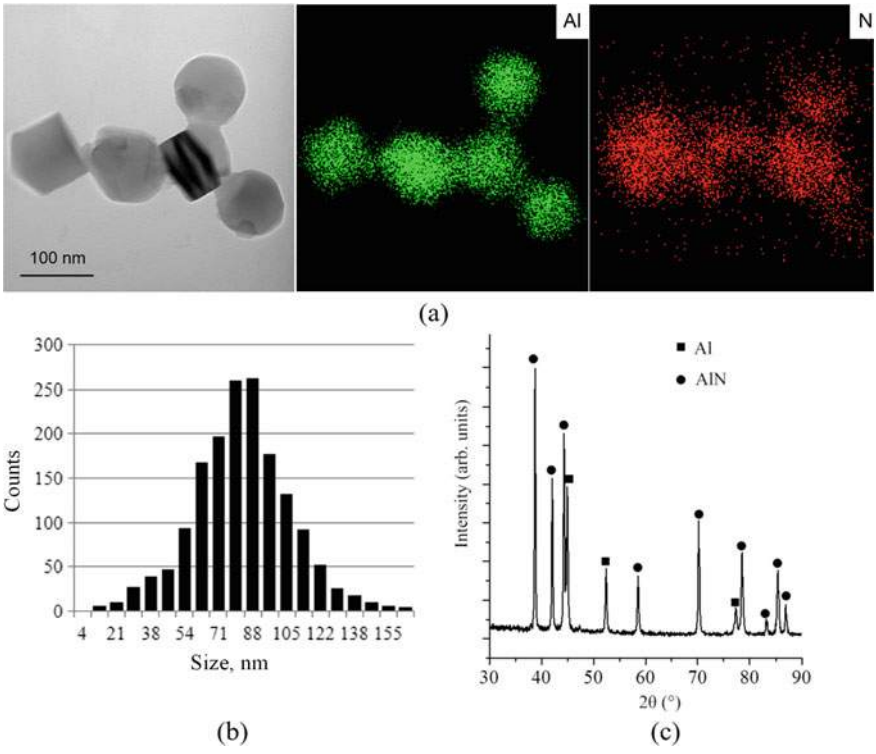


Fig. 1 TEM image and EDS mapping (**a**), particle size distribution and XRD pattern (**c**) of AlN/Al nanoparticles

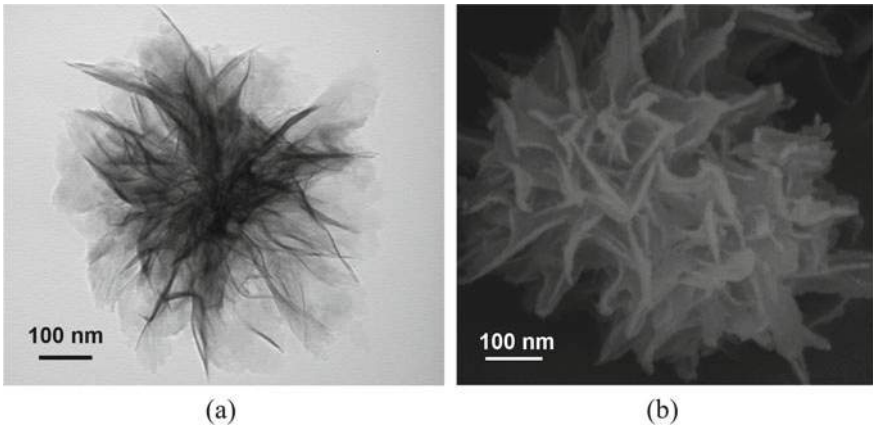


Fig. 2 TEM (**a**) and SEM (**b**) images of agglomerates of crumpled nanosheets. Reproduced with permission from ACS [38]

presence of metallic aluminum in them. The average particle size is 85 nm (Fig. 1b). According to the X-ray diffraction (XRD) analysis, the main peaks in the XRD pattern correspond to the Al and AlN phases (Fig. 1c).

As the result of AlN/Al nanoparticles oxidation in water at 60 °C the agglomerates of size up to 2 μm, consisting of nanosheets of size up to 200 nm and thickness of about 2–5 nm, are formed (Fig. 2).

According to the X-ray phase analysis (Fig. 3), the main peaks of the AlN/Al oxidation products correspond to crystalline boehmite. Both the shift of the reflection in the (020) plane toward smaller angles and the broadening of the peaks indicate the absence of long-range order in the arrangement of atoms and the amorphous structure, which is characteristic for pseudo-boehmite.

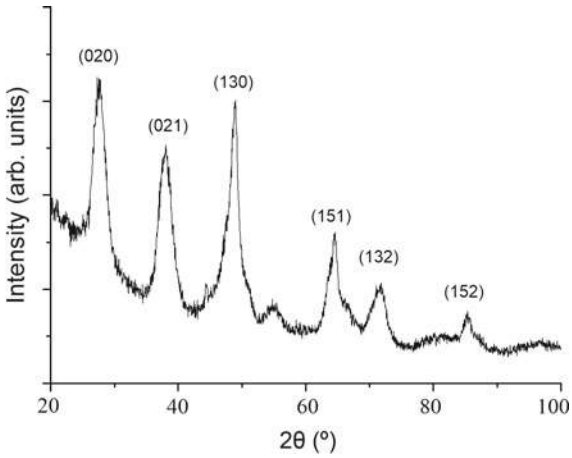


Fig. 3 XRD pattern of the agglomerates of AlOOH crumpled nanosheets

Nitrogen N_2 adsorption-desorption isotherms obtained for AlOOH nanosheet agglomerates have a hysteresis loop in the region of capillary condensation at a relative pressure of P/Po ≥ 0.4. Such isotherms are typical for mesoporous materials with slit-shaped pores. The maximum of pore size distribution, determined by the method of Barrett, Joyner and Halenda, is within the region of 4 nm. The specific surface area, determined by Brunauer-Emmett-Teller method was 286 m^2/g. The slit-like structure of the mesopores and the large specific surface area of the AlOOH nanosheets agglomerates provide them with high adsorption activity (Fig. 4).

Figure 5 shows synthesized nanostructures zeta-potential dependency on pH level of the medium. The nanostructures have a large positive zeta-potential of about 30 mV at pH ~6.5–8.0 and 37 °C. The pH of the zero charge point pH_{IEP} of AlOOH nanostructures was estimated as 9.53.

The above results suggest nonequilibrium morphology and structural state of AlOOH nanosheets agglomerates. The crumpled structure causes local deformations and multiple micro-stresses, which contribute to the formation of a larger number of surficial active centers due to surface defect zones. It should be noted that the resulting folded crumpled nanostructures due to the presence of a large number of (active) hydroxyl groups per unit surface area can acquire a significant pH-dependent electric charge. Such aluminum nanomaterial is an amphoteric compound, which in the case of an acidic and even neutral medium relatively easily releases the OH-groups into the solution. Dissociation of OH^- groups is facilitated in the areas of defects of the nanomaterial, similarly to the case of hydroxyl groups deprotonation, observed in cationic clays [39]. Therefore, it can be assumed that AlOOH nanosheets with a

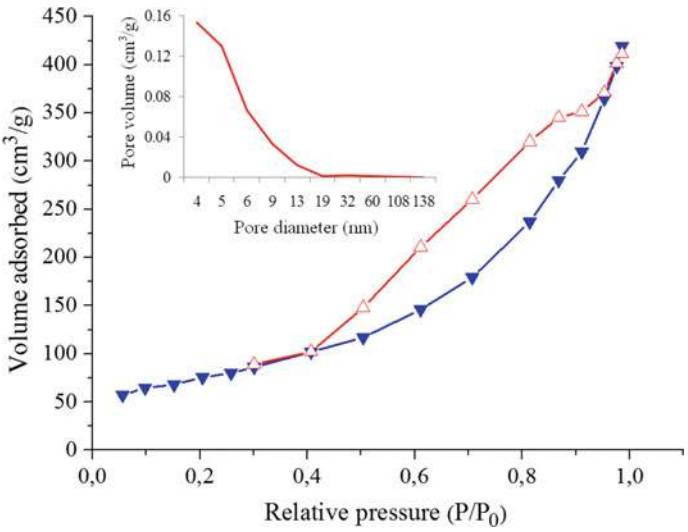


Fig. 4 The N_2 adsorption/desorption isotherms and pore size distributions of AlOOH crumpled nanosheets

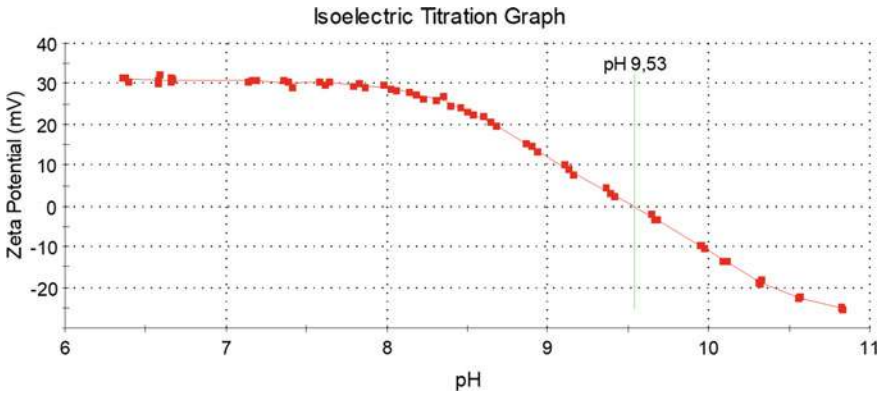


Fig. 5 Dependency of zeta potential of AlOOH crumpled nanosheets on pH of the medium at 37 °C

disordered defective surface structure will have a larger positive charge in comparison with regular aluminum oxyhydroxide.

3 Anticancer Activity of Radially Assembled Al Hydroxide Crumpled Nanosheets

Applied to tumor microenvironment, interaction of the positively charged crumpled nanosheets with extracellular anionic species can lead to disturbance of extracellular ion concentration, and thus result in anti-cancer activity. To investigate the anticancer activity of the agglomerates of radially assembled crumpled AlOOH nanosheets, termed Alohene ("Aloohene") referring to its chemical formula AlOOH, we have therefore performed a series of in vitro and in vivo studies using different cancerous cell lines and a mouse melanoma tumor model, respectively.

3.1 Effect of Aloohene on Tumor Cells Viability and Proliferation in Vitro

First, we have investigated the effect of the synthesized nanomaterial on the viability of three human cancer cell lines (MCF-7, UM-SCC-14C and Hela) cells. After 24 h incubation of Aloohene, a significant decrease in tumor cell viability was measured for all tested cells compared to the non-treated controls (Fig. 6a). Furthermore, a 30–37% decrease in proliferation of the tested cells was detected using the BrdU (5-bromo-2′-deoxyuridine) assay (Fig. 6b). Taken together, these data indicate that Aloohene affects tumor cells viability, most likely through inhibition of their proliferation.

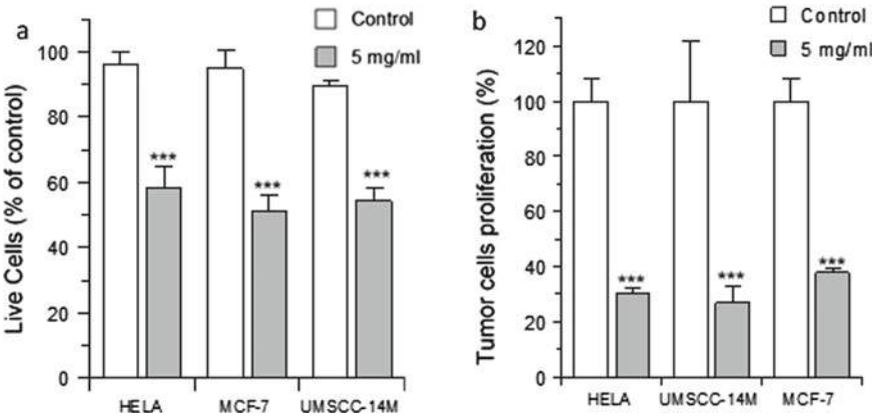


Fig. 6 Analysis of cellular toxicity of Aloohene. **a** FACS analysis of Annexin V and PI staining of Hela, MCF-7 and UM-SCC-14C cells with or without co-culture with 5 mg/ml of Aloohene at 37 °C for 24 h. Fluorescence intensity was measured by flow cytometry and data were analyzed by the Cell Quest software. **b** Proliferation of Hela, UM-SCC-14C and MCF-7 cells, as measured by BrDU assay. Cell cultures were incubated with Aloohene (5 mg/ml) at 37 °C for 24 h. Fluorescence intensity was measured 48 h after BrDU labeling at excitation and emission wavelengths of 370 nm and 470 nm, respectively. Results are means of 3 independent experiments. ***$p < 0.001$. Reproduced with permission from ACS [38]

3.2 Evaluation of Antitumor Activity of Aloohene in Mouse Model of Cancer

We next investigated whether the in vitro cancer cell growth inhibition effect of Aloohene would translate into antitumor efficacy in vivo in a melanoma mouse cancer model based on intradermal administration of 5×10^4 B16F10 cells into the C56BL/6J mice. When the tumor volume reached 70 mm^3, mice were treated with Aloohene, doxorubicin and a combination of both reagents. Aloohene was administered at a dose of 10 mg/ml weekly via two intratumoral injections and doxorubicin was injected intraperitoneally at a single dose of 10 mg/kg. The horizontal and vertical tumor diameters were measured by a digital calliper every second day until the end of treatment and volume was calculated using the formula $V = \frac{\pi}{6}ab^2$ where a and b are the longer and shorter diameter of the tumour, respectively. After two weeks of treatment, Aloohene treatment resulted in a significant decrease of tumor growth as compared to the vehicle treated control animals (Fig. 7). A combination therapy with doxorubicin, the standard-of-care anticancer drug, and Aloohene demonstrated a much stronger suppressive effect on tumor growth, also surpassing the anti-tumor effect of doxorubicin alone (Fig. 7).

To address the mechanism of tumor growth inhibition by Aloohene, we have measured the markers of tumor cells proliferation, cell death and vascularization. Notably, a significant decrease in the proliferation rate for all treatment regimens as compared to the control group was measured by immunohistochemical (IHC)

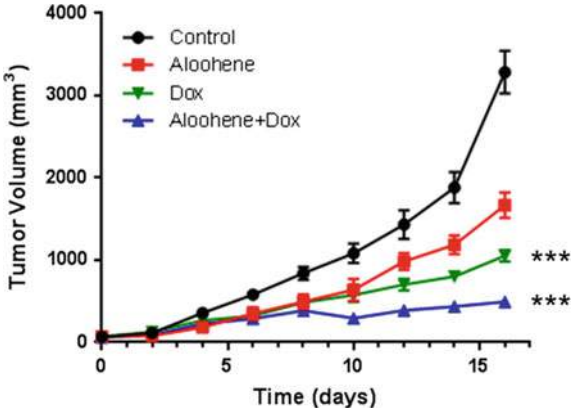


Fig. 7 Antitumor effect of Aloohene in vivo in a mouse melanoma model. **a** Effect of treatment with Aloohene alone and in combination with doxorubicin. Mice were treated with 10 mg/ml of Aloohene, 10 mg/kg of doxorubicin and their combination, and tumor volumes were measured twice a week. Data are presented as mean tumor volume ± standard errors of mean (n = 10 per treatment group). The statistical significance of differences between the groups was assessed by Student's t-test. ***$p < 0.001$ compared with the control group. Reproduced with permission from ACS [38]

staining for the proliferation marker Ki67 (Fig. 8), with the most profound effect detected after the combinatorial treatment of doxorubicin and Aloohene.

Further, the level of tumor necrosis, which could be indicative of treatment effect, was assessed by TUNEL (terminal deoxynucleotidyl transferase dUTP nick end labelling) staining. Large areas of dead cells were detected by this method in tumors treated with Aloohene, doxorubicin and their combination unlike in the control group ($p < 0.05$) (Fig. 9).

No difference in distribution of the endothelial cell marker CD31 was detected in the analyzed tumor sections, suggesting that tested treatments had no effect on tumor vascularization (Fig. 10). Taken together, these results demonstrate a significant in vivo antitumor effect of Aloohene and its potential to evolve into a novel strategy of cancer treatment, possibly in combination with established chemotherapy drugs, such as doxorubicin.

4 Summary

Collectively, our data demonstrates that AlOOH nanosheets with a disordered defective surface structure that have a large positive charge are capable of disturbing tumor-microenvironment extracellular ion balance, therefore representing a novel class of inorganic materials exhibiting strong antitumor effect. Here we report that Al hydroxide nanostructures in the form of agglomerates of crumpled and radially assembled nanosheets (Aloohene) trigger cancer cell death in vitro and inhibit

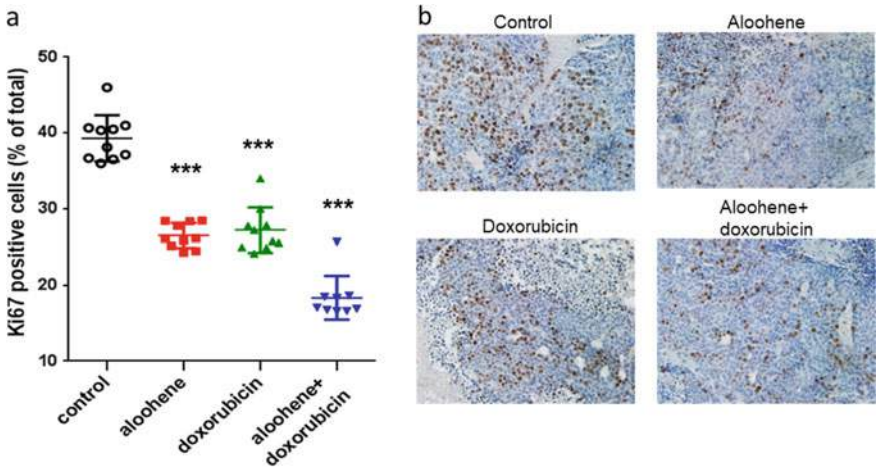


Fig. 8 Effect of Aloohene on tumor proliferation in vivo in mouse melanoma model. **a** Cell proliferation in primary melanomas determined by immunodetection of Ki67 in non-treated mice and mice treated with doxorubicin, Aloohene or their combination. Ki67-positive cells calculated as percentage of total cells. **b** Corresponding illustrative images of Ki67 staining in tumors of all tested groups. ***$p < 0.001$ compared with the control group. Reproduced with permission from ACS [38]

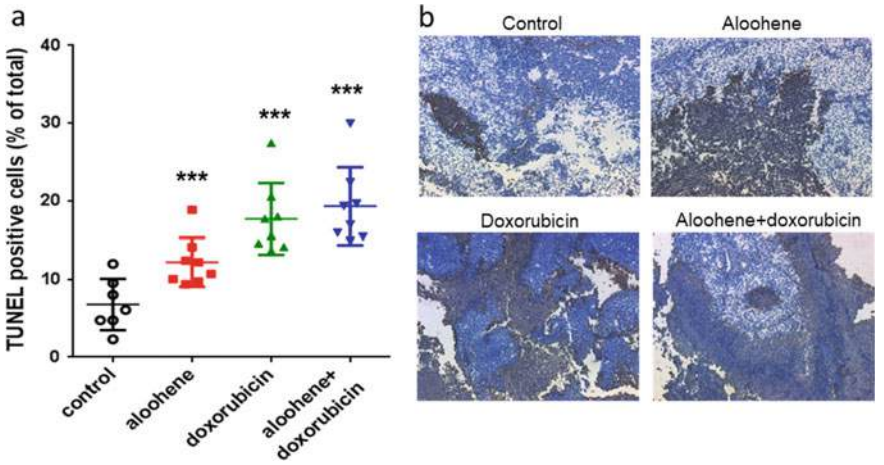


Fig. 9 Effect of Aloohene on tumor necrosis in vivo in mouse melanoma model. **a** The percentage of areas of dead cells was determined on high-power fields of primary melanomas of non-treated mice and mice treated with doxorubicin, Aloohene or their combination, based on immunohistochemical detection by terminal dUTP nick-end labeling staining (brown areas). The percentage of proliferative, necrotic and apoptotic cells are presented as means and standard errors, n = 10. **b** Representative images are shown for control mice and mice treated with Aloohene, doxorubicin, and their combination. The statistical significance of differences between the groups was assessed by Student's t-test. ***$p < 0.001$ compared with the control group. Reproduced with permission from ACS [38]

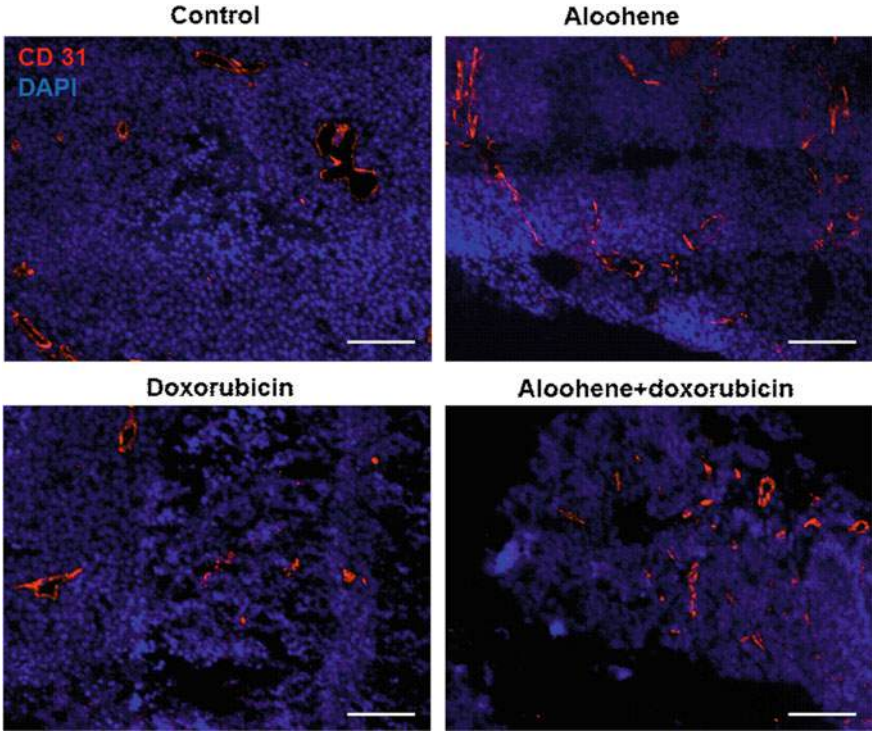


Fig. 10 Vascularization of skin melanoma tumors after treatment with Aloohene, doxorubicin or their combination. Representative images of immunofluorescence staining of the endothelial cell specific marker CD31 (red staining) in cryopreserved tumor sections. The scale bar corresponds to 200 μm. Reproduced with permission from ACS [38]

tumor growth in vivo. The disturbing effect of Aloohene on ion balance in the tumor-microenvironment is supported by our direct molecular dynamics simulation [38]. Furthermore, our results show that Aloohene nanostructures can potentiate the anticancer action of the cytotoxic agent doxorubicin and thus could improve the efficacy of state-of-the-art chemotherapy when used in combination. The findings of the present research highlight the important role of Aloohene, a novel class of tumor-microenvironment dysregulating nanomaterial based on AlOOH nanosheets agglomerates, in the development of more effective anticancer strategies.

Acknowledgements This work was inspired and motivated by Professor S. Psakhie. His visionary ideas for innovative scientific approaches layered foundation for multiple projects we have been collaborating over the years.

Author Contributions AL investigated the patterns of synthesis and the physicochemical characteristics of crumpled aluminum hydroxide nanostructures. G. M. and O. V. evaluated the effects of Aloohene on tumor cell growth inhibition in vitro and in vivo.

References

1. Yu MF, Files BS, Arepalli S, Ruoff RS (2000) Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties. Phys Rev Lett 84(24):5552–5555
2. Marie X, Urbaszek B (2015) 2D materials: ultrafast exciton dynamics. Nat Mater 14(9):860–861
3. Hayamizu Y, So CR, Dag S, Page TS, Starkebaum D, Sarikaya M (2016) Bioelectronic interfaces by spontaneously organized peptides on 2D atomic single layer materials. Sci Rep 6:33778
4. Wang ZL (2004) Nanostructures of zinc oxide. Materialstoday 7(6):26–33
5. Mamalis AG (2007) Recent advances in nanotechnology. J Mater Process Technol 181(1):52–58
6. Tallinen T, Åström JA, Timonen J (2009) The effect of plasticity in crumpling of thin sheets. Nat Mater 8(1):25–29
7. Zang J, Ryu S, Pugno N, Wang Q, Tu Q, Buehler MJ, Zhao X (2013) Multifunctionality and control of the crumpling and unfolding of large-area graphene. Nat Mater 12(4):321–325
8. Wang Q, Zhao X (2015) A three-dimensional phase diagram of growth-induced surface instabilities. Sci Rep 5:8887
9. Holmes DP, Ursiny M, Crosby AJ (2008) Crumpled surface structures. Soft Matter 4(1):82–85
10. Lozhkomoev AS, Glazkova EA, Bakina OV, Lerner MI, Gotman I, Gutmanas EY, Kazantsev SO, Psakhie SG (2016) Synthesis of core–shell AlOOH hollow nanospheres by reacting Al nanoparticles with water. Nanotechnology 27(20):205603
11. Cai W, Chen S, Yu J, Hu Y, Dang C, Ma S (2013) Template-free solvothermal synthesis of hierarchical boehmite hollow microspheres with strong affinity toward organic pollutants in water. Mater Chem Phys 138(1):167–173
12. Tsukanov AA, Psakhie SG (2016) Energy and structure of bonds in the interaction of organic anions with layered double hydroxide nanosheets: a molecular dynamics study. Sci Rep 6:19986
13. Bakina OV, Svarovskaya NV, Glazkova EA, Lozhkomoev AS (2015) Flower-shaped ALOOH nanostructures synthesized by the reaction of an AlN/Al composite nanopowder in water. Adv Powder Technol 26(6):1512–1519
14. Mikhaylov G, Mikac U, Magaeva AA, Itin VI, Naiden EP, Psakhye I, Babes L, Reinheckel T, Peters C, Zeiser R, Bogyo M, Turk V, Psakhye SG, Turk B, Vasiljeva O (2011) Ferri-liposomes as an MRI-visible drug-delivery system for targeting tumours and their microenvironment. Nat Nanotechnol 6(9):594–602
15. Thakor AS, Gambhir SS (2013) Nanooncology: the future of cancer diagnosis and therapy. CA Cancer J Clin 63(6):395–418
16. Allegra A, Penna G, Alonci A, Rizzo V, Russo S, Musolino C (2011) Nanoparticles in oncology: the new theragnostic molecules. Anti-Cancer Agents Med Chem (Formerly Curr Med Chem-Anti-Cancer Agents) 11(7):669–686
17. Joseph D, Tyagi N, Geckeler C, Geckeler KE (2014) Protein-coated pH-responsive gold nanoparticles: microwave-assisted synthesis and surface charge-dependent anticancer activity. Beilstein J Nanotechnol 5(1):1452–1462
18. Akhtar MJ, Ahamed M, Kumar S, Khan MM, Ahmad J, Alrokayan SA (2012) Zinc oxide nanoparticles selectively induce apoptosis in human cancer cells through reactive oxygen species. Int J Nanomed 7:845–857
19. Bisht G, Rayamajhi S (2016) ZnO nanoparticles: a promising anticancer agent. Nanobiomedicine (Rij) 3:9
20. Kulkarni M, Mazare A, Gongadze E, Perutkova Š, Kralj-Iglič V, Milošev I, Schmuki P, Iglič A, Mozetič M (2015) Titanium nanostructures for biomedical applications. Nanotechnology 26(6):062002
21. Elhissi AMA, Ahmed W, Hassan IU, Dhanak VR, D'Emanuele A (2011) Carbon nanotubes in cancer therapy and drug delivery. J Drug Deliv 2012, Article ID 837327
22. Theodosopoulos GV, Bilalis P, Sakellariou G (2015) Polymer functionalized graphene oxide: a versatile nanoplatform for drug/gene delivery. Curr Org Chem 19(18):1828–1837

23. Kumar VB, Kumar K, Gedanken A, Paik P (2014) Facile synthesis of self-assembled spherical and mesoporous dandelion capsules of ZnO: efficient carrier for DNA and anti-cancer drugs. J Mater Chem B 2(5):3956–3964
24. Lin N, Sun JM, Hsiao J-H, Hwu Y (2009) Spontaneous emergence of ordered phases in crumpled sheets. Phys Rev Lett 103(26):263902
25. Wang T, Wan Y, Zhanqiang L (2016) Fabrication of hierarchical micro/nanotopography on bio-titanium alloy surface for cytocompatibility improvement. J Mater Sci 51(21):9551–9561
26. Ulanova M, Tarkowski A, Hahn-Zoric M, Hanson LÅ (2001) The common vaccine adjuvant aluminum hydroxide up-regulates accessory properties of human monocytes via an interleukin-4-dependent mechanism. Infect Immun 69(2):1151–1159
27. Dong E, Wang Y, Yang S-T, Yuan Y, Nie H, Chang Y, Wang L, Liu Y, Wang H (2011) Toxicity of nano gamma alumina to neural stem cells. J Nanosci Nanotechnol 11(9):7848–7856
28. Xie Y, Kocaefe D, Kocaefe Y, Cheng J, Liu W (2016) The effect of novel synthetic methods and parameters control on morphology of nano-alumina particles. Nanoscale Res Lett 11(1):1–11
29. Lerner MI, Svarovskaya NV, Psakhie SG, Bakina OV (2009) Production technology, characteristics, and some applications of electric-explosion nanopowders of metals. Nanotechnol Russ 4(11):741–757
30. Svarovskaya NV, Bakina OV, Glazkova EA, Lerner MI, Psakh'e SG (2010) The formation of nanosheets of aluminum oxyhydroxides from electroexposive nanopowders. Russ J Phys Chem A 84(9):1566–1569
31. Yang Y, Gai W-Z, Deng Z-Y, Zhou J-G (2014) Hydrogen generation by the reaction of Al with water promoted by an ultrasonically prepared $Al(OH)_3$ suspension. Int J Hydrogen Energy 39(33):18734–18742
32. Gai W-Z, Deng Z-Y (2014) Effect of initial gas pressure on the reaction of Al with water. Int J Hydrogen Energy 39(25):13491–13497
33. Rosenband V, Gany A (2010) Application of activated aluminum powder for generation of hydrogen from water. Int J Hydrogen Energy 35(20):10898–10904
34. Deng Z-Y, Ferreiraw JMF, Tanaka Y, Ye JH (2007) Physicochemical mechanism for the continuous reaction of gamma-Al_2O_3-modified aluminum powder with water. J Am Ceram Soc 90(5):1521–1526
35. Sarathi R, Sankar B, Chakravarthy SR (2010) Influence of nano aluminium powder produced by wire explosion process at different ambience on hydrogen generation. J Electr Eng-Elektrotechnicky Cas 61(4):215–221
36. Ivanov VG, Safronov MN, Gavrilyuk OV (2001) Macrokinetics of oxidation of ultradisperse aluminum by water in the liquid phase. Combust Explosion Shock Waves 37(2):173–177
37. Bakina OV, Svarovskaya NV, Glazkova EA, Lozhkomoev AS, Khorobraya EG, Lerner MI (2015) Flower-shaped ALOOH nanostructures synthesized by the reaction of an AlN/Al composite nanopowder in water. Adv Powder Technol 26(6):1512–1519
38. Lerner MI, Mikhaylov G, Tsukanov AA, Lozhkomoev AS, Gutmanas E, Gotman I, Bratovs A, Turk V, Turk B, Psakhye SG, Vasiljeva O (2018) Crumpled aluminum hydroxide nanostructures as a microenvironment dysregulation agent for cancer treatment. Nano Lett 18(9):5401–5410
39. Choy J-H, Park M (2004) Clay surfaces: fundamentals and applications. Academic Press, Cambridge, pp 403–424 (Cationic and anionic clays for biological applications)

Influence of Lattice Curvature and Nanoscale Mesoscopic Structural States on the Wear Resistance and Fatigue Life of Austenitic Steel

Viktor E. Panin, Valery E. Egorushkin, and Natalya S. Surikova

Abstract The gauge dynamic theory of defects in a heterogeneous medium predicts the nonlinearity of plastic flow at low lattice curvature and structural turbulence with the formation of individual dynamic rotations at high curvature of the deformed medium. The present work is devoted to the experimental verification of the theoretical predictions. Experimentally studied are the influence of high-temperature radial shear rolling and subsequent cold rolling on the internal structure of metastable Fe–Cr–Mn austenitic stainless steel, formation of nonequilibrium ε- and α′-martensite phases, appearance of dynamic rotations on fracture surfaces, fatigue life in alternating bending, and wear resistance of the material. Scratch testing reveals a strong increase in the damping effect in the formed hierarchical mesosubstructure. The latter is responsible for a nanocrystalline grain structure in the material, hcp ε martensite and bcc α′ martensite in grains, a vortical filamentary substructure on the fracture surface as well as for improved high-cycle fatigue and wear resistance of the material. This is related to a high concentration of nanoscale mesoscopic structural states, which arise in lattice curvature zones during high-temperature radial shear rolling combined with smooth-roll cold rolling. These effects are explained by the self-consistent mechanical behavior of hcp ε-martensite laths in fcc austenite grains and bcc α′-martensite laths that form during cold rolling of the steel subjected to high-temperature radial shear rolling.

Keywords Gauge dynamic theory of defects · Nanoscale mesoscopic structural states · Lattice curvature · Damping effect · Dynamic rotations · Fatigue failure · Wear resistance

V. E. Panin (✉) · V. E. Egorushkin · N. S. Surikova
Institute of Strength Physics and Materials Science SB RAS, Tomsk 634021, Russia
e-mail: paninve@ispms.tsc.ru

N. S. Surikova
e-mail: surikova@ispms.ru

V. E. Panin
National Research Tomsk Polytechnic University, Tomsk 634050, Russia

National Research Tomsk State University, Tomsk 634050, Russia

G.-P. Ostermeyer et al. (eds.), *Multiscale Biomechanics and Tribology of Inorganic and Organic Systems*, Springer Tracts in Mechanical Engineering,
https://doi.org/10.1007/978-3-030-60124-9_11

1 Introduction

Mechanical behavior of metals were subject of intensive theoretical and experimental studies over many decades. Conventionally, mechanisms of mechanical behavior under various loading conditions are associated with various-scale strain-induced defects, including cracks, in a translation-invariant crystal lattice. However, in most cases, the translational invariance of the crystal lattice in a deformed solid is strongly violated, which does not allow for a correct description of plastic deformation and fracture within the linear approach of Newtonian mechanics. In this work, we thoroughly study the mechanical behavior of Fe–Cr–Mn austenitic steel, in which complex structural phase transformations of the initial fcc lattice with the formation of hcp and bcc martensite phases occur in uniaxial tension. According to [1, 2], crystal lattice transformations are related to the appearance of nanoscale mesoscopic structural states at the interstices of lattice curvature zones. Thus, the account for lattice curvature provides a basis for nonlinear solid mesomechanics.

In this work, crystal lattice curvature throughout the Fe–Cr–Mn austenitic steel specimen is produced using the complex treatment by high-temperature radial shear rolling + cold rolling at room temperature. This complex treatment reveals important new regularities in the mechanical behavior of austenitic steel.

2 Gauge Dynamic Theory of Defects in the Heterogeneous Medium

2.1 Basic Equations of the Gauge Theory

A deformable solid as a multilevel hierarchically organized system requires a self-consistent description at the nano-, micro-, meso- and macroscopic levels. Gauge dynamic theory of defects in the heterogeneous medium provides a framework for such description. In this theory, fluxes of strain-induced defects J and their density α are described by the equations

$$\frac{\partial}{\partial x_\alpha} J_\mu^\alpha = -\frac{\partial \ln u_\alpha(x,t)}{\partial t}, \tag{1}$$

$$\varepsilon_{\mu\chi\delta} \frac{\partial J_\delta^\alpha}{\partial x_\chi} = -\frac{\partial \alpha_\mu^\alpha}{\partial t}, \tag{2}$$

$$\frac{\partial \alpha_\mu^\alpha}{\partial x_\alpha} = 0, \tag{3}$$

$$\varepsilon_{\mu\chi\delta} \frac{\partial \alpha_\delta^\alpha}{\partial x_\chi} = \frac{1}{c^2} \frac{\partial J_\mu^\alpha}{\partial t} + \frac{\partial \ln u_\beta(x,t)}{\partial x_\nu} \frac{C_{\alpha\beta}^{\mu\nu}}{E} - P_\nu^\beta \frac{C_{\alpha\beta}^{\mu\nu}}{E}, \tag{4}$$

$$\frac{1}{c^2}\frac{\partial^2 \ln u_\alpha(x,t)}{\partial t^2} + \frac{\partial^2 \ln u_\beta(x,t)}{\partial x_\mu \partial x_\nu}\frac{C^{\mu\nu}_{\alpha\beta}}{E} = -P^\beta_\nu \frac{C^{\mu\nu}_{\alpha\beta}}{E} \tag{5}$$

Equations (1)–(5) have the following sense and use the following notations:

(1) is the continuity equation for a defected medium, which indicates that the source of plastic flow causes the defect flow rate;
(2) is the condition of plastic-strain compatibility, the time variation of the medium density is determined in this case by the flux rotor, i.e. its heterogeneity, rather than by the divergence;
(3) is the condition of continuity of defects, which corresponds to the absence of charges of the rotational component of the plastic strain field $\left(\alpha^\beta_\chi = \varepsilon_{\chi\mu\nu}\partial_\mu P^\beta_\nu\right)$;
(4) is the governing equation for the medium with plastic flow;
(5) is the equation of quasi-elastic equilibrium, which presents the known continuum mechanics equation. Along with the elastic strain, it contains plastic distortions on the right-hand side. In fact, this summand corresponds to the nucleation of strain-induced defects in local zones of hydrostatic tension due to stress concentrators.

Expression (4) is applicable only to the medium with plastic flow. It relates the time variation in plastic flow to the anisotropic spatial variation in the defect density $\varepsilon_{\mu\chi\delta}\partial\alpha^\alpha_\delta/\mathrm{d}x$ and sources $\left(\sigma^\alpha_\mu - P^\beta_\nu C^{\mu\nu}_{\alpha\beta}/E\right)$. Equations (4) and (5) differ from the corresponding elastic equations in that the time variation in plastic strain rate is determined by the stresses themselves rather than by $\partial\sigma^\alpha_\mu/\partial x$, as in the elastic case. In addition, the right-hand side of (4) includes plastic distortion $P^\beta_\nu(x,t)$ taken as sources, which indicates the duality of defects as field sources.

From the system of Eqs. (1)–(5), wave equations can be found for the dimensionless quantities of defect flux J and density α:

$$\frac{1}{c^2}\frac{\partial^2 J^\mu_\alpha}{\partial t^2} - \frac{\partial^2 J^\mu_\alpha}{\partial x^2} = \frac{\partial}{\partial t}\left\{\frac{\partial \ln u_\alpha(x,t)}{\partial x_\mu} - \frac{1}{E}\frac{\partial \ln u_\beta}{\partial x_\nu}C^{\mu\nu}_{\alpha\beta} - \frac{1}{E}P^\beta_\nu C^{\mu\nu}_{\alpha\beta}\right\} \tag{6}$$

$$\frac{1}{c^2}\frac{\partial^2 \alpha^\mu_\alpha}{\partial t^2} - \frac{\partial^2 \alpha^\mu_\alpha}{\partial x^2_\nu} = \varepsilon_{\mu\chi\sigma}\left\{\frac{\partial^2 \ln u_\beta(x,t)}{\partial x_\chi \partial x_\nu}C^{\mu\nu}_{\alpha\beta} - \frac{\partial P^\beta_\nu}{\partial x_\chi}C^{\mu\nu}_{\alpha\beta}\right\}\frac{1}{E}, \tag{7}$$

subject to the source compatibility condition

$$\frac{\partial N_\mu}{\partial t} + \varepsilon_{\mu\ell m}\frac{\partial M_\mu}{\partial x_\ell} = 0, \tag{8}$$

where M is the right-hand side of expression (6), N is the right-hand side of expression (7), and $u(x,t)$ is the inelastic displacements in the wave of inelastic localized strain.

The right-hand side of Eq. (6) characterizes defect flow sources. They are determined by the rate of quasi-elastic strain $\frac{\partial}{\partial t}\left(E^\alpha_\mu E - E^\beta_\nu C^{\mu\nu}_{\alpha\beta}\right)\frac{1}{E}$. Parenthesized is the difference between internal compressive (tensile) stresses and shear stresses

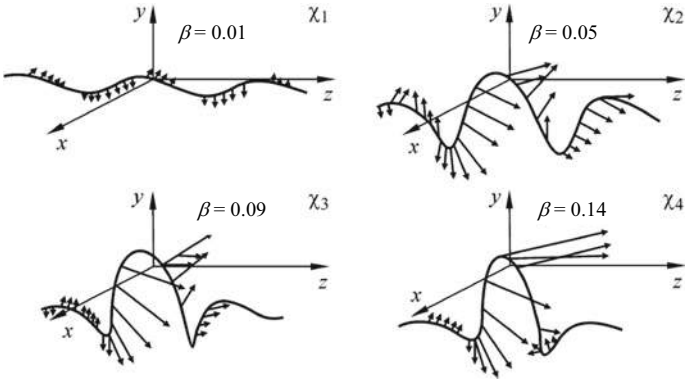


Fig. 1 Shape and velocity of plastic deformation depending on the curvature of the deformed region

associated with the stress distribution in the stress concentration zone. Relaxation processes of defect rearrangement (such as various atomic configurations or their conglomerates) are presented in (6) by the term $P_{\nu}^{\beta} C_{\alpha\beta}^{\mu\nu} / E$.

The right-hand side of Eq. (7) expresses a source of the strain-induced defect density. This is vorticity $\varepsilon_{\mu\chi\delta} \frac{\partial}{\partial x} \left(E_{\nu}^{\beta} - P_{\nu}^{\beta} \right) \frac{C_{\alpha\beta}^{\mu\nu}}{E}$ of shear strain induced by the shear stress relaxation in local zones of hydrostatic tension during the defect formation.

The wave pattern of strain-induced defect flows is defined by the right-hand side of Eqs. (6) and (7). Plastic distortion $P_{\nu}^{\beta}(x, t)$ plays a major role in the wave pattern of localized plastic flow.

Equation (6) for the strain-induced defect flux along the direction L (at $r < L$) is solved as

$$\overrightarrow{J} = \frac{b_1 - b_2}{4\pi} \chi(s, t) \overrightarrow{b}(s, t) \left(\ln \frac{2L}{r} - 1 \right) - \nabla f, \tag{9}$$

where $\overrightarrow{b}$ is the binormal vector in the local coordinate system, $\vec{n}$ is the normal, $\vec{t}$ is the tangent, χ is the curvature variation in a deformed region due to external loading, s is the current length of the region, b_1 and b_2 are the Burgers vector moduli of the bulk translational and subsurface rotational inconsistency, respectively, ∇f is the gradient part of the flux due to third-party sources.

The expression for the defect flux includes curvature χ of the deformed medium. Let us analyse the role of this factor in the behavior of the strain-induced defect flow.

2.2 Structural Turbulence at Severe Lattice Curvature

The wave pattern of localized plastic flow in Eqs. (6) and (7) is generally shown in Fig. 1. It demonstrates the dependence of the wave profile of localized plastic flow on the curvature of the deformed region $\chi(x,t) = 4\beta\,\mathrm{sech}[2\beta(x + 4\upsilon t)]$.

In Fig. 1, it can be seen that an increase in the curvature of the deformed region greatly changes the shape and velocity of plastic deformation. The experiment confirms the theoretical prediction [1].

At a high curvature of the deformed medium, plastic distortion $P_{\nu}^{\beta}(x,t)$ strongly grows, too. This means that the right-hand sides of Eqs. (6) and (7) increase significantly. The wave pattern of plastic flow cannot be preserved in this case, and plastic deformation becomes turbulent, breaking up into individual dynamic rotations. Such a theoretical prediction of structural turbulence in a deformed solid was impossible under the conditions of translational invariance of the crystal. The violation of translational invariance and account of curvature in a deformed medium are fundamentally new. Their experimental confirmation was earlier obtained [2].

Severe curvature of plastic flow was previously achieved by introducing titanium carbonitride nanoparticles into a deformed solid [2]. Unlike carbides, which have a spherical configuration, carbonitrides have a cubic structure [3, 4]. Due to plastic flow in the presence of titanium carbonitride nanoparticles, a severe curvature is formed in the deformed medium, especially under shock loading. Figure 2 shows the fracture surface of low-carbon steel 09Mn2Si when measuring impact toughness with 0.15% titanium carbonitride nanoparticles introduced into the steel. The New View profilometer discovers pronounced dynamic rotations of turbulent flow on the surface in the form of individual vortical protrusions. The scratch test shows a significant decrease in the groove size in the presence of nanoparticles (Fig. 3b), in contrast to

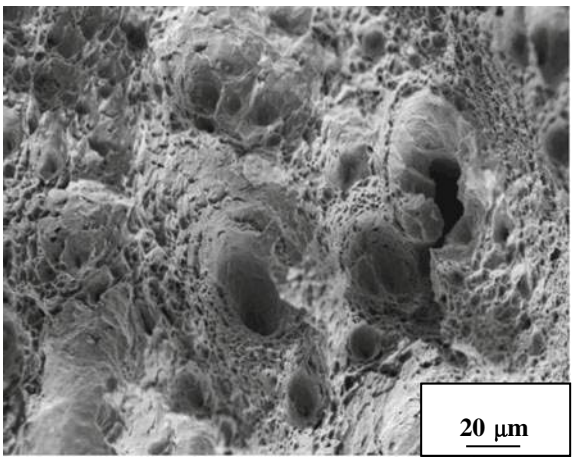


Fig. 2 Structural turbulence on the fracture surface of low-carbon steel 09Mn2Si, with 0.15% TiCN nanoparticles introduced into it

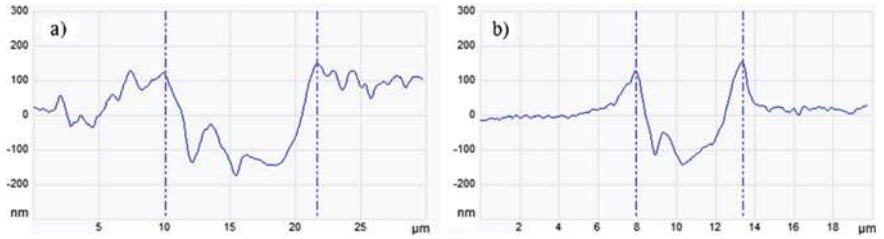


Fig. 3 Cross profiles of the scratch groove in the 09Mn2Si specimens: initial (**a**) and with 0.15% TiCN nanoparticles (**b**)

a deep groove forming in the absence of nanoparticles (Fig. 3a).

Structural turbulence in a solid is characterized by relay-race transfer of momentum from particle to particle. This is what makes it different from classical turbulence characterized by the Avogadro number [5, 6].

Structural turbulence is even more pronounced on the fracture surface of a chevron-notched specimen shown in Fig. 4. Counter shear loads from the specimen notches induce couple stresses that form dynamic rotations (Fig. 5). A filamentary structure appears around the notches, which bears witness not to translational invariance but to the formation of special structural states in lattice curvature zones. These states also arise during the formation of dynamic rotations (Fig. 6). In other words, at the crystal structure curvature the electronic subsystem forms special structural states that are responsible for dynamic rotations by the mechanism of plastic distortion.

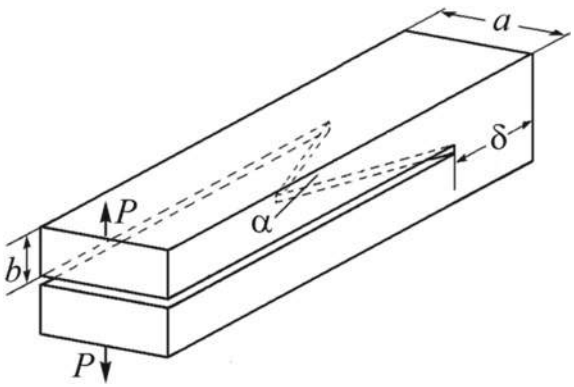


Fig. 4 Geometry of a chevron-notched specimen

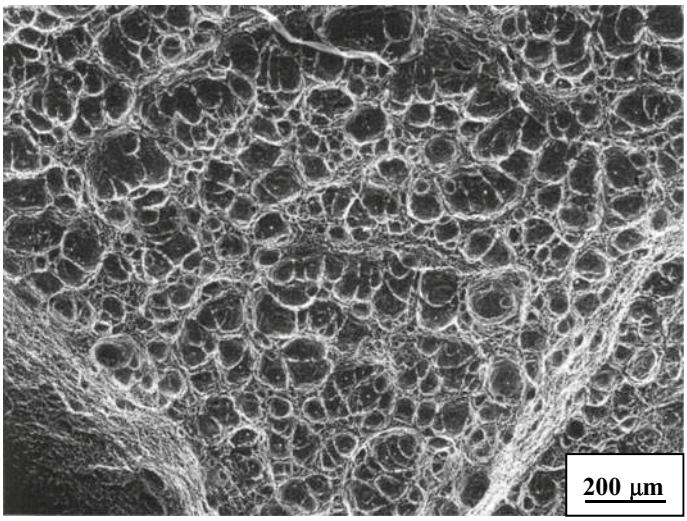


Fig. 5 Formation of a filamentary structure in the vicinity of the edge-cut notches in the chevron-notched 09Mn2Si steel specimen in fracture

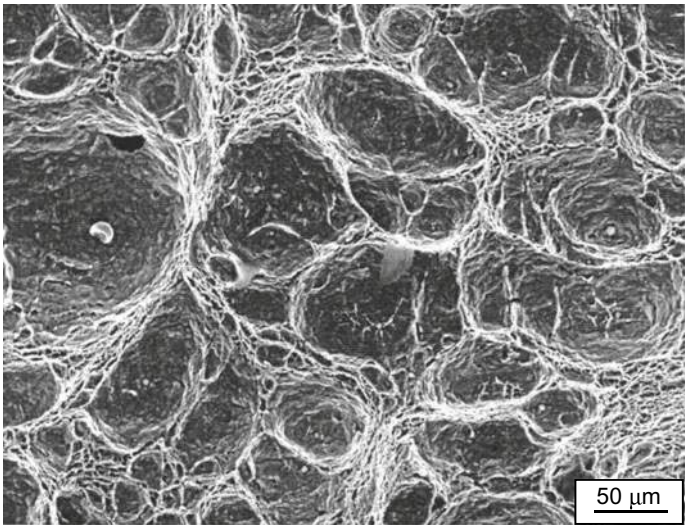


Fig. 6 Formation of noncrystallographic boundaries due to interstitial mesoscopic structural states in turbulent rotations on the fracture surface of a chevron-notched 09Mn2Si steel specimen

3 Role of Lattice Curvature in the Mechanical Behavior of Austenitic Steel

3.1 Influence of High-Temperature Radial Shear Rolling and Subsequent Smooth-Roll Cold Longitudinal Rolling on the Austenitic Steel Microstructure

High-temperature radial shear rolling exerts no effect on the phase composition of austenitic steel. It forms a layered structure with varying degrees of the crystal lattice curvature-torsion and refines the grain structure. In the near-surface layers, the material has a globular structure with the average grain size $d \approx 0.57$ μm; grains in the near-axial zone are elongated along the rolling direction. A small-angle substructure is well developed within the grains. The fraction of special twin boundaries with a misorientation of ~60° is significantly reduced. The yield stress $\sigma_{0.2}$ and ultimate strength σ_B increase, respectively, from 400 to 620 MPa and from 850 to 1050 MPa. The ductility decreases from 90 to 55% (the true strain $\varepsilon = 0.7$).

Radial shear and cold rolling to the cold strain $\varepsilon = 1.8$ changes significantly the austenitic steel microstructure. In austenite grains, a two-phase (ε martensite + α' martensite) nanocrystalline mesosubstructure is formed. Grains containing α' martensite with a distorted bcc structure prevail. According to X-ray structural estimates, the volume fraction of α martensite after rolling to $\varepsilon_{\text{true}} \sim 1.8$ comprises ≈85.6%, and the size of coherent scattering regions is 40 nm. Cold rolling forms a heterogeneous structure. In the specimen areas with a high plastic strain, the grain size is 40–100 nm. The specimen areas with a lower strain (Fig. 7) contain austenite grains 200–400 nm in size (Fig. 7b) and grains with ferrite zones 30–400 nm in size (Fig. 7d).

Fragmentation processes during cold rolling begin with the formation of fine stacking faults and ε-martensite nuclei (Fig. 8a) or α′-martensite plates (Fig. 8b) in various grains of the material.

The results presented in Figs. 7 and 8 indicate that ε-martensite and α′-martensite laths are important intermediate phases for the structural transformation of fcc austenite to bcc ferrite. Of special note is a strong texture of the intragranular structure. In austenite crystals, the direction and plane of rolling are close to $\langle 111 \rangle_\gamma$ and $\{110\}_\gamma$, respectively; in ferrite zones, to $\langle 110 \rangle_\alpha$ and $\{111\}\alpha_\gamma$.

Thus, within a nanostructured austenitic steel subjected to the complex treatment by radial shear rolling + smooth-roll cold rolling a multiscale hierarchically organized structure is formed that allows translational-rotational modes of plastic deformation from macro to nanoscale levels. Let us consider the mechanisms of such plastic deformation in fracture by uniaxial tension.

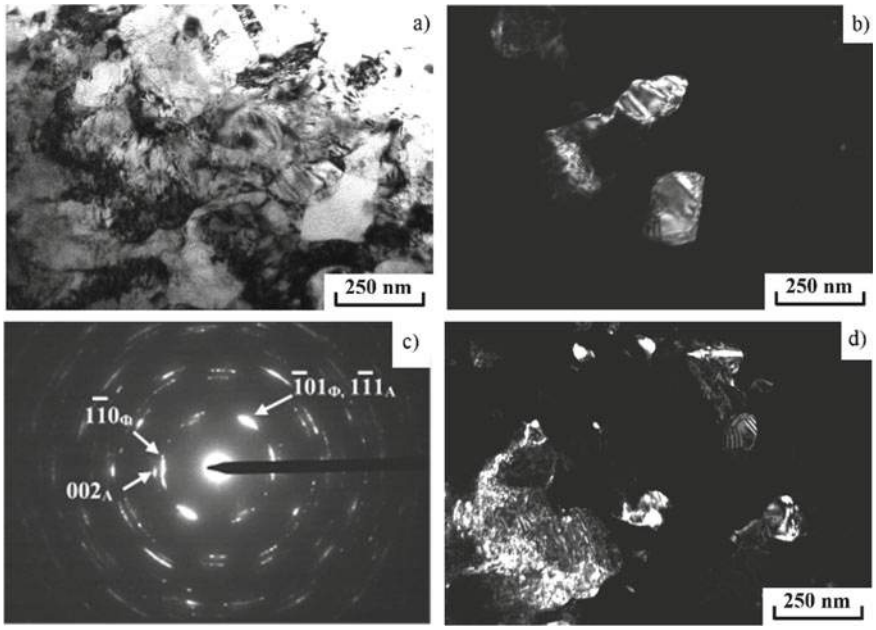


Fig. 7 Structure of steel 12Cr15Mn9NiCu after stepwise radial shear hot rolling to ε ~ 65% and subsequent cold rolling to $\varepsilon \sim 80\%$. **a** The bright field; **c** the selected-area diffraction image, shown are the two zones of reflections with azimuthal misorientations: [110] austenite zone and [111] ferrite zone; **b** the dark-field image in the 002_{A} austenite reflection; **d** the dark-field image in the $1\bar{1}0_{\Phi}$ ferrite reflection

3.2 *Fracture Surface in Uniaxial Tension of Austenitic Steel Specimens After Various Treatments*

The fracture mode in tension of austenitic steel in the initial state and after the complex treatment is schematized in Fig. 9. In the initial steel, fracture starts with the propagation of an opening mode crack (Fig. 9a) and ends with the propagation of a tearing mode crack. Specimens subjected to radial shear rolling are fractured only by the propagation of a tearing mode crack (Fig. 9b).

The fracture surfaces of the steel specimens in the initial state (Figs. 10, 11 and 12) bears witness to the spatial stress state in fracture. After the treatment by radial shear rolling + cold rolling, the specimens under tension are fractured in the plane stress state (Fig. 13).

The fracture pattern of the initial steel specimens in the scale hierarchy of the grain structure is shown in Figs. 10, 11 and 12. Grain-boundary sliding of grain conglomerates (Fig. 10) is accompanied by stochastic microcracking and initiates individual accommodation rotations of individual grains with the formation of micropores in them (Fig. 11). A pronounced dimple relief on the opening mode surface (Fig. 12a) accompanied by the formation of microporosity is a sign of ductile fracture [7, 8].

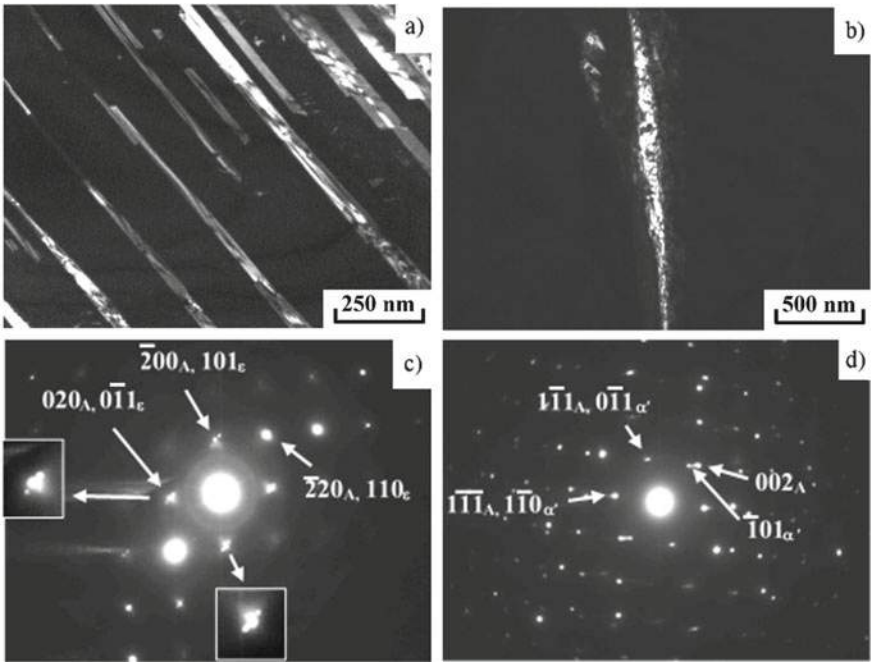


Fig. 8 **a** The dark-field image of ε-martensite laths in the $0\bar{1}1_{\varepsilon}$ reflection. **c** The selected-area diffraction image for (**a**), shown are reflections in the two zones: [001] austenite zone and [$\bar{1}$11] ε-martensite zone; the arrangement and form of the reflections are given at higher magnification on the left and at the bottom of the image. **b** The dark-field of α′-martensite laths in the $10\bar{1}_{\alpha}$ reflection. **d** The selected-area diffraction image for (**b**), shown are reflections in the two parallel zones: [110] austenite zone and [111] α′-martensite zone

The micropore formation is associated with the coalescence of vacant lattice sites under the conditions of plastic distortion at the lattice curvature interstices [9, 10].

Quasi-elastic cleavages in the zones of tearing mode cracks (Figs. 11b and 12b) also exhibit microporosity. This shows the important role of accommodation processes of plastic distortion at the nanoscale level at various fracture mechanisms of the initial austenitic steel. In other words, the self-consistency of rotational deformation modes in a wide range of scales, from macro to nano, causes the high ductility $\delta = 90\%$ in uniaxial tension of the initial austenitic steel.

A high concentration of transition martensite phases in fcc austenite grains with ε martensite and in grains with bcc ferrite zones in α′ martensite, which is due to radial shear plus cold rolling of the steel, provides a means for dynamic rotations as a mechanism of structural turbulence (Fig. 13). The phenomenon of structural turbulence in a deformed solid is detailed elsewhere [5, 6]. The presence of nanoscale mesoscopic structural states in the lattice curvature interstices makes it possible to synthesize vortical nanofilaments of a material under the structural turbulence condition. A hierarchy of dynamic rotations on fracture surfaces was numerously

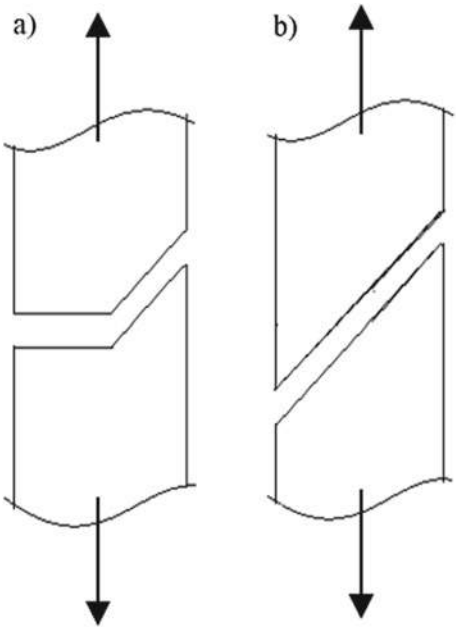


Fig. 9 Schematic of the propagation of the main crack in fracture by uniaxial tension of the specimens of the initial austenite steel (**a**) and after the complex treatment by radial shear rolling + cold rolling (**b**)

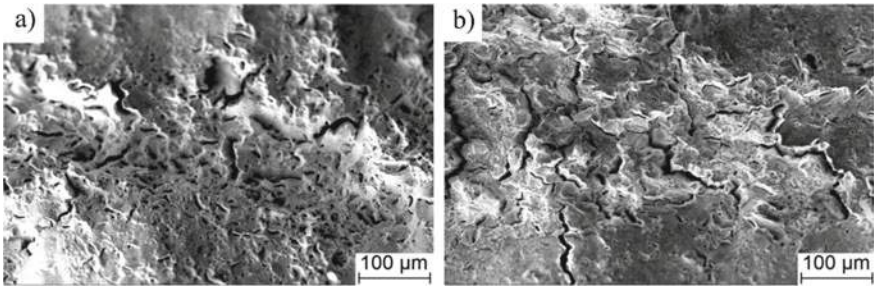


Fig. 10 Stochastic cracks on the fracture surface in tension of the specimens made of the initial austenitic steel: **a** opening mode zone; **b** tearing mode zone

observed [9, 11–13]. However, the mechanism of their formation and their relation to structural turbulence of plastic flow of atoms in lattice curvature zones have not been discussed yet.

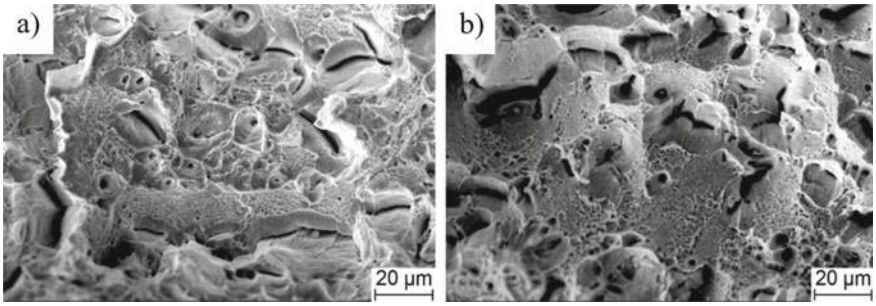


Fig. 11 Rotational mode of a grain conglomerate in the initial steel in the opening mode zone accommodated by rotations of individual grains with the formation of micropores in them (**a**); cleavages of grain conglomerates with the formation of micropores in the tearing mode zone (**b**)

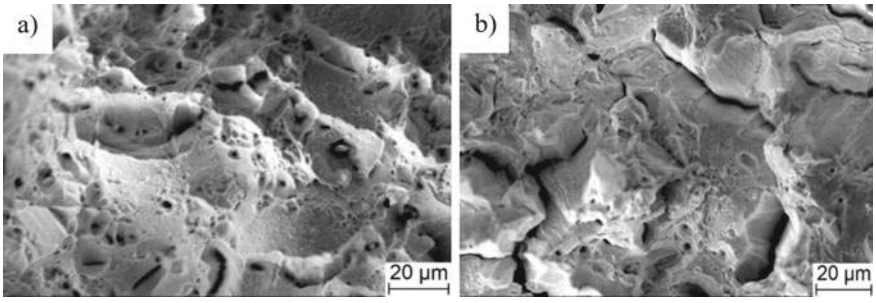


Fig. 12 Ductile dimple fracture of the initial steel in the opening mode zone (**a**); quasi-elastic cleavages of grain conglomerates of the initial steel in the tearing mode zone (**b**)

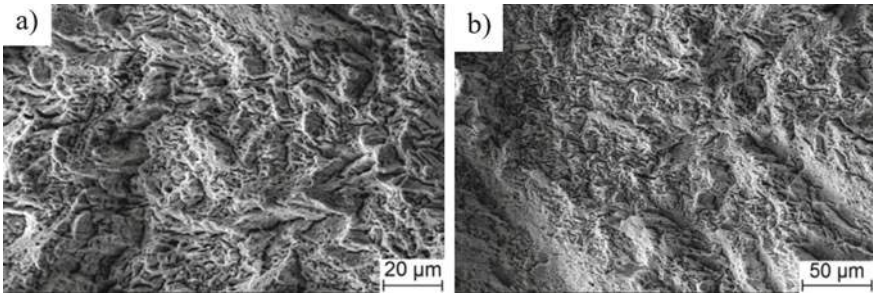


Fig. 13 Dynamic rotations with the formation of a vortical filamentary structure on the fracture surface of the steel specimens after radial shear and cold rolling: initial (**a**) and final (**b**) zones of the main tearing mode crack