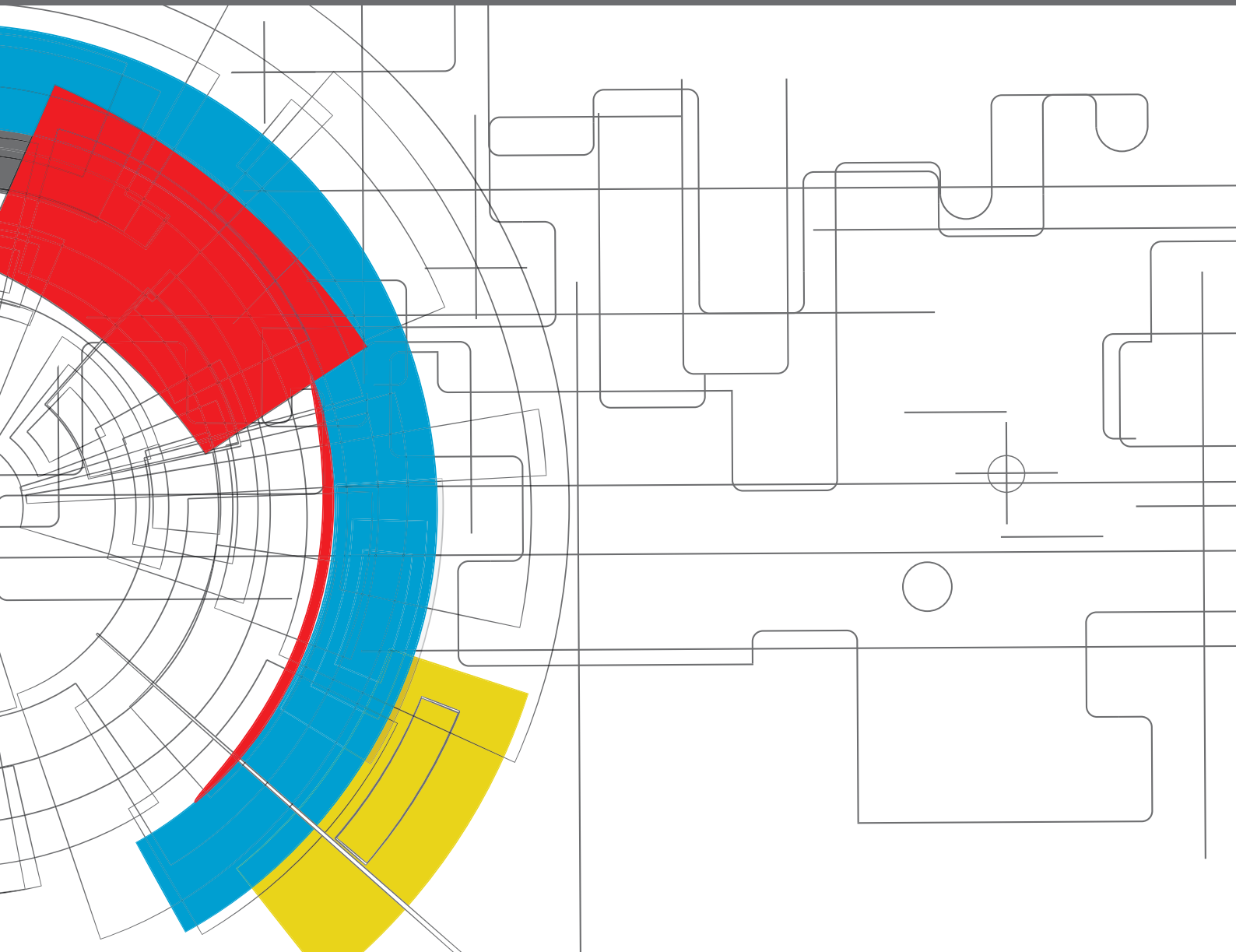


CONTACT MECHANICS PERSPECTIVE OF TRIBOLOGY

EDITED BY: Irina Goryacheva, Marco Paggi and Valentin L. Popov
PUBLISHED IN: Frontiers in Mechanical Engineering





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-761-1

DOI 10.3389/978-2-88966-761-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CONTACT MECHANICS PERSPECTIVE OF TRIBOLOGY

Topic Editors:

Irina Goryacheva, Institute for Problems in Mechanics (RAS), Russia

Marco Paggi, IMT School for Advanced Studies Lucca, Italy

Valentin L. Popov, Technical University of Berlin, Germany

Citation: Goryacheva, I., Paggi, M., Popov, V. L., eds. (2021). Contact Mechanics Perspective of Tribology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-761-1

Table of Contents

06 Editorial: Contact Mechanics Perspective of Tribology

Irina G. Goryacheva, Marco Paggi and Valentin L. Popov

I. INTERACTIONS ON ATOMIC SCALE AND SUPERLUBRICITY

11 Influence Factors on Mechanisms of Superlubricity in DLC Films: A Review

Qingyuan Yu, Xinchun Chen, Chenhui Zhang and Jianbin Luo

28 How Thermal Fluctuations Affect Hard-Wall Repulsion and Thereby Hertzian Contact Mechanics

Yunong Zhou, Anle Wang and Martin H. Müser

42 Contacts With Negative Work of “Adhesion” and Superlubricity

Valentin L. Popov

46 Numerical Model for Hydration-Lubricated Contact and Its Friction Behavior at Nanoscale

Yanfei Fang, Liran Ma, Xianzhang Wang and Jianbin Luo

II. ANALYTICAL AND SEMI-ANALYTICAL METHODS IN CONTACT MECHANICS

53 Analysis of Elastic Normal Contact of Surfaces With Regular Microgeometry Based on the Localization Principle

Irina G. Goryacheva and Ivan Y. Tsukanov

63 Stress Tensor and Gradient of Hydrostatic Pressure in the Half-Space Beneath Axisymmetric Bodies in Normal and Tangential Contact

Fabian Forsbach

74 Ratio of Loss and Storage Moduli Determines Restitution Coefficient in Low-Velocity Viscoelastic Impacts

Emanuel Willert

81 On Applications of Semi-Analytical Methods of Contact Mechanics

Vladislav V. Aleshin

III. FFT-ASSISTED BOUNDARY ELEMENT METHOD

92 FFT-Based Methods for Computational Contact Mechanics

Q. Jane Wang, Linlin Sun, Xin Zhang, Shuangbiao Liu and Dong Zhu

114 Calculation of the BEM Integrals on a Variable Grid With the FFT

Justus Benad

122 Non-adhesive Contacts With Different Surface Tension Inside and Outside the Contact Area

Qiang Li and Valentin L. Popov

131 Boundary Element Calculations for Normal Contact of Soft Materials With Tensed Surface Membrane

Weike Yuan and Gangfeng Wang

IV. CONTACT RESPONSE OF ROUGH SURFACES

139 Multi-Scale Surface Roughness Optimization Through Genetic Algorithms

Paolo Cinat, Giorgio Gnecco and Marco Paggi

- 153** *Artificial Neural Network Architecture for Prediction of Contact Mechanical Response*
Kalle Kalliorinne, Roland Larsson, Francesc Pérez-Ràfols, Marcus Liwicki and Andreas Almqvist
- 164** *Probabilistic, Fractal, and Related Techniques for Analysis of Engineering Surfaces*
Feodor M. Borodich, Xiaoqing Jin and Andrey Pepelyshev
- 177** *Effect of Structural Parameters on the Relative Contact Area for Ideal, Anisotropic, and Correlated Random Roughness*
Yunong Zhou and Martin H. Müser

V. INFLUENCING FACTORS OF FRICTION

- 191** *Continuum Model Analysis of QCM Nanotribological Data to Obtain Friction Coefficients for 304SS Contacts Lubricated by Water and TiO₂ Nanoparticle Suspensions*
Caitlin M. Seed, Biplav Acharya and Jacqueline Krim
- 201** *Meso–Macro Coupled Analysis of Pressure-Dependent Friction of Rubber*
Shingo Ozaki, Keishi Mieda, Satoru Maegawa and Ken Nakano
- 213** *Transient and Steady Sliding Friction of Elastomers: Impact of Vertical Lift*
Ken Nakano and Masaharu Kono
- 224** *The Influence of Vibration on Friction: A Contact-Mechanical Perspective*
Mikhail Popov
- 232** *Friction Influenced by Vibrations: A Refined Contact-Mechanics View on Lateral and Rotational Oscillations*
Roman Pohrt

VI. WEAR

- 240** *Metal Transfer and Wear*
James A. Greenwood
- 246** *Comparative Analysis of Error Sources in the Determination of Wear Volumes of Oscillating Ball-on-Plane Tests*
Manuel Reichelt and Brunero Cappella
- 256** *Contact Geometry Adaptation in Fretting Wear: A Constructive Review*
Ivan Argatov and Young Suck Chai
- 266** *Theoretical Study of Fretting Wear Rate Evolution in Axi-Symmetrical Elastic Contact Using the Method of Dimensionality Reduction*
Andrey V. Dimaki

VII. ADHESION

- 272** *Adhesion Between Rigid Indenter and Soft Rubber Layer: Influence of Roughness*
Iakov A. Lyashenko and Roman Pohrt
- 282** *Viscoelastic and Adhesion Properties of New Poly(Ether-Urethane) Pressure-Sensitive Adhesives*
Mónica Fuensanta and José Miguel Martín-Martínez
- 292** *Effect of Surface Roughness on Adhesive Instabilities for the Elastic Layer*
Junki Joe, M. D. Thouless and J. R. Barber

300 Onset of Sliding of Elastomer Multicontacts: Failure of a Model of Independent Asperities to Match Experiments

Julien Scheibert, Riad Sahli and Michel Peyrard

VIII. BIOTRIBOLOGY

309 Hallmarks of Life in Single Cell Contact Mechanics: Outstanding Challenges and Perspectives

Susana Moreno-Flores

333 Macroscopic Modeling of Fingerpad Friction Under Electroadhesion: Possibilities and Limitations

Markus Heß and Fabian Forsbach

349 Milestones in Natural Lubrication of Synovial Joints

Alessandro Ruggiero

355 Review on the Bubble Dynamics Based Cavitation Dynamics for the Negative Squeeze Motion in Lubricated Contacts

Thomas Geike

IX. INFLUENCE OF PLASTICITY AND INTERNAL STRESSES

361 Strengthening and Weakening Effects in Bilayer Coated Spherical Contact

Kurien S. Parel, Zhou Chen and Izhak Etsion

371 Studies on the Influence of Residual Stresses on the Fatigue Life of Rolling Bearings in Dependence on the Production Processes

Florian Pape, Timm Coors and Gerhard Poll

380 Mechanical Aspects of Deformation-Induced Surface Roughening in the Presence of Inclusions in a Subsurface Layer. Numerical Modeling

Varvara Romanova, Ekaterina Dymnich, Ruslan Balokhonov and Olga Zinovieva



Editorial: Contact Mechanics Perspective of Tribology

Irina G. Goryacheva^{1*}, Marco Paggi^{2*} and Valentin L. Popov^{3*}

¹Ishlinsky Institute for Problems in Mechanics RAS, Moscow, Russia, ²IMT School for Advanced Studies Lucca, Lucca, Italy, ³Technische Universität Berlin, Berlin, Germany

Keywords: contact, adhesion, friction, wear, viscoelastic media, superlubricity, biotribology, boundary element method

Editorial on the Research Topic Contact Mechanics Perspective of Tribology

A time to throw stones and a time to gather stones
(Ecclesiastes 3:5)

It is a time to gather stones. . . The Research Topic (RT) “Contact Mechanics Perspective of Tribology” was planned as a comprehensive overview of recent developments in the areas of contact mechanics and friction. Much has changed in this field in the last few decades. Contact mechanics expanded to qualitatively new fields of application which are at the forefront of the global development tendencies of technology and society, in particular micro- and nanotechnology as well as biology and medicine. The last decade was the time when vital numerical tools for simulating complex contacts, such as the FFT-based boundary element method, were created. The goal of the Research Topic was to review the recently established concepts, tools, and research activities and to outline the most important open issues for future investigations.

The main conceptual idea behind this Research Topic was to show to what extent one can understand tribology with macroscopic contact mechanics. By “macroscopic contact mechanics” we mean any approach based on continuum mechanics, including the corresponding elastic problems, viscoelasticity, adhesion, hydrodynamic and elasto-hydrodynamic lubrication, etc. In short, the papers of this Research Topic address the question: What can and what cannot be described in the framework of the macroscopic continuum mechanics approaches? Certainly, macroscopic approaches do not go to the ultimate (atomic) scale. Based on the advances of contact mechanics achieved during the last decades, it should be tracked where such limits are.

In the following, we briefly discuss the 36 papers comprising the RT.

INTERACTIONS ON ATOMIC SCALE AND SUPERLUBRICITY

The article collection is opened by a review “Influence Factors on Mechanisms of Superlubricity in DLC Films” by Yu et al.. It is devoted to one of the most fundamental achievements of tribology over the last 25 years (Erdemir and Martin, 2007). It clearly shows that a pure mechanical analysis of a tribological system is not sufficient to explain superlubricity of DLC films, since chemical aspects play a major role: passivation of the dangling bonds by hydrogen and formation of an easy-shearing sp²-C rich interface film. However, even these properties can be partially understood macroscopically as the transfer film still has a thickness of 5–40 nm, which can be considered as “macroscopic” as compared to the atomic scale. The excellent self-lubrication properties of DLC films seem to relate to the ability

OPEN ACCESS

Edited and reviewed by:

Giuseppe Carbone,
Politecnico di Bari, Italy

*Correspondence:

Irina G. Goryacheva
goryache@ipmnet.ru
Marco Paggi
marco.paggi@imtlucca.it
Valentin L. Popov
v.popov@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 05 January 2021

Accepted: 11 February 2021

Published: 19 March 2021

Citation:

Goryacheva IG, Paggi M and Popov VL
(2021) Editorial: Contact Mechanics
Perspective of Tribology.
Front. Mech. Eng 7:649792.
doi: 10.3389/fmech.2021.649792

of carbon to produce both very hard (diamond) and very soft (graphite) phases as well as self-organization of these phases in the region of high load and shear.

In the second paper of RT, Zhou et al. discuss the fundamental question of whether it is possible to use macroscopic contact mechanics even at the level when thermal fluctuations start playing an important role. The paper is based on the Green-functions method extended to account for thermal fluctuations using the fluctuation-dissipation theorem. The result is simple and very impressive: for the most usual conditions which can be awaited in the praxis, the fluctuations are almost equivalent to a constant shift of the surface. This is approximately 1/1.5 of the sqrt fluctuation of the free surface. This means that thermal fluctuations contribute to a sort of repulsion of surfaces which may essentially influence friction, in particular contributing to the effect of liquid superlubricity.

The opinion-paper “Contacts with negative work of “adhesion” and superlubricity” is an attempt to take a general look on the problem of “adhesion and superlubricity” (Ge et al., 2019). The main conclusion of this paper could be formulated as “Ernest Rabinowicz was right!” In his famous book on “Friction and wear of materials” (Rabinowicz, 1995), he stated that there are two main properties, which determine friction: adhesion and “compatibility” of materials. The less the compatibility (that means the tendency to form alloys) and the less the adhesion, the smaller the friction. The above opinion paper continues: best of all, the adhesion should be negative, then one has the effect of “superlubricity”! Very interesting is also the Rabinowicz’s notion of “compatibility.” According to the well-known Hume-Rothery rules, the materials are incompatible if their lattices are incommensurate, which, however, is a well-established rule for minimizing friction at the atomic scale! Thus, the concept of Rabinowicz looks now really “trivial”: incommensurate (and thus non-intermixable) materials should be combined with the lowest attraction (or even better, repelling!). The paper of RT by Fang et al. illustrates this mechanism in the special case of hydration-lubricated contacts, which e.g., provide the uniquely low coefficient of friction in natural joints [Klein (2013)].

ANALYTICAL AND SEMI-ANALYTICAL METHODS IN CONTACT MECHANICS

Despite the rapid development of numerical simulation methods, analytical and semi-analytical solutions remain of immense importance. They are employed as benchmark for numerical methods, to achieve the “analytical understanding” of problems, or for empirically capturing numerical results in multi-dimensional parameter spaces. An example of a combination of exact analytical solutions with a “nearly exact” approximation provides the *localization principle* (Goryacheva, 1998) which is exploited in the paper by Goryacheva and Tsukanov for the analysis of a great variety of surfaces with regular microgeometry.

The paper by Forsbach shows simple yet numerically efficient way to determine the complete stress tensor and the hydrostatic pressure gradient in the half-space beneath the contact of an arbitrary axisymmetric indenter (under normal and/or tangential

load). The calculation method is based on the same “trick” as is used in the Method of Dimensionality Reduction [see e.g., Heß (2011) and Popov et al. (2019)].

Impact tests are an important tool to analyze dynamic material properties of viscoelastic media in technology and biology. In this context, rigorous contact mechanical models of the collision problem are necessary to adequately interpret data from impact experiments. Willert shows theoretically that the coefficient of restitution in this type of testing is mainly a function of one specific material property, namely, the ratio between the loss and storage moduli of the viscoelastic probe at the characteristic timescale of impact. This, for the first time, gives a concise, comprehensive statement about what (regarding material properties) is actually measured in low-velocity viscoelastic impact or rebound tests.

For the pure normal contact problem, a closed analytical solution is possible. For tangentially loaded contacts, on the contrary, semi-analytical methods must be applied. A review and application possibilities of such methods is given by Aleshin. The semi-analytical solutions, such as the Method of Memory Diagrams or MDR, allows to calculate hysteretic responses to extremely complex loading histories, such as random vibrations. As an example, the author shows that a single elastic body with a frictional contact to the substrate demonstrates rich dynamic behavior when excited even by a simple harmonic signal.

FFT-ASSISTED BOUNDARY ELEMENT METHOD

In the last years, the FFT based Boundary Element Method became the standard tool for contact simulations of real surfaces both in research and industry. The basic development of the FFT-based BEM took place in the late 1990s and early 2000s and is associated first of all with the groups of Leon M. Keer and Q. Jane Wang from the Northwestern University. The paper “FFT-Based Methods for Computational Contact Mechanics” is a concise and instructive review of the basics and the current state of the FFT-based BEM. The FFT-based BEM is, however, restricted to macroscopically plane geometries. Benad shows a way in which the BEM can be generalized to arbitrary shapes of contacting bodies without increasing the computational complexity.

Soft matter attracts ever greater attention of researchers. The softer the body, the more visible becomes the influence of the *surface tension* of the contacting bodies (which is neglected in the classical contact mechanics of “stiff bodies”). In the paper by Yuan and Wang, a BEM-version is formulated accounting for the surface tension. Li and Popov generalize this approach further by considering a non-adhesive contact with different surface tension inside and outside the contact area.

CONTACT RESPONSE OF ROUGH SURFACES

Contact mechanics of rough surfaces was one of the most debated topics in contact mechanics of the last decades. These debates

seem to come to a consensus summarized in the “contact challenge”-paper by Müser et al. (2017). The central tool which is now accepted by the most researchers is the FFT-based BEM (see previous Section). But even this extremely efficient simulation method is computationally expensive when the identification of optimal roughness and texturing of surfaces is a matter of research, to achieve desired tribological responses for industrial applications. One further possibility to shorten the simulation time may be the use of genetic algorithms (Cinat et al.) or artificial neural network architectures (Kalliorinne et al.) that open new perspectives in data-driven identification and machine learning techniques applied to contact mechanics.

While the BEM allows simulating contact with an arbitrary shaped surface (as long as the half-space-approximation for the response is valid), most studies on rough surfaces focused on the so-called randomly rough surfaces. Borodich et al. stress in their paper that randomly rough surfaces are only a small sub-class of all surfaces and correlated roughness may completely change the contact mechanical properties even if the power density of roughness remains unchanged. A comparative study of the effect of structural parameters on the contact area for randomly rough, anisotropic and correlated random surfaces is undertaken by Zhou and Müser. In particular, the authors show that the famous parameter κ (coefficient connecting the real contact area with normal load and rms slope) is not a function of the Nayak parameter alone [Nayak (1971)]. The study also attracts attention of researchers to the fact that at small contact area the statistics of the contact profile is not the same as that of the full rough surface. This dependence maybe the reason for the weak dependency of the coefficient of friction of elastomer materials on the normal load [Popov et al. (2018)].

INFLUENCING FACTORS OF FRICTION

One of the pioneers of the use of Quartz Crystal Microbalances (QCM) for studies of friction, J. Krim, examines together with her co-authors the limits of macroscopic continuum mechanics, by probing where continuum methods break down as the atomic scale is approached. They find that contacts with water solutions are very well described by macroscopic theory up to nanometer scale, but the model failed to adequately describe contacts lubricated with the suspension of nanoparticles with radius 20 nm. However, its continuum nature did not appear to be the dominant factor underlying failure, which is most probably due to other factors not correctly described in the model. Further investigations are necessary.

Ozaki et al. suggest a loop-type coupled analysis scheme to bridge the mesoscale and macroscale domains of friction analysis. Specifically, the mesoscale multipoint contact model was linked with the macroscale finite element analysis model via the rate-, state-, and pressure-dependent friction model previously proposed by the authors.

Kinetics of elastomer friction is an important phenomenon for dynamic applications. Nakano and Kono show that the character and the very existence of kinetics do depend on how stiff the body is fixed normally to the contact surface, which means that the

friction coefficient is not a material constant even in simple sliding contacts.

An old problem of fundamental importance for the understanding of friction and its applications is the interplay between oscillations and friction. Vibrations can strongly influence friction, which in turn often leads to vibrational instabilities [Bowden and Tabor (1950)]. Thus, friction should always be understood as the interplay of dynamics and friction on different spatial and temporal scales. This interplay has many aspects which have been studied intensively in the past decades. The paper by M. Popov gives a high-level overview of active control of friction by normal, lateral and transverse vibrations and derives the main properties of friction under oscillation from a purely macroscopic contact-mechanical model. In addition to classic influencing factors such as frequency and amplitude, the key role of contact stiffness is emphasized and the effectiveness of various oscillation waveforms is discussed. Pohrt extends the discussion by employing a BEM model to include the effect of partial slip. Using the same approach, he introduces rotational oscillations and finds favorable properties to reduce the observable macroscopic friction.

WEAR

The name of James Greenwood is known to every tribologist, first of all due to his works on contact mechanics of rough surfaces [Greenwood and Williamson (1966)]. Even more attention deserves his appeal “Stop studying the purely normal contact of rough surfaces!” In his opinion-paper Metal Transfer and Wear,” he writes: “These thoughts are offered as a reminder that Tribology is not all about the normal contact of fractal surfaces, and indeed, not all about elastic contact of rubber and polymers, or even about dry contact.”

This recalls that wear is one of the most complicated, yet unexplained, phenomena in tribology. The difficulties start already by the definition of this notion. Reichelt and Cappella examine various wear definitions and various experimental methods to measure wear and estimate the error sources.

Just as in frictional processes, wear can be influenced by oscillations—both in the cases of partial slip and gross slip. Oscillations may lead to a specific kind of wear called fretting. On the other hand, they may lead also to the damage of material in form of fretting fatigue. In the review paper by Argatov and Chai, both partial and gross slip situations are considered, and the authors provide a review of recent works on fretting wear. They consider such aspects as the force-controlled steady-state regime, wearing-in period, wear of functionally graded materials, limiting profile, wear accumulation. The authors also discuss various analytical approaches including the MDR-based approach.

While the very often used Archard wear law is a very strong simplification which is not always confirmed by experiments [Meng and Ludema (1995)], it is still used for qualitative estimation of wear processes in a limited range of parameters, as done in the paper by Dimaki. The author provides an analysis of the wear rate as function of time and compares theoretical prediction with available experimental data.

ADHESION

Adhesion has been one of the most debated topics in contact mechanics in the last years. This interest was driven by many factors including the bio-medical applications and general interest to “soft matter.” Adhesion has been incorporated into the FFT-based BEM. The most simple and straightforward implementation is available for JKR-type adhesion [Popov et al. (2017)]. In this case, the inclusion of adhesion not even influence the computation time. However, adhesive contacts have been studied mostly theoretically and experimental results are still sparse. Lyashenko and Pohrt contribute to closing this gap providing a detailed experimental study of the influence of roughness on adhesion between rigid indenter and soft rubber layer.

In the context of contact mechanics, only very simple models of adhesion are considered. For industrial applications, however, material parameters are important which cannot be reduced to the work of adhesion but take also into account the propensity of a material to detach due to cavitation. Material aspects are considered in the paper by Fuensanta and Martín-Martínez.

Adhesion can also lead to elastic instabilities and formation of regular patterns with a characteristic wavenumber. Joe et al. analyze the effect of surface roughness on adhesive instabilities for the elastic layer and find that for moderate to large RMS amplitudes, roughness exerts a stabilizing effect, but low RMS roughness can trigger the instability in ranges where the uniform layer would be stable.

Another much debated topic related to adhesion is adhesive contact under tangential loading. This is an example of a topic where there are as much opinions as are authors. In-depth experimental studies with accompanying theoretical modeling started only very recently. The processes occurring at the onset of sliding of elastomer multicontacts are studied theoretically and experimentally by Scheibert et al..

BIOTRIBOLOGY

Contact mechanics of cells, in particular their adhesion on various surfaces has become an important and intriguing part of contact mechanics. In her review paper “Hallmarks of Life in Single Cell Contact Mechanics: Outstanding Challenges and Perspectives,” Moreno-Flores places the focus on cultured mammalian cells and on experimental techniques that rely on contact mechanics, in particular scanning-probe and traction force microscopies. The importance of time as key variable in theory and experiment is highlighted, together with the cellular structures and organelles needed to correctly understand the mechanobiology of cells.

Electrovibration is one of the key technologies in surface haptics. By inducing controlled electrostatic forces, the friction within a sliding contact between the human finger and a capacitive screen is modulated, which in turn gives effective tactile feedback to the user. Such powerful haptic displays can be built into mobile phones, tablets, navigation devices, games consoles and many other devices of consumer electronics. However, due to the layered structure and complex material of

human skin, the underlying contact mechanical processes have not yet been fully understood. Heß and Forsbach develop a macroscopic model for sliding friction of a fingerpad over a smooth surface under electroadhesion, which is based on new theoretical approaches as well as finite element simulations using the Ogden model to account for skin stiffening. It provides reasonable results for all contact mechanical quantities. In particular, the predicted friction force and friction coefficient show excellent agreement with experimental data over the entire range of relevant voltages and applied normal forces.

In his mini review “Milestones in Natural Lubrication of Synovial Joints,” Ruggiero focuses on the history of research related to lubrication of human synovial joints. He carefully examines existing models of synovial joint lubrication and their fundamental mechanisms: boundary, weeping, elastohydrodynamic, squeeze-film, as well as boosted lubrication and ultrafiltration. The rheology of the synovial fluid which plays a key role in natural joints, is also described; various existing laws of its description are given in the historical vein both in the Newtonian model and in the non-Newtonian one. As often, such historical view allows highlighting the main “milestones” of the research which remain the “pillars” of our present understanding and contribute to scientific cooperation between tribology, biology and medicine in general, as well as to a deep understanding of the complex phenomena acting in biological tribosystems.

One of the central processes in lubrication is cavitation. As a matter of fact, without cavitation no bearing capacity can exist. However, cavitation belongs to the most complicated processes for mathematical modeling. In his mini review, Geike describes the current state of research in this area.

INFLUENCE OF PLASTICITY AND INTERNAL STRESSES

Stiff hard coatings, such as TiN or CrN on metallic substrates, are often used to enhance tribological properties of components. A mismatch of the Young’s moduli at the coating/substrate interface can lead to additional stresses in the coated system. To reduce this effect, functionally graded materials can be used. An even simpler solution is the usage of a single middle-layer, which has an intermediate Young’s modulus between the substrate and coating. Parel et al. show in their paper that both strengthening and weakening effects can be observed in bilayer coated spherical contacts, depending on material and geometric parameters of layers.

Pape et al. study the influence of residual stresses on the fatigue life of rolling bearings in dependence on the production processes.

Very often, formation of the surface topography during frictional sliding is attributed to wear. However, another mechanism of surface roughening is just the volume plastic deformation of a spatially heterogeneous material. Romanova et al. simulate how the internal heterogeneity of a material leads to surface roughening during deformation.

The editors thank all authors of this excellent book which impressively documents the current state of contact mechanics

and friction theory and discusses the prospects for the future. It shows clearly that contact mechanics and friction is not only important but also beautiful branch of science of high technological and esthetical value.

REFERENCES

- Bowden, F. P., and Tabor, D. (1950). *The friction and lubrication of solids*. Oxford, United Kingdom: Oxford University Press.
- Erdemir, A., and Martin, J.-M. (Editors) (2007). *Superlubricity* (Amsterdam, Netherlands: Elsevier). doi:10.1016/B978-0-444-52772-1.X5029-X
- Ge, X., Li, J., and Luo, J. (2019). Macroscale superlubricity achieved with various liquid molecules: a Review. *Front. Mech. Eng.* 5, 2. doi:10.3389/fmech.2019.00002
- Goryacheva, I. G. (1998). *Contact mechanics in tribology*. New York, NY: Kluwer Academic Publishers.
- Greenwood, J. A., and Williamson, J. B. P. (1966). Contact of nominally flat surfaces. *Proc. R. Soc. A: Math. Phys. Eng. Sci.* 295 (1442), 300–319. doi:10.1098/rspa.1966.0242
- Heß, M. (2011). *Über die exakte Abbildung ausgewählter dreidimensionaler Kontakte auf Systeme mit niedrigerer räumlicher Dimension*. Göttingen, Germany: Cuvillier-Verlag.
- Klein, J. (2013). Hydration lubrication. *Friction* 1, 1–23. doi:10.1007/s40544-013-0001-7
- Meng, H. C., and Ludema, K. C. (1995). Wear models and predictive equations: their form and content. *Wear* 181–183, 443–457. doi:10.1016/0043-1648(95)90158-2
- Müser, M. H., Dapp, W. B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T. A., et al. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65, 4. doi:10.1007/s11249-017-0900-2
- Nayak, P. R. (1971). Random process model of rough surfaces. *J. Lubricat. Technol.* 93, 398–407. doi:10.1115/1.3451608
- Popov, V. L., Heß, M., and Willert, E. (2019). “Handbook of contact mechanics,” in *Exact solutions of axisymmetric contact problems*. Berlin Heidelberg: Springer-Verlag. doi:10.1007/978-3-662-58709-6
- Popov, V. L., Pohrt, R., and Li, Q. (2017). Strength of adhesive contacts: influence of contact geometry and material gradients. *Friction* 5, 308–325. doi:10.1007/s40544-017-0177-3
- Popov, V. L., Voll, L., Kusche, S., Li, Q., and Rozhkova, S. V. (2018). Generalized master curve procedure for elastomer friction taking into account dependencies on velocity, temperature and normal force. *Tribology Int.* 120, 376–380. doi:10.1016/j.triboint.2017.12.047
- Rabinowicz, E. (1995). *Friction and wear of materials*. 2nd Edn. New York, NY: John Wiley & Sons. 1st Edn 1965.

AUTHOR CONTRIBUTIONS

All authors have contributed equally to this work and share first authorship.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Goryacheva, Paggi and Popov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Influence Factors on Mechanisms of Superlubricity in DLC Films: A Review

Qingyuan Yu, Xinchun Chen, Chenhui Zhang* and Jianbin Luo

State Key Laboratory of Tribology, Tsinghua University, Beijing, China

As a kind of self-lubricating material, diamond-like carbon (DLC) film is famous for its excellent tribological properties. Superlubricity state with nearly-vanishing friction achieved with DLC film has enormous potential applications in future mechanical systems. It is pointed out that its superlubricity state is highly related to both the inherent properties of the DLC film and external sliding conditions. Moreover, the underlying mechanisms of the superlubricity are complicated, posing uncertainties on their engineering application. This review provides an overview of the influence factors, including film composition, ambient, temperature, normal load, and sliding velocity and their correlations with the anti-friction behaviors of DLC films. These understandings will enable a more effective engineering application of self-lubricating carbon films with excellent tribological properties.

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Lars Pastewka,
University of Freiburg, Germany
Wenling Zhang,
University of Alberta, Canada

*Correspondence:

Chenhui Zhang
chzhang@tsinghua.edu.cn

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 21 May 2020

Accepted: 06 July 2020

Published: 11 August 2020

Citation:

Yu Q, Chen X, Zhang C and Luo J
(2020) Influence Factors on
Mechanisms of Superlubricity in DLC
Films: A Review.
Front. Mech. Eng. 6:65.
doi: 10.3389/fmech.2020.00065

Keywords: diamond-like carbon, superlubricity, friction, influence factor, mechanism

INTRODUCTION

Friction is one of the decisive factors affecting the efficiency and service life of a mechanical system. Statistically, 23% of global energy consumption is due to the friction-related activities, of which 40% can be prevented through application of advanced surface, material and lubrication technologies (Holmberg and Erdemir, 2017). With the increasing demand on higher efficiency, power and life of mechanical systems, studies on reducing friction and wear are becoming increasingly important both from the economic and environmental perspectives. The discovery of superlubricity provides a possibility to reduce tribological consumption to an extremely low level, and has triggered increasingly extensive attention from researchers around the world over the past three decades.

The concept of superlubricity was firstly brought out by Hirano in a theoretical prediction based on calculations that the friction force could reach zero level when two crystal surfaces sliding in an incommensurable condition (Hirano and Shinjo, 1990). And it was soon verified by nanoscale friction test on cleaved mica surfaces with different contact angles (Hirano et al., 1991). Now this ideal friction vanishing state is generally redefined as structural superlubricity, and the concept of superlubricity is generally accepted as a sliding state with a kinetic friction coefficient below 0.01 (Erdemir and Eryilmaz, 2007). Around 2000s, with the rapid increase of research investment, more superlubricity phenomena have been revealed, which can be generally divided into solid and liquid superlubricity, respectively. In nanoscale or microscale, superlubricity is achievable via a variety of solid materials such as layered materials like graphite, graphene (Dienwiebel et al., 2004), BN (Song et al., 2018), and MoS₂ (Martin et al., 1993), atomically smoothed crystals including covalent crystals like Si (001) (Hirano et al., 1997), ionic crystals (Socoliuc et al., 2004) like NaCl, metallic crystals like Ag (Goto and Honda, 2004). To date, various carbon-based materials have been found to be capable of achieving superlubricity (Chen and Li, 2020), for instance, C₆₀ intercalated graphite

films (Miura et al., 2005), carbon nanotubes (Zhang et al., 2013), graphite or graphene, ultra-nanocrystalline diamond (Kumar et al., 2011), onion-like/fullerene-like carbon (Gong et al., 2017), and diamond-like carbon (Erdemir and Eryilmaz, 2014). Nevertheless, most of these superlubricity behaviors are based on the incommensurable contact of the ultra-smooth crystalline surfaces, which is still far from being applicable for engineering due to the diversified interferences in macroscopic tribo-systems.

Diamond-like carbon (DLC) film is a class of carbon-based amorphous coating, which exists in the form of a disordered covalent network of sp^1 , sp^2 , and sp^3 hybridized carbon atoms, and has the ability to introduce other elements such as hydrogen into the film. According to the differences in structure and doped materials, DLC can be generally divided into amorphous carbon (a-C), hydrogenated amorphous carbon (a-C:H), tetrahedral amorphous carbon (ta-C), and hydrogenated tetrahedral amorphous carbon (ta-C:H) (Robertson, 2002). The synthesis methods of DLC films can be divided into physical vapor deposition (PVD) and chemical vapor deposition (CVD). With the development of these technologies, DLC coatings can be deposited flexibly on a variety of materials such as silicon, ceramic, glass, metal, and rubber. Meanwhile, by changing the deposition conditions, their structure and composition are adjustable to obtain superior mechanical and tribological properties for different application conditions.

DLC was first synthesized around the 1950s. Systematic research on DLC started in the 1970s (Aisenberg and Chabot, 1971). Later in 1981, Enke reported the superlubricity behavior of DLC films with a friction coefficient of 0.005 in ultrahigh vacuum (Enke, 1981). In the 1990s, superlubricious capacity of a-C:H in vacuum was systematically studied by Donnet et al. (1994) and Donnet and Grill (1997). In 2000, an ultralow friction coefficient of 0.002 was achieved with self-mated a-C:H tested in nitrogen and argon atmosphere (Erdemir et al., 2000a), and an ultra-long superlubricity lifetime of 32 days was achieved (Erdemir et al., 2000b), as shown in **Figure 1A**. Recently, an

extremely low friction coefficient of 0.0001 under heavy load was achieved with a-C:H in hydrogen atmosphere (Nosaka et al., 2017), as shown in **Figure 1B**. These results indicate that a-C:H films are capable of achieving stable superlubricity under dry inert atmosphere, vacuum and hydrogen, and the underlying mechanism is closely related to friction-induced structural transformation of the contact area and tribo-chemical interaction between the sliding surfaces. Further researches have shown that introducing silicon (Chen et al., 2014, 2017) or sulfur (Freyman et al., 2006) into a-C:H films can significantly reduce its sensitivity to oxygen or water vapor, leading to superlubricity in humid air via novel pathways. Fullerene-like hydrogenated amorphous carbon (FL-C:H) films also exhibit superlubricious properties in both nitrogen and air (Wang et al., 2008). Kato et al. suggested that carbon nitride (CN_x) can achieve superlubricity in nitrogen atmosphere without the presence of hydrogen (Kato et al., 2003; Adachi and Kato, 2008). **Table 1** shows the timeline of the research progress on solid superlubricity achieved with DLC films. As for liquid lubrication, researches showed that hydrogen-free ta-C can achieve superlubricity with the assistance of ester additive (Kano, 2006), glycerol (De Barros Bouchet et al., 2007), unsaturated fatty acids (Kuwahara et al., 2019). DLC lubricated with nano boron nitride particle dispersed in PAO is also available to achieve superlubricity (Zeng et al., 2013).

To date, a growing interest in superlubricious DLC is aroused, and diversified superlubricity pathways have been discovered. Nevertheless, most of these superlubricity phenomena are achieved in strictly controlled lab conditions. The influencing mechanisms of environment and working conditions are complicated, and are still not fully comprehended, which poses uncertainties on the effective manipulation of superlubricious DLC films in mechanical engineering. In recent decades, numerous researches have been done, aiming at uncovering these mechanisms and optimizing the performance of DLC films in complicated environments. In the following sections, a review of superlubricity of DLC is provided in the perspective

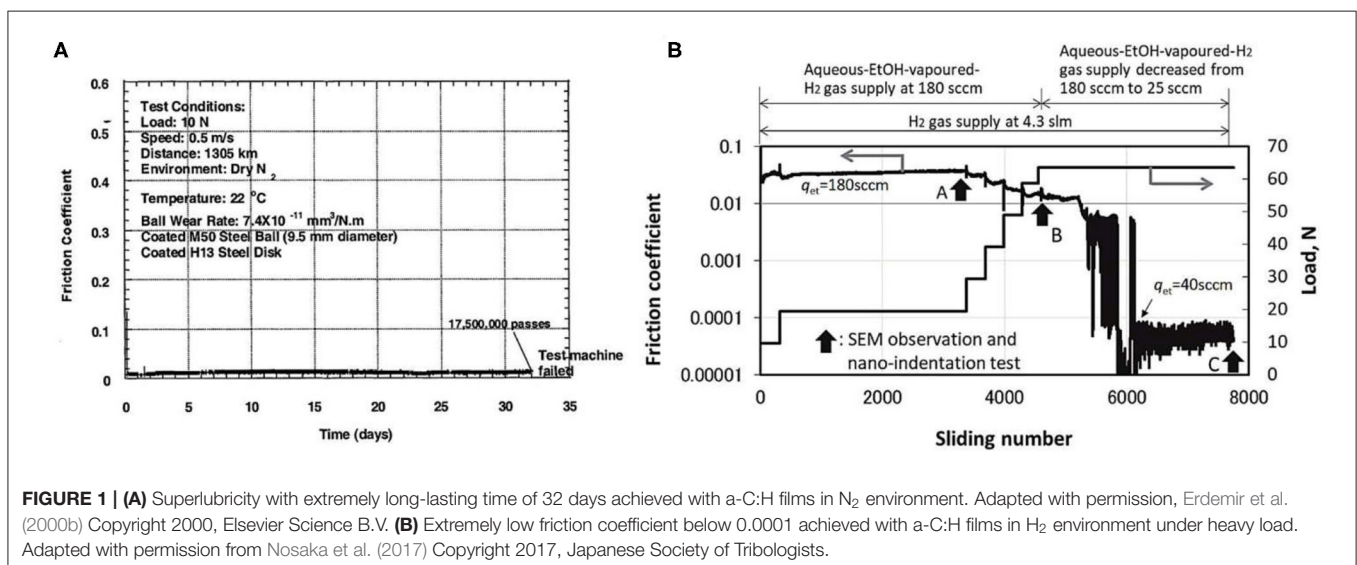


FIGURE 1 | (A) Superlubricity with extremely long-lasting time of 32 days achieved with a-C:H films in N₂ environment. Adapted with permission, Erdemir et al. (2000b) Copyright 2000, Elsevier Science B.V. **(B)** Extremely low friction coefficient below 0.0001 achieved with a-C:H films in H₂ environment under heavy load. Adapted with permission from Nosaka et al. (2017) Copyright 2017, Japanese Society of Tribologists.

TABLE 1 | Research progress on the solid superlubricity achieved with different types of DLC films under various environments.

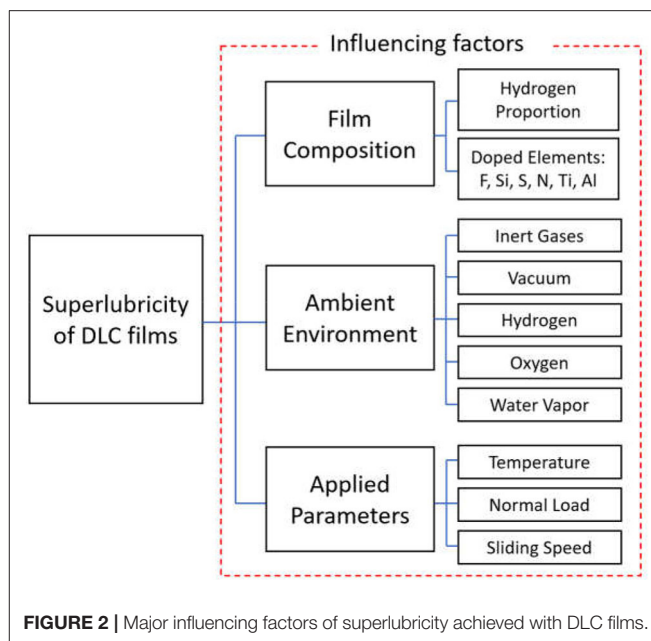
Published year	Film type	Test environment	Friction coefficient	References used
1981	a-C:H	Vacuum	0.005	Enke, 1981
1994	a-C:H	Vacuum	0.006	Donnet et al., 1994
2000	a-C:H	Nitrogen	0.002	Erdemir et al., 2000a
2003	CN _x	Nitrogen	0.009	Kato et al., 2003
2006	a-C:H:S	Humid air	0.004	Freyman et al., 2006
2014	a-C:H:Si	Humid air	0.004	Chen et al., 2014
2017	a-C:H	Hydrogen	0.0001	Nosaka et al., 2017

of influence factors as shown in **Figure 2**. Film composition is one of the key factors for the establishment of superlubricity. The influences of different elements, including hydrogen, fluorine, silicon, nitrogen, and metallic elements are discussed in section Influence of Film Elements. Stable superlubricity achieved with DLC films usually needs dry inert gaseous atmosphere or hydrogen environment. Though the friction of DLC films in vacuum is extremely low, their wear rate is very high. And it is still a challenge for DLC to achieve superlubricity in oxygen and water containing environments. The environmental sensitiveness and its mechanisms and controlling methods are discussed in section Influence of Environments. From the perspective of engineering applications, it is important to understand the influences of working conditions such as temperature, normal load and sliding velocity on the superlubricity of DLC films. The influence rules and their underlying mechanisms are introduced in section Influence of Testing Parameters. Finally, section Discussion and Conclusion presents the most prevailing theories for the mechanisms of superlubricity in the perspective of origin of friction. The fundamental restrictions of applying superlubricious DLC films in engineering and the approaches to combat these deficiencies are summarized.

INFLUENCE OF FILM ELEMENTS

The proper composition of DLC films is generally considered to be the precondition for achieving superlubricity state. Understanding the influencing mechanism of the film elements can enable a more effective manipulation of DLC with excellent tribological properties. In this section, the influence of film elements, for example, surface passivation induced by hydrogen and fluorine and low humidity sensitiveness enabled by silicon and sulfur, are presented. Superlubricity achieved via novel mechanisms in carbon nitride and metal-containing DLC films are also introduced.

Hydrogen is the most common element in DLC film, and its proportion can be controlled by changing deposition conditions such as the bias voltage (Chen and Kato, 2014) and the



proportion of gas source (Erdemir et al., 2000c). Most of the hydrogen atoms exist in the form of bonded C-H, while some of hydrogen atoms are trapped in the interstice of carbon framework with the forms of unbonded atoms and not fully decomposed gas source molecules (Erdemir, 2004). During dry sliding contact, they are released in the forms of H₂, CH₄, and so on (Nevshupa et al., 2019). Studies have shown that the friction property of a-C:H is closely related with the proportion of hydrogen (Donnet et al., 1998, 1999; Erdemir, 2001; Fontaine et al., 2005). Generally, in vacuum or inert environments, hydrogen-rich a-C:H films, or polymer-like carbon (PLC) films (Casiraghi et al., 2005) with a hydrogen proportion higher than 40% are most favorable for achieving superlubricity (Donnet et al., 1994, 1999; Gao et al., 2003). Hydrogen-free DLC such as ta-C or a-C exhibits a very high friction coefficient up to 1 when tribo-tested in vacuum (Andersson et al., 2003). However, implanting hydrogen ion into the surface of hydrogen-free DLC can also lead to superlubricity (Eryilmaz and Erdemir, 2008). It is believed that introducing hydrogen into the film matrix will increase its void density and sp³ proportion, thereby reducing the residual stress in the film and softening the material (Chen and Li, 2020). Most importantly, hydrogen is speculated to passivate the carbon dangling bonds at the sliding interface, reducing tribochemical wear and friction caused by the formation of cross-interfacial carbon bonds and forming a none-adhesive sliding surface (Erdemir, 2001), as shown in **Figure 3**, which is proved by simulations (Pastewka et al., 2008, 2010; Schall et al., 2010). Meanwhile, simulations (Dag and Ciraci, 2004) also suggest that due to the low electronegativity of hydrogen, both hydrogenated surfaces are positively charged and generate repulsive force, thus reducing friction and wear (Eryilmaz and Erdemir, 2008). In addition to controlling the hydrogen proportion, introducing other elements is also an effective approach to improve the

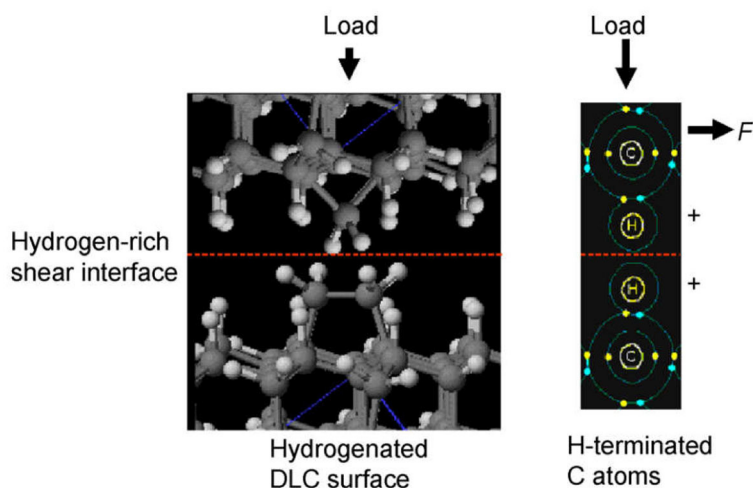


FIGURE 3 | Schematic of hydrogen-passivated none-adhesive sliding surface. Adapted with permission from Erdemir (2001) Copyright 2001, Elsevier Science B.V.

mechanical and tribological performance of DLC films. Fluorine is a case in point. Fluorine-doped a-C:H with lower hydrogen proportion (5 at.% H, 18 at.% F) could also achieve superlubricity with a friction coefficient of 0.005 in ultrahigh vacuum (Fontaine et al., 2004). They also exhibited higher thermal stability, even after annealing at a maximum temperature of 500°C (Nobili and Guglielmini, 2013). Fluorine, as another monovalent element like hydrogen, has the highest electronegativity and can effectively eliminate the dangling bond on the surfaces, providing negatively charged surfaces and repulsive force, leading to ultralow friction coefficients.

Silicon is another common dopant for DLC films, which could reduce their humidity sensitiveness and improve the thermal stability. Similar with a-C:H, a-C:H:Si films also exhibit superlubricous properties in vacuum (Sugimoto and Miyake, 1990), nitrogen, and hydrogen environments (Chen et al., 2014). As shown in **Figures 4A,B**, the lowest friction coefficient in nitrogen could reach 0.001 for self-mated a-C:H:Si (31.9 at.% H, 9.3 at.% Si), similar with a-C:H (~40 at.% H) (Chen et al., 2017). Moreover, under certain conditions, it could also achieve superlubricity in humid air (Chen et al., 2013), as shown in **Figure 4C**, and even under liquid water (Zhao et al., 2009). The superlubricity behavior of a-C:H:Si films is closely related to the formation of a softer polymerlike tribofilm in the contact area (Chen et al., 2017). Silicon is capable of reducing sp^2 proportion and stabilizing sp^3 carbon network, improving thermal stability, and reducing residual stress. During dry sliding contact, silicon is expected to promote the phase transformation and the tribo-softening of the interface and involved in the evolution process of the tribofilm (Chen and Li, 2020). In humid conditions, the formation of a silica-like tribo-film and the hydroxyl groups attached to silicon atoms on the surface is supposed to be an important factor for their superlubricity behaviors, as they can directionally absorb water molecules and form layered-like structures under appropriate conditions.

To some extent, sulfur is similar to silicon in reducing humid sensitivity of DLC films. Experiments showed that a-C:H:S films (5.0 at. % S) are capable to realize superlubricity in humid environments with RH ranging from 0 to 50% (Freyman et al., 2006). Researchers suggested that sulfur can be involved in the formation of thiol-like (-C-S-H) groups on the sliding surfaces, leading to a weaker binding energy between water molecules and the sliding surface, and hence a low friction coefficient in humid environments.

Nitrogen doped DLC or CN_x , could achieve superlubricity via a different pathway without the assistance of hydrogen. The friction coefficient between CN_x (12 at.% N, 7 at.% O) and Si_3N_4 could reach a superlubricity state when tested in nitrogen environment, but lost this capacity in other inert ambient such as CO_2 , He and vacuum (Kato et al., 2003). The mechanism is still not fully comprehended, and the formation of a transfer layer on the counterpart material and the establishment of a chemical inert surface terminated by nitrogen are thought to be the key factors.

The incorporation of metal elements such as titanium, chromium, and tungsten in DLC films could improve their hardness and toughness (Voievodin and Zabinski, 2000). However, superlubricity is difficult to achieve in these cases, mainly due to the abrasive wear effect caused by the metallic carbide nanocrystals. Generally, in order to achieve superlubricity, metallic elements should be controlled within a low proportion to limit the formation of nanocrystals in the film. Titanium-doped DLC was reported to achieve superlubricity with a friction coefficient of 0.005 in humid air with a relative humidity of 40%, and the authors suggested it was attributed to the anti-oxidation effect of the TiC-rich surface and the formation of fullerene-like structure with TiO_2 cores (Zhao et al., 2016). Titanium and silicon co-doped a-C:H films are also capable of achieving superlubricity in air with a friction coefficient of 0.007 (Jiang et al., 2010). Another study showed that aluminum and silicon co-doped a-C:H films could realize

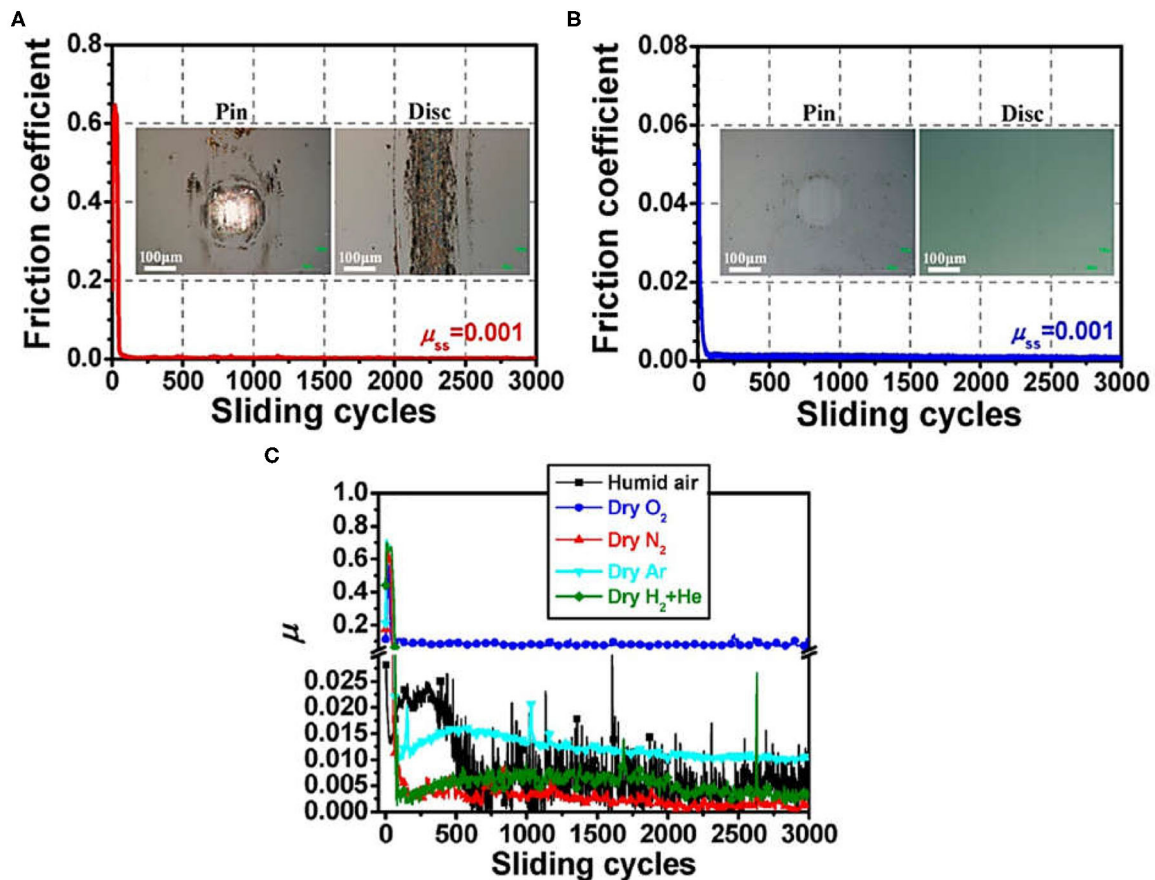


FIGURE 4 | Contact area and extremely low friction coefficient of 0.001 achieved with a-C:H:Si (A) and (B) a-C:H. (C) Superlubricity achieved with a-C:H:Si in humid air, dry Ar, dry N₂, and Dry H₂+He. Adapted with permission from Chen et al. (2014) Copyright 2014, American Chemical Society.

a superlow friction coefficient of 0.001 in high vacuum. Self-assembled nanostructure network and fullerene-like structures were observed to be the main factors for the superlubricity (Liu et al., 2012).

INFLUENCE OF ENVIRONMENTS

Previous researches have shown that a-C:H films can reach ultralow friction in clean, inert environments (such as nitrogen, carbon dioxide, noble gases, and vacuum) and hydrogen, but it is much challenging in oxygen or humid air (Erdemir et al., 2000c; Chen et al., 2014). Although DLC films are capable of achieving superlubricity in macroscale, their environmental sensitiveness is still a barrier to limit their engineering application. In this section, the experimental studies on superlubricious DLC films in dry inert gaseous atmosphere, vacuum, hydrogen, oxygen, or water containing environments are presented, respectively, and the influence mechanisms of these environments are discussed.

Dry Inert Gaseous Atmosphere

Dry nitrogen is the most frequently used environment for achieving superlubricity with DLC films. Generally, dry nitrogen

is considered as a kind of inert gas that can exclude the interference of oxygen and water molecules in lab condition, which is beneficial for the self-lubrication effect of DLC films. The same is true with carbon dioxide and noble gases like argon. Simultaneously, the wear rate of a-C:H in dry inert gaseous environments is very low, sometimes even unmeasurable. Nevertheless, though superlubricity is available in nitrogen, carbon dioxide, and argon, the friction coefficient in argon is higher than that in nitrogen and carbon dioxide (Erdemir et al., 1991; Zhang et al., 2002; Ji et al., 2009), as shown in **Figure 4C**. The researchers suggest that at an appropriate sliding speed, due to the interaction between π orbital-lone pair electrons on the film and lone pair of electrons of the molecules, N₂ and CO₂ molecules are able to be absorbed on the surfaces to form a molecule layer with another lone pair of electrons pointing outside, generating coulomb repulsion forces and preventing the π - π^* interactions between the two surfaces (Ji et al., 2009).

Vacuum

Vacuum is another important environment for superlubricious DLC films due to their superior properties for space applications

(Vanhulsel et al., 2007). Researchers have shown that hydrogen-free DLC film exhibit a drastic friction and wear in vacuum with a friction coefficient as high as 1 (Erdemir and Eryilmaz, 2014), while hydrogen-rich DLC films are able to achieve superlubricity of 0.006–0.008 in vacuum from 10^{-7} to 10^{-1} Pa (Donnet et al., 1994). However, compared with that in dry inert gaseous environment and hydrogen, the wear rate of a-C:H films is generally high in vacuum (Fontaine et al., 2001), which is still a barrier for their effective application in aerospace. Compared with the inert gas ambient, the vacuum environment mainly has three distinct characteristics theoretically (Xue and Wang, 2012). Firstly, the protective effects of absorbed gas molecules mentioned above do not exist, leading to a higher chance for π - π^* interactions and hence tribochemical interactions between two surfaces. Secondly, convection heat transfer enabled by gas medium is not available, resulting in a higher flash temperature of the contact area, which facilitates their tribochemical reactions and hydrogen loss. Thirdly, the main free path of molecules in vacuum is significantly higher than that in gas ambient, which could increase the diffusion and loss speed of hydrogen on the contacting surfaces, and resulting in a higher wear rate. Improving the hydrogen proportion is an available approach to prolong the lasting time of superlubricity state in vacuum. For example, experiments showed that a-C:H films with a hydrogen proportion of 34% could only maintain the durability of about 40 cycles, while under the same circumstances a-C:H films with a higher hydrogen content up to 40% could maintain more than 500 cycles (Fontaine et al., 2001). In another experiment, the superlubricity state of DLC film with 50% of hydrogen lasted for about 34,000 cycles (Vanhulsel et al., 2007). In addition, another approach to prolong their lifetime in vacuum is introducing other elements into the film. For instance, sulfur-containing DLC films exhibit a higher sustainability but a relatively large friction coefficients in vacuum (Moolsradoo and Watanabe, 2010). Titanium and silver co-doping can also prolong the superlubricity lifetime of DLC films in vacuum. It has been observed that Ag was agglomerated on the sliding surface due to cold welding effect, and the anti-friction property of Ag contributed to the superlubricity behavior of the DLC film (Wang et al., 2012).

Hydrogen Ambient

As mentioned above, hydrogen contained in the film matrix or implanted at the surface is important for achieving superlubricity due to its passivation effect. Similarly, hydrogen gaseous ambient is also an advantageous environment for achieving superlubricity. The lasting time and friction coefficient of superlubricity state achieved with same DLC films in hydrogen is significantly superior than that in helium or vacuum (Fontaine et al., 2001), and the lasting time increased with the elevating pressure of hydrogen gas (Fontaine et al., 2005). Furthermore, hydrogen environment could significantly lower the friction coefficient of hydrogen-free ta-C films or a-C films, and even reach the superlubricity state (Erdemir and Eryilmaz, 2014; Okubo et al., 2015). In the hydrogen environment, an extremely low friction coefficient of 0.0001 was achieved with the tribo-couple of hydrogen-rich DLC films and ZrO₂ counterpart under

a heavy load of 2.6 GPa, as shown in **Figure 1B** (Nosaka et al., 2017). The researchers illustrated that hydrogen molecules in ambient are able to be involved in the tribochemical interactions at the surface with the assistance of shear force and flash temperature in the contact area, leading to the formation of a hydrogen passivated surface (Chen and Li, 2020). The researchers suggested that ZrO₂ could act as a catalyzer to promote the disintegration of H₂ into dissociative H, which facilitated the tribochemical process and contributed to the extremely low friction (Nosaka et al., 2017). It should be point out that during dry sliding contact, the film itself was observed to release gaseous molecules such as hydrogen, methane, carbonic oxide, as the molecules were squeezed out from the carbon framework or tribo-chemically generated during sliding contact (Nevshupa et al., 2019). The released gas molecules also took part in the tribochemical interaction in the contact area, and some researchers speculated that these released gaseous molecules were confined in the contact area, which was possible to act as gas bearing due to the elasto-hydrostatic effect and reduced the friction coefficient (Kato et al., 2018).

Oxygen and Water Containing Ambient

Although DLC films are capable to achieve ultralow friction and wear in dry inert environment, it is still a challenge to optimize their tribological properties in humid atmosphere. In ambient air, hydrogen-rich DLC films generally exhibit higher friction coefficients and it is much harder to achieve superlubricity, while hydrogen-free DLC films exhibit lower friction coefficient in humid environment (Ronkainen and Holmberg, 2008). This difference is mainly induced by the interruption of oxygen and water molecules in tribochemical-induced structural evolution on the surface. Experiments have shown that from ultra-high vacuum to several thousand Pa of dry oxygen, the superlubricity behaviors of a-C:H films are not significantly influenced (Donnet et al., 1998). However, when the pressure of oxygen reaches above this threshold, the friction coefficient increases with it obviously (Kim et al., 2006). Some researchers believed that an oxide layer with a depth of several nanometers would be formed on the top of films due to the exposure to oxygen, which was one of the factors explaining the high friction coefficient in the running-in process (Al-Azizi et al., 2015). According to classical molecular dynamics simulations, oxygen molecules can cause the breakdown of carbon chains step by step and finally convert them into carbon dioxide (Moras et al., 2011). Meanwhile, the formation of C-O-C bonds bridging two contact surfaces can also induce adhesive effect and result in higher friction coefficients (Zeng et al., 2018).

Compared to oxygen, water molecules have more significant impacts on the superlubricity of DLC films, and the corresponding mechanisms are more complicated. For hydrogen-free DLC films, humidity usually has a positive influence, which can reduce the friction coefficient from 0.7 to a relatively low value below 0.2 (Fontaine et al., 2008). Some researches even showed that cleaved (111) diamond could realize superlubricity at 85% relative humidity when sliding perpendicular to their cleaved lines (Liu et al., 2007). However, the impacts of water molecules on hydrogen-rich DLC films are

quite otherwise. When humidity increases above a threshold, the superlubricity state will be interrupted, and their friction coefficient would increase above 0.05, approaching that of hydrogen-free DLC films (Fontaine et al., 2008). More detailed experiments showed that the threshold of water pressure was around 10 Pa, which was two orders of magnitude lower than that of oxygen, as shown in Kim et al. (2006). Several hypotheses have been proposed to explain this phenomenon. Firstly, water molecules are involved in the tribochemical interactions in the contact area. Calculations suggest that water molecules are energetically favorable to chemically absorb on the carbon chains (Moras et al., 2011). For hydrogen-free DLC films, this process can provide hydrogen and hydroxyl terminations, which could passivate the contact area. However, for hydrogen-rich DLC films, the oxygen-containing groups would cause hydrogen bond interactions, leading to higher adhesion force between the sliding surfaces. Energy dissipation during physical and chemical interactions between water molecules and the film also possibly contribute to the increase of friction. Secondly, higher humidity is also a barrier for the formation of a compact, robust carbonaceous transferfilm on the counterpart material (Donnet et al., 1998), which is considered to be a pivotal factor for DLC films to establish a non-adhesive sliding surface and realize their self-lubrication function. Thirdly, the physically absorbed water layer in humid ambient is also an influential factor. Experiments showed that about 1.5 layers of water molecules will be absorbed on a-C:H surface when the pressure of water was around 700 Pa (Shukla et al., 2003), and the adhesive force and friction force of a-C:H films would significantly increase when the thickness of absorbed water was higher than 1 monolayer (Tagawa et al., 2004). It is believed that the dipole-dipole interactions, the viscous drag effect, and the capillary effect of water layer absorbed on the film ultimately lead to a higher friction coefficient. As mentioned above, to combat the adverse impact of water molecules, introducing other elements such as silicon (Chen et al., 2013), sulfur (Freyman et al., 2006), or titanium (Zhao et al., 2016) is an effective approach. In these cases, water layers absorbed on the surfaces are utilized to establish the easy-shearing sliding surface. As shown in **Figure 5**, when sliding in moderate humid environment, silicon atoms were involved in the formation of a soft silica-like tribofilm and provided hydroxylated sliding interfaces with Si-OH termination and attracted water molecules directionally to arrange on the surface by hydrogen bond (Chen et al., 2013). Molecular dynamics simulations suggested that the layered water structure was robust in a wide range of normal loads and was capable of providing low shearing stress via boundary lubrication effect (Washizu et al., 2007).

In summary, the physical and chemical interactions between gas molecules and surfaces play an important role in the self-lubricating process of DLC films. The physical influences of environment are relative moderate, mainly through energy dissipation during the adsorption and desorption of gas molecules, heat transformation of contact area, elastic fluid dynamic pressure effect of gas in high-speed condition. The chemical reaction between surrounding molecules and surface material plays a most important role in the formation of

tribolayers and passivation of dangling bonds, which governs the superlubricity properties of DLC films.

INFLUENCE OF TESTING PARAMETERS

Understanding the performance of DLC films in different working conditions are important for the engineering application of these self-lubricating materials with superlubricity potential. In this section, the influence of temperature, normal load and sliding velocity are presented, respectively, and the most prevailing theories explaining their impacts on superlubricity are discussed.

Temperature

One of the advantages of solid lubricants is that they are applicable under both cryogenic and high temperature conditions where liquid lubrication cannot work. DLC is a case in point, which is capable of achieving ultralow friction coefficient in a wide range of temperatures. Temperature is also a core factor influencing physicochemical reaction, phase transformation and evolution of tribofilm, which fundamentally affects the superlubricity behavior of DLC films. DLC films exhibit superior tribological properties with a decreasing friction coefficient from cryogenic temperature to room temperature in ultrahigh vacuum (Aggleton et al., 2009). This trend continues in ambient air until reaching the lowest friction coefficient around 200–300°C, and then friction increases with elevated temperature (Lee and Wei, 2006; Zeng et al., 2015). Generally, in a moderate extent, the increase of temperature facilitates tribochemical reaction rate, which is conducive to friction-induced evolution of tribofilm and self-lubrication of DLC films. Specifically, more graphitization was observed (Zeng et al., 2015) and hydrogen emission rate was improved, promoting a more effective establishment of easy-shearing, highly passivated tribofilms. Meanwhile, the suppressed adsorption of gas or liquid layer at higher temperature may also be responsible for lower friction. Nevertheless, the heat resistance property of DLC films is confronted with a threshold. As the temperature is being higher, hydrogen spontaneous emission will become significant around 400–450°C (Liu et al., 1996), obvious thermal-induced graphitization will occur at around 500°C (Fink et al., 1983; Zeng et al., 2015), and thermal decomposition of C-H bond will start around 700°C (Su and Lin, 1998). These processes will result in an overquick hydrogen loss and more intensive π - π interactions, thus leading to higher friction coefficients. The deterioration caused by oxidation in ambient air also becomes a dominant factor, resulting in more extensive bond cleavage, and surface abrasion. As mentioned above, doping silicon or silicon-oxide into DLC films is an effective method to improve their thermal stability (Mangolini et al., 2018). Research shows that the friction coefficient of a-C:Si film is lower than 0.02 at a temperature ranging from 250 to 450°C (Jantschner et al., 2014). Adversely, in cryogenic condition, the tribochemical interaction and evolution of self-lubricated tribofilm are slowed down, and the flexibility of carbon network will be limited, leading to a higher friction coefficient. It should be pointed out that the friction coefficient of DLC films as well as other materials may

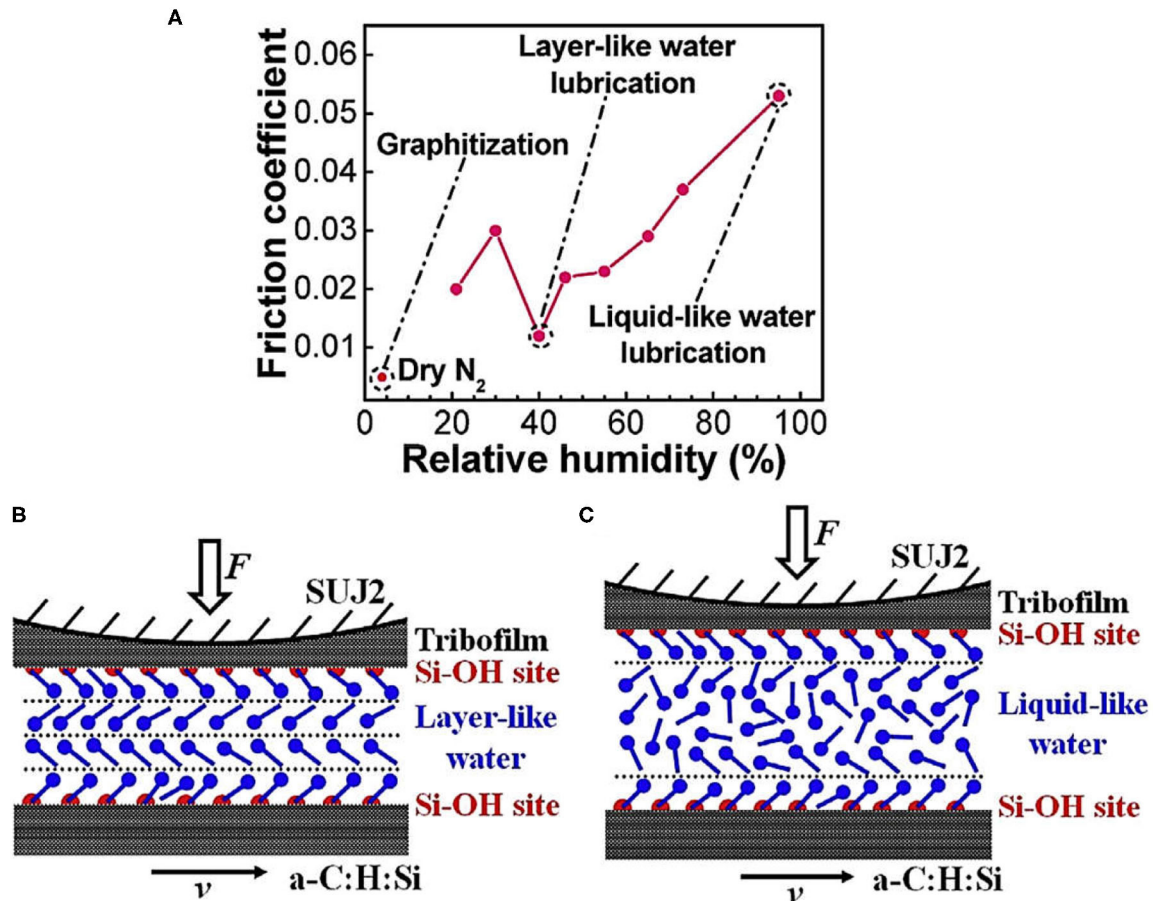


FIGURE 5 | Friction coefficient of a-C:H:Si under different humidities **(A)** and its lubricating mechanism: **(B)** layer-like water absorption on the hydroxylated surfaces at moderate relative humidity and **(C)** liquid-like water absorption under high relative humidity. Adapted with permission from Chen et al. (2013) Copyright 2013, IOP Publishing Ltd.

have a peak around 100 K, which is considered to be a common character possessed by various materials. This phenomenon is governed by the competition process between thermally activated formation and rupturing of atomic contact (Barel et al., 2010).

Friction induced temperature rise, or flash temperature of the contact area should also be considered. In some cases, the temperature rise could reach above 200°C (Yamamoto et al., 2019). Theoretically, temperature rise is depending on the friction energy and the thermal conductivity of the tribo-pairs, and is interrelated with other factors, such as applied normal load, contact diameter, and sliding velocity, which can be estimated by the following formula (Rabinowicz, 2013):

$$\Delta T = \mu F_n v / 4J (K_1 + K_2) a \quad (1)$$

where μ is the friction coefficient, F_n is the normal load, v is the sliding velocity, a is the radius of contact area, J is a constant, K_1 and K_2 are the thermal conductivities of ball and disk, respectively. Higher friction energy or sliding velocity could induce more significant temperature rises, while larger contact area would restrain temperature rise. Thus, it's possible for other

working condition parameters (such as normal load, sliding velocity) to influence the friction properties via thermal effect.

Normal Load

Normal load is an important factor influencing the superlubricity behavior of DLC films. Researches have shown that a-C:H films generally exhibit lower friction properties under higher normal load both in air (Zhang et al., 2004; Liu Y. et al., 2019) and inert gaseous ambient such as argon and nitrogen (Kunze et al., 2014; Chen et al., 2017). As shown in **Figure 6A**, the friction coefficient of a-C:H films decreased from 0.008 to 0.001 as the normal load increased from 2 N to 10 N. In vacuum, higher normal load can significantly shorten the lifetime of DLC films (Kunze et al., 2014). The influencing mechanism of normal load is multi-faceted.

Theoretically, the friction coefficient of dry sliding can be expressed as:

$$\mu = \frac{F_f}{F_n} = \frac{A_r \tau}{F_n} \quad (2)$$

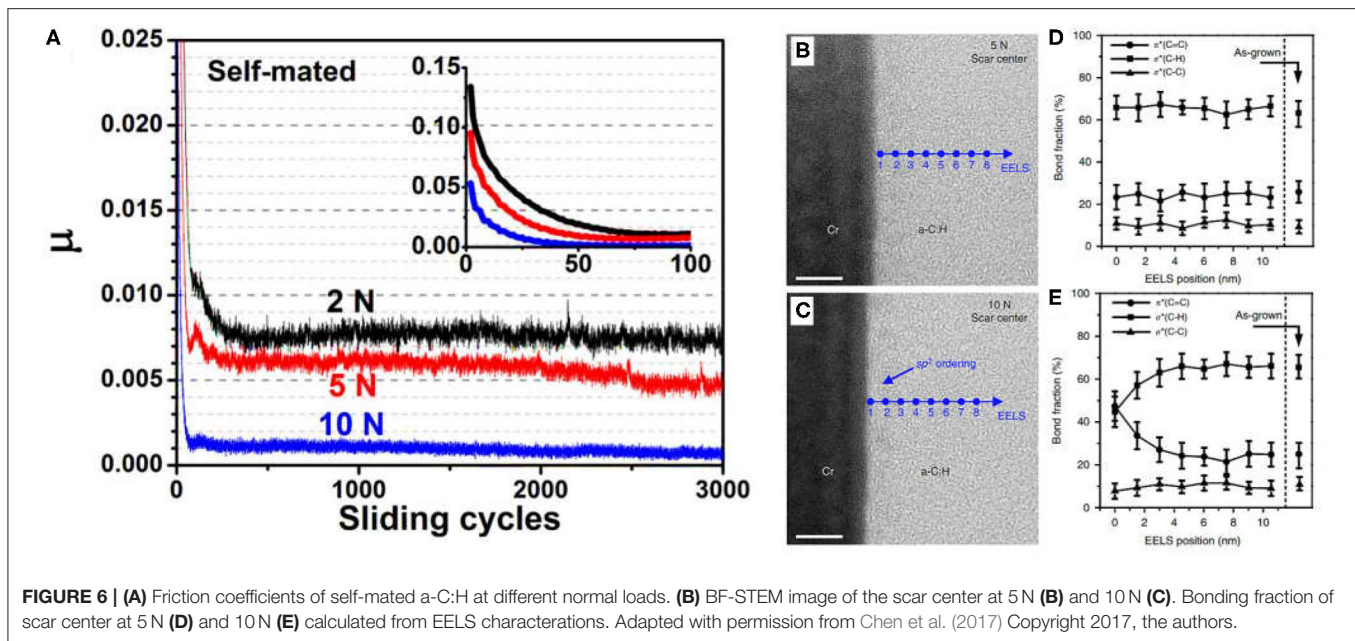


FIGURE 6 | (A) Friction coefficients of self-mated a-C:H at different normal loads. (B) BF-STEM image of the scar center at 5 N (B) and 10 N (C). Bonding fraction of scar center at 5 N (D) and 10 N (E) calculated from EELS characterizations. Adapted with permission from Chen et al. (2017) Copyright 2017, the authors.

where F_f is friction force, F_n is normal load, A_r is real contact area, and τ is shear stress (Fontaine et al., 2008). A_r can be approximately replaced by Hertz contact area which is defined as:

$$A_r = kA = k\pi \left(\frac{3R}{4E} \right)^{\frac{2}{3}} F_n^{\frac{2}{3}} \quad (3)$$

where k is the ratio of the real contact area to the apparent contact area, A is the Hertz contact area, R and E are the effective radius and modulus of two surface materials, respectively. Then, the friction coefficient μ can be expressed as:

$$\mu = k\pi \left(\frac{3R}{4E} \right)^{\frac{2}{3}} \tau F_n^{-\frac{1}{3}} \quad (4)$$

When the normal load increases, with the hypothesis that the DLC surface remains highly passivated (no structure change) and the shearing force τ generally stays constant, the correlation between normal load and friction coefficient is negative, which is believed to be one of the important factor for the lower friction under heavy load (Xue and Wang, 2012). However, as k increases with F_n , the negative influences of F_n on μ induced by contact area should be slighter than $F_n^{-\frac{1}{3}}$, which is lower than the tendency shown in experiments, indicating that the shearing force τ is negatively influenced by the increasing normal load.

As mentioned above, friction-induced thermal effect may be another possible reason for the negative correlation between normal load and friction coefficient. At the same circumstance, higher normal load would induce higher friction energy, which may result in a more significant temperature rise. However, according to Hertz contact theory, the contact diameter also increases with the normal load, which would partly restrain the

temperature rise due to more effective thermal conduction. In formula (1), if we replace a with Hertz contact radius and replace μ with formula (4), the temperature rise can be expressed as:

$$\Delta T = \frac{\pi k}{4J(K_1 + K_2)} \left(\frac{3R}{4E} \right)^{\frac{1}{3}} \tau F_n^{\frac{1}{3}} \nu \quad (5)$$

Regarding that the shearing stress τ is also negatively affected by normal load for DLC films, the temperature rise induced by higher normal load should be lower. And it should be pointed out that temperature rising is more obvious during the high-friction state or running-in period. But for a typical superlubricity state, the friction energy dissipation is every low (for example, no more than 0.005 W in the experiment of Figure 6), and the temperature rise is supposed to be very small, which is not enough to individually induce significant structural change of DLC films. Normal load induced structural evolution of tribolayers should be the key factor for more superior properties under heavy load.

Experiments and simulations have indicated that friction induced structural evolution occurred during sliding contact, which could lead to a lower shearing force τ . During this process, normal load is a key factor. As shown in Figure 6, an extremely low friction of 0.001 was achieved with the highest normal load of 10 N. BF-STEM images and EELS results indicated that in this heavy-load condition, significant phase transformations of carbon bonds from sp² to sp³ and local ordering appeared in the outer-most 3 nm region (Chen et al., 2017). These evidences suggest that higher normal load could facilitate the graphitization and hydrogen emission process in the contact area, leading to a nanostructured sliding surface with lower shearing forces. As shown in Figure 7, simulation with a-C films shows a similar trend that a higher normal load could induce more significant bond rehybridization, reorientation, and structural ordering within a localized shearing band, and friction state changes from

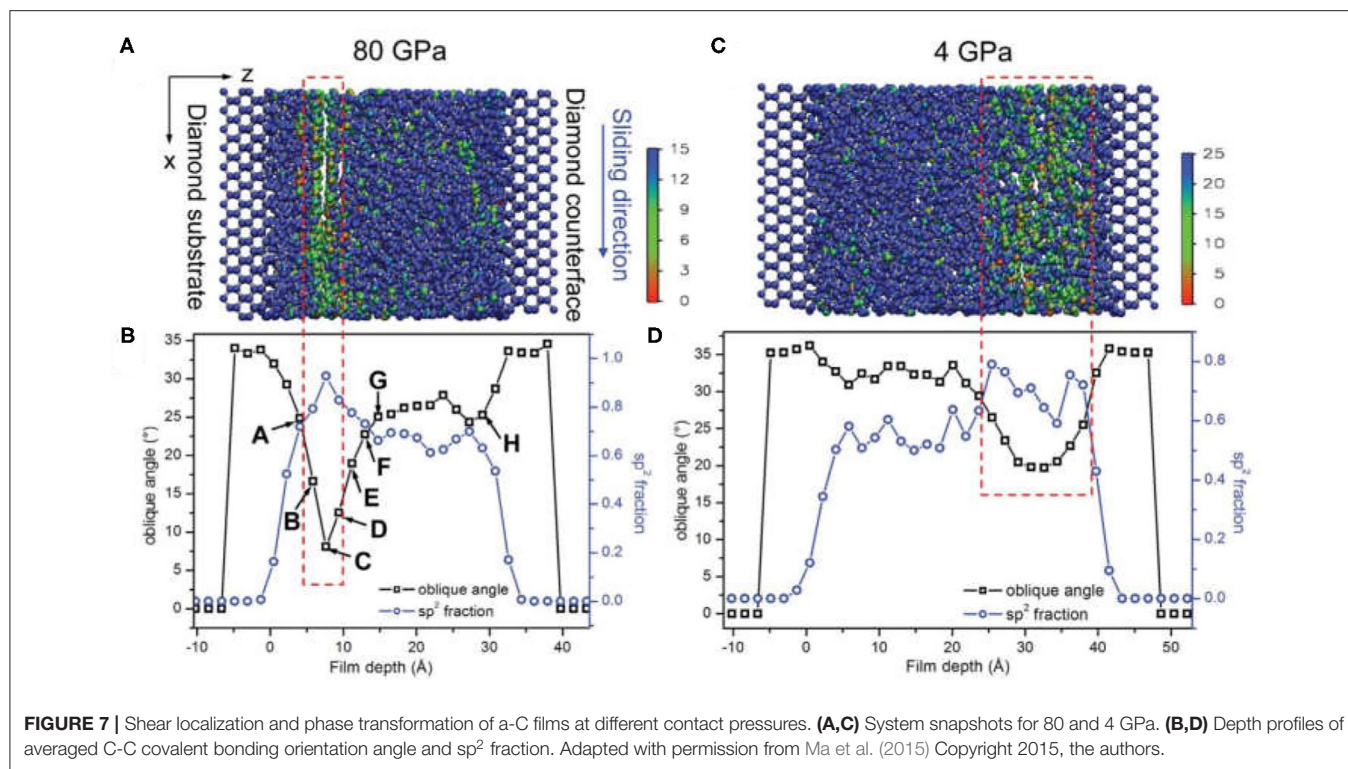


FIGURE 7 | Shear localization and phase transformation of a-C films at different contact pressures. **(A,C)** System snapshots for 80 and 4 GPa. **(B,D)** Depth profiles of averaged C-C covalent bonding orientation angle and sp^2 fraction. Adapted with permission from Ma et al. (2015) Copyright 2015, the authors.

stick-slip to continuous sliding with ultra-low friction (Ma et al., 2015).

When DLC films are coupled with other materials, the establishment of a robust carbonaceous transfer film on the counterpart material is a fundamental factor for superlubricity of DLC film, and this process is also influenced by the applied normal load, especially in the cases that the bonding strength between the counterpart material and carbon is relatively low (Xue and Wang, 2012). As shown in **Figure 8**, when Al_2O_3 ball was sliding against a-C:H films in vacuum at a normal load of 1 N, a stable ultralow friction coefficient of 0.002 was easily achieved and continuous carbonaceous transfer layers were formed on the ball. However, as the applied load increased to 3 N, the duration of superlubricity became shorter and ended up with a high friction state without transferfilm observed on the ball surface. Nevertheless, when these experiments were conducted in air, the stable state friction coefficient decreased from 0.15 to 0.06 with the normal load varied from 1 to 5 N, and an increasingly condensed transferfilm was developed on the ball scar (Liu Y. et al., 2019).

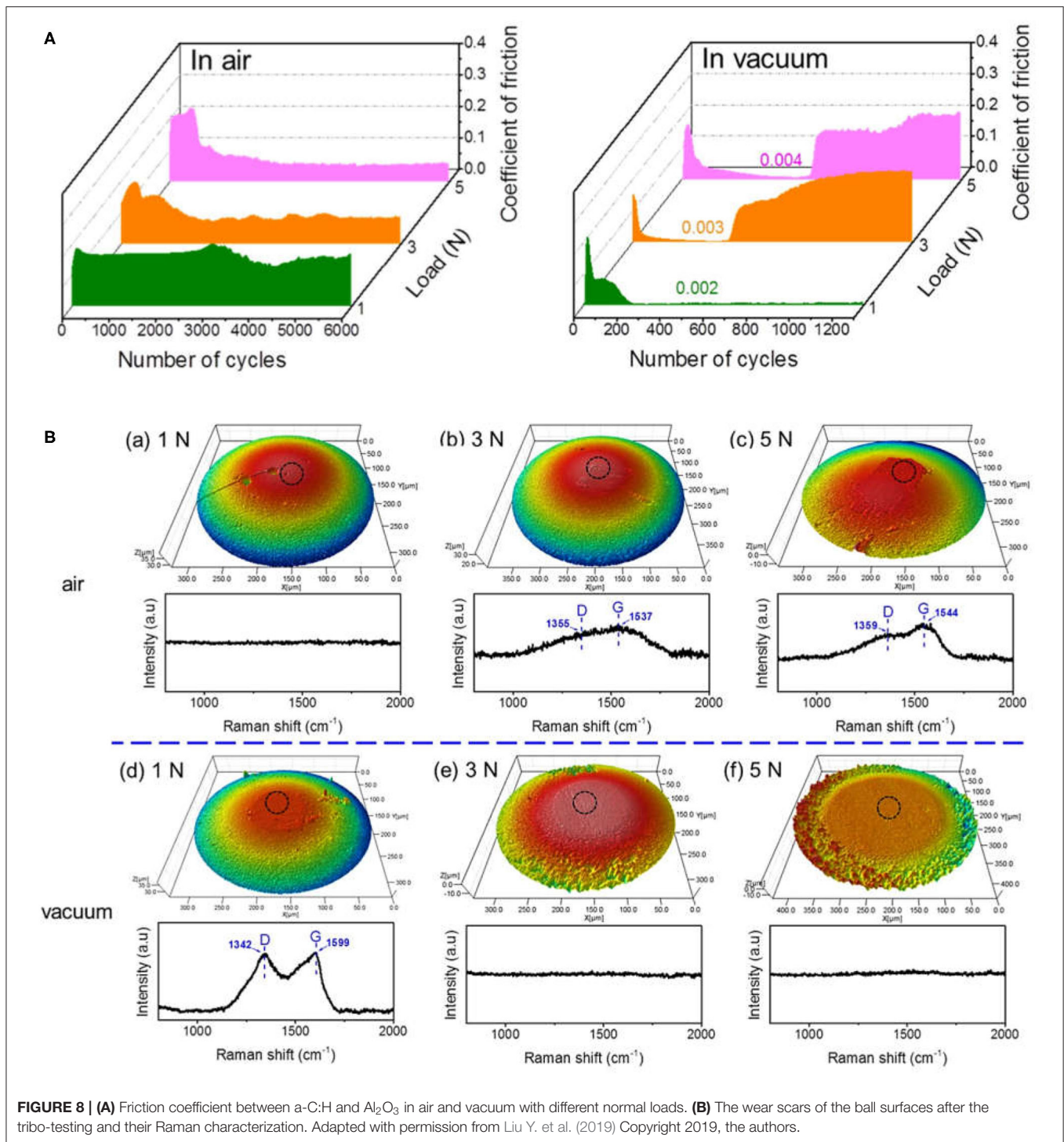
Sliding Velocity

Experiments have shown that sliding velocity also has a strong effect on superlubricity of DLC films. Generally, in ambient air or water vapor, higher sliding velocity will induce a decrease of friction coefficient (Kim et al., 2006; Xue and Wang, 2012). In inert gaseous ambient such as N_2 , similar trends were observed (Heimberg et al., 2001). To explain this phenomenon, multiple theories have been put forward. Friction induced

thermal effect, gas absorption, and establishment of transfer-films are considered to be the influencing mechanisms for this trend.

As mentioned above, higher sliding speed will induce a more intense temperature rise due to higher friction energy, which is possible to promote the self-lubrication properties of DLC films. As shown in Formula (5), compared with normal load, sliding velocity has a more significant influence on temperature rise. Nevertheless, in most experiments, thermal effect is still not high enough to induce dramatic structural changes of tribofilm, as calculations suggest that for a high-friction running-in state ($\mu = 0.22$, $F_n = 1$ N, $v = 700$ mm/s), the temperature rise ΔT is only 55.3°C (Liu et al., 2018), let alone the low energy dissipation of stable superlubricity state with friction coefficients lower than 0.01.

The most important factor for velocity dependence of superlubricous a-C:H is gas absorption. In gaseous environments, a chemical or physical absorbed layer exists on the surface of tribo-pairs, which will be removed during each sliding contact, causing energy dissipation, and occurrence of friction. According to the experiments coupled with the model based on kinetics of gas absorption (Borodich et al., 2007), higher sliding velocity will shorten the time interval between two slides of a single region, shortening the exposure time for gas absorption and reducing the energy cost for their remove. The influencing extent of sliding speed is closely related to the characteristics and density of ambient gases. For example, in ambient air (RH = 40%), when sliding speed increased from 262 to 2,093 mm/s, the friction coefficient was gradually reduced from 0.16 to 0.07. In contrast, their friction coefficients in



nitrogen ($\text{RH} = 5\%$) stayed constant around 0.02 for the same condition due to the relatively inert environment (Xue and Wang, 2012). Experiments under different pressures of water vapor showed that friction coefficient dropped faster with an increasing speed at relatively low pressure (Kim et al., 2006). Friction test with different reciprocating sliding distances at the same speed indicated that the exposure time dependence

of superlubricity of DLC in nitrogen or carbon dioxide was significantly higher than that in argon. Moreover, with the increase of exposure time, their friction coefficient first decreased and then increased. Researchers suggested that there existed an optimized exposure time for nitrogen or carbon dioxide directionally absorbing with π bonds and leaving lone electron pair pointing outside, which could reduce friction due to

repulsive force effect, inducing a minimum friction coefficient (Ji et al., 2009).

As mentioned above, when a-C:H coated planes slide against other materials, the establishment of a compact and continuous carbonaceous transfer layer on the ball side is an important factor for superlubricity. Likewise, researches have shown that when Al₂O₃ ball slides against a-C:H in vacuum, superlubricity state with friction coefficient of 0.003 can be achieved, as shown in **Figure 9A**. However, fluctuation of friction force increased with elevated sliding speed from 30 to 700 mm/s, and the proportion of superlubricity state decreased from 78.1 to 3.1%. According to SEM and Raman characteristics of the contact area, as shown in **Figures 9B,C**, the failure of superlubricity is always accompanied by the loss of transfer films. The authors believe that for this condition, excessive sliding speed will result in insufficient tribochemical interactions as well as intensified impact and vibration of friction interfaces, making it harder for transferlayers to be generated and tightly attached to the ball site, and resulting in the failure of superlubricity (Liu et al., 2018).

DISCUSSION AND CONCLUSION

The influencing factors and the mechanisms of superlubricity achieved with a-C:H films are multifaceted and interrelated. To date, diversified theories focusing on mechanical properties (Erdemir, 2004; Fontaine et al., 2004), establishment of transfer film (Chen et al., 2017; Liu et al., 2018; Liu Y. et al., 2019), surface passivation (Erdemir and Eryilmaz, 2014), and graphitization (Ma et al., 2015) have been brought up to explain the mechanism of solid superlubricity achieved with DLC films, and it seems that these factors are not working independently. Ultimately, the superlubricity state originates from the establishment of a sliding surface with low energy dissipation, meaning that the major interaction pathways between sliding surfaces are somehow sufficiently shielded. It is generally believed that the origin of friction force between solid surfaces can be qualitatively ascribed to three fundamental phenomena: abrasion, shearing, and adhesion (Fontaine et al., 2008), as shown in **Figure 10**. Abrasion is the plowing or scratching effect caused by micro-asperities or hard debris trapped between the sliding surfaces. Shearing is the energy consumption origins from the plastic or viscous flow of the material between the friction pairs. Adhesion is the resistance to break the micro-junctions bridging two sliding interfaces, mainly deriving from bonding force, electrostatic force, capillary force, and polarization forces. The realization of superlubricity needs all these friction contributors simultaneously to be limited at extremely low levels. DLC films exactly have the potential to accomplish this task owing to their numerous unique properties. Here, we can use superlubricous hydrogen-rich a-C:H film as an example for discussing.

From the mechanical perspective, DLC films exhibit special advantages in avoiding abrasion. Compared with other self-lubricating soft coatings, DLC films have higher hardness. For example, a hardness of 13.4 GPa for hydrogen-rich a-C:H film (Chen et al., 2017), is higher than that of common bearing steel or tool steel, which is generally around 10 GPa (Fontaine

et al., 2008). Meanwhile, the surface roughness of DLC film is generally in accordance with their substrate, meaning that it can be controlled at an extremely low level, even atomically smooth. For example, the surface roughness of a-C:H film deposited on silicon wafer was 0.10 nm (Chen et al., 2017). The high hardness and ultrasurface of DLC films can reduce the chance of abrasion in most cases. Moreover, due to the non-porous structure, hydrogen-rich DLC films also have unusual viscoelastic properties (Fontaine et al., 2004). As implied by the slow recovery of indents observed in nanoindentation experiments, a-C:H films may have some “healing” properties against scratches (Fontaine et al., 2008). These properties of DLC films can minimize the abrasive effect during sliding, paving the way for the establishment of stabilized, low-energy dissipation sliding interfaces.

When a-C:H film reaches a stable superlubricity state, the shearing behavior generally occurs inside an *in-situ* formed tribofilm or transferfilm with different mechanical properties, rather than in the film matrix (Fontaine et al., 2008). As shown in **Figure 11**, after friction against bare SUJ2 ball and Si₃N₄ ball, tribolayers of 27 and 5 nm thickness were generated on the balls, respectively. The establishment of these continuous, stably-attached carbonaceous tribofilms, together with the easy shearing nanostructure, will be a key factor for the superlubricity of DLC films (Liu et al., 2018; Liu Y. et al., 2019). High proportion of sp² rehybridized carbon phase can be observed within the tribofilms, and obvious reorientation will be taking place under heavy normal loads, forming layer-like structure (Chen and Li, 2020). Simulations suggest that the significant drop of friction during running-in process is also accompanied with shearing localization, limiting the occurrence of sliding in a thin region (Ma et al., 2015). Meanwhile, near the sliding surface, the as-formed carbon rings with higher sp² proportion are rotated parallel to the sliding direction, forming an interface with low proportion of out-of-plane sp³ bonds (Pastewka et al., 2010). The formation of easy-shearing sp²-C rich sliding interface (or called graphitization) and locally-ordered nanostructures, as well as the localization of shearing region, may be the key factors to keep the shearing resistance in a very low state, contributing to the stable superlubricous friction state.

Tribological tests with DLC films in microscale have shown that the friction vs. normal load curve does not go through the origin, indicating that adhesion cannot be ignored (Shi et al., 2020). For a well-established carbon-based sliding interface, the adhesion phenomenon caused by the establishment and breaking of micro junctions across sliding surfaces is supposed to be the main contributor for the friction force (Fontaine et al., 2008). Firstly, σ dangling bonds generated during mechanical and tribochemical interactions have potential to form strong covalent bonds bridging the sliding surfaces, causing high adhesive friction and wear (Schall et al., 2010). This is the reason for the high friction of hydrogen-free DLC films in dry inert environment or vacuum and its prevention of friction via hydrogen gaseous environment or hydrogen contained in the film. The hydrogen species could react and passivate the dangling bonds generated at sliding surface timely, and hence eliminate the high friction caused by cross-linking of the interfacial bridging

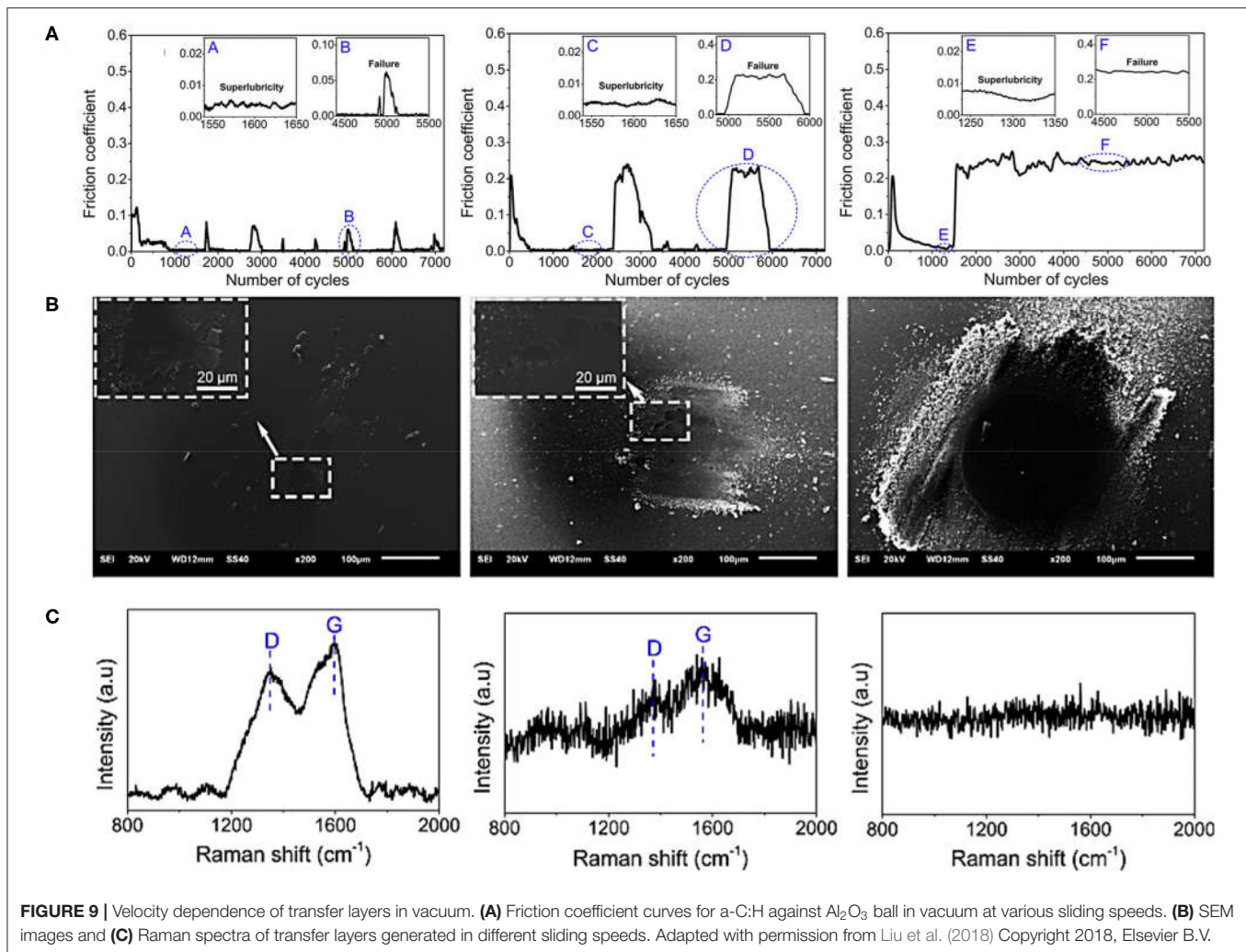


FIGURE 9 | Velocity dependence of transfer layers in vacuum. **(A)** Friction coefficient curves for a-C:H against Al₂O₃ ball in vacuum at various sliding speeds. **(B)** SEM images and **(C)** Raman spectra of transfer layers generated in different sliding speeds. Adapted with permission from Liu et al. (2018) Copyright 2018, Elsevier B.V.

(Robertson, 2002). A self-terminated structure with carbon rings could be established during running-in process, leading to a continuous passivated interface with fewer dangling bonds exposed (Pastewka et al., 2010). Secondly, the interaction between out-of-plane π -orbitals of sp^2 bonded carbon atoms, or π - π interaction, is another factor contributing to the friction force of DLC films. Likewise, this interaction can also be shielded by hydrogen, as they are reactive to carbon double bonds. Absorption of N₂ and CO₂ molecules can also restrain this phenomenon. Thirdly, for hydrogen-rich a-C:H films tested in humid environment, the hydrogen bond, and capillary force (mainly in micro-tribological experiments) induced by absorbed water layer are also contributors of friction force. Dry inert gaseous environments can prevent this phenomenon, thus superlubricity of a-C:H films can be achieved. Nevertheless, for hydrogen-free DLC, water can also act as a passivator preventing the strong π - π interactions and σ dangling bonds, leading to a relatively low friction, but has difficulty in achieving superlubricity. Fourthly, the van der Waals force is the weakest form for adhesion, and is the dominant factor for superlubricity state achieved with passivated DLC surfaces sliding in dry and

inert environment. The repulsive coulombic force between two positively-charged C-H terminated interfaces can also account for the ultralow friction of DLC films.

To sum up, superlubricity of a-C:H films achieved in dry inert environment mainly originates from three unique characteristics of the film: mechanical properties such as high hardness and viscoplastic properties and extremely low roughness, abilities in forming easy-shearing tribofilm or transferfilm through friction-induced phase transformation and bonding reorientation, abilities to passivate the sliding surface through self-contained hydrogen. All these factors work together to eliminate abrasion, shearing and adhesion effects to extremely low levels simultaneously, leading to superlubricity state. Basing on these understandings, involving other elements into the film is an effective approach to optimize their tribological properties since they could change the evolution process of tribofilms and state of sliding surface. As for gaseous ambient molecules, in addition to chemical involvement in the establishment of transfer film and passivation of sliding surface, they also influence friction force physically via thermal conduct and energy dissipation during desorption. Although the influence mechanisms of

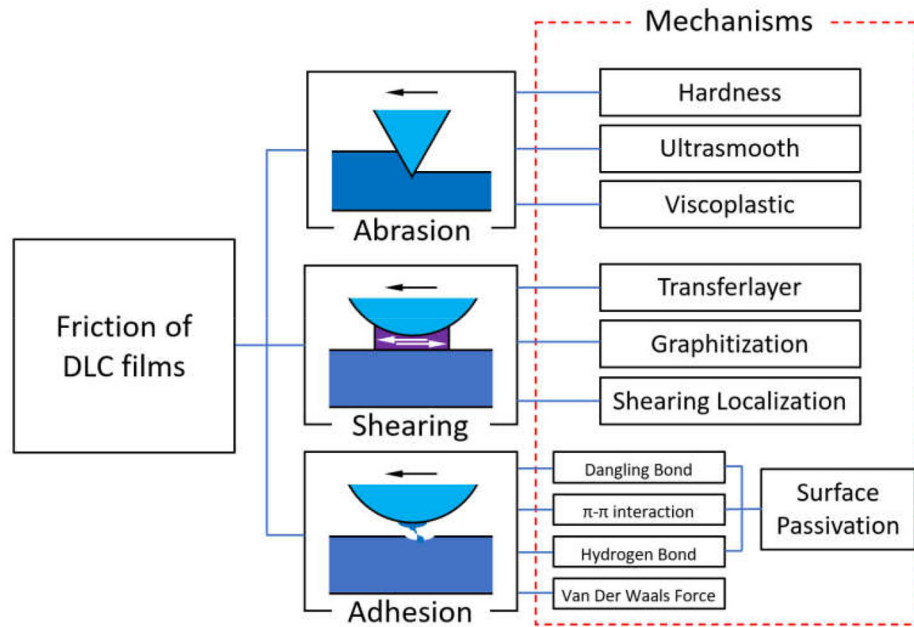


FIGURE 10 | The origins of friction for DLC films and mechanisms for their elimination.

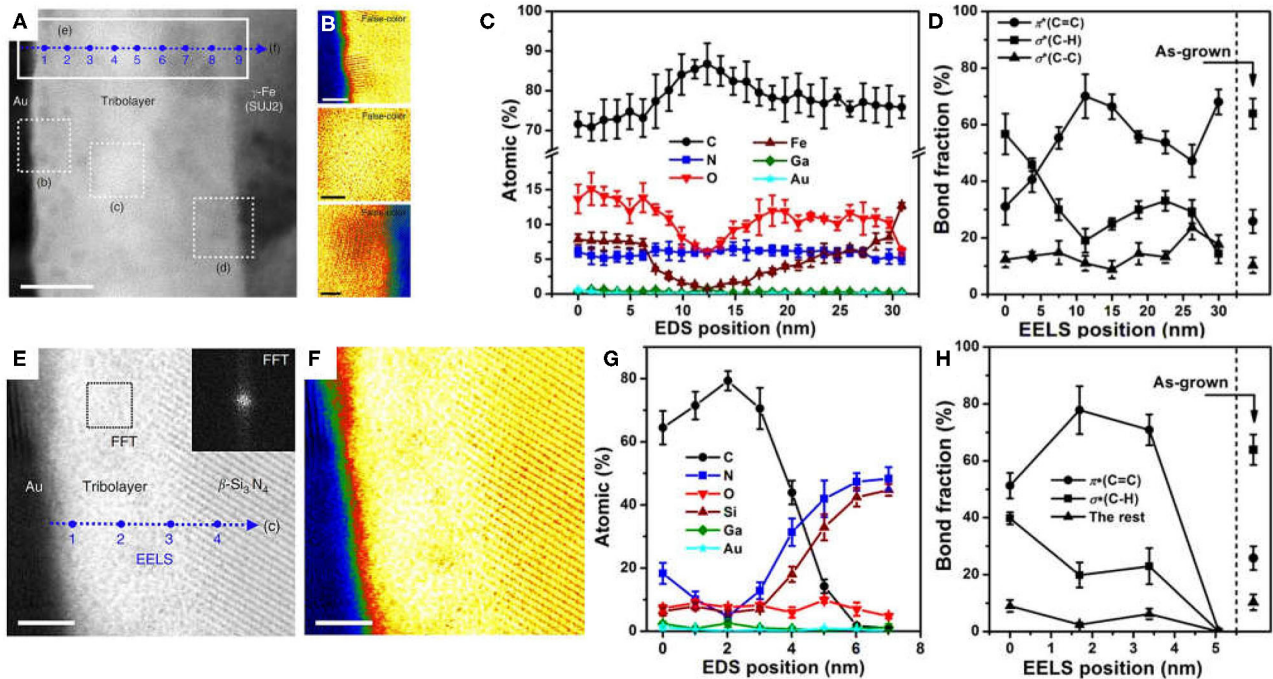


FIGURE 11 | (A–D) Characterization of tribolayer generated on the ball surface for the friction pair of bare SUJ2 ball and a-C:H films. **(A)** BF-STEM image showing a tribolayer of ~27 nm generated on the SUJ2 ball. **(B)** False-color displayed BF-STEM images of three individual sublayers marked in **(A)**. **(C)** The evolution of elemental composition across the tribolayer marked in **(A)**. **(D)** Evolution of the calculated EELS C-bonds fractions across the tribolayer as marked in **(A)**. **(E–H)** Characterization of tribolayer generated on ball surface for the friction pair of bare Si₃N₄ ball and a-C:H film. **(E)** BF-STEM image showing a tribolayer of ~5 nm generated on the Si₃N₄ ball. **(F)** False-color displayed BF-STEM images of **(E)**. **(G)** The evolution of elemental composition across the tribolayer marked in **(E)**. **(H)** Evolution of the calculated EELS C-bonds fractions across the tribolayer as marked in **(E)**. Adapted with permission from Chen et al. (2017) Copyright 2017, the authors.

temperature, normal load, and sliding velocity are multifaceted, under the precondition that they are not crossing the threshold of the mechanical or thermal capacity of the film, their influencing trends mainly depend on whether they promote or deteriorate the evolution process of robust easy-shearing tribofilms and passivation of sliding surfaces.

Since the discovery of DLC films, their application as a self-lubricating and protective coating in mechanical engineering is increasingly broadened. For example, their applications in gearings (Michalczewski et al., 2019), bearings (Okamura et al., 2019), engine piston rings (Kumar et al., 2019), oil-well tubings (Liu L. et al., 2019), rubber seals (Liu J. Q. et al., 2019), and friction pairs in spacecraft (Donnet et al., 1999) have been testified to be effective and valuable. Superlubricity achieved with DLC coatings can provide a new concept to meet the increasing demand for lower energy consumption, longer life time, and higher power density of future mechanical systems, which has an enormous potential value from both economic and environmental perspectives. Nevertheless, it is still a challenge for engineering applications of superlubricious DLC films. More efforts should be concentrated on preventing their environmental

sensitivity in humid condition and prolonging their service life in vacuum. Also, their performance still needs to be optimized in harsh working conditions such as extremely high or low temperature, extremely high pressure and velocity, and contaminated environment. Further researches on their superlubricity behaviors and mechanism can enable a more effective manipulation of DLC coatings with excellent self-lubrication properties.

AUTHOR CONTRIBUTIONS

JL, CZ, and XC gave the guidance on the structure and content of this review. QY completed this review. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (Grant Nos. 51975314, 51925506, and 51527901).

REFERENCES

- Adachi, K., and Kato, K. (2008). "Tribology of carbon nitride coatings" in *Tribology of Diamondlike Carbon Films: Fundamentals and Applications*, eds C. Donnet and A. Erdemir (New York, NY: Springer), 339–361.
- Aggleton, M., Burton, J. C., and Taborek, P. (2009). Cryogenic vacuum tribology of diamond and diamond-like carbon films. *J. Appl. Phys.* 106:013504. doi: 10.1063/1.3158339
- Aisenberg, S., and Chabot, R. (1971). Ion-beam deposition of thin films of diamondlike carbon. *J. Appl. Phys.* 42, 2953–2958. doi: 10.1063/1.1660654
- Al-Azizi, A. A., Eryilmaz, O., Erdemir, A., and Kim, S. H. (2015). Surface structure of hydrogenated diamond-like carbon: origin of run-in behavior prior to superlubricious interfacial shear. *Langmuir* 31, 1711–1721. doi: 10.1021/la504612c
- Andersson, J., Erck, R. A., and Erdemir, A. (2003). Frictional behavior of diamondlike carbon films in vacuum and under varying water vapor pressure. *Surface Coat. Technol.* 163–164, 535–540. doi: 10.1016/S0257-8972(02)00617-5
- Barel, I., Urbakh, M., Jansen, L., and Schirmeisen, A. (2010). Multibond dynamics of nanoscale friction: the role of temperature. *Phys. Rev. Lett.* 104:066104. doi: 10.1103/PhysRevLett.104.066104
- Borodich, F. M., Korach, C. S., and Keer, L. M. (2007). Modeling the tribochemical aspects of friction and gradual wear of diamond-like carbon films. *J. Appl. Mech.* 74, 23–30. doi: 10.1115/1.2172267
- Casiraghi, C., Ferrari, A. C., and Robertson, J. (2005). Raman spectroscopy of hydrogenated amorphous carbons. *Phys. Rev. B* 72:85401. doi: 10.1103/PhysRevB.72.085401
- Chen, X., and Kato, T. (2014). Growth mechanism and composition of ultrasmooth a-C:H:Si films grown from energetic ions for superlubricity. *J. Appl. Phys.* 115, 2143–2201. doi: 10.1063/1.4863123
- Chen, X., Kato, T., Kawaguchi, M., Nosaka, M., and Choi, J. (2013). Structural and environmental dependence of superlow friction in ion vapour-deposited a-C: H: Si films for solid lubrication application. *J. Phys. D* 46:255304. doi: 10.1088/0022-3727/46/25/255304
- Chen, X., Kato, T., and Nosaka, M. (2014). Origin of superlubricity in a-C:H:Si films: a relation to film bonding structure and environmental molecular characteristic. *ACS Appl. Mater. Inter.* 6, 13389–13405. doi: 10.1021/am502416w
- Chen, X., and Li, J. (2020). Superlubricity of carbon nanostructures. *Carbon* 158, 1–23. doi: 10.1016/j.carbon.2019.11.077
- Chen, X., Zhang, C., Kato, T., Yang, X., Wu, S., Wang, R., et al. (2017). Evolution of tribo-induced interfacial nanostructures governing superlubricity in a-C:H and a-C:H:Si films. *Nat. Commun.* 8:1675. doi: 10.1038/s41467-017-01717-8
- Dag, S., and Ciraci, S. (2004). Atomic scale study of superlow friction between hydrogenated diamond surfaces. *Phys. Rev. B* 70:241401. doi: 10.1103/PhysRevB.70.241401
- De Barros Bouchet, M. I., Matta, C., Le-Mogne, T., Martin, J. M., Sagawa, T., Okuda, S., et al. (2007). Improved mixed and boundary lubrication with glycerol-diamond technology. *Tribol. Mater. Surf. Interfaces* 1, 28–32. doi: 10.1179/175158407x181507
- Dienwiebel, M., Verhoeven, G. S., Pradeep, N., and Frenken, J. W. M. (2004). Superlubricity of graphite. *Phys. Rev. Lett.* 92:126101. doi: 10.1103/PhysRevLett.92.126101
- Donnet, C., Belin, M., Augé, J. C., Martin, J. M., Grill, A., and Patel, V. (1994). Tribochemistry of diamond-like carbon coatings in various environments. *Surf. Coat. Technol.* 68–69, 626–631. doi: 10.1016/0257-8972(94)90228-3
- Donnet, C., Fontaine, J., Le Mogne, T., Belin, M., Héau, C., Terrat, J. P., et al. (1999). Diamond-like carbon-based functionally gradient coatings for space tribology. *Surf. Coat. Technol.* 120–121, 548–554. doi: 10.1016/S0257-8972(99)00432-6
- Donnet, C., and Grill, A. (1997). Friction control of diamond-like carbon coatings. *Surf. Coat. Technol.* 94–95, 456–462. doi: 10.1016/S0257-8972(97)00275-2
- Donnet, C., Mogne, T. L., Ponsonnet, L., Belin, M., Grill, A., Patel, V., et al. (1998). The respective role of oxygen and water vapor on the tribology of hydrogenated diamond-like carbon coatings. *Tribol. Lett.* 4, 259–265. doi: 10.1023/A:1019140213257
- Enke, K. (1981). Some new results on the fabrication of and the mechanical, electrical and optical properties of i-carbon layers. *Thin Solid Films* 80, 227–234. doi: 10.1016/0040-6090(81)90226-1
- Erdemir, A. (2001). The role of hydrogen in tribological properties of diamond-like carbon films. *Surf. Coat. Technol.* 146–147, 292–297. doi: 10.1016/S0257-8972(01)01417-7
- Erdemir, A. (2004). Genesis of superlow friction and wear in diamondlike carbon films. *Tribol. Int.* 37, 1005–1012. doi: 10.1016/j.triboint.2004.07.018
- Erdemir, A., and Eryilmaz, O. (2014). Achieving superlubricity in DLC films by controlling bulk, surface, and tribochemistry. *Friction* 2, 140–155. doi: 10.1007/s40544-014-0055-1

- Erdemir, A., and Eryilmaz, O. L. (2007). "Superlubricity in diamondlike carbon films" in *Superlubricity*, eds A. Erdemir and J. Martin (Amsterdam: Elsevier Science B.V.), 253–271.
- Erdemir, A., Eryilmaz, O. L., and Fenske, G. (2000c). Synthesis of diamondlike carbon films with superlow friction and wear properties. *J. Vacuum Sci. Technol.* A 18, 1987–1992. doi: 10.1116/1.582459
- Erdemir, A., Eryilmaz, O. L., Nilufer, I. B., and Fenske, G. R. (2000a). Effect of source gas chemistry on tribological performance of diamond-like carbon films. *Diam. Relat. Mater.* 9, 632–637. doi: 10.1016/S0925-9635(99)00361-1
- Erdemir, A., Eryilmaz, O. L., Nilufer, I. B., and Fenske, G. R. (2000b). Synthesis of superlow-friction carbon films from highly hydrogenated methane plasmas. *Surf. Coat. Technol.* 133–134, 448–454. doi: 10.1016/S0257-8972(00)00968-3
- Erdemir, A., Switala, M., Wei, R., and Wilbur, P. (1991). A tribological investigation of the graphite-to-diamond-like behavior of amorphous carbon films ion beam deposited on ceramic substrates. *Surf. Coat. Technol.* 50, 17–23. doi: 10.1016/0257-8972(91)90187-2
- Eryilmaz, O. L., and Erdemir, A. (2008). On the hydrogen lubrication mechanism(s) of DLC films: an imaging TOF-SIMS study. *Surf. Coat. Tech.* 203, 750–755. doi: 10.1016/j.surfcoat.2008.06.156
- Fink, J., Müller-Heinzerling, T., Pflüger, J., Bubenzer, A., Koidl, P., and Crecelius, G. (1983). Structure and bonding of hydrocarbon plasma generated carbon films: An electron energy loss study. *Solid State Commun.* 47, 687–691. doi: 10.1016/0038-1098(83)90635-X
- Fontaine, J., Donnet, C., and Erdemir, A. (2008). "Fundamentals of the tribology of DLC coatings" in *Tribology of Diamondlike Carbon Films: Fundamentals and Applications* (New York, NY: Springer), 139–154.
- Fontaine, J., Donnet, C., Grill, A., and LeMogne, T. (2001). Tribochemistry between hydrogen and diamond-like carbon films. *Surf. Coat. Technol.* 146–147, 286–291. doi: 10.1016/S0257-8972(01)01398-6
- Fontaine, J., Le Mogne, T., Loubet, J. L., and Belin, M. (2005). Achieving superlow friction with hydrogenated amorphous carbon: some key requirements. *Thin Solid Films* 482, 99–108. doi: 10.1016/j.tsf.2004.11.126
- Fontaine, J., Loubet, J. L., Mogne, T. L., and Grill, A. (2004). Superlow friction of diamond-like carbon films: a relation to viscoplastic properties. *Tribol. Lett.* 17, 709–714. doi: 10.1007/s11249-004-8077-x
- Freyman, C. A., Chen, Y., and Chung, Y. (2006). Synthesis of carbon films with ultra-low friction in dry and humid air. *Surf. Coat. Technol.* 201, 164–167. doi: 10.1016/j.surfcoat.2005.11.075
- Gao, G. T., Mikulski, P. T., Chateaufort, G. M., and Harrison, J. A. (2003). The effects of film structure and surface hydrogen on the properties of amorphous carbon films. *J. Phys. Chem. B* 107, 11082–11090. doi: 10.1021/jp034544+
- Gong, Z., Shi, J., Zhang, B., and Zhang, J. (2017). Graphene nano scrolls responding to superlow friction of amorphous carbon. *Carbon* 116, 310–317. doi: 10.1016/j.carbon.2017.01.106
- Goto, M., and Honda, F. (2004). Film-thickness effect of Ag lubricant layer in the nano-region. *Wear* 256, 1062–1071. doi: 10.1016/S0043-1648(03)00530-1
- Heimberg, J. A., Wahl, K. J., Singer, I. L., and Erdemir, A. (2001). Superlow friction behavior of diamond-like carbon coatings: time and speed effects. *Appl. Phys. Lett.* 78, 2449–2451. doi: 10.1063/1.1366649
- Hirano, M., and Shinjo, K. (1990). Atomistic locking and friction. *Phys. Rev. B* 41, 11837–11851. doi: 10.1103/PhysRevB.41.11837
- Hirano, M., Shinjo, K., Kaneko, R., and Murata, Y. (1991). Anisotropy of frictional forces in muscovite mica. *Phys. Rev. Lett.* 67, 2642–2645. doi: 10.1103/PhysRevLett.67.2642
- Hirano, M., Shinjo, K., Kaneko, R., and Murata, Y. (1997). Observation of superlubricity by scanning tunneling microscopy. *Phys. Rev. Lett.* 78, 1448–1451. doi: 10.1103/PhysRevLett.78.1448
- Holmberg, K., and Erdemir, A. (2017). Influence of tribology on global energy consumption, costs and emissions. *Friction* 5, 263–284. doi: 10.1007/s40544-017-0183-5
- Jantschner, O., Field, S. K., Music, D., Terziyska, V. L., Schneider, J. M., Munnik, F., et al. (2014). Sputtered Si-containing low-friction carbon coatings for elevated temperatures. *Tribol. Int.* 77, 15–23. doi: 10.1016/j.triboint.2014.04.006
- Ji, L., Li, H., Zhao, F., Quan, W., Chen, J., and Zhou, H. (2009). Effects of environmental molecular characteristics and gas-surface interaction on friction behaviour of diamond-like carbon films. *J. Phys. D* 42:135301. doi: 10.1088/0022-3727/42/13/135301
- Jiang, J., Hao, J., Wang, P., and Liu, W. (2010). Superlow friction of titanium/silicon codoped hydrogenated amorphous carbon film in the ambient air. *J. Appl. Phys.* 108:033510. doi: 10.1063/1.3462469
- Kano, M. (2006). Super low friction of DLC applied to engine cam follower lubricated with ester-containing oil. *Tribol. Int.* 39, 1682–1685. doi: 10.1016/j.triboint.2006.02.068
- Kato, K., Umehara, N., and Adachi, K. (2003). Friction, wear and N₂-lubrication of carbon nitride coatings: a review. *Wear* 254, 1062–1069. doi: 10.1016/S0043-1648(03)00334-X
- Kato, T., Matsuoka, H., Kawaguchi, M., and Nosaka, M. (2018). Possibility of elasto-hydrostatic evolved-gas bearing as one of the mechanisms of superlubricity. *Proc. Instit. Mech. Eng. Part J* 233, 532–540. doi: 10.1177/1350650117746025
- Kim, H. I., Lince, J. R., Eryilmaz, O. L., and Erdemir, A. (2006). Environmental effects on the friction of hydrogenated DLC films. *Tribol. Lett.* 21, 51–56. doi: 10.1007/s11249-005-9008-1
- Kumar, N., Sharma, N., Dash, S., Popov, C., Kulisch, W., Reithmaier, J. P., et al. (2011). Tribological properties of ultrananocrystalline diamond films in various test atmosphere. *Tribol. Int.* 44, 2042–2049. doi: 10.1016/j.triboint.2011.09.003
- Kumar, V., Sinha, S. K., and Agarwal, A. K. (2019). Wear evaluation of engine piston rings coated with dual layer hard and soft coatings. *J. Tribol.* 141:031301. doi: 10.1115/1.4041762
- Kunze, T., Posselt, M., Gemming, S., Seifert, G., Konicek, A. R., Carpick, R. W., et al. (2014). Wear, plasticity, and rehybridization in tetrahedral amorphous carbon. *Tribol. Lett.* 53, 119–126. doi: 10.1007/s11249-013-0250-7
- Kuwahara, T., Romero, P. A., Makowski, S., Weihnacht, V., Moras, G., and Moseler, M. (2019). Mechano-chemical decomposition of organic friction modifiers with multiple reactive centres induces superlubricity of ta-C. *Nat. Commun.* 10:151. doi: 10.1038/s41467-018-08042-8
- Lee, K. Y., and Wei, R. (2006). Tribological characteristics of DLC-coated alumina at high temperatures. *J. Tribol.* 128, 711–717. doi: 10.1115/1.2345395
- Liu, E., Ding, Y. F., Li, L., Blanpain, B., and Celis, J. P. (2007). Influence of humidity on the friction of diamond and diamond-like carbon materials. *Tribol. Int.* 40, 216–219. doi: 10.1016/j.triboint.2005.09.012
- Liu, J. Q., Wu, Z. Y., Cao, H. T., Wen, F., and Pei, Y. T. (2019). Effect of bias voltage on the tribological and sealing properties of rubber seals modified by DLC films. *Surf. Coat. Technol.* 360, 391–399. doi: 10.1016/j.surfcoat.2018.12.100
- Liu, L., Wu, Z., Cui, S., An, X., Ma, Z., Shao, T., et al. (2019). Abrasion and erosion behavior of DLC-coated oil-well tubings in a heavy oil/sand environment. *Surf. Coat. Technol.* 357, 379–383. doi: 10.1016/j.surfcoat.2018.09.081
- Liu, X., Yang, J., Hao, J., Zheng, J., Gong, Q., and Liu, W. (2012). A near-frictionless and extremely elastic hydrogenated amorphous carbon film with self-assembled dual nanostructure. *Adv. Mater.* 24, 4614–4617. doi: 10.1002/adma.201200085
- Liu, Y., Chen, L., Zhang, B., Cao, Z., Shi, P., Peng, Y., et al. (2019). Key role of transfer layer in load dependence of friction on hydrogenated diamond-like carbon films in humid air and vacuum. *Materials* 12:1550. doi: 10.3390/ma12091550
- Liu, Y., Erdemir, A., and Meletis, E. I. (1996). An investigation of the relationship between graphitization and frictional behavior of DLC coatings. *Surf. Coat. Tech.* 86–87, 564–568. doi: 10.1016/S0257-8972(96)03057-5
- Liu, Y., Yu, B., Cao, Z., Shi, P., Zhou, N., Zhang, B., et al. (2018). Probing superlubricity stability of hydrogenated diamond-like carbon film by varying sliding velocity. *Appl. Surf. Sci.* 439, 976–982. doi: 10.1016/j.apsusc.2018.01.048
- Ma, T., Wang, L., Hu, Y., Li, X., and Wang, H. (2015). A shear localization mechanism for lubricity of amorphous carbon materials. *Sci. Rep.* 4:3662. doi: 10.1038/srep03662
- Mangolini, F., Krick, B. A., Jacobs, T. D. B., Khanal, S. R., Streller, F., McClimon, J. B., et al. (2018). Effect of silicon and oxygen dopants on the stability of hydrogenated amorphous carbon under harsh environmental conditions. *Carbon* 130, 127–136. doi: 10.1016/j.carbon.2017.12.096
- Martin, J. M., Donnet, C., Mogne, T. L., and Epicier, T. (1993). Superlubricity of molybdenum disulfide. *Phys. Rev. B Cond. Matter* 48, 10583–10586. doi: 10.1103/PhysRevB.48.10583
- Michalczewski, R., Kalbarczyk, M., Mankowska-Snopczynska, A., Osuch-Słomka, E., Piekoszewski, W., Snarski-Adamski, A., et al. (2019). The effect of a gear oil on abrasion, scuffing, and pitting of the DLC-coated 18CrNiMo7-6 steel. *Coatings* 9:2. doi: 10.3390/coatings9010002

- Miura, K., Tsuda, D., and Sasaki, N. (2005). Superlubricity of C60 intercalated graphite films. *E J. Surf. Sci. Nanotechnol.* 3, 21–23. doi: 10.1380/ejssnt.2005.21
- Moolsradoo, N., and Watanabe, S. (2010). Modification of tribological performance of DLC films by means of some elements addition. *Diam. Relat. Mater.* 19, 525–529. doi: 10.1016/j.diamond.2010.01.010
- Moras, G., Pastewka, L., Gumbsch, P., and Moseler, M. (2011). Formation and oxidation of linear carbon chains and their role in the wear of carbon materials. *Tribol. Lett.* 44, 355–365. doi: 10.1007/s11249-011-9864-9
- Nevshupa, R., Caro, J., Arratibel, A., Bonet, R., Rusanov, A., Ares, J. R., et al. (2019). Evolution of tribologically induced chemical and structural degradation in hydrogenated a-C coatings. *Tribol. Int.* 129, 177–190. doi: 10.1016/j.triboint.2018.08.023
- Nobili, L., and Guglielmini, A. (2013). Thermal stability and mechanical properties of fluorinated diamond-like carbon coatings. *Surf. Coat. Technol.* 219, 144–150. doi: 10.1016/j.surfcoat.2013.01.018
- Nosaka, M., Morisaki, Y., Fujiwara, T., Tokai, H., Kawaguchi, M., and Kato, T. (2017). The run-in process for stable friction fade-out and tribofilm analyses by SEM and nano-indenter. *Tribol. Online* 12, 274–280. doi: 10.2474/trol.12.274
- Okamura, Y., Suzuki, D., Ikoma, K., Nagatomo, T., and Utsunomiya, H. (2019). Effects of segment-structured DLC film on the fretting wear of railway axle journal bearings. *Mech. Eng.* 6:18-00446. doi: 10.1299/mej.18-00446
- Okubo, H., Tsuboi, R., and Sasaki, S. (2015). Frictional properties of DLC films in low-pressure hydrogen conditions. *Wear* 340–341, 2–8. doi: 10.1016/j.wear.2015.03.018
- Pastewka, L., Moser, S., and Moseler, M. (2010). Atomistic insights into the running-in, lubrication, and failure of hydrogenated diamond-like carbon coatings. *Tribol. Lett.* 39, 49–61. doi: 10.1007/s11249-009-9566-8
- Pastewka, L., Moser, S., Moseler, M., Blug, B., Meier, S., Hollstein, T., et al. (2008). The running-in of amorphous hydrocarbon tribocoatings: a comparison between experiment and molecular dynamics simulations. *Int. J. Mater. Res.* 99, 1136–1143. doi: 10.3139/146.101747
- Rabinowicz, E. (2013). *Friction and Wear of Materials, 2nd Edn.* New York, NY: Wiley.
- Robertson, J. (2002). Diamond-like amorphous carbon. *Mater. Sci. Eng. R Rep.* 37, 129–281. doi: 10.1016/S0927-796X(02)00005-0
- Ronkainen, H., and Holmberg, K. (2008). “Environmental and thermal effects on the tribological performance of DLC coatings,” in *Tribology of Diamond-Like Carbon: Fundamentals and Applications*, eds C. Donnet and A. Erdemir (New York, NY: Springer), 155–200.
- Schall, J. D., Gao, G., and Harrison, J. A. (2010). Effects of adhesion and transfer film formation on the tribology of self-mated DLC contacts. *J. Phys. Chem. C* 114, 5321–5330. doi: 10.1021/jp904871t
- Shi, P., Sun, J., Yan, W., Zhou, N., Zhang, J., et al. (2020). Roles of phase transition and surface property evolution in nanotribological behaviors of H-DLC: Effects of thermal and UV irradiation treatments. *Appl. Surf. Sci.* 514:145960. doi: 10.1016/j.apsusc.2020.145960
- Shukla, N., Svedberg, E., van der Veerdonk, R. J. M., Ma, X., and Gellman, A. J. (2003). Water adsorption on lubricated aCH x in humid environments. *Tribol. Lett.* 15, 9–14. doi: 10.1023/A:1023503904927
- Socoliuc, A., Bennewitz, R., Gnecco, E., and Meyer, E. (2004). Transition from stick-slip to continuous sliding in atomic friction: entering a new regime of ultralow friction. *Phys. Rev. Lett.* 92:134301. doi: 10.1103/PhysRevLett.92.134301
- Song, Y., Davide, M., Oded, H., Michael, U., Ma, M., and Zheng, Q. (2018). Robust microscale superlubricity in graphite/hexagonal boron nitride layered heterojunctions. *Nat. Mater.* 17, 894–899. doi: 10.1038/s41563-018-0144-z
- Su, C., and Lin, J. C. (1998). Thermal desorption of hydrogen from the diamond C(100) surface. *Surf. Sci.* 406, 0–166. doi: 10.1016/s0039-6028(98)00107-1
- Sugimoto, I., and Miyake, S. (1990). Oriented hydrocarbons transferred from a high performance lubricative amorphous C:H:Si film during sliding in a vacuum. *Appl. Phys. Lett.* 56, 1868–1870. doi: 10.1063/1.103072
- Tagawa, M., Ikemura, M., Nakayama, Y., and Ohmae, N. (2004). Effect of water adsorption on microtribological properties of hydrogenated diamond-like carbon films. *Tribol. Lett.* 17, 575–580. doi: 10.1023/B:TRIL.0000044507.44022.13
- Vanhulsel, A., Velasco, F., Jacobs, R., Eersels, L., Havermans, D., Roberts, E. W., et al. (2007). DLC solid lubricant coatings on ball bearings for space applications. *Tribol. Int.* 40, 1186–1194. doi: 10.1016/j.triboint.2006.12.005
- Voevodin, A. A., and Zabinski, S. J. (2000). Supertough wear-resistant coatings with ‘chameleon’ surface adaptation. *Thin Solid Films* 370, 223–231. doi: 10.1016/S0040-6090(00)00917-2
- Wang, C., Yang, S., Wang, Q., Wang, Z., and Zhang, J. (2008). Super-low friction and super-elastic hydrogenated carbon films originated from a unique fullerene-like nanostructure. *Nanotechnology* 19:225709. doi: 10.1088/0957-4484/19/22/225709
- Wang, Y., Wang, J., Zhang, G., Wang, L., and Yan, P. (2012). Microstructure and tribology of TiC(Ag)/a-C:H nanocomposite coatings deposited by unbalanced magnetron sputtering. *Surf. Coat. Technol.* 206, 3299–3308. doi: 10.1016/j.surfcoat.2012.01.036
- Washizu, H., Sanda, S., Hyodo, S., Ohmori, T., Nishino, N., and Suzuki, A. (2007). Molecular dynamics simulations of elasto-hydrodynamic lubrication and boundary lubrication for automotive tribology. *J. Phys.* 89:012009. doi: 10.1088/1742-6596/89/1/012009
- Xue, Q., and Wang, L. (2012). *Nanoscience and Technology: Diamond-Like Carbon Film Materials*. Beijing: Science Press (in Chinese).
- Yamamoto, S., Liskiewicz, T., Fujimura, K., Tashiro, K., and Takai, O. (2019). Temperature rise of diamond-like carbon during sliding: consideration of the real contact area. *Tribol. Int.* 131, 496–507. doi: 10.1016/j.triboint.2018.09.022
- Zeng, Q., Cao, Q., Erdemir, A., Li, S., and Zhu, J. (2018). Current development situation of superlow friction behavior of DLC films. *China Surf. Eng.* 31, 1–19. doi: 10.11933/j.issn.1007-9289.20180202002 (in Chinese).
- Zeng, Q., Eryilmaz, O., and Erdemir, A. (2015). Superlubricity of the DLC films-related friction system at elevated temperature. *RSC Adv.* 5, 93147–93154. doi: 10.1039/C5RA16084G
- Zeng, Q., Yu, F., and Dong, G. (2013). Superlubricity behaviors of Si3N4/DLC films under PAO oil with nano boron nitride additive lubrication. *Surf. Interface Anal.* 45, 1283–1290. doi: 10.1002/sia.5269
- Zhang, R., Ning, Z., Zhang, Y., Zheng, Q., Chen, Q., Xie, H., et al. (2013). Superlubricity in centimetres-long double-walled carbon nanotubes under ambient conditions. *Nat. Nanotechnol.* 8, 912–916. doi: 10.1038/nnano.2013.217
- Zhang, S., Wagner, G., Medyanik, S. N., Liu, W., Yu, Y., and Chung, Y. (2004). Experimental and molecular dynamics simulation studies of friction behavior of hydrogenated carbon films. *Surf. Coat. Technol.* 177–178, 818–823. doi: 10.1016/j.surfcoat.2003.06.022
- Zhang, W., Tanaka, A., Wazumi, K., and Koga, Y. (2002). Effect of environment on friction and wear properties of diamond-like carbon film. *Thin Solid Films* 413, 104–109. doi: 10.1016/s0040-6090(02)00351-6
- Zhao, F., Li, H., Ji, L., Wang, Y., Liu, X., Zhou, H., et al. (2016). Effect of microstructural evolution on mechanical and tribological properties of Ti-doped DLC films: how was an ultralow friction obtained? *J. Vacuum Sci. Technol.* 34:031504. doi: 10.1116/1.4944053
- Zhao, F., Li, H. X., Ji, L., Mo, Y. F., Quan, W. L., Du, W., et al. (2009). Superlow friction behavior of Si-doped hydrogenated amorphous carbon film in water environment. *Surf. Coat. Technol.* 203, 981–985. doi: 10.1016/j.surfcoat.2008.09.025

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yu, Chen, Zhang and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



How Thermal Fluctuations Affect Hard-Wall Repulsion and Thereby Hertzian Contact Mechanics

Yunong Zhou, Anle Wang and Martin H. Müser*

Lehrstuhl für Materialsimulation, Department of Materials Science and Engineering, Saarland University, Saarbrücken, Germany

OPEN ACCESS

Edited by:

Julien Scheibert,
CNRS/Ecole Centrale de Lyon, France

Reviewed by:

Valentin L. Popov,
Technische Universität Berlin,
Germany
Alexander Filippov,
Dnetsk Institute for Physics and
Engineering, Ukraine

*Correspondence:

Martin H. Müser
martin.mueser@mx.uni-saarland.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 08 November 2019

Accepted: 22 November 2019

Published: 06 December 2019

Citation:

Zhou Y, Wang A and Müser MH
(2019) How Thermal Fluctuations
Affect Hard-Wall Repulsion and
Thereby Hertzian Contact Mechanics.
Front. Mech. Eng. 5:67.
doi: 10.3389/fmech.2019.00067

Contact problems as they occur in tribology and colloid science are often solved with the assumption of hard-wall and hard-disk repulsion between locally smooth surfaces. This approximation is certainly meaningful at sufficiently coarse scales. However, at small scales, thermal fluctuations can become relevant. In this study, we address the question how they render non-overlap constraints into finite-range repulsion. To this end, we derive a closed-form analytical expression for the potential of mean force between a hard wall and a thermally fluctuating, linearly elastic counterface. Theoretical results are validated with numerical simulations based on the Green's function molecular dynamics technique, which is generalized to include thermal noise while allowing for hard-wall interactions. Applications consist of the validation of our method for flat surfaces and the generalization of the Hertzian contact to finite temperature. In both cases, similar force-distance relationships are produced with effective potentials as with fully thermostatted simulations. Analytical expressions are identified that allow the thermal corrections to the Hertzian load-displacement relation to be accurately estimated. While these corrections are not necessarily small, they turn out surprisingly insensitive to the applied load.

Keywords: contact mechanics, statistical mechanics and classical mechanics e.t.c., molecular dynamics simulation, boundary element method, modeling and simulation, Hertzian contact analysis

1. INTRODUCTION

One of several drawbacks when applying continuum theory to small-scale contact problems, as they occur, for example, in contact mechanics or in colloid science, is that continuum theories often ignore the effect of thermal fluctuations. This can lead to noticeable errors of continuum-theory based predictions for the dependence of displacement or indentation on load when two objects are pressed against each other (Luan and Robbins, 2005, 2006). Temperature can affect mechanical contacts and their interpretation in numerous other ways. For example, the presence of thermal noise generally impedes an unambiguous definition of contact area (Mo et al., 2009; Cheng et al., 2010; Mo and Szlufarska, 2010; Eder et al., 2011; Jacobs and Martini, 2017). In addition, large standard deviations of experimentally measured depinning forces of atomic-force microscope tips have been observed, which were accompanied by unexpectedly large reductions of the depinning force with increasing temperature (Pinon et al., 2016). It is possible that thermal surface fluctuations, which were not included in the modeling of temperature effects on tip depinning, are responsible for a significant reduction of effective surface energy and thereby for

a reduction of the depinning force. In fact, it has been shown that thermal fluctuations limit the adhesive strength of compliant solids (Tang et al., 2006). Finally, in the context of colloid science, it may well be that thermal corrections have a non-negligible effect on the surprisingly complex phase diagram of Hertzian spheres (Pàmies et al., 2009). It is therefore certainly desirable to model the effect of thermal fluctuations in a variety of contact and colloid problems.

While thermal fluctuations can be incorporated into simulations with so-called thermostats (Allen and Tildesley, 1987; Frenkel and Smit, 2002), proper sampling can require a significant computational overhead. In addition, some contact solvers do not appear amenable to thermostating. This concerns in particular those contact-mechanics approaches that optimize the stress field, as done with the classical solver by Polonsky and Keer (Polonsky and Keer, 1999; Müser et al., 2017), rather than the displacement fields in the Green's function molecular dynamics (GFMD) method (Campaña and Müser, 2006; Zhou et al., 2019).

The just-mentioned issues motivated us to investigate how thermal noise affects the mean force F (per unit area) between surfaces as a function of their interfacial separation, or, gap g . The pursued idea is to integrate out the internal degrees of freedom, whereby an areal free-energy density can be defined. The procedure is similar in spirit to the construction of interatomic potentials, for which the (quantum-mechanical ground-state) fluctuations of electrons are integrated out rather than the (thermal) fluctuations of internal elastic degrees of freedom.

In our first attempt on constructing effective surface interactions, we restrict our attention to the oldest, and arguably most commonly used model for the interactions between surfaces, namely a non-overlap constraint. Depending on context and dimension, it can also be called hard-wall, hard-disk, or hard-sphere repulsion, which, by definition is infinitesimally short ranged. Since atoms fluctuate about their equilibrium sites in solids, thermal fluctuations automatically make repulsion effectively adopt a finite range.

The central goal of this study is to quantify the just-described effects and to ascertain if constitutive laws obtained for flat walls can be applied to other systems, in particular to a Hertzian contact. A secondary goal is to identify an analytical expression for the thermal corrections to the load-displacement relation in a Hertzian contact.

2. MODEL AND NUMERICAL METHOD

2.1. Definition of the Model and Nomenclature

The model consists of a homogeneous, semi-infinite, elastic solid with an originally flat bottom surface, which is pressed down against a continuous, perfectly rigid substrate being fixed in space. The latter, which will also be called indenter, is either perfectly flat, i.e., $h(\mathbf{r}) = 0$, or parabola, in which case $h(\mathbf{r}) = -r^2/(2R_c)$, where R_c is the radius of curvature. In order to reduce finite-size effects and to simplify both analytical and numerical

treatments, periodic boundary conditions are assumed by default within the quadratic, interfacial plane.

The elastic surface is subjected not only to an external load per particle, l , squeezing it down against the indenter but also to thermal fluctuations, as they would occur in thermal equilibrium at a finite temperature T . We restrict our attention to frictionless contacts and small counterface slopes. This allows us to consider only displacements of the elastic surface normal to the interface. As such, the elastic energy of the surface can be written as a functional of the field $u(\mathbf{r})$ according to

$$U_{\text{ela}}[u(\mathbf{r})] = \frac{E^*A}{4} \sum_{\mathbf{q}} |\tilde{u}(\mathbf{q})|^2. \quad (1)$$

Here, $u(\mathbf{r})$ states the z -coordinate of the elastic solid's bottom surface as a function of the in-plane coordinate $\mathbf{r} = (x, y)$. E^* is the contact modulus, A the (projected) interfacial area, \mathbf{q} an in-plane wave vector, and q its magnitude.

$$\tilde{u}(\mathbf{q}) = \frac{1}{A} \int d^2r e^{-i\mathbf{q}\cdot\mathbf{r}} u(\mathbf{r}) \quad (2)$$

denotes the Fourier transform of $u(\mathbf{r})$. The short-hand notation $u_0 = \tilde{u}(\mathbf{q} = 0)$ will be used for the center-of-mass coordinate.

For flat indenters, only u_0 will be used to denote the mean distance, or gap, between indenter and the solid surface. Here, we define the displacement d as a function of temperature and load according to

$$d(T, L) \equiv h_{\text{ind}}(r = 0) - \langle u(T, L, r \rightarrow \infty) \rangle, \quad (3)$$

where $\langle u(T, L, r \rightarrow \infty) \rangle$ is the thermal expectation value that the field $u(\mathbf{r})$ would have (infinitely) far away from the top if the simulation cell were infinitely large. d is sometimes also called interference, as it states an effective penetration of the indenter into the elastic solid.

It is discussed in the literature (Müser, 2014) how to extrapolate accurately $u(L, r)$ to $r \rightarrow \infty$ for all those cases, in which an indenter acts relatively localized in the center of a finite simulation cell. However, in the current work, we are interested mostly in the temperature-induced reductions of d , i.e., in the term d_T defined in the expression

$$d = d_0 - d_T, \quad (4)$$

where d_0 denotes the displacement for an ideal, athermal Hertzian indenter at a given load. In the current work, we compute d_T through the following approximation

$$d_T \approx \langle u(T, L, \mathbf{r}_X) \rangle - u(0, L, \mathbf{r}_X), \quad (5)$$

where \mathbf{r}_X is the point that is the most distant from the center of the Hertzian indenter. We found that the first three to four digits are accurate in this estimate if the athermal Hertzian contact radius is less than one quarter of the simulation cell's linear dimension. This is because the (true) surface displacement fields converge quite quickly to their asymptotic $1/r$ form

outside the (original) contact radius in the case of short-ranged potentials and because the finite-size corrections to the true surface displacements are not very sensitive to temperature.

The interaction with a counterface is modeled within the Derjaguin approximation (Derjaguin, 1934) so that the surface energy density depends only on the local interfacial separation, or, gap, $g(\mathbf{r}) = u(\mathbf{r}) - h(\mathbf{r})$, between the surfaces, i.e., the interaction potential is obtained via an integration over the surface energy density $\gamma(g)$ via

$$U_{\text{int}} = \int_A d^2r \gamma\{g(\mathbf{r})\}, \quad (6)$$

In the full microscopic treatment, hard-wall repulsion is assumed, i.e.,

$$\gamma(g) = \begin{cases} \infty & \text{if } g < 0 \\ 0 & \text{else.} \end{cases} \quad (7)$$

Finally, the probability of a certain configuration to occur is taken to be proportional to the Boltzmann factor, i.e.,

$$\text{Pr}[u(\mathbf{r})] \propto e^{-\beta(U_{\text{ela}} + U_{\text{int}})}, \quad (8)$$

where $\beta = 1/k_B T$ is the inverse thermal energy.

One central “observable” in this work is the distance dependence of the mean force per atom, $f(u_0)$, for flat surfaces and finite temperatures. One might want to interpret this function as a cohesive-zone model, or, in the given context better as a repulsive-zone model. Because of the so-called equivalence of ensembles, which is valid for sufficiently large, systems, it does not matter if the separation is fixed and the force measured, or, vice versa.

Note that we will go back and forth between continuous and discrete descriptions of displacement fields. For the discrete description, the elastic solid is partitioned into atoms, which are arranged on a square lattice with the lattice constant Δa . This was done for reasons of simplicity, even if other discretizations are possible, e.g., into a triangular lattice (Campaña and Müser, 2006). Transitions between discrete and continuous representations in real space can be achieved with the substitutions

$$\sum_n \dots \leftrightarrow \frac{1}{\Delta a^2} \int_A d^2r \dots, \quad (9)$$

while transitions between summations and integrals in the wavevector domain can be achieved with

$$\sum_{\mathbf{q}} \dots \leftrightarrow \frac{A}{(2\pi)^2} \int d^2q \dots. \quad (10)$$

To simplify the analytical evaluation of integrals, the square Brillouin zone (BZ) of the surface will be approximated with a circular domain. In this case, the upper cutoff for q is chosen to be $q_{\text{max}} = \sqrt{4\pi}/\Delta a$ as to conserve the number of degrees of freedom with respect to the original BZ.

2.2. Thermal GFMD

GFMD is a method allowing a linearly elastic boundary-value problem to be solved efficiently (Campaña and Müser, 2006; Venugopalan et al., 2017; Zhou et al., 2019). The (discretized) surface displacement field reflects the dynamical degrees of freedom. Elastic interactions are described in terms of appropriate elastic Green's functions, which—in the case of in-plane spatial homogeneity and infinitely large (or periodically repeated) systems—are (block) diagonal in the Fourier representation. The simplest case, which is considered here, is a frictionless contact and a semi-infinite elastic substrate. The equations to be solved in GFMD—using the regular tricks of the trade—are

$$m_q \ddot{u}(\mathbf{q}) + \eta_q \dot{u}(\mathbf{q}) + \frac{qE^*}{2} \tilde{u}(\mathbf{q}) = \tilde{F}(\mathbf{q}, t), \quad (11)$$

where $\tilde{F}(\mathbf{q}, t)$ is the Fourier transform of all external forces acting on the surface atoms. The terms m_q and η_q represent inertia and damping coefficients of different surface modes, which may depend on the wave vector. For isotropic systems, these terms only depend on q but not on the direction of \mathbf{q} .

The effect of thermal fluctuations can be cast as random forces, which have to satisfy the fluctuation-dissipation theorem (FDT) (Kubo, 1957). In the given formalism, random forces must have a zero mean, while their second moments must satisfy,

$$\langle \Gamma(\mathbf{q}, t) \Gamma(\mathbf{q}', t') \rangle = 2 \eta_q k_B T \delta_{\mathbf{q}, \mathbf{q}'} \delta(t - t'), \quad (12)$$

assuming discrete atoms, finite domains but continuous times. Here, $\delta(\dots)$ is the Dirac delta function, which can be replaced with $\delta_{t,t'}/\Delta t$ in a molecular dynamics (MD) simulation, in which the time t is discretized into steps of size Δt .

At this point, GFMD is only used to generate the correct distribution of configurations, which—in a classical system—does not depend on the choice of inertia. As such, the m_q can be chosen at will as far as static observables are targeted. However, in order to reproduce realistic dynamics, appropriate choices for m_q (see also the discussion on quantum effects in section 3.3) and η_q have to be made. In fact, realistic dynamics require the treatment of damping and random noise to have “memory,” as discussed in Kajita et al. (2010). When being interested in fast equilibration, the m_q are better chosen such that the usually slowly equilibrating long-wavelength modes are made light so that characteristic times for different modes coincide as closely as possible (Zhou et al., 2019). In this context, it is also worth mentioning that significant progress has been made recently on GFMD to properly reflect not only true (rather than efficient) dynamics of crystalline solids (Kajita, 2016) but also for truly visco-elastic materials with broad relaxation functions (van Dokkum and Nicola, 2019).

2.3. Hard-Wall Interactions in Thermal GFMD

Non-overlap constraints can be implemented in athermal GFMD by placing any atom, predicted to have penetrated the rigid solid, back onto its surface. This procedure no longer works at finite

temperatures. It violates the FDT because the damping that is effectively imposed by this algorithm, is not compensated by a conjugate random force.

The standard way of treating hard-wall or hard-disk interactions is to make it make the time step so large that the next collision between two hard sphere occurs at the end of it. Before proceeding with the time stepping, an ideal, elastic collision is then assumed. This course of action does not appear to be viable for contact mechanics, because it would lead to prohibitively small time steps for large-scale contacts, where several (hundred) thousands of grid points are usually classified as being in contact. Specifically, when doubling the system size N , the typically allowed time step will have to be halved on average so that the asymptotic computational effort would scale with N^2 rather than with N or $N \ln N$.

2.3.1. Effective Hard-Wall Potentials

An alternative to the standard ways of implementing non-overlap constraints is to allow its violation in a controlled fashion. For example, the true hard-wall interaction can be replaced with a finite-range energy density penalty of the form

$$\gamma(g) = \frac{\kappa_o E^* \Delta a}{n} \left(\frac{-g}{\Delta a} \right)^n \Theta(-g) \quad (13)$$

where Θ is the Heavyside step function and κ_o and n are dimensionless parameters. In loose analogy to a Richardson extrapolation, an observable of interest O can be computed for a fixed exponent n but different values of κ_o . Finally, the results can be extrapolated to hard-wall interactions by investigating the asymptotics of $O(1/\kappa)$ in the limit of $1/\kappa \rightarrow 0$. Large values of κ_o will limit the time step Δt . However, these limits do not depend on system size. Thus, the numerical effort will scale with $O(1/N)$ rather than with $O(1/N^2)$ as is the case when dynamics are based on the more accurate, flexible time-step collision dynamics.

Good numbers for the exponent n and the dimensionless hard-wall stiffness κ_o need to be chosen. In order for the effective hard-wall potential to have a minimal effect on Δt , the (non-negative) exponent n should be as small as possible. However, we would like the force to be a continuous function, for reasons explained at length in any better text book on molecular dynamics (Allen and Tildesley, 1987; Frenkel and Smit, 2002). While these arguments can be somewhat academic when the discontinuities are small, we are going to send κ_o to large numbers resulting in significant force discontinuities. Thus, n must be chosen greater equal two. This appears to make $n = 2$ the optimal choice.

The next question to be answered is: Given a time step Δt and an exponent of $n = 2$, what is a good value for κ_o ? Here, it is useful to keep in mind that we do not need very accurate dynamics in the “forbidden” overlap zone. The main purpose of the stiff harmonic potential is to eliminate overlap as quickly as possible, i.e., to effectively realize a collision of the particles with the position of the (original) hard wall. However, the stiffness should remain (well) below a critical value above which energy conservation is violated in the absence of a thermostat even when a symplectic integrator, such as the Verlet algorithm, is used. For

Verlet, the critical time step for a harmonic oscillator is $\Delta t_c = T/\pi$, where T is the oscillator period, i.e., for $\Delta t < \Delta t_c$, the trajectory may be inaccurate, but the energy is conserved (except for round-off errors). This can be achieved by setting the overlap stiffness to

$$k_o = \nu_o \pi^2 \frac{m}{dt^2} - k_s, \quad (14)$$

where $k_s = \Delta u^2/(k_B T)$, while m is the inertia of the considered degree of freedom. ν_o is a numerical factor, which must be chosen less than unity. At and above the critical value of $\nu_o = 1$, energy conservation would be no longer obeyed in the absence of a thermostat. At the same time, dynamics but also static distribution functions are very inaccurate, even if a thermostat prevents the system from blowing up.

The optimum value for k_o certainly depends on the specific investigated problem. However, the analysis of simple models can provide useful preliminary estimates. This will be done in section 2.3.3.

2.3.2. Approximate Collision Rules

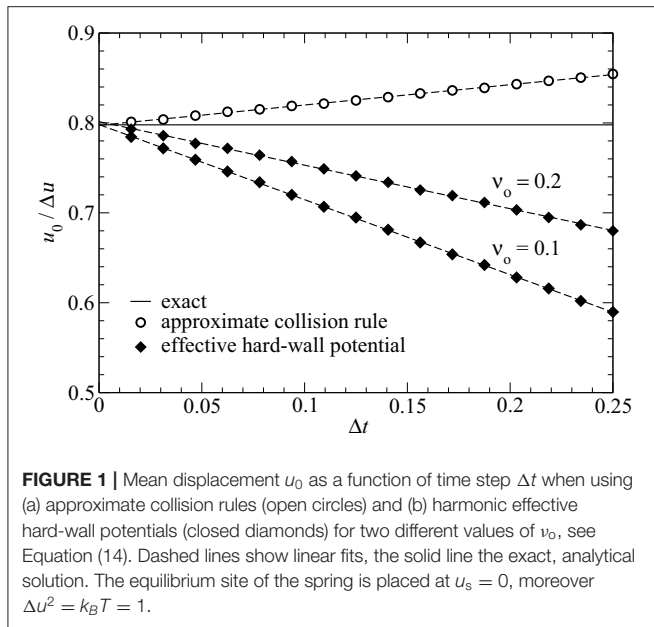
A second possibility to avoid the poor efficiency of exact collision dynamics is to use approximate collision rules and to control the error of the imprecision with the time step. A simple possibility would be to keep Δt fixed in a simulation and to make the deflection of the atom after the regular time stepping. For example, when using velocity Verlet, the following pseudo code could be invoked after a regular time step, in which the constraint was ignored:

```
if (z violates constraint) then
  z = 2zconstr - z
  vz = -vz (velocity Verlet)
  zold = 2zconstr - zold (standard Verlet)
end if
```

Note that this approach requires extra care to be taken when dynamics are formulated in a wavevector representation, which is usually the case in efficient boundary-element methods. If implemented the following overhead would have to be realized: old positions (or velocities) in real space will then have to be kept in memory. Moreover, two additional Fourier transforms will have to be invoked in each time step, which would double the number of the (asymptotically) most expensive function calls. Since approximate collision dynamics turn out to show similar scaling with Δt in simple models as effective hard-wall repulsion, see section 2.3.3, we did not pursue approximate collision rules further at this point of time in the full contact-mechanics simulations.

2.3.3. Numerical Case Studies

To explore the relative merit of the two proposed hard-wall methods, we investigate the following single-particle problem: an originally free, harmonic oscillator with a (thermal) variance of Δu^2 . This harmonic oscillator is then constrained to have no negative deflections from its mechanical equilibrium site. The analytical solution to this problem stating the force F needed



to realize a given constraint is contained in the mean-field approximation to the full elastic problem, which is presented in section 3.2.2. The given constraint of the spring sitting exactly on the hard wall corresponds to a value, where $\langle u_0 \rangle$ crosses over from its short-range to its long-range asymptotic behavior. Therefore, we see this case as being representative for both scaling regimes.

In essence, the problem we investigate corresponds to the choice where $k_B T$, k , and m are used to define the unit system, which makes Δu^2 being unity (in units of $k_B T/k$) as well. The default time step that we use for the free oscillator is $2\pi/30$, i.e., 30 time steps per period. The damping coefficient is chosen to be $\gamma = 1$, whereby the free harmonic oscillator is slightly underdamped. While this choice is not necessarily ideal, it still tends to be effective for a fast equilibration, irrespective of whether the temperature is zero or finite. Results for the convergence of how the estimate for the mean displacement u_0 approaches the exact value with decreasing time step Δt are shown in **Figure 1**.

At a given value of Δt , the approximate collision rules clearly outperform the approximate hard-wall interactions. However, u_0 has leading-order corrections of order Δt in both approaches. With the choice $v_0 = 0.1$, the asymptotic result for the parabolic, effective hard-wall potential has an accuracy of better than 1%, which should be accurate enough for most purposes. In both approaches, simulations must be run at two different values of Δt , say e.g., at $\Delta t = 0.25$ and $\Delta t = 0.15$ in order to perform a meaningful $\Delta t \rightarrow 0$ extrapolation. In a full contact-mechanics simulation, the number of required Fourier transforms doubles when using the approximate collision rules, which in turn leads to increased stochastic errors given a fixed computing time contingent. For this reason, but also because approximate collision rules require significantly more coding—in particular when averaging wall-surface forces from collisions when using wavevector dependent inertia—we decided to use the harmonic, effective hard-wall potential for the full contact-mechanics simulations.

3. THEORY

The main purpose of this section is to identify an analytical expression for the thermal expectation value of an interfacial force per atom $f(u_0)$ as a function of their mean separation u_0 in the case of a hard wall. This will be done by defining the partition function $Z(N, \beta, u_0)$ of a fluctuating surface in front of a wall, which is linked to the free energy through the relation $\mathcal{F}(k_B T, u_0) = -k_B T \ln Z(\beta, u_0)$. The mean force between hard wall and elastic surface can then be calculated from

$$f = -\frac{1}{N} \frac{\partial \mathcal{F}(N, k_B T, u_0)}{\partial u_0}. \quad (15)$$

Minor errors in the treatment presented below appear in numerical coefficients that result, for example, by having approximated the Brillouin zone of a square with a sphere, or, by having replaced a discrete set of wave vectors (finite system) with a continuous set (infinitely large system). However, these and related approximations are controlled, because errors resulting from them can be estimated and they could even be corrected systematically.

3.1. The Statistical Mechanics of a Free Surface

Since the free surface is the reference state, we start with its discussion. An important quantity, in particular in a mean-field approach, is the variance of atomic displacements due to thermal noise. For a fixed center-of-mass coordinate, it is defined as the following thermal expectation value:

$$\Delta u^2 \equiv \langle \{u(\mathbf{r}) - \tilde{u}(0)\}^2 \rangle. \quad (16)$$

It can be evaluated in its wavevector representation in a straightforward manner. Specifically,

$$\Delta u^2 = \sum_{\mathbf{q}'} \langle |\tilde{u}(\mathbf{q}')|^2 \rangle \quad (17)$$

$$\approx \frac{A}{(2\pi)^2} \int d^2 q \frac{2 k_B T}{q E^* A} \quad (18)$$

$$\approx \frac{2}{\sqrt{\pi}} \frac{k_B T}{E^* \Delta a}, \quad (19)$$

where we made use of equipartition for harmonic modes, see also Equation (29).

Of course, up to the prefactor of $2/\sqrt{\pi} \approx 1.1284$, which is very close to unity, Equation (19) follows directly from dimensional analysis. However, in a quantitative theory, we wish to know and perhaps to understand its precise value. A numerical summation over a square BZ assuming a square real-space domain with N atoms reveals that Δu^2 can be described by

$$\Delta u^2 = \left(1.1222 - \frac{1.24}{\sqrt{N}} \right) \frac{k_B T}{E^* \Delta a}, \quad (20)$$

with more than three digits accuracy if $\sqrt{N} > 512$. This result is fairly close to the analytical result based on a BZ, which is approximated as sphere.

Assuming discretization down to the atomic scale of $\Delta a \approx 2.5 \text{ \AA}$ yields a root-mean square (rms) height of

$$\Delta u \approx 1.5 \sqrt{\text{GPa}/E^*} \text{ \AA} \quad (21)$$

at room temperature. Thus, for soft-matter systems, the effect of thermal fluctuations is not necessarily non-negligible at room temperature. The dominant restoring forces to height fluctuation at short scales will then be due to surface tension rather than due to elasticity (Xu et al., 2014). However, it might be possible to suppress those effects when immersing the surfaces into an appropriate liquid, e.g., crosslinked polyethylene glycol (PEG) into uncrosslinked PEG.

An outcome of Equation (19) is that the fluctuations are dominated by the small scales. In the simplest approximation, which can be made in direct association with the Einstein model of solids, each surface atom is coupled harmonically to its lattice site with a spring of stiffness $k_E = k_B T / \{(N-1) \Delta u^2\}$. In reality, i.e., in less than infinite dimensions, there is always a correlation of thermal height fluctuations.

To deduce an estimate for the distance over which height fluctuations are correlated, we calculate the thermal displacement autocorrelation function (ACF) $C_{uu}(r)$. It can be defined and evaluated to obey:

$$C_{uu}(\Delta r) = \langle u(\mathbf{r}) u(\mathbf{r} + \Delta \mathbf{r}) \rangle \quad (22)$$

$$\approx \frac{1}{2\pi^2} \frac{k_B T}{q E^*} \int_0^{\sqrt{4\pi}/\Delta a} dq \int_0^{2\pi} d\varphi e^{iqr \cos \varphi} \quad (23)$$

$$= \frac{1}{\pi} \frac{k_B T}{r q E^*} \int_0^{\sqrt{4\pi} r / \Delta a} d(qr) J_0(qr) \quad (24)$$

$$= \frac{2 k_B T}{q E^*} \frac{\sqrt{4\pi} r}{\Delta a} {}_1F_2 \left(\frac{1}{2}; 1, \frac{3}{2}; -\frac{\pi r^2}{\Delta a^2} \right) \quad (25)$$

$$\approx \begin{cases} \frac{2 k_B T}{\sqrt{\pi} E^* \Delta a} + \mathcal{O}(r^2) & \text{for } r \rightarrow 0 \\ k_B T / (\pi q E^* r) & \text{for } r \rightarrow \infty, \end{cases} \quad (26)$$

where $J_0(x)$ is the Bessel function of the first kind and ${}_1F_2(\dots)$ is a generalized hypergeometric function. Unfortunately, the result obtained analytically this way shows Helmholtz ringing at intermediate values of r (i.e., within a substantial range of Δu), which is why the exact analytical solution for $C_{uu}(r)$ is of little practical use, except in the two limiting cases $r = 0$ and $r \rightarrow \infty$. Helmholtz ringing is generally a consequence of sharp cutoffs in the wave vector domain. Interestingly, it persists even for a square BZ when the exact expectation values for $|\tilde{u}(\mathbf{q})|^2$ are used and the correlation function $C_{uu}(\mathbf{r})$ is extended to the continuous domain between the lattice positions. The validity of these claims is demonstrated in **Figure 2**.

A quite reasonable approximation or rather generalization of $C_{uu}(r)$ to a continuous function can be made by constructing the simplest expression with the correct asymptotic behaviors summarized in Equation (26):

$$C_{uu}(r) \approx \frac{2}{\sqrt{\pi}} \frac{k_B T}{E^*} \frac{1}{(\Delta a^2 + 4\pi r^2)^{1/2}}. \quad (27)$$

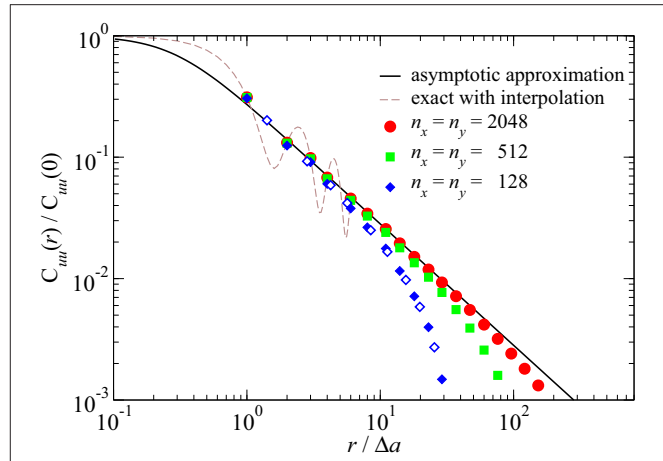


FIGURE 2 | The radial displacement ACF $C_{uu}(r)$ —normalized to its value at $r = 0$ —as a function of distance r : asymptotic approximation given in Equation (27) (black line), exact correlation function along the [10] direction with interpolation between non-lattice sites (dashed brown line), numerically exact results for systems of size $2,048 \times 2,048$ (red circles), 512×512 (green squares), and 128×128 (blue diamonds). They were also obtained for the [10] direction, except for the open symbols, which refer to the [11] direction.

As can be seen in **Figure 2**, this asymptotic approximation is quite reasonable already at a nearest-neighbor spacing of $r = \Delta a$ and has errors of less than 5% (in the limit of large N) for larger values of r . While numerical results for finite systems in **Figure 2** include predominantly data for \mathbf{r} parallel to $[1, 0]$, similar results are obtained for other directions as well, as demonstrated exemplarily for the $[1, 1]$ direction of the $N = 128 \times 128$ lattice.

The asymptotic ACF has decayed to approximately 30% of its maximum value at the nearest-neighbor distance. This means that the displacements of adjacent lattice sites are essentially uncorrelated.

The last property of interest used in the subsequent treatment is the partition function of a free surface (fs):

$$Z_{\text{fs}}(\beta) = \prod_{\mathbf{q}} \frac{\lambda_{\mathbf{q}}}{\sqrt{2\pi \Delta u^2(q)}} \quad (28)$$

with

$$\Delta u^2(q) = \frac{2 k_B T}{q E^* A}. \quad (29)$$

$\lambda_{\mathbf{q}} = h / \sqrt{2 m_{\mathbf{q}} k_B T}$ represents the thermal de Broglie wavelength of a surface mode. It reflects the ideal-gas contribution of the momenta conjugate to the $\tilde{u}(\mathbf{q})$ to the partition function. As long as E^* is small compared to the ambient pressure and as long as temperature is kept constant, the sole purpose of including $\lambda_{\mathbf{q}}$ into the calculation is to render the partition function dimensionless. This is why a precise determination of $m_{\mathbf{q}}$ is not needed at this point, even if it might be an interesting topic in itself and of relevance for a quantum-mechanical treatment, which is discussed in section 3.3.

In the mean-field (Einstein solid) approximation, the partition function simplifies to

$$Z_{\text{mf}}(\beta) = \left(\frac{\lambda_{\text{mf}}}{\sqrt{2\pi \Delta u^2}} \right)^N, \quad (30)$$

with Δu having been introduced in Equation (19) and λ_{mf} being a mean-field de Broglie wavelength.

3.2. Interaction of a Thermal, Elastic Surface With a Flat Wall

In this section, we investigate the statistical mechanics of an elastic surface in front of a flat, hard wall. To this end we derive expressions for the partition function of the system, from which the mean force between surface and wall (at fixed mean separation) can be derived in a straightforward fashion. Different mean-field strategies will be pursued toward this end. They turn out to be quite accurate in different asymptotic limits of the full problem.

3.2.1. First Mean-Field Approximation

The arguably simplest analytical approach to the contact problem is an adaptation of the so-called Einstein solid, which was already alluded to in section 3.1, to surface atoms. We first do it such that a degree of freedom is a hybrid of an atom in real space and a delocalized, ideal sine wave. Specifically, we first assume that elastic energy of an individual atom reads

$$v_{\text{ela}}^{\text{mf}}(u) = \frac{k_B T}{2 \Delta u^2} u^2. \quad (31)$$

In order to maintain a zero expectation value of u , it is furthermore assumed that the interaction energy with a counterface placed at a distance u_0 from the atom's mean position is given by

$$v_{\text{sub}}^{\text{mf}}(u) = \frac{\Delta a^2}{2\pi} \int_0^{2\pi} d\varphi \gamma(u_0 + u \cos \varphi). \quad (32)$$

This means, an oscillation of an atom entails an undulation. With this assumption, u_0 automatically corresponds to the atom's mean position.

The excess free energy per particle $\Delta \mathcal{F}/N$ for a fixed center-of-mass position satisfies

$$e^{-\beta \mathcal{F}/N} = \frac{1}{\sqrt{2\pi \Delta u^2}} \int_{-\infty}^{\infty} du e^{-\beta \{v_{\text{ela}}^{\text{mf}}(u) + v_{\text{sub}}^{\text{mf}}(u)\}}, \quad (33)$$

where the term “excess” refers to the change of the free energy relative to that of a free surface. For hard-wall interactions, the integral in Equation (33) can be evaluated to be

$$\begin{aligned} e^{-\beta \mathcal{F}/N} &= \frac{1}{\sqrt{2\pi \Delta u^2}} \int_{-u_0}^{u_0} du e^{-\beta v_{\text{ela}}(u)} \\ &= \text{erf} \left(\frac{u_0}{\sqrt{2\Delta u}} \right). \end{aligned} \quad (34)$$

Hence,

$$\frac{\mathcal{F}}{N k_B T} = -\ln \left\{ \text{erf} \left(\frac{u_0}{\sqrt{2\Delta u}} \right) \right\} \quad (35)$$

$$\approx \begin{cases} -\ln \left(\sqrt{\frac{2}{\pi}} \frac{u_0}{\Delta u} \right) & \text{for } u_0 < \Delta u/2 \\ \frac{\Delta u}{\sqrt{\pi} u_0} e^{-u_0^2/(2\Delta u^2)} & \text{for } u_0 > 2\Delta u. \end{cases} \quad (36)$$

For reasons of completeness, the force predicted from this first mean-field approximation is stated as:

$$f_{\text{mf}}(u_0) = \sqrt{\frac{2}{\pi}} \frac{k_B T}{\Delta u} \frac{\exp\{-u_0^2/(2\Delta u^2)\}}{\text{erf}\{u_0/(\sqrt{2\Delta u})\}}. \quad (37)$$

In the limit of $u_0 \rightarrow 0$, repulsion diverges proportionally with $1/u_0$, while it decays slightly quicker than exponentially in u_0^2 for separations $u_0 \gg \Delta u$. Both limiting behaviors are confirmed in the results section, albeit, with a prefactor of a little less than one half for large separations.

3.2.2. Second Mean-Field Approximation

Another mean-field approach would be to abandon the evaluation of the interaction in terms of an undulation and to introduce a Lagrange parameter, i.e., an external force f divided by the thermal energy, ensuring u to adopt the desired value of u_0 . Thus, the probability of a displacement u to occur satisfies

$$\text{Pr}(u) \propto e^{-(u-u_0)^2/(2\Delta u^2) - \beta f(u-u_0)} \Theta(u), \quad (38)$$

where f needs to be chosen such that $\langle u \rangle = u_0$ so that the lattice position of the particle u_{eq} is situated at $u_{\text{eq}} = u_0 + \beta f \Delta u^2$. At u_{eq} , there is no elastic restoring force in the spring. The requirement $\langle u \rangle = u_0$ automatically leads to the following self-consistent equation for f :

$$\beta f \Delta u = \sqrt{\frac{2}{\pi}} \frac{\exp \left\{ -\frac{(\beta f \Delta u^2 - u_0)^2}{2\Delta u^2} \right\}}{1 - \text{erf} \left(\frac{\beta f \Delta u^2 - u_0}{\sqrt{2\Delta u}} \right)}. \quad (39)$$

This line of attack leads to similar results for the $f(u_0)$ at small u_0 as the first mean-field approach. However, for large u_0 the predicted force turns out half that of the first mean-field approximation. In fact, the second mean-field theory turns out to be a quite reasonable approximation to the numerical data for any value of Δu , see the results and discussion presented in section 4.

3.2.3. Probabilistic Approach

The exact expression for the excess free energy of an elastic body in front of a hard wall can be defined by a path integral,

$$e^{-\beta \mathcal{F}(u'_0)} = \frac{1}{Z_A} \int \mathcal{D}[u(\mathbf{r})] \delta(u'_0 - u_0) e^{-\beta v_{\text{tot}}[u(\mathbf{r})]}, \quad (40)$$

where $\mathcal{D}[u(\mathbf{r})]$ denotes an integral over all possible displacement realizations and

$$Z_A = \int \mathcal{D}[u(\mathbf{r})] \delta(u_0 - u'_0) e^{-\beta v_{\text{ela}}[u(\mathbf{r})]}. \quad (41)$$

In the case of hard-wall repulsion, the r.h.s. of Equation (40) is easy to interpret: It represents the relative number of configurations that are produced with the thermal equilibrium distribution of a free surface (fs), whose maximum displacement is less than u_0 , i.e.,

$$e^{-\beta \mathcal{F}(u_0)} = \langle \Pr(u_{\max} < u_0) \rangle_{\text{fs}}, \quad (42)$$

This insight defers the problem of having to solve the path integral in Equation (40) to an exercise in probability theory: determine the likelihood of $N' = N\Delta a^2/\Delta A_c$ independent Gaussian random number with mean zero and variance Δu^2 to be less than u_0 . Here ΔA_c is the correlation area for the displacements. Given that $C_{uu}(\Delta r)$ has decayed to a few 10% at nearest-neighbor distances, it can only be marginally larger than Δa^2 .

For large values of N' , the distribution of maximum values $u_{\max} = \max\{u(\mathbf{r})\}$ converges to the Gumbel distribution, also known as the generalized extreme value (gev) distribution type-I (David and Nagaraja, 2003). It is given by

$$\Pr(u_{\max}) = \frac{1}{\beta_{\text{gev}}} e^{-(e^{-z})} \quad (43)$$

with

$$z = \frac{u_{\max} - \mu_{\text{gev}}}{\beta_{\text{gev}}}, \quad (44)$$

where μ_{gev} is the mode of the Gumbel distribution, i.e., the most likely value for u_{\max} to occur, and β_{gev} a parameter determining the shape of the distribution. For a normal Gaussian distribution $\Phi_G(u/\Delta u)$, they are given by

$$\frac{\mu_{\text{gev}}}{\Delta u} = \sqrt{2} \operatorname{erf}^{-1} \left(1 - \frac{2}{N'} \right) \quad (45)$$

$$\frac{\beta_{\text{gev}}}{\Delta u} = \frac{1}{N' \cdot \Phi_G(\mu_{\text{gev}}/\Delta u)} \quad (46)$$

in the limit of large N' . Here $\operatorname{erf}^{-1}(\dots)$ stands for the inverse function of the error function (David and Nagaraja, 2003).

In fact, **Figure 3** shows that the distribution of u_{\max} as produced with GFMD and by taking the maximum value of $N' = 0.92N$ independent random numbers are essentially identical and that both can be approximated quite well with the Gumbel distribution. If setting $N' = N$, the (open) symbols in **Figure 3** would shift by roughly half their symbol size to the right. As expected, discrepancies between the Gumbel distribution and the numerical data decrease with increasing N' .

Rather than relying on the Gumbel distribution, one might as well write down the exact probability of one positive Gaussian random variable (grv) to be less than u_0 and take the result into the $N'/2$ -th power. (On average, there are $N'/2$ positive grv's, whose value may not exceed u_0 . The negative grv's are irrelevant with respect to the violation of the violation of the non-overlap constraint.) In this approximation,

$$\Pr(u_{\max} < u_0) = \left\{ \operatorname{erf} \left(\frac{u_0}{\sqrt{2}\sigma} \right) \right\}^{N'/2}. \quad (47)$$

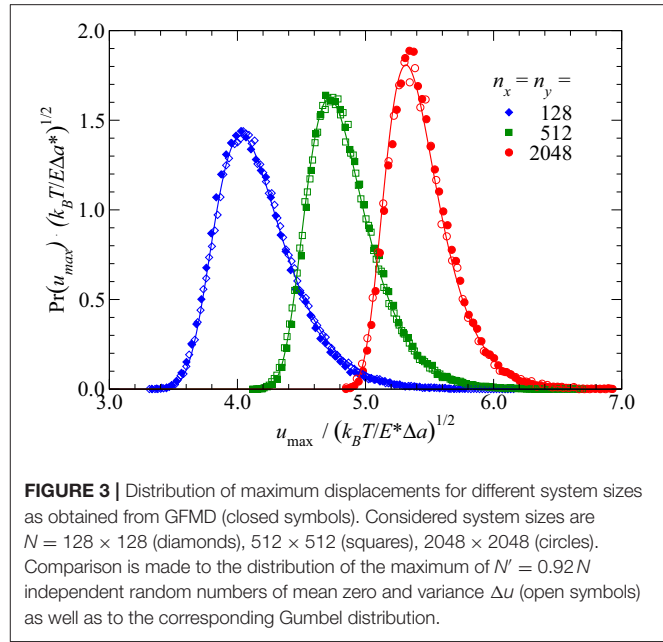


FIGURE 3 | Distribution of maximum displacements for different system sizes as obtained from GFMD (closed symbols). Considered system sizes are $N = 128 \times 128$ (diamonds), 512×512 (squares), 2048×2048 (circles). Comparison is made to the distribution of the maximum of $N' = 0.92N$ independent random numbers of mean zero and variance Δu (open symbols) as well as to the corresponding Gumbel distribution.

and therefore

$$\Delta \mathcal{F} = -\frac{N' k_B T}{2} \ln \left\{ \operatorname{erf} \left(\frac{u_0}{\sqrt{2}\Delta u} \right) \right\}. \quad (48)$$

This result turns out to apply to large separations, that is, to $u_0/\Delta u \gg 1$. The functional form of $\mathcal{F}(u)$ is identical to the one obtained in the first mean-field variant, except for the prefactor, which is reduced by a little more than a factor of two.

3.3. Handling Quantum Effects

Throughout this paper, it is assumed that thermal vibrations are classical. In reality, atoms are quantum mechanical, which enhances their fluctuations about their equilibrium sites. Differences between classical and quantum systems can matter when the Debye temperature is clearly larger than the ambient temperature. In this section, we briefly sketch how the quantum-mechanical fluctuations could be modeled rigorously but also suggest an alternative approach. The latter is easily implemented and should be reasonably accurate except for very large squeezing forces.

A rigorous treatment could be based on path-integral techniques (Berne and Thirumalai, 1986; Müser, 2002), in which quantum-mechanical (QM) point particles are represented in terms of classical ring polymers. The course of action would be an acquisition of the proper Green's function with similar fluctuation formulae as in the original GFMD paper (Campaña and Müser, 2006), while simulating (half) solids and acquiring elastic tensor or stiffness elements as done, for example, in Schöffel and Müser (2001). For a harmonic system, the stiffness of the various modes would be identical in the classical and the quantum system so that the most important variable to be determined would be the q -dependent inertia m_q of the surface modes. It would have to be selected such that it yields the correct zero-point vibration in a path-integral augmented

GFMD simulation. The latter would benefit from replacing the free-particle propagator in so-called imaginary time with one that is symmetry-adopted for hard walls as done in Müser and Berne (1997).

In classical systems, Δu is dominated by the shortest wavelength modes. This will be even more so for quantum systems, because modes show greater quantum effects at short than at long wavelengths. In other words, the model of an Einstein solid should provide a reasonable approximation for the true, quantum-mechanical variance Δu_{QM}^2 of a free elastomer. In this approximation, the effective stiffness of a spring coupling the z -coordinate of a surface atom to its lattice site is $E^* \Delta a$, divided by a factor very close to 1.12, which we consider negligible in the present discussion. If m is the mass of a surface atom, the associated eigenfrequency would be $\omega_0 = \sqrt{E^* \Delta a / m}$. Thus, the temperature-dependent internal energy of an Einstein mode is obtained as $U = \hbar \omega_0 \coth\{\hbar \omega_0 / (2k_B T)\} / 2$, where \hbar is the reduced Planck constant, see any textbook on statistical mechanics. Since $U = 2\langle V_{\text{pot}} \rangle$ for the quantum or classical harmonic oscillator, it can be deduced that

$$\Delta u_{\text{QM}}^2 \approx \frac{\Delta u^2}{2 k_B T \tanh\{\hbar \omega_0 / (2k_B T)\}}. \quad (49)$$

If quantum effects need to be included, the value of Δu_{QM}^2 would have to replace that of Δu^2 in any application of the method. The treatment would not be exact, because the wavefunction and thus the quantum-mechanical probability density go to zero at the hard-wall constraint. This would lead to enhanced repulsion, in particular at large compression, for which we expect repulsion to diverge more quickly than with $1/u_0$ because of the uncertainty principle. Yet, using Δu_{QM}^2 instead of the classical Δu^2 would be roughly analogous to a first-order perturbation theory and thereby represent quantum effects accurately for separations $u_0 \gtrsim \Delta u_{\text{QM}}$.

3.4. Thermal Hertzian Contacts

3.4.1. Preliminary Considerations

At small temperatures, the relative leading-order corrections to the zero-temperature displacement $u_0(T=0)$ can be expected to depend on powers of the variables defining the problem, i.e.,

$$\frac{d_T}{d_0} \propto \left(\frac{R_c}{\Delta a}\right)^\alpha \left(\frac{E^* R_c^2}{L}\right)^\beta \left(\frac{k_B T}{E^* R_c^3}\right)^\gamma, \quad (50)$$

where the contact modulus E^* and the contact radius R_c were effectively used to define the units of pressure and length, respectively. With the help of a further dimensional analysis, which can be conducted in a similar fashion as that in Müser (2014), the sum rule

$$\alpha + 3\beta - 5\gamma = 0 \quad (51)$$

follows immediately for the exponents introduced on the r.h.s. of Equation (50). This relation is valid for a quadratic tip shape, linear elasticity, assuming the interfacial stress is a function of $u(r)/\Delta u$ with $\Delta u \propto \sqrt{T}$.

The r.h.s. of Equation (50) and the sum rule for exponents in Equation (51) can also be valid at high-temperatures. However, different exponents will apply. At intermediate temperatures, an expansion over terms such as those discussed so far are the only possibility to conform to the dimensional analysis.

3.4.2. Low-Temperature Approximation

At very small temperatures, the stress profile can be expected to differ only marginally from that of the athermal contact. In a perturbative approach to the problem, one could therefore assume that the most important correction to the original Hertzian gap $g_H(r)$ is a constant shift by d_T . The latter can be determined by minimizing the thermal excess energy per atom

$$e_T = -d_T L + \frac{1}{\Delta a^2} \int d^2 r \mathcal{F}_{\text{pa}} \{g_H(r) + d_T\} \quad (52)$$

$$\approx -d_T L + \frac{2\pi}{\Delta a^2} \int_0^{a_c} dr r \mathcal{F}_{\text{pa}}(d_T), \quad (53)$$

where $\mathcal{F}_{\text{pa}} \equiv \mathcal{F}/N$ denotes the hard-wall, free-energy normalized to the atom. The approximation in Equation (53) is motivated by the expectation that the dominant contribution to e_T resides within the original contact area. Minimization of e_T w.r.t. d_T leads to

$$L = \frac{\pi a_c^2}{\Delta a^2} f(d_T) \quad (54)$$

$$\approx \frac{\pi a_c^2}{\Delta a^2} \sqrt{\frac{2}{\pi}} \frac{k_B T}{\Delta u} \frac{\exp(-u_0^2/2\Delta u^2)}{\text{erf}(u_0/\sqrt{2}\Delta u)} \quad (55)$$

where the last approximation is only valid at small temperatures. Taylor expanding this last expression leads to

$$\frac{d_T}{d_0} \approx \frac{T}{T^*} \quad (56)$$

with

$$T^* = \frac{L \Delta a^2}{\pi k_B R_c}. \quad (57)$$

3.4.3. High-Temperature Approximation

At very large temperatures, d_T is in excess of d_0 so that deformations of the elastic solids are very small. In a first-order perturbative approach, it then makes sense to assume the displacement field to be a constant, i.e., to be d_T . In that approximation, individual forces can be simply summed up with a mean gap of $d_T + r_n^2/(2R_c)$. Recasting the sum as an integral yields

$$L \approx \frac{N'}{2N} \frac{1}{\Delta a^2} \int d^2 r f_{\text{mfl}} \left(d_T + \frac{r^2}{2R_c} \right) \quad (58)$$

$$\approx L_0 \frac{\Delta u}{d_T} e^{-d_T^2/(2\Delta u^2)} \quad (59)$$

with

$$L_0 = \sqrt{\frac{2}{\pi}} \frac{N'}{N} \frac{k_B T R_c}{\Delta a^2}. \quad (60)$$

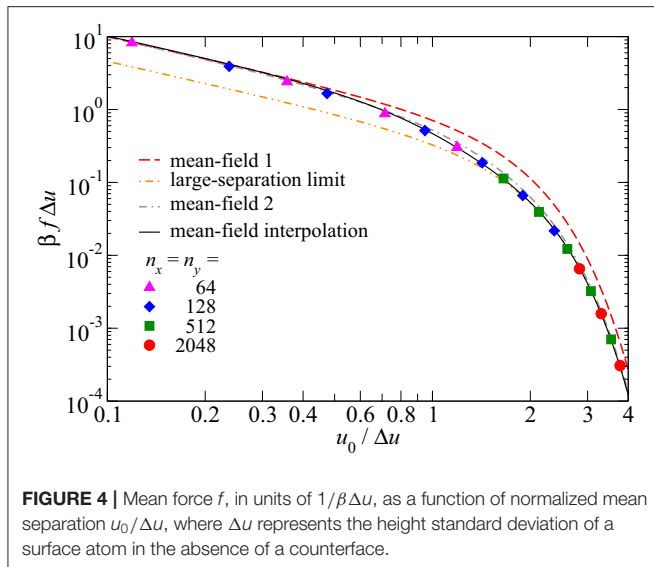


FIGURE 4 | Mean force f , in units of $1/\beta\Delta u$, as a function of normalized mean separation $u_0/\Delta u$, where Δu represents the height standard deviation of a surface atom in the absence of a counterface.

Equation (59) can be solved for d_T with the help of the Lambert W function $W(x) \approx \ln x - \ln \ln x$ for $x \gg 1$:

$$\frac{d_T}{\Delta u} \approx \sqrt{W\left(\frac{L_0^2}{L^2}\right)}. \quad (61)$$

4. RESULTS

4.1. Potential of Mean Force for a Flat Hard Wall

In this section, we investigate to what extent the three approaches introduced in section 3.1 reproduce accurate, numerical results for the thermal repulsive-zone model. To this end, we chose units such that $E^* = 1$ and $\Delta a = 1$ and consider different values of $u_0/\Delta u$, which is the only dimensionless variable for the given problem besides the system size, which is varied as well.

Figure 4 compares GFMD data to the various approximative approaches introduced in section 3. The first mean-field approach appears to be asymptotically exact for small u_0 , while the approach based on the law of large numbers seems to be asymptotically exact for large u_0 . In between these two regimes, there is a smooth transition between them. This transition is reflected quite well by the second mean-field approach. Unfortunately, we did not identify a closed-form analytical expression for it, which would nevertheless be nice to have when implementing a potential of mean force into a simulation. However, as is demonstrated in **Figure 4**, simple switching functions introduced next allow one to approximate numerical data reasonably well.

Since both force-distance asymptotic dependencies have the same functional form and since the transition between them is quite continuous, it is relatively easy to come up with switching functions allowing the numerically determined free energy to be approximated reasonably well. Defining \mathcal{F}_{mfl} through the

free-energy expression in Equation (35), this is done via

$$\mathcal{F}(u_0) \approx w_1(u_0)\mathcal{F}_{\text{mfl}}(u_0) + w_2(u_0)\Delta\mathcal{F} \quad (62)$$

with the weighting functions

$$w_1(u_0) = \frac{1}{2} \left\{ \frac{N'}{N} + \left(2 - \frac{N'}{N} \right) e^{-u_0^2/\Delta u^2} \right\} \quad (63)$$

$$w_2(u_0) = e^{-u_0^2/\Delta u^2} \{ 1 - \tanh(u_0/\Delta u) \} \quad (64)$$

The numerical value for $\Delta\mathcal{F}$ turned out to be $\Delta\mathcal{F} = -N'k_B T/2$. The forces $f(u)$ in a coarse-grained description are obtained as negative derivative by differentiating the r.h.s. of Equation (62). The resulting expression corresponds to the numerical GFMD data for systems with $n_x = n_y \geq 128$ with maximum errors less than 10%, at least when taking the exact value for Δu^2 .

In terms of an efficient implementation of the method, we recommend to use tabulated expressions for $f(u)$ for intermediate values of u and the asymptotic expressions for $u \ll \Delta u$ and $u \gg \Delta u$.

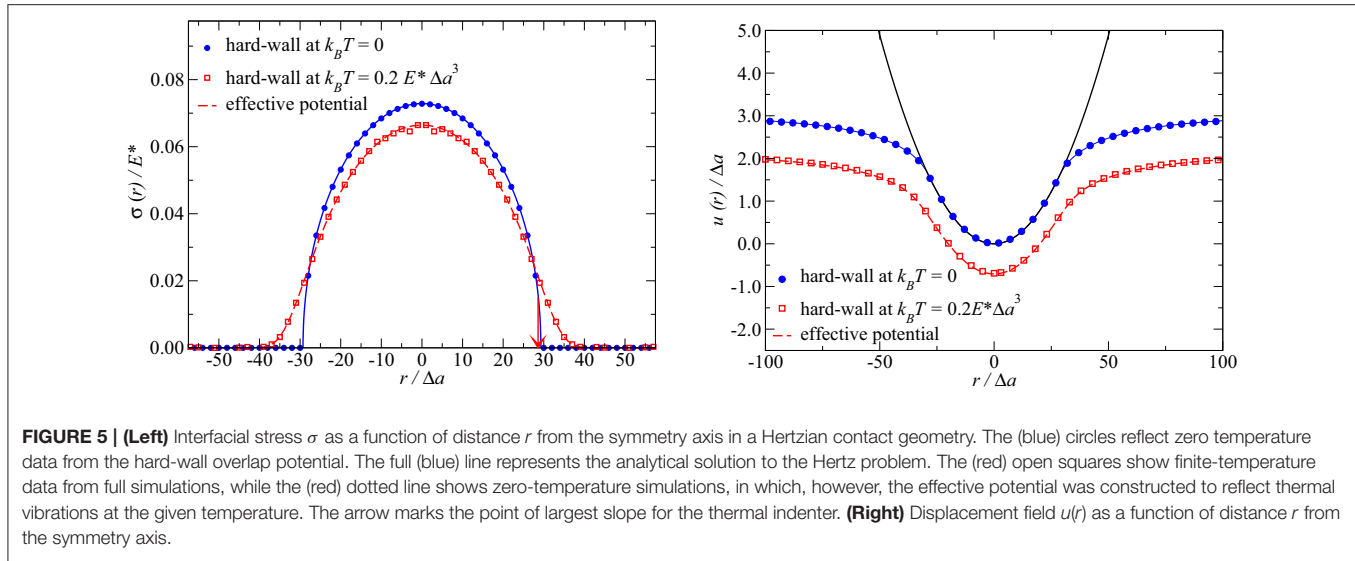
4.2. Hertzian Indenter

We now consider a Hertzian indenter as transferability test for our effective potential. In addition, the effects that thermal fluctuations have on the load displacement relation are explored along with an analysis of how to meaningfully define a contact area in the presence of thermal fluctuations.

The solution of the continuous displacement field has no dimensionless number if the contact radius a_c is taken to be the unit of length. However, $a_c/\Delta a$ starts to matter as soon as it is no longer large compared to unity. Since discreteness effects are a different topic discussed elsewhere (Müser, 2019), $a_c/\Delta a$ is chosen sufficiently large so that the discrete problem reflects the continuous Hertz contact reasonably well.

To test the applicability of the thermal repulsive-zone model in the realm of Hertzian contact mechanics, the following parameters were chosen as useful defaults after some trial and error: $R_c = 256 \Delta a$ and a normal load of $L = 131 E^* \Delta a^2$ leading to $a_c \approx 30 \Delta a$ within regular Hertzian contact mechanics. In the athermal Hertzian contact, the mean contact pressure turns for these parameters is $p \approx 0.049 E^*$. Results for the stress profile at a temperature of $k_B T = 0.2 E^* \Delta a^3$ are shown in **Figure 5**.

An interesting but perhaps also obvious outcome of the data presented in **Figure 5** is that there is no abrupt transition from finite to zero contact stress, once thermal fluctuations are finite. This observation is of relevance when discussing the concept of “true contact area.” Since collisions in a hard-wall potential are instantaneous, the probability of observing two (finite) surfaces to be in contact has a statistical measure of zero, so that the instantaneous contact area could be argued to be (almost) always zero. Contact exists only in the isolated moments of time at which collisions take place. However, during these isolated moments of time, the forces between surfaces is infinitely large such that time averaging yields a distribution which resembles the well-known Hertzian stress profile; the smaller the temperature the closer the stress profiles between original and finite-temperature stress profiles.



The question of how to meaningfully define (repulsive) contact area when repulsion has a finite range and adhesion is neglected arises naturally. In a recent paper (Müser, 2019), it was proposed to define the contact line (or edge) to be located, where the gradient of the normal stress has a maximum slope. In the current example, this leads to a reduction of the contact radius of order 1%, which is significantly less than the reduction of approximately 30% of the normal displacement in the given case study.

In contrast to contact radii, force and displacement can be defined unambiguously. Thermal noise will reduce the interference d by d_T due to the effectively finite range of the repulsion, as discussed in the definition of the model in section 2.1. Since the description for an athermal Hertzian contact is scale free—in the sense that the functional form for stress and displacement are independent of any parameter defining a Hertzian contact—the function $f(T) \equiv d_T/d_0$ must have a universal shape if $\Delta a \ll a_c$. This is because the thermal repulsive zone model for hard-wall repulsion is a scale-free function of the gap divided by Δu . **Figure 6** reveals that results on the thermal displacement for different Hertzian contact realizations can indeed be collapsed quite closely onto a single master curve $\Xi(T/\tilde{T})$ defined through

$$d_T = \tilde{d}_0 \Xi(T/\tilde{T}), \quad (65)$$

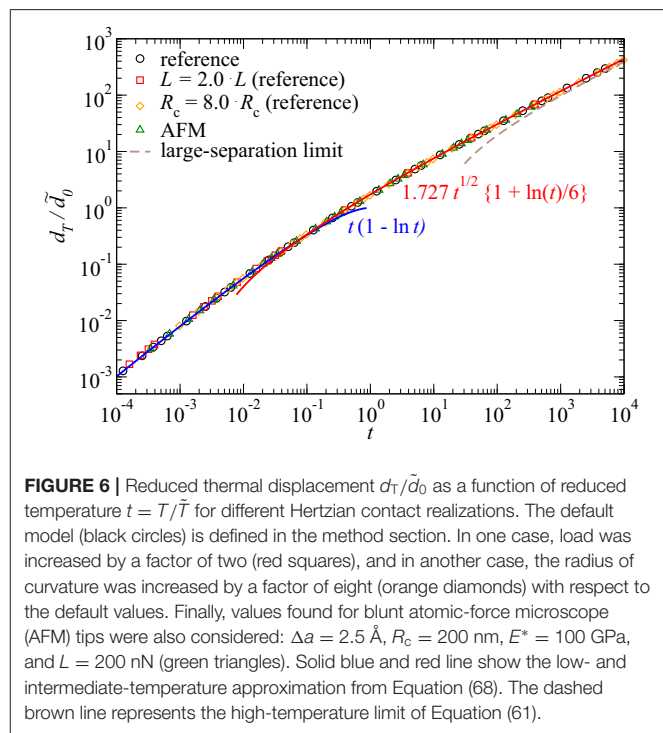
where

$$\tilde{d}_0 = \left(\frac{R_c L}{E^* \Delta a^3} \right)^{-\frac{1}{3}} d_0 \quad (66)$$

and

$$\tilde{T} = \left(\frac{L}{E^* R_c^2} \right)^{\frac{2}{3}} \frac{E^* \Delta a^3}{k_B}. \quad (67)$$

The master curve shown in **Figure 6** reveals asymptotic regimes at low and at high temperatures, respectively. They can be



approximated with power laws. However corrections logarithmic in temperature need to be made at low temperature to obtain quantitative agreement over broad temperature ranges. We find numerically that

$$\Xi(t) \approx \begin{cases} t(1 - \ln t) & \text{for } t \ll 1 \\ 1.727 \sqrt{t} \{1 + \ln(t)/6\} & \text{for } 0.1 < t < 10^4 \end{cases} \quad (68)$$

Inserting the low-temperature approximation of the master curve into Equation (65) and reshuffling terms yields

$$\frac{d_T}{d_0} \approx \frac{T}{T^*} \left(1 - \ln \frac{T}{T^*} \right) \quad (69)$$

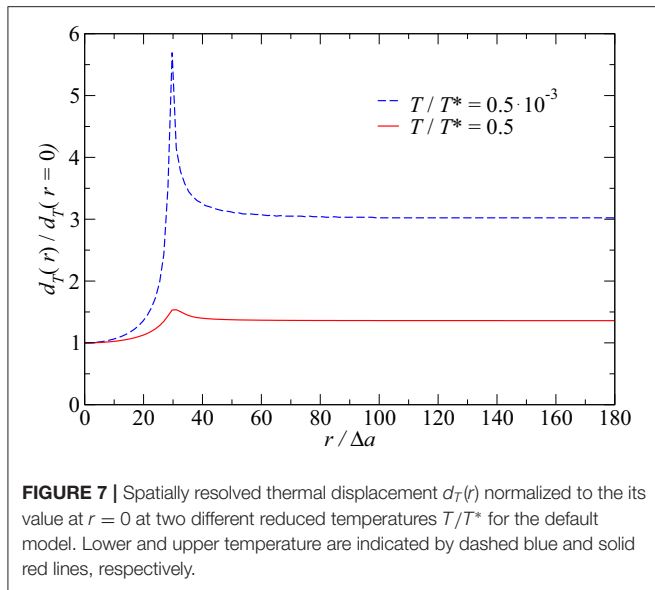


FIGURE 7 | Spatially resolved thermal displacement $d_T(r)$ normalized to its value at $r = 0$ at two different reduced temperatures T/T^* for the default model. Lower and upper temperature are indicated by dashed blue and solid red lines, respectively.

for $T \ll \tilde{T}$. This means that the low-temperature treatment presented in section 3.4.2 obtained the correct linear term, but failed to predict the logarithmic corrections, which become very large at small ratios T/\tilde{T} . Before discussing the origin of those corrections, we wish to emphasize that there are indeed two characteristic temperatures for the Hertzian contact, namely T^* and \tilde{T} .

The suspicion that significantly better results at small T/\tilde{T} are obtained when extending the integration domain in Equation (53) back to radii beyond the athermal contact radius turns out incorrect. The main reason for the deviations lies in the assumption of a constant thermal shift of the thermal displacement. **Figure 7** reveals that the thermal shift far away from the indenter is noticeably larger than at $r = 0$ and that discrepancies grow (logarithmically) with decreasing temperature. Since the simple treatment allows one to rationalize why d_T is (roughly) linear in temperature, we decided to keep the discussion of the low-temperature limit.

Before investigating the magnitude of thermal displacements in real units and not just in reduced units, we briefly comment on the intermediate-temperature behavior. Most importantly, we wish to emphasize that the approximation made in Equation (68) for $t > 0.1$ is only valid on the shown domain and that it does *not* extend to $t \rightarrow \infty$.

However, from a practical point of view, it appears virtually impossible to design a real-laboratory experiment, in which the asymptotic high-temperature regime of $t > 10^3$ could ever be reached. The only possible exception coming to our minds would involve the use of hagfish slime. It has extraordinarily small elastic moduli of order 0.02 Pa (Ewoldt et al., 2011), though the values of Δa to be used in a continuum model would be clearly in excess of the atomic scale, because hagfish slime stops being homogeneous well above the atomic scale. Since the contact mechanics of hagfish slime and related systems is somewhat of a niche application, we would argue that the analytical solution given in Equation (61) is merely a nice mathematical result and

that the $t > 0.1$ approximation made in Equation (68) can be considered the high-temperature limit for all other purposes.

One may wonder how the master curve shown in **Figure 6** translates into a $d(T)$ dependence when real units rather than reduced units are used. To answer that question, the expansions obtained previously are represented again for a moderately hard-matter ($E^* = 1$ GPa) and a soft-matter ($E^* = 50$ MPa) system, see **Figure 8** and further validated by additional GFMD simulations. In both cases, a radius of curvature of $R_c = 50$ nm was assumed and the load was chosen such that the ratio of maximum Hertz pressure to E^* was in the order of 0.1%, i.e., a load where plastic deformation can be assumed to be minor.

Figure 8 reveals that both studied systems qualify as being clearly in the intermediate-temperature regime at room temperature. Relative corrections of the normal displacement for the stiffer system are rather minor but non-negligible for the soft-matter system. This observation brings us to the next and final question, which is addressed in **Figure 9**, namely to what extent do thermal correction affect the load-displacement relation? After all, most indentation experiments are done at constant temperature and varying load rather than at constant load and varying temperature. Combining Equations (65–67) with the intermediate-temperature expansion of Equation (68) and the analytical solution for the displacement-load relation in a Hertzian contact, leads to the following relation:

$$d_T = d_T^{\text{ref}} \left\{ 1 - \frac{1}{9} \ln \left(L/L^{\text{ref}} \right) \right\} \quad (70)$$

with $d_T^{\text{ref}} \approx 1.426 \Delta u$ and

$$L^{\text{ref}} = \left(\frac{k_B T}{E \Delta a^3} \right)^{3/2} E^* R_c^2. \quad (71)$$

In other words, the elastomer surface is effectively shifted by a little less than 1.5 times the thermal standard deviation of its smallest-scale surface fluctuations. The effects of load are minuscule as they enter only logarithmically in the ninth' root of the load.

Figure 9 confirms that the thermal fluctuation in most real Hertzian contacts should lead to corrections that appear as almost constant shifts to the eye, even for soft-matter systems, for which the absolute shifts can be relatively large. In the case study presented in **Figure 9**, the thermal shift reads $d_T \approx 1.2$ at a load of $L \approx 16$ nN and barely more at a much reduced load $d_T \approx 1.7$ at a load as small as $L \approx 0.16$ nN. In order for the d_T correction to acquire twice the value compared to that at 16 nN, the compressive force in our example would have to be as small as $L \approx 20$ fN, which is scarcely measurable. For the reasons of completeness, we state that the range of validity of the intermediate-temperature approximation of $0.1 < t < 10^4$ demonstrated in **Figure 6** translates to a range of loads of $0.15 < L/\text{nN} < 1.5 \cdot 10^4$ for the specific examples studied here. Upper and lower limits are well beyond loads that could be meaningfully applied or measured experimentally for the system of question while measuring the normal displacement with high resolution.

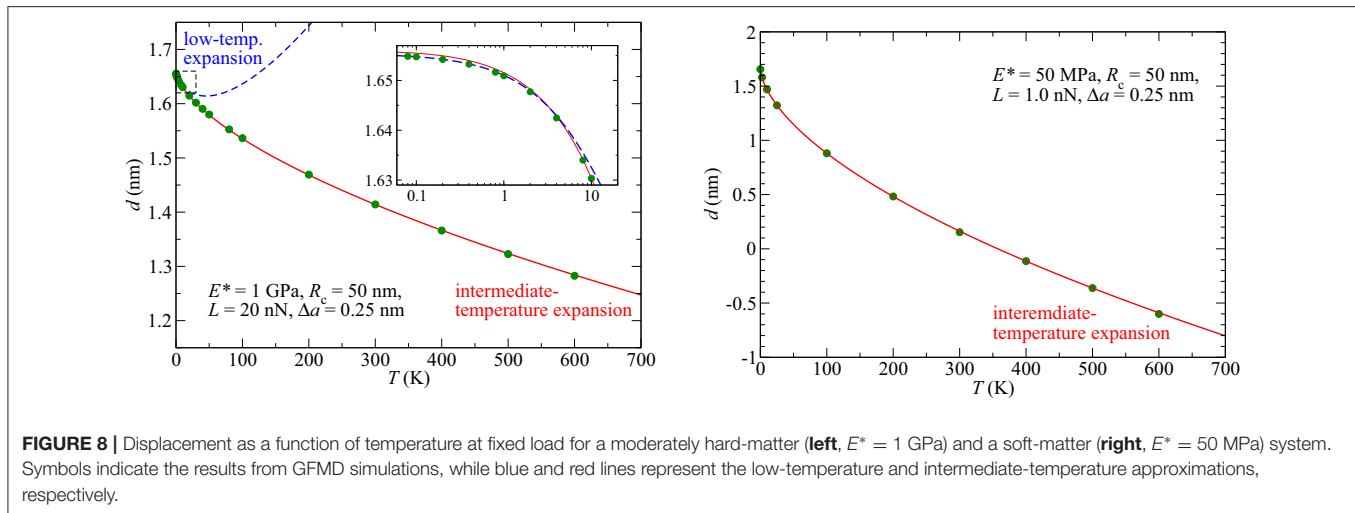


FIGURE 8 | Displacement as a function of temperature at fixed load for a moderately hard-matter (left, $E^* = 1$ GPa) and a soft-matter (right, $E^* = 50$ MPa) system. Symbols indicate the results from GFMD simulations, while blue and red lines represent the low-temperature and intermediate-temperature approximations, respectively.

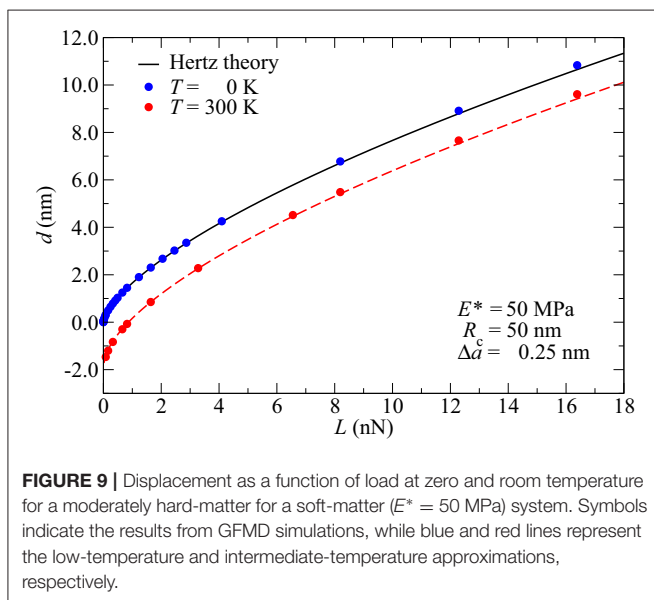


FIGURE 9 | Displacement as a function of load at zero and room temperature for a moderately hard-matter for a soft-matter ($E^* = 50$ MPa) system. Symbols indicate the results from GFMD simulations, while blue and red lines represent the low-temperature and intermediate-temperature approximations, respectively.

5. SUMMARY

In this work, we analyzed the effect that thermal fluctuations can have on contact mechanics in the case of hard-wall interactions. To this end, we first demonstrated that thermal surface fluctuations are dominated by short wavelengths undulations. They smear out the originally infinitesimally short-range repulsion to a finite range of $\Delta u \approx \sqrt{k_B T / (E^* \Delta a)}$. The functional form of the repulsive force was derived analytically and shown to diverge inversely proportionally with the interfacial separation u_0 at small u_0 but to decay slightly more quickly than exponentially in $-u_0^2 / (2\Delta u^2)$ at separations clearly exceeding Δu .

To come to these results, the Green's function molecular dynamics (GFMD) technique was generalized to include thermal noise. Particular emphasis was placed on the question how to handle (the original) hard-wall interactions in the simulations.

We found that replacing the hard-wall overlap constraint with a stiff harmonic potential produces satisfactory results if simulations are done at different values for the stiffness and extrapolation is made to infinite stiffness. The GFMD results are described very well with different mean-field approximations to the problem, which made it possible to identify a highly-accurate, closed-form analytical expression for the distance-force relation a flat, thermal elastomer interacting with a flat, rigid substrate.

It may be important to note that each microscopic interaction law requires the coarse-graining to be done for that particular interaction. For example, if thermal fluctuations were to be treated in a Dugdale model (Dugdale, 1960) (e.g., hard-wall constraint plus a constant adhesive stress acting up to a finite range), results for the hard-wall constraint cannot be simply incorporated, but a new parametrization of thermal effects has to be done.

Application of our methodology to Hertzian contacts revealed that thermal fluctuations can induce non-negligible shifts in the normal displacement. As a zero-order approximation, it can be assumed that the thermally induced shift of a Hertzian indenter is a little less than 1.5 times the thermal standard deviation of surface positions of a free, unconstrained surface. Corrections turn out to depend only logarithmically on the ninth root of the normal load. Thus, thermal noise leads to a shift of the load-displacement curve that is roughly equal to the root-mean-square fluctuation of surface atoms but almost independent of the load. As a referee of this manuscript noticed: *This picture is simple and easily understandable intuitively a posteriori, but by no means trivial and understandable in advance.* The result of an essentially load-independent displacement shift may, in part, explain why Hertzian contact models often apply all the way down to the nanoscale: Essentially constant shifts remain unnoticed.

We expect similar results for randomly rough and other hard-wall indenters as for Hertzian contacts, because the thermal shift for the Hertz contact is relatively insensitive to the radius of curvature. However, the effect of thermal fluctuations

will be more important in the case of short-range adhesion. Given the results from this study, quite noticeable adhesion reductions may be expected when its range is in the order of or less than the thermal displacement Δu . Future studies are ongoing elucidating the reduction of adhesion due to thermal vibrations.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

REFERENCES

- Allen, M., and Tildesley, D. (1987). *Computer Simulation of Liquids*. Oxford: Clarendon.
- Berne, B. J., and Thirumalai, D. (1986). On the simulation of quantum systems: path integral methods. *Annu. Rev. Phys. Chem.* 37, 401–424. doi: 10.1146/annurev.pc.37.100186.002153
- Campaná, C., and Müser, M. H. (2006). Practical Green's function approach to the simulation of elastic semi-infinite solids. *Phys. Rev. B* 74:075420. doi: 10.1103/PhysRevB.74.075420
- Cheng, S., Luan, B., and Robbins, M. O. (2010). Contact and friction of nanoasperities: effects of adsorbed monolayers. *Phys. Rev. E* 81:016102. doi: 10.1103/PhysRevE.81.016102
- David, H. A., and Nagaraja, H. N. (2003). *Order Statistics*. 2003. New York, NY: Wiley.
- Derjaguin, B. (1934). Untersuchungen über die Reibung und Adhäsion, IV. *Kolloid-Zeitschrift* 69, 155–164. doi: 10.1007/BF01433225
- Dugdale, D. (1960). Yielding of steel sheets containing slits. *J. Mech. Phys. Solids* 8, 100–104. doi: 10.1016/0022-5096(60)90013-2
- Eder, S., Vernes, A., Vorlaufer, G., and Betz, G. (2011). Molecular dynamics simulations of mixed lubrication with smooth particle post-processing. *J. Phys.* 23:175004. doi: 10.1088/0953-8984/23/17/175004
- Ewoldt, R. H., Winegard, T. M., and Fudge, D. S. (2011). Non-linear viscoelasticity of hagfish slime. *Int. J. Non-Linear Mech.* 46, 627–636. doi: 10.1016/j.ijnonlinmec.2010.10.003
- Frenkel, D., and Smit, B. (2002). *Understanding Molecular Simulation: From Algorithms to Applications, volume 1 of Computational Science Series, 2nd Edn*. San Diego, CA: Academic Press.
- Jacobs, T. D. B., and Martini, A. (2017). Measuring and understanding contact area at the nanoscale: a review. *Appl. Mech. Rev.* 69:060802. doi: 10.1115/1.4038130
- Kajita, S. (2016). Green's function nonequilibrium molecular dynamics method for solid surfaces and interfaces. *Phys. Rev. E* 94:033301. doi: 10.1103/PhysRevE.94.033301
- Kajita, S., Washizu, H., and Ohmori, T. (2010). Approach of semi-infinite dynamic lattice Green's function and energy dissipation due to phonons in solid friction between commensurate surfaces. *Phys. Rev. B* 82:115424. doi: 10.1103/PhysRevB.82.115424
- Kubo, R. (1957). Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.* 12, 570–586. doi: 10.1143/JPSJ.12.570
- Luan, B., and Robbins, M. O. (2005). The breakdown of continuum models for mechanical contacts. *Nature* 435, 929–932. doi: 10.1038/nature03700
- Luan, B., and Robbins, M. O. (2006). Contact of single asperities with varying adhesion: comparing continuum mechanics to atomistic simulations. *Phys. Rev. E* 74:026111. doi: 10.1103/PhysRevE.74.026111
- Mo, Y., and Szlafarska, I. (2010). Roughness picture of friction in dry nanoscale contacts. *Phys. Rev. B* 81:035405. doi: 10.1103/PhysRevB.81.035405
- Mo, Y., Turner, K. T., and Szlafarska, I. (2009). Friction laws at the nanoscale. *Nature* 457, 1116–1119. doi: 10.1038/nature07748
- Müser, M. (2002). On new efficient algorithms for PIMC and PIMD. *Comput. Phys. Commun.* 147, 83–86. doi: 10.1016/S0010-4655(02)00221-7
- Müser, M. H. (2014). Single-asperity contact mechanics with positive and negative work of adhesion: influence of finite-range interactions and a continuum description for the squeeze-out of wetting fluids. *Beilstein J. Nanotechnol.* 5, 419–437. doi: 10.3762/bjnano.5.50
- Müser, M. H. (2019). Elasticity does not necessarily break down in nanoscale contacts. *Tribol. Lett.* 67:57. doi: 10.1007/s11249-019-1170-y
- Müser, M. H., and Berne, B. J. (1997). Circumventing the pathological behavior of path-integral monte carlo for systems with coulomb potentials. *J. Chem. Phys.* 107, 571–575. doi: 10.1063/1.474442
- Müser, M. H., Dapp, W. B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T. A., et al. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65:118.
- Pàmies, J. C., Cacciuto, A., and Frenkel, D. (2009). Phase diagram of Hertzian spheres. *J. Chem. Phys.* 131:044514. doi: 10.1063/1.3186742
- Pinon, A. V., Wierez-Kien, M., Craciun, A. D., Beyer, N., Gallani, J. L., and Rastei, M. V. (2016). Thermal effects on van der waals adhesive forces. *Phys. Rev. B* 93:035424. doi: 10.1103/PhysRevB.93.035424
- Polonsky, I., and Keer, L. (1999). A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. *Wear* 231, 206–219. doi: 10.1016/S0043-1648(99)00113-1
- Schöffel, P., and Müser, M. H. (2001). Elastic constants of quantum solids by path integral simulations. *Phys. Rev. B* 63:224108. doi: 10.1103/PhysRevB.63.224108
- Tang, T., Jagota, A., Chaudhury, M. K., and Hui, C.-Y. (2006). Thermal fluctuations limit the adhesive strength of compliant solids. *J. Adhes.* 82, 671–696. doi: 10.1080/00218460600775781
- van Dokkum, J. S., and Nicola, L. (2019). Green's function molecular dynamics including viscoelasticity. *Model. Simulat. Mater. Sci. Eng.* 27:075006. doi: 10.1088/1361-651X/ab3031
- Venugopalan, S. P., Müser, M. H., and Nicola, L. (2017). Green's function molecular dynamics meets discrete dislocation plasticity. *Model. Simulat. Mater. Sci. Eng.* 25:065018. doi: 10.1088/1361-651X/aa7e0e
- Xu, X., Jagota, A., and Hui, C.-Y. (2014). Effects of surface tension on the adhesive contact of a rigid sphere to a compliant substrate. *Soft Matt.* 10, 4625–4632. doi: 10.1039/C4SM00216D
- Zhou, Y., Moseler, M., and Müser, M. H. (2019). Solution of boundary-element problems using the fast-inertial-relaxation-engine method. *Phys. Rev. B* 99:144103. doi: 10.1103/PhysRevB.99.144103

AUTHOR CONTRIBUTIONS

MM, YZ, and AW developed the code and analyzed the results. YZ run the simulations. MM performed the analytical calculations. MM and YZ wrote the paper.

FUNDING

Salaries of AW and YZ were paid in parts by the DFG through grant MU 1694/5-2 and in parts by Saarland University while the research was conducted. A University grant will cover the APF.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhou, Wang and Müser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Contacts With Negative Work of “Adhesion” and Superlubricity

Valentin L. Popov*

Department of System Dynamics and Friction Physics, Technische Universität Berlin, Berlin, Germany

Keywords: negative work of adhesion, electrohaptics, superlubricity, van der Waals forces, Hertzian contact

INTRODUCTION

Van der Waals forces between solids in vacuum are always attractive and are considered as the main source of adhesion. However, in the presence of an intermediate medium, they can also be repelling (Dzyaloshinskii et al., 1961) which means that the “work of adhesion” becomes negative. Similarly to the case of adhesion, the interaction range of these forces can be either comparable (or larger) than the minimum characteristic length scale of the contact problem or it can be negligible compared with all characteristic length scales. We call this latter case the “JKR-approximation,” as the JKR theory of adhesion (Johnson et al., 1971) is also valid in this limit. The repelling interaction can also be due to the presence (and squeezing out) of a thin fluid layer between solids as considered in Müser (2014). In the papers Popov and Hess (2018) and Heß and Popov (2019), it was shown that the contact of two oppositely charged surfaces at a constant voltage is equivalent to the adhesive contact with an effective van der Waals interaction. Similarly, the contact of the bodies with the same charge would be equivalent to repelling van der Waals forces with a negative work of adhesion. Further kinds of repelling forces may be solvation, structural, and hydration forces (Israelachvili, 2011). In the following, we speak about van der Waals forces, but they are thought as representative for a larger class of long range repelling forces.

We argue that in the JKR approximation, the Hertz’ solution of the contact problem with a repelling van der Waals interaction, remains practically unchanged. However, the contact area falls apart into the area of “weak (van der Waals) interaction” and “strong (rigid wall) interaction.” It is speculated that if the normal force is smaller than a critical value at which the core region of strong interaction disappears, a macroscopic superlubricity state of the contact may be observed.

ATTRACTIVE AND REPELLING VAN DER WAALS FORCES

The interaction force between neutral molecules is often modeled as a superposition of the very sharp increasing “core repulsion” $\propto 1/r^{13}$ and a weaker van der Waals “tail” $\propto r^{-7}$ (the corresponding potential is known as Lennard-Jones-Potential), **Figure 1A**, left:

$$F(r) = \frac{12 \cdot w}{r_0} \left[\left(\frac{r_0}{r} \right)^{13} - \left(\frac{r_0}{r} \right)^7 \right] \text{ for attractive van der Waals forces,} \quad (1)$$

where r_0 is the equilibrium distance, at which the core force and the van der Waals force become equal and w is the work of adhesion (the work needed to separate the molecules starting with their equilibrium position).

For two bodies with a plane surface at distance z , Equation (1) is modified to an equation of the interaction stress:

$$\sigma(z) = \frac{8 \cdot \Delta\gamma}{3z_0} \left[\left(\frac{z_0}{z} \right)^9 - \left(\frac{z_0}{z} \right)^3 \right] \text{ for attractive van der Waals forces} \quad (2)$$

OPEN ACCESS

Edited by:

Yu Tian,
Tsinghua University, China

Reviewed by:

Liran Ma,
Tsinghua University, China
Noshir Sheriar Pesika,
Tulane University, United States

*Correspondence:

Valentin L. Popov
v.popov@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 05 December 2019

Accepted: 24 December 2019

Published: 22 January 2020

Citation:

Popov VL (2020) Contacts With
Negative Work of “Adhesion” and
Superlubricity. *Front. Mech. Eng.* 5:73.
doi: 10.3389/fmech.2019.00073

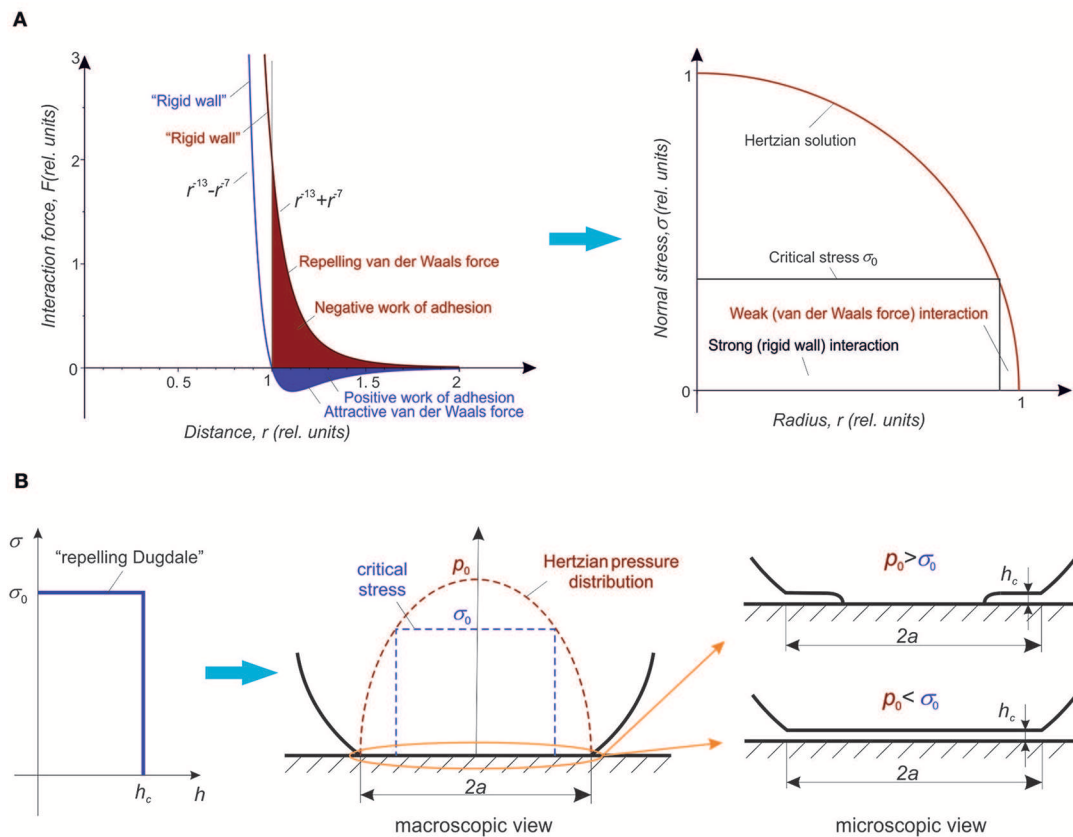


FIGURE 1 | (A) Left: Model intermolecular force in the case of attractive van der Waals force (blue line) and repelling van der Waals force (red line), in relative units. The distance $r = 1$ corresponds to the equilibrium in the case of attractive van der Waals forces. At distances smaller than $r = 1$ the repelling force increases steeply: At these distances, one can qualitatively speak about a “rigid wall.” At distances $r > 1$, there exist weak long-range van der Waals force, either positive or negative. The work needed to separate the surfaces starting from the “rigid wall” position $r = 1$ are shown by filled areas (blue filled area—positive work of adhesion in an adhesive contact, red filled area—“negative work of adhesion” in the repelling case). **(A)** Right: The normal stress (pressure) distribution in a Hertzian contact. If the range of action of both core “rigid wall” force and van der Waals force are negligible compared with all characteristic length scales of the contact problem, the repelling van der Waals forces do not influence the contact problem and do not change the stress distribution. However, on the microscopic scale, in the regions where the stress is larger than a critical stress σ_0 (see the text of paper for details), the surfaces are in “direct rigid wall contact” and feel strong atomic corrugation potential. In the regions where the stress is smaller than σ_0 , they “levitate” due to van der Waals force and see only weak corrugation. By decreasing the normal force, one can achieve the state in which the normal stress is smaller than σ_0 in the whole contact area. This state corresponds to the state of superlubricity. **(B)** Schematic illustration of the contact configuration on the example of a simple model adhesive stress of “Dugdale type” (Dugdale, 1960) with a constant repelling stress σ_0 up to the distance h_c (scheme on the left). In the center: Macroscopic shape of a soft sphere in contact with a rigid surface (black line) and the Hertzian stress distribution (brown dashed line). On the right: Microscopic view of the contact gap in the undercritical and overcritical cases. From the macroscopic point of view, the gap in both cases is zero, from the microscopic point of view, the bodies can be either in direct rigid wall contact (where the local Hertzian pressure is larger σ_0 or “levitate” in the distance h_c where the Hertzian pressure is smaller σ_0).

where z_0 is the equilibrium distance between the bodies (of the order of r_0), and $\Delta\gamma$ is the (positive) specific work of adhesion (work of separation of two surfaces per unit area).

While in vacuum the van der Waals forces are always attractive, in the presence of an intermediate medium between two bodies, they can also become repelling—if the dielectric constant of the intermediate medium lies between the dielectric constants of the contacting bodies (Dzyaloshinskii et al., 1961). In this case, Equation (1) for the interaction force is modified by changing the sign of the van der Waals force, **Figure 1A**, left:

$$F(r) = \frac{12 \cdot w}{r_0} \left[\left(\frac{r_0}{r} \right)^{13} + \left(\frac{r_0}{r} \right)^7 \right] \quad \text{for repelling van der Waals forces.} \quad (3)$$

In this equation, r_0 loses its meaning of the equilibrium position (which without external force does not exist anymore), but can still be considered as a distance characterizing the transition from the “core potential” to the “van der Waals potential.” At smaller distances, the repelling force increase very steeply and can be qualitatively considered as a “rigid wall,” while at larger distances it describes long range repelling van der Waals force. For bodies with plane surfaces, the corresponding modification of Equation (2) reads

$$\sigma(z) = \frac{8 \cdot \Delta\gamma}{3z_0} \left[\left(\frac{z_0}{z} \right)^9 + \left(\frac{z_0}{z} \right)^3 \right] \quad \text{for repelling van der Waals forces.} \quad (4)$$

Therein, $\Delta\gamma$ is not the work of adhesion anymore but has to be considered just formally as a coefficient determining the amplitude of the interaction. However, if we calculate the “work needed to separate” the bodies starting with the distance z_0 (the presumable position of the “rigid wall”), we get the specific “negative work of adhesion”

$$\Delta\gamma_{\text{repelling}} = -(5/3) \Delta\gamma. \quad (5)$$

In the past, there were only a few attempts to study contact mechanics in the presence of explicit surface interaction potential (“soft walls”) (e.g., Hughes and White, 1979; Vinogradova and Feuillebois, 2003; Müser, 2014). However, these works were focused on the normal interaction while we would like to discuss the implications of the surface interactions to the tangential force (friction).

INFLUENCE OF REPELLING VAN DER WAALS FORCES ON CONTACT AND FRICTION

As Martin Müser writes in Müser (2014), “For repulsive contacts,..., there is obviously no finite contact radius at zero normal load ... The repelled rigid tip simply “hovers” at (infinitely) large distance over an undeformed elastic manifold ...”. Let us consider this absolutely correct statement more closely. It is correct that the repelling van der Waals forces will keep the surfaces at “infinite large distance” which physically means at “very large distance.” However, from the macroscopic point of view, this “very large distance” may be smaller than any other characteristic length of the contact problem and thus can be considered as being zero (JKR-approximation). In the contact with repelling forces, the smallest characteristic length is the indentation depth, so the range of van der Waals interactions is assumed to be smaller than the indentation depth.

Whether the “very large distance” is zero or infinite—depends on the quantities, which we are interested in. If we consider the contact problem itself and the range of repelling van der Waals forces can be neglected, then they have no influence on the contact problem at all. In particular, all displacements and stress distributions will remain the same as in the classical “rigid wall” Hertz contact problem. However, if we consider the tangential forces (caused by the microscopic tangential corrugation potential), the “very large distance” can again be considered as infinite which leads to a vanishing force of friction. In order for the macroscopic frictional force to disappear, it is not even necessary that the tangential corrugation potential becomes zero; it is enough that it becomes smaller than some critical value, as has been shown theoretically and experimentally in Socoliuc et al. (2004).

When two surfaces approach each other, the interaction stress (4) increases monotonously and accepts at the distance $z = z_0$ the value $\sigma_0 = (16/5) (\Delta\gamma_{\text{repelling}}/z_0)$. If the local elastic stress in the material is larger than this critical stress, then, roughly speaking, the bodies are in “rigid wall contact.” If the local stress is much smaller than this critical stress, then the surfaces “hover” at “infinitely large distance.” In a Hertzian contact, all parts of the

contact where the stress is larger than the critical one, will be in “direct rigid wall contact,” and feel strong corrugation potential, while the areas where the stress is essentially smaller than σ_0 will be held apart by the repelling van der Waals forces and feel only a very weak corrugation potential. This is illustrated in **Figure 1A**, right for the example of a Hertzian contact. The contact area is divided into two parts: the inner part of “rigid wall contact” and high friction and the outer part of levitation due to repelling van der Waals forces and a weak (or zero) friction force. **Figure 1B** schematically illustrates the macroscopic and microscopic views of the contact for the case of a simplified repelling stress of “Dugdale type” (Dugdale, 1960).

The most interesting conclusion is that if the Hertzian stress is smaller than the critical stress needed to bring the surfaces into the “rigid wall contact,” then they “levitate” in the whole contact area. This inevitably should lead to a small or vanishing macroscopic frictional force—the macroscopic state of superlubricity. The critical stress depends on particular mechanism of repulsion force. A very rough estimation can be made by assuming the value of $\Delta\gamma \sim 4 \cdot 10^{-2} \text{ J/m}^2$ for the surface energy [which is “typical” for polymers and fluids (Popov, 2017)] and $\Delta l \sim 10^{-9} \text{ m}$ as the “levitation” distance needed for suppressing the tangential corrugation potential. The critical stress will then be on the order of $\sigma_c \sim \Delta\gamma/\Delta l \approx 40 \text{ MPa}$. In Ge et al. (2019a), the values up to 600 MPa were reported.

CONCLUSION

Let us briefly summarize the main points of the above-sketches picture:

1. If one does not look at the contact so closely, the “adhesive” repulsion looks just like a hard wall, similarly to the actual hard wall of the Pauli principle.
2. Therefore, in this “JKR-limit” nothing changes in the solution of the contact problem.
3. Nevertheless, if one looks more closely, the adhesive repulsion (as opposed to the hard wall) has some reach, so this long-range tail of the repelling force can hold the surfaces apart.
4. There is (almost) no friction in the areas that are levitating due to “negative adhesion.”
5. If the maximum stress in the whole contact area is smaller than the “critical stress of levitation,” the system transits into the state of very low (or vanishing) friction—state of superlubricity.

The key prerequisite of the described mechanism of superlubricity is the presence of repelling long range interaction forces which are able to hold the surfaces apart so that they do not feel the corrugation potential. This repulsion can be achieved in different ways:

- As repelling van der Waals force due to an intermediate medium with dielectric constant lying between the dielectric constancies of both contacting bodies. This mechanics can only be active for a contact of two bodies having different dielectric constants.

- As thermodynamic repulsion due to a layer of free or grafted macromolecules between the bodies so that the entropy of the intermediate layer decreases at small distances due to stronger confinement.
- As electrical repulsion due to external electrical voltage (so that both surfaces receive the charge of the same sign).
- As repulsion due to electrical double layer (Guldbrand et al., 1984).
- Possible is also an effective repulsion due to thermal fluctuations (Müser et al., 2019).

The necessity of an intermediate layer for achieving negative work of adhesion leads to the conclusion that the kinetic friction will also be essentially dependent on the rheology or viscosity of this intermediate medium. The importance of the local pressure brings the problematics of the surface roughness in play. The flattening of roughness and lowering local stresses may be one of the reasons of the necessity of the wearing-in process for achieving the superlubricity state (Ge et al., 2019a).

In my opinion, the above simple picture can help a lot for both qualitative and quantitative physical understanding of the

transition into the superlubricity state—both in the case of fluid superlubricity (Ge et al., 2019b) and solid state superlubricity (Erdemir and Eryilmaz, 2014) as well of the tuning of friction by electric fields (Krim, 2019; **Figure 1**).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This work was funded by the German Research Foundation (DFG PO 810/55-1).

ACKNOWLEDGMENTS

I am grateful to Emanuel Willert and Qiang Li for discussions. I acknowledge support from the German Research Foundation and the Open Access Publication Funds of TU Berlin.

REFERENCES

- Dugdale, D. S. (1960). Yielding of steel sheets containing slits. *J. Mech. Phys. Solids* 8, 100–104. doi: 10.1016/0022-5096(60)90013-2
- Dzyaloshinskii, I. E., Lifshitz, E. M., and Pitaevskii, L. P. (1961). General theory of van der Waals' forces. *Sov. Phys. Usp.* 4, 153–176. doi: 10.1070/PU1961v004n02ABEH003330
- Erdemir, A., and Eryilmaz, O. (2014). Achieving superlubricity in DLC films by controlling bulk, surf ace, and tribochemistry. *Friction* 2, 140–155. doi: 10.1007/S40544-014-0055-1
- Ge, X., Li, J., and Luo, J. (2019b). Macroscale superlubricity achieved with various liquid molecules: a review. *Front. Mech. Eng.* 5:2. doi: 10.3389/fmech.2019.00002
- Ge, X., Li, J., Wang, H., Zhang, C., Liu, Y., and Luo, J. (2019a). Macroscale superlubricity under extreme pressure enabled by the combination of graphene-oxide nanosheets with ionic liquid. *Carbon* 151, 76–83. doi: 10.1016/j.carbon.2019.05.070
- Guldbrand, L., Jönsson, B., Håkan Wennerström, H., and Linse, P. (1984). Electrical double layer forces. A Monte Carlo study. *J. Chem. Phys.* 80:2221. doi: 10.1063/1.446912
- Heß, M., and Popov, V. L. (2019). Voltage-induced friction with application to electrovibration. *Lubricants* 7:102. doi: 10.3390/lubricants7120102
- Hughes, B. D., and White, L. R. (1979). 'Soft' contact problems in linear elasticity. *Q. J. Mech. Appl. Math.* 32, 445–471. doi: 10.1093/qjmath/32.4.445
- Israelachvili, J. N. (2011). *Intermolecular and Surface Forces*, 3rd Edn. Waltham, MA; San Diego, CA; Oxford; Amsterdam: Elsevier. doi: 10.1016/C2009-0-21560-1
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. R. Soc. Lond. A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Krim, J. (2019). Controlling friction with external electric or magnetic fields: 25 examples. *Front. Mech. Eng.* 5:22. doi: 10.3389/fmech.2019.00022
- Müser, M. H. (2014). Single-asperity contact mechanics with positive and negative work of adhesion: influence of finite-range interactions and a continuum description for the squeeze-out of wetting fluids. *Beilstein J. Nanotechnol.* 5, 419–437. doi: 10.3762/bjnano.5.50
- Müser, M. H., Zhou, Y., and Wang, A. (2019). How thermal fluctuations affect hard-wall repulsion and thereby Hertzian contact mechanics. *Front. Mech. Eng.* 5:67. doi: 10.3389/fmech.2019.00067
- Popov, V. L. (2017). *Contact Mechanics and Friction: Physical Principles and Applications*, 2nd Edn. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-10803-7
- Popov, V. L., and Hess, M. (2018). Voltage induced friction in a contact of a finger and a touchscreen with a thin dielectric coating. *arXiv[Preprint].arXiv:1805.08714*
- Socoliuc, A., Bennewitz, R., Gnecco, E., and Meyer, E. (2004). Transition from stick-slip to continuous sliding in atomic friction: entering a new regime of ultralow friction. *Phys. Rev. Lett.* 92:134301. doi: 10.1103/PhysRevLett.92.134301
- Vinogradova, O. I., and Feuillebois, F. J. (2003). Interaction of elastic bodies via surface forces 2. Exponential decay. *Colloid Interface Sci.* 268, 464–475. doi: 10.1016/j.jcis.2003.09.00

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Popov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Numerical Model for Hydration-Lubricated Contact and Its Friction Behavior at Nanoscale

Yanfei Fang, Liran Ma*, Xianzhang Wang and Jianbin Luo

State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing, China

In this study, we propose a numerical model to investigate the contact and friction behaviors of hydration-lubricated contact in an aqueous solution confined in a gap of several molecular diameters. A force–distance curve was obtained under different surface forces, and friction behavior was determined using the modified Eyring thermal activation model. The results showed that the normal boundary contact pressure and contact gaps were obtained using the proposed numerical model. In addition, friction force increased with the logarithm of sliding velocity. Hence, the Eyring model could predict the friction behavior of hydration lubrication. Moreover, it could capture the underlying physical principles of the contact and friction behaviors of hydration lubrication.

Keywords: hydration lubricated contact, surface force, friction behaviors, Eyring model, liquid thin film, numerical model

OPEN ACCESS

Edited by:

Alessandro Ruggiero,
University of Salerno, Italy

Reviewed by:

Valentin L. Popov,
Technical University of
Berlin, Germany
Alexander Filippov,
Dnesh Institute for Physics and
Engineering, Ukraine

*Correspondence:

Liran Ma
maliran@tsinghua.edu.cn

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 22 May 2020

Accepted: 17 August 2020

Published: 25 September 2020

Citation:

Fang Y, Ma L, Wang X and Luo J
(2020) Numerical Model for
Hydration-Lubricated Contact and Its
Friction Behavior at Nanoscale.
Front. Mech. Eng. 6:564756.
doi: 10.3389/fmech.2020.564756

INTRODUCTION

In nanofilm lubrication, where film thickness is in the range of a few nanometers, the physical, and chemical properties of surfaces strongly affect the friction behavior of liquid thin films, particularly of aqueous liquids. The hydration water layer is a particular form of water molecules tightly bound to ionic or hydrophilic surfaces (Kim et al., 2013, 2015). It plays an important role in specific fields, such as biolubrication (Hansen et al., 2013), hydration polymer brushes (Raviv et al., 2003; Chen et al., 2009), and nanofriction.

In the past few decades, numerous studies have investigated the physical processes of nanoscale thin film lubrication (TFL). The improvement of modern measurement techniques has considerably facilitated the study of TFL with aqueous liquids. The surface force apparatus (SFA) and atomic force microscopy (AFM) have significantly improved understanding of the tribological behavior of thin films by providing techniques for measurement of nanoscale liquid film thickness and normal and tangential forces. Jadhao and Robbins (2017), Israelachvili and Pashley (1983), Israelachvili and Wennerstroem (1990), and Israelachvili and Wennerström (1996) measured the short-range hydration repulsive force in water containing ionic or polar species. They found that a water film separates contacting surfaces, and it is not completely squeezed out even under high pressures. Additionally, they resolved several contradictions and findings that were inconsistent with hydration models. Horn et al. (2006) and Manica et al. (2008) investigated the dynamic forces maintaining relative separation between two deformable droplets. They reported that hydrodynamic and disjoining pressures strongly influence the deformation of droplets. Raviv and Klein (2002) measured the shear forces between solid surfaces sliding in aqueous salt solutions. They found strong repulsive hydration forces in these systems. Ma et al. (2015) used surface force balance to study the hydration lubrication between atomically smooth charged surfaces and found viscous losses. Diao and Espinosa-Marzal (2018) used AFM to investigate the role of water in

fault lubrication. They found that a lubricating film remains confined under high pressure and friction is significantly reduced at low sliding velocities. Experimental techniques have shown that the thickness of a water film is in the nanometer range. Additionally, the film is stabilized by the hydration repulsion between hydrophilic surfaces rather than by hydrodynamic lift forces. In nanoconfined water, surface interactions consist of long-range van der Waals attraction, short-range repulsion due to hydration force, and electrostatic repulsion with an intermediate action range (longer than that of hydration repulsion but shorter than that of van der Waals attraction). These repulsive forces are effective for achieving negative adhesion (Popov, 2020).

The above literature review shows that numerous studies have experimentally examined hydration-lubricated contact. However, there is no effective numerical model that describes the hydration contact characteristics and energy dissipation during sliding. The main experimental difficulty remains the determination of separation of surfaces under hydration repulsion because AFM experiments cannot provide accurate information about surface separation and the shear rate. In this study, we propose a numerical model for the boundary friction due to hydration lubrication. This numerical model may provide a promising method for further improvement of investigation of hydration lubrication.

GOVERNING EQUATIONS

Contact Theory Considering Hydration Effect

There are four interacting surface forces based on the interaction range: electrical double layer (EDL) force, hydration force, van der Waals force, and the repulsive force between two solid hydrophilic surfaces under nanoscale separation. Derjaguin and Landau (1941) firstly proposed the theory that surface force is the sum of EDL force and van der Waals force under long-distance interaction. Later, this theory was independently proposed by Verwey and Overbeek, and it is known as the Derjaguin–Landau–Verwey–Overbeek (DLVO) theory [also refer to Israelachvili (2011)]. The DLVO theory can be applied to accurately describe long-distance interaction forces in correlated systems. In contrast with the DLVO force, hydration repulsive force decays exponentially in the subnanometer range. Hydration repulsive force has a significant effect at subnanometer distances, and it decreases rapidly with surface separation. Therefore, different surface forces can prevail depending on the thickness of the thin film of an aqueous solution.

Based on the Lifshitz–Hamaker approach (Lifshitz, 1956; Israelachvili, 2011), van der Waals interaction and repulsive contact interaction can be obtained as the Derjaguin approximation by the integration of the Lennard-Jones interatomic potential. The contribution of van der Waals attractive interaction to overall disjoining pressure can be expressed as

$$p_{vdW}(h) = -\frac{A_H}{6\pi h(x, y)^3} \quad (1)$$

The integration of the repulsive part of the Lennard-Jones potential yields

$$p_{rep}(x, y) = \frac{A_H l_0^6}{90\pi h(x, y)^9} \quad (2)$$

where A_H is the Hamaker constant of a material, l_0 is the equilibrium separation when the interatomic force between two molecules is zero (Yu and Polycarpou, 2004), and $h(x, y)$ is the separation of contact surfaces.

Typically, two similarly charged surfaces electrostatically repel each other in a solution. However, they may interact at small separations of several molecular diameters under certain conditions. The EDL interactions of ionic aqueous liquids can be expressed as follows:

$$p_{EDL}(h) = p_{EDL0} \tanh^2(ze\varphi_0/4kT) e^{-\kappa h} \quad (3)$$

where p_{EDL0} is the initial EDL pressure defined by $p_{EDL0} = 64kT\rho_\infty[M]$, k is the Boltzmann constant, T is the absolute temperature, ρ_∞ is the concentration of alkali metal ions, φ_0 is the electrostatic potential at the midplane, and κ^{-1} is the Debye length.

When two hydrophilic surfaces or particles approach each other up to a distance of a few nanometers, water is arranged in an ordered manner in the vicinity of the surfaces. This structure can generate an effective repulsive force to support the external load between the surfaces. Hydration repulsive force (also referred to as non-DLVO force) becomes extremely strong in the short range. Hydration repulsive pressure can be described as a simple exponential expression (Faraudo and Bresme, 2005).

$$p_{hyd}(x, y) = p_{hyd0} e^{-h(x, y)/\lambda_0} \quad (4)$$

where p_{hyd0} is the initial constant hydration pressure and λ_0 is the characteristic decay length of hydration.

As analyzed above, the local distribution of the total contact pressure between two contact surfaces has four terms, namely, hydration pressure, EDL pressure, van der Waals attractive pressure, and repulsive contact pressure.

$$p_{sf}(x, y) = p_{hyd}(x, y) + p_{EDL}(x, y) + p_{vdW}(x, y) + p_{rep}(x, y) \quad (5)$$

This study investigates the contact between an elastic sphere with radius R and a smooth flat surface, as shown in **Figure 1**. The contact profile under load can be expressed as

$$h(x, y) = h_0 + \frac{x^2 + y^2}{2R} + u(x, y) \quad (6)$$

where h_0 is the initial gap of contact surfaces and $u(x, y)$ is the elastic deformation of the contact region. The elastic deformation can be calculated using the Boussinesq integral (Johnson, 1987) as

$$\delta(x, y) = \frac{2}{\pi E'} \iint_{\Omega_c} \frac{p_{sf}(x', y')}{\sqrt{(x - x')^2 + (y - y')^2}} dx' dy' \quad (7)$$

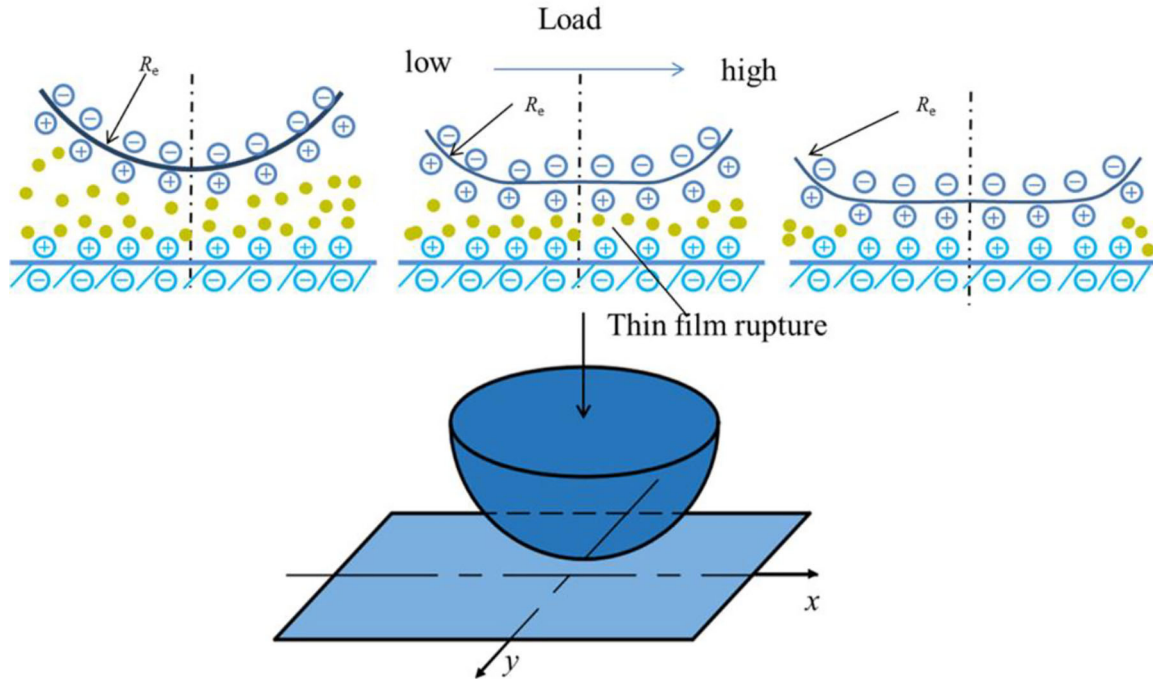


FIGURE 1 | Hydration-lubricated contact and its rupture under increasing load.

where E' is the effective elastic modulus of contacting bodies, $\frac{1}{E'} = \frac{1}{2}(\frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2})$ with E_1, E_2 being the Young modulus, and ν_1, ν_2 the Poisson's ratio of two contacting bodies, respectively.

The total external force balanced with the surface pressure on the contact region is given by,

$$F = \iint p_{sf}(x, y) dx dy \quad (8)$$

Friction Evaluation Based on the Eyring Model

If contact surfaces are subjected to relative tangential displacement, the molecules of a hydration lubrication layer must be rearranged. This rearrangement is supported by thermal fluctuations. The Eyring model (Eyring, 1935, 1936) describes the transition state theory of the reaction rate. It is used to evaluate and analyze the friction force of hydration-lubricated contact.

The Eyring model assumes that, in the absence of tangential loading, the heights of the energy barrier for the hopping of molecules to adjacent positions have the same value (E_0) in opposite directions; thus, the net flow is zero. The application of shear force reduces the height of the energy barrier by ΔE in the direction of flow and increases it by the same amount in the opposite direction, as shown in **Figure 2**. When the fluid is under shear stress τ , the change in the energy barrier can be approximated as $V^*\tau$, where V^* is an effective stress-activated volume. The flow rates in the forward and backward directions

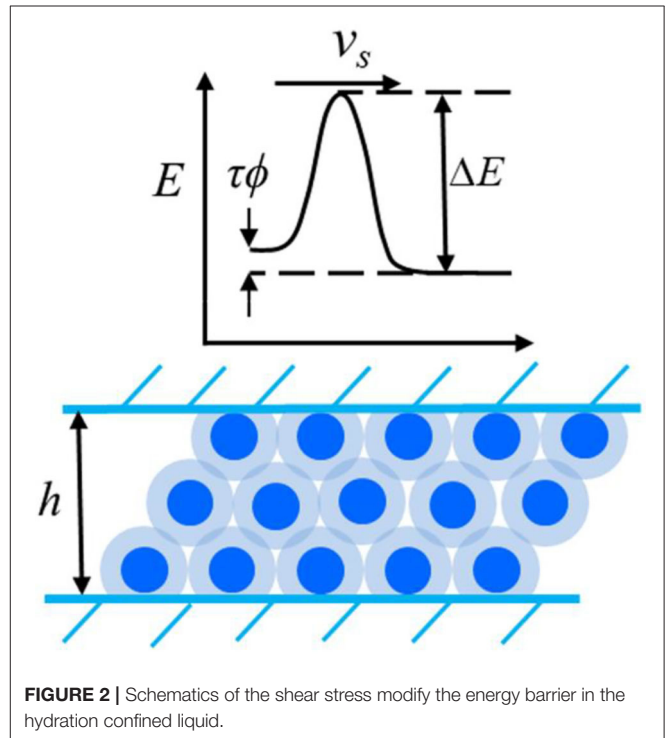
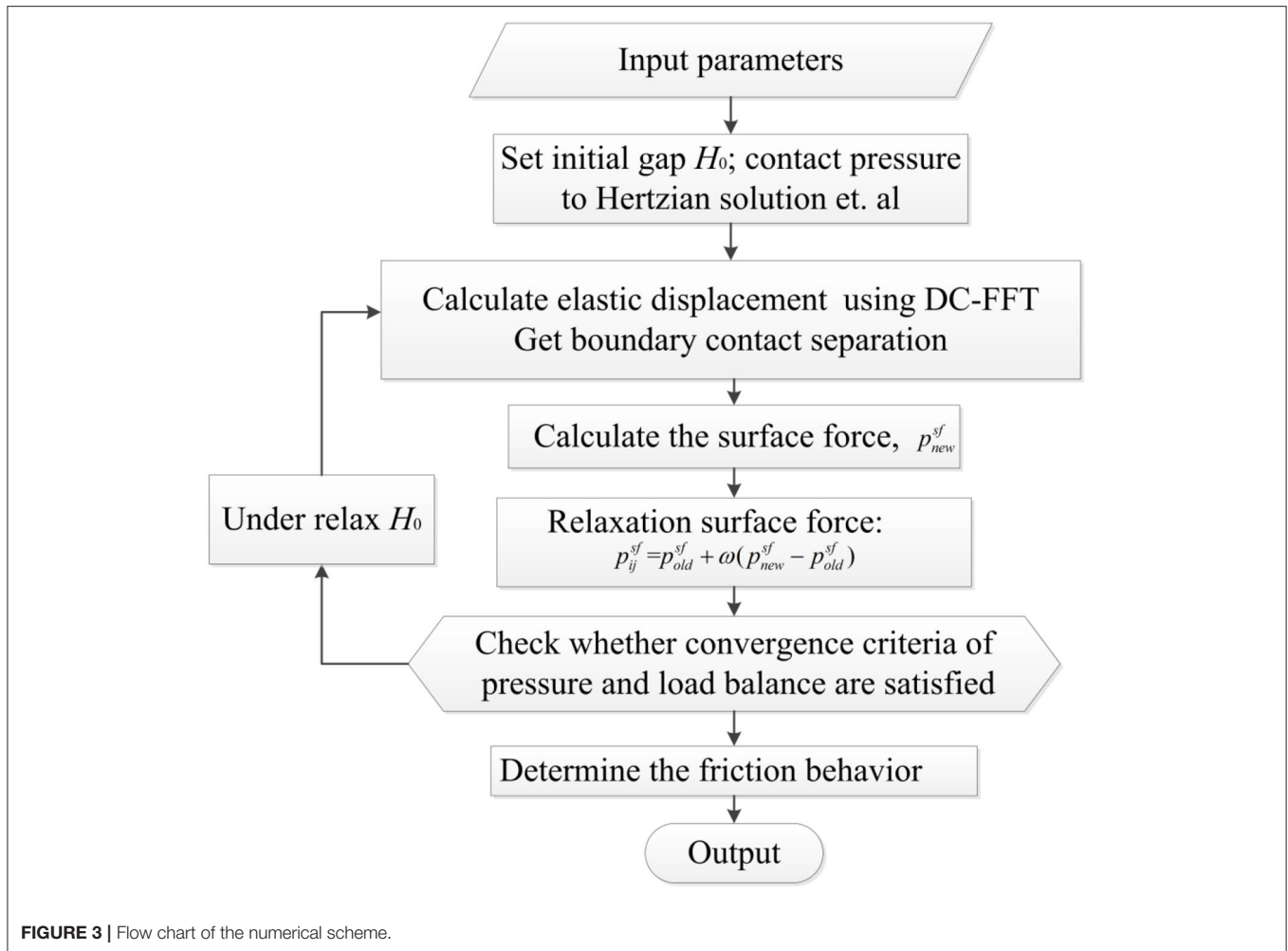


FIGURE 2 | Schematics of the shear stress modify the energy barrier in the hydration confined liquid.

under shear stress are expressed as follows (Spikes, 2018):

$$v_f = v_0 e^{-(\Delta E - V^*\tau)/kT} \quad (9a)$$

$$v_b = v_0 e^{-(\Delta E + V^*\tau)/kT} \quad (9b)$$



where v_0 is the frequency of the attempted rearrangements of a hydration shell. In laminar flow, the rate of relative molecular motion to a neighboring layer is given by

$$\dot{\gamma} = (v_f - v_b) = v_0 \exp^{-\Delta E/k_B T} \sinh\left(\frac{V^* \tau}{k_B T}\right) \quad (10)$$

It is assumed that the shear strain rate is proportional to the rate of activated molecular motion. Further, based on the definitions of the Eyring stress, $\tau_E = k_B T/V^*$, and the Newtonian viscosity, $\eta_N = \frac{\tau_E}{2c v_0} \exp(\Delta E/k_B T)$, the following expression can be obtained (Jadhao and Robbins, 2017, 2019):

$$\dot{\gamma} = \frac{\tau_E}{\eta_N} \sinh\left(\frac{\tau}{\tau_E}\right) \quad (11)$$

At a large value of τ/τ_E , $\sinh(\tau/\tau_E) \approx 0.5 \exp(\tau/\tau_E)$. Taking logarithms on both sides of Equation 11 and rearranging the terms, the following equation can be obtained:

$$\tau = \tau_E \ln\left(2 \frac{\eta_N \dot{\gamma}}{\tau_E}\right) \quad (12)$$

Experimental investigations suggest that this equation should be extended by an additional term τ_0

$$\tau(x, y) = \tau_0 + \tau_E \ln\left(2 \frac{\eta_N \dot{\gamma}}{\tau_E}\right) \quad (13)$$

where τ_0 is initial shear stress.

Therefore, when the shear rate is determined, shear stress can be calculated using Equation 13. The shear force under hydration-lubricated contact can be expressed as,

$$f = \iint \tau(x, y) dx dy = \iint \tau_E \ln\left(2 \frac{\eta_N \dot{\gamma}}{\tau_E}\right) dx dy + \text{constant} \quad (14)$$

NUMERICAL SCHEME

In the numerical scheme, control equations were discretized and expressed in the dimensionless form using dimensionless variables as follows: $X = \frac{x}{a}$, $Y = \frac{y}{a}$, $H = \frac{hR}{a^2}$, and $P_{sf} = \frac{P_{sf}}{P_H}$. Here, a is the Hertz contact radius, and P_H is the Hertz contact pressure.

Surface repulsion force, the contact profile, and load balance were numerically calculated to determine the contact pressure of

surface force and the contact separation between the two surfaces. Elastic deformation was determined by utilizing surface force; this was the most time-consuming step. The discrete convolution and fast Fourier transform (DC-FFT) technique (Liu et al., 2000) was used to improve the efficiency of calculating elastic deformation. The relaxation method was applied to ensure the robustness of the numerical scheme. The simulated area was five times larger than the Hertz diameter in the x and y directions. It was discretized into 257×257 nodes. The convergence criteria for pressure and load were $< 10^{-4}$. The converged solutions of thickness and pressure distributions were obtained after the convergence criteria were satisfied. A flow chart of the numerical scheme is shown in Figure 3.

RESULTS AND DISCUSSION

We obtained important data on hydration capacity from a previous study (Ma et al., 2015). Then, we analyzed the surface force of a thin film of a 0.1 mol NaCl water solution using following parameters: $p_{edl0} = 15.9 \times 10^6$ Pa, $\kappa^{-1} = 1.36$ nm, $\varphi = 70$ mV, $\lambda_0 = 0.2$ nm, $p_{hyd0} = 1.00 \times 10^9$ Pa, $A_H = -2 \times 10^{-20}$ J, and $l_0 = 0.3$ nm. The plots of interface pressure as a function of surface separation h were obtained using Equation 5, and they are shown in Figure 4. It can be observed that hydration repulsive pressure and EDL pressure have remarkable characteristics in the short and long ranges. Hydration repulsive pressure has a significant effect in the range of 0.5–1.5 nm, while EDL pressure may have a significant effect in the range of 2–5 nm in this case. Further, for the ultrathin gap, interface pressure increases rapidly and exceeds 1 GPa when the gap is < 0.3 nm. The sharp increase in repulsive interaction can be interpreted as a sign of the “direct contact” of the surfaces that are no longer separated by a lubricant layer. The pressure–distance curve shows that hydration pressure and EDL pressure could work synergistically to separate the contact surface, so the liquid film would be not squeezed out in contact region to produce beneficial results for TFL. These repulsive pressures could support external load so that contact surface would not contact directly.

We simulated hydration-lubricated contact under the conditions of an experiment performed using the SFA. Atomically smooth mica surfaces were bent in the form of two crossed cylinders with a radius of curvature (R) of ~ 1 cm in salt solutions. The elastic modulus of mica is ~ 60 GPa (Israelachvili et al., 2004). The contact profile and surface pressure were obtained using the numerical model, and they are shown in Figure 5. The surface contact profile is negligibly affected by elastic deformation under low contact pressure mainly caused by the EDL effect. The values of pressure are $< \sim 2.5$ MPa. As contact load increases, contact separation decreases and becomes < 1 nm when contact load is 10 mN. The maximum contact pressure is 25 MPa. It can be inferred that the hydration effect plays a dominant role in contact deformation. This implies that the hydration effect generates a force to support external load.

Figure 6 shows normal force as a function of surface separation under different initial constant hydration pressure

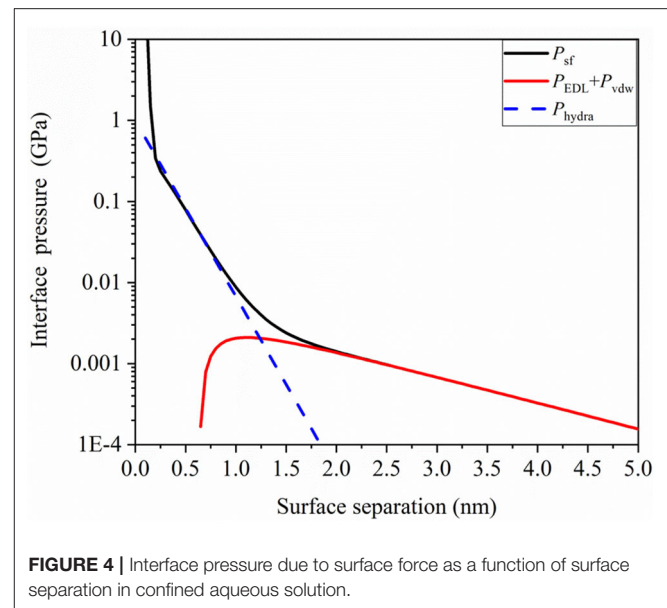


FIGURE 4 | Interface pressure due to surface force as a function of surface separation in confined aqueous solution.

changing from 1GP to 0.3GP. As surface separation decreases, normal force sharply increases at several molecular diameters. This implies that hydration load decreases exponentially when surface separation is $< \sim 1$ nm. In addition, the surface separations for the same normal force approach each other as the hydration effect decreases. The repulsive pressure generated from the interaction of solid atoms also supports a part of external load. When the separation distance of surfaces is > 2 nm, the surface force is mainly caused by EDL force and van der Waals attractive force to support the external force.

After the separation, h , between contact surfaces is calculated, the friction force due to the relative tangential motion of surfaces can be determined. In the following example, the friction behavior is shown for a load of 1 mN. The initial hydration pressure is 0.3 GPa. Other parameters remain unchanged. The minimum separation is 0.63 nm. Based on Figure 6, it can be inferred that there is hydration-lubricated contact. The effective stress-activated volume is $V^* = 2 \times 10^{-23}$ m³, and the Newtonian viscosity of the liquid is 1 mPa s⁻¹. The constant of friction force is set to 6.5 μ N in Equation 14. It should be noted that the parameters of the Eyring shear force can change in different contact stages. In this case study, we only applied the parameters given in Ma et al. (2015). Figure 7 shows the dependence of friction force on sliding velocity under hydration-lubricated contact. Friction force increases with the logarithm of sliding velocity. Energy dissipation predominantly occurs under hydration-lubricated contact. The maximum friction coefficient is ~ 0.003 at a sliding velocity of 3 μ m/s. This reveals that, to achieve ultralow friction movement, it is extremely important to utilize surface forces such as hydration force and EDL force. Friction force mainly depends on the Eyring shear forces when rubbing surfaces are separated by a liquid hydrated structure. The hydration ions are not squeezed out and remain localized on charged surface, once the shear force applied on the hydration

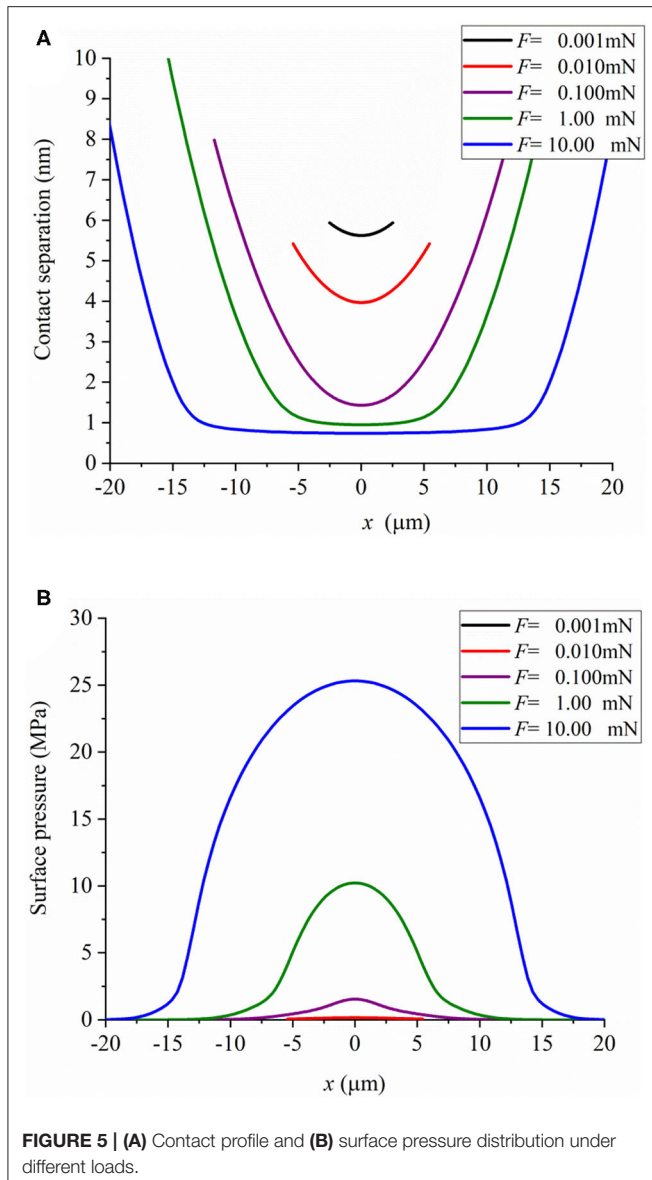


FIGURE 5 | (A) Contact profile and **(B)** surface pressure distribution under different loads.

shells reduces the energy barrier, and thereby, promotes the thermally activated slip (Ma et al., 2015; Diao and Espinosa-Marzal, 2018). Conversely, AFM shows a new regime of ultralow friction in the transition from stick-slip to continuous sliding (Socoliuc et al., 2004). Based on this phenomenon, in future, we will consider the coupling of the proposed model with the Tomlinson model to investigate surface forces and their potential influence on friction behavior via AFM.

CONCLUSIONS

A numerical model of hydration lubrication was proposed to evaluate contact and friction behavior at the nanoscale. The effects of hydration lubrication on contact pressure, surface profiles, and friction force were discussed. Preliminary numerical

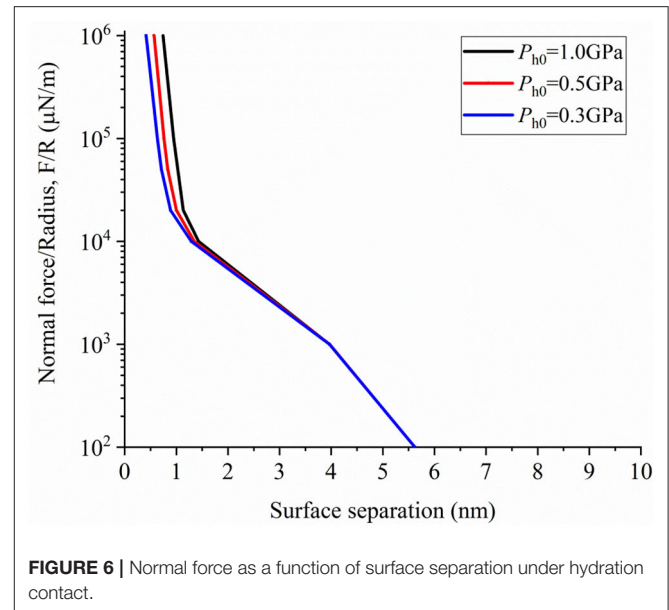


FIGURE 6 | Normal force as a function of surface separation under hydration contact.

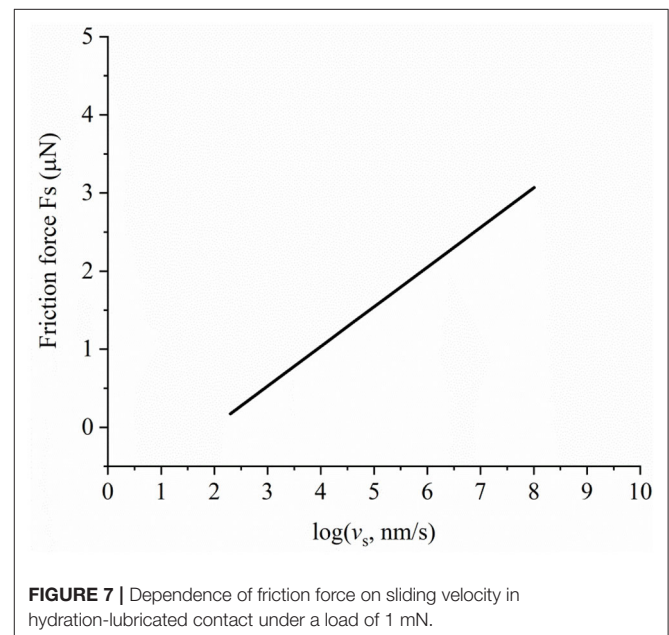


FIGURE 7 | Dependence of friction force on sliding velocity in hydration-lubricated contact under a load of 1 mN.

studies showed that surface forces significantly influenced boundary lubrication. Friction behavior was determined using the modified Eyring thermal activation model. It was observed that friction force increased with the logarithm of sliding velocity. This model is a preliminary investigation of the behavior of hydration-lubricated contact. The model may be further expanded to consider other factors that affect hydration lubrication, such as surface roughness, interface viscosity, and the hydrodynamic effect. The proposed numerical model may provide a new method of investigating hydration-lubricated contact at the nanoscale considering surface forces. This may be a key for understanding superlubricity in ultrathin film-lubricated contact at the nanoscale. One of the limitations of this study is

that the influence of the atomic corrugation and the roughness of surfaces were not considered explicitly. In the future work, we plan to take these factors into account and expand the prediction of friction under different surface force effect.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

YF formulated the numerical problem, carried out calculations, and drafted the manuscript. YF, LM, and XW discussed the

calculation and revised the manuscript. LM and JL provided guidance on the structure and content of this paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the National Natural Science Foundation of China (51527901 and 51675297).

ACKNOWLEDGMENTS

The authors would like to thank Prof. Valentin L. Popov for his constructive suggestions and the fruitful discussion for this study.

REFERENCES

- Chen, M., Briscoe, W. H., Armes, S. P., and Klein, J. (2009). Lubrication at physiological pressures by Polyzwitterionic brushes. *Science* 323, 1698–1701. doi: 10.1126/science.1169399
- Derjaguin, B. V., and Landau, L. D. (1941). Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solutions of electrolytes. *Prog. Surf. Sci.* 43, 30–59. doi: 10.1016/0079-6816(93)90013-L
- Diao, Y., and Espinosa-Marzal, R. M. (2018). The role of water in fault lubrication. *Nat. Commun.* 9:2309. doi: 10.1038/s41467-018-04782-9
- Eyring, H. (1935). The activated complex in chemical reactions. *J. Chem. Phys.* 3, 107–115. doi: 10.1063/1.1749604
- Eyring, H. (1936). Viscosity, plasticity, and diffusion as examples of absolute reaction rates. *J. Chem. Phys.* 4, 283–291. doi: 10.1063/1.1749836
- Faraudo, J., and Bresme, F. (2005). Origin of the short-range, strong repulsive force between ionic surfactant layers. *Phys. Rev. Lett.* 94:077802. doi: 10.1103/PhysRevLett.94.077802
- Hansen, Y. V., Gekle, S., and Netz, R. R. (2013). Anomalous anisotropic diffusion dynamics of hydration water at lipid membranes. *Phys. Rev. Lett.* 111:118103. doi: 10.1103/PhysRevLett.111.118103
- Horn, R. G., Asadullah, M., and Connor, J. N. (2006). Thin film drainage: hydrodynamic and disjoining pressures determined from experimental measurements of the shape of a fluid drop approaching a solid wall. *Langmuir* 22, 2610–2619. doi: 10.1021/la052314b
- Israelachvili, J., and Wennerström, H. (1996). Role of hydration and water structure in biological and colloidal interactions. *Nature* 379, 219–225. doi: 10.1038/379219a0
- Israelachvili, J. N. (2011). *Intermolecular and Surface Forces*. 3rd Edn. Burlington, MA: Academic Press.
- Israelachvili, J. N., Alcantar, N. A., Maeda, N., Mates, T. E., and Ruths, M. (2004). Preparing contamination-free mica substrates for surface characterization, force measurements, and imaging. *Langmuir* 20, 3616–3622. doi: 10.1021/la0352974
- Israelachvili, J. N., and Pashley, R. M. (1983). Molecular layering of water at surfaces and origin of repulsive hydration forces. *Nature* 306, 249–250. doi: 10.1038/306249a0
- Israelachvili, J. N., and Wennerstroem, H. (1990). Hydration or steric forces between amphiphilic surfaces. *Langmuir* 6, 873–876. doi: 10.1021/la00094a028
- Jadhao, V., and Robbins, M. O. (2017). Probing large viscosities in glass-formers with nonequilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7952–7957. doi: 10.1073/pnas.1705978114
- Jadhao, V., and Robbins, M. O. (2019). Rheological properties of liquids under conditions of elastohydrodynamic lubrication. *Tribol. Lett.* 67:66. doi: 10.1007/s11249-019-1178-3
- Johnson, K. L. (1987). *Contact Mechanics*. Cambridge: Cambridge University Press.
- Kim, B., Kim, Q., Kwon, S., An, S., Lee, K., and Lee, M., et al. (2013). Unified stress tensor of the hydration water layer. *Phys. Rev. Lett.* 111:246102. doi: 10.1103/PhysRevLett.111.246102
- Kim, B. H., Kwon, S., Lee, M., Kim, Q., An, S., and Jhe, W. (2015). Probing nonlinear rheology layer-by-layer in interfacial hydration water. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15619–15623. doi: 10.1073/pnas.1515033112
- Lifshitz, E. M. (1956). The theory of molecular attractive forces between solids. *Sov. Phys. J.* 2, 73–83
- Liu, S., Wang, Q., and Liu, G. (2000). A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear* 243, 101–111. doi: 10.1016/S0043-1648(00)00427-0
- Ma, L., Gaisinskaya-Kipnis, A., Kampf, N., and Klein, J. (2015). Origins of hydration lubrication. *Nat. Commun.* 6:6060. doi: 10.1038/ncomms7060
- Manica, R., Connor, J. N., Dagastine, R. R., Carnie, S. L., Horn, R. G., and Chan, D. Y. (2008). Hydrodynamic forces involving deformable interfaces at nanometer separations. *Phys. Fluids* 20:32101. doi: 10.1063/1.2839577
- Popov, V. L. (2020). Contacts with negative work of “adhesion” and superlubricity. *Front. Mech. Eng.* 5:73. doi: 10.3389/fmech.2019.00073
- Raviv, U., Giasson, S., Kampf, N., Gohy, J., Jerome, R., and Klein, J. (2003). Lubrication by charged polymers. *Nature* 425, 163–165. doi: 10.1038/nature01970
- Raviv, U., and Klein, J. (2002). Fluidity of bound hydration layers. *Science* 297, 1540–1543. doi: 10.1126/science.1074481
- Socoliuc, A., Bennewitz, R., Gnecco, E., and Meyer, E. (2004). Transition from stick-slip to continuous sliding in atomic friction: entering a new regime of ultralow friction. *Phys. Rev. Lett.* 92:134301. doi: 10.1103/PhysRevLett.92.134301
- Spikes, H. A. (2018). Stress-augmented thermal activation: Tribology feels the force. *Friction* 6, 1–31. doi: 10.1007/s40544-018-0201-2
- Yu, N., and Polycarpou, A. A. (2004). Adhesive contact based on the Lennard-Jones potential: a correction to the value of the equilibrium distance as used in the potential. *J. Colloid Interf. Sci.* 278, 428–435. doi: 10.1016/j.jcis.2004.06.029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Fang, Ma, Wang and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Analysis of Elastic Normal Contact of Surfaces With Regular Microgeometry Based on the Localization Principle

Irina G. Goryacheva and Ivan Y. Tsukanov*

Ishlinsky Institute for Problems in Mechanics RAS, Moscow, Russia

OPEN ACCESS

Edited by:

Mitjan Kalin,
University of Ljubljana, Slovenia

Reviewed by:

Xi-Qiao Feng,
Tsinghua University, China
Michele (Mike) Ciavarella,
Politecnico di Bari, Italy

*Correspondence:

Ivan Y. Tsukanov
tsukanov@ipmnet.ru

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 16 March 2020

Accepted: 02 June 2020

Published: 07 July 2020

Citation:

Goryacheva IG and Tsukanov IY
(2020) Analysis of Elastic Normal
Contact of Surfaces With Regular
Microgeometry Based on the
Localization Principle.
Front. Mech. Eng. 6:45.
doi: 10.3389/fmech.2020.00045

Formulations of periodic contact problems for an elastic half-plane and an elastic half-space interacting with a rigid body, having regular microgeometry, and a method for their approximate solution based on the localization principle are proposed. General relations, connecting contact characteristics of the interface (contact pressure distribution and dependence of the real contact area on the nominal pressure) with a single asperity shape and the distance between them, are obtained. The examples, illustrating the use of the obtained approximate relations for the contact characteristics analysis in the case of wavy and wedged profiles, are presented. The comparison of the obtained results with the available exact solutions is carried out. It was established that the approximate dependences coincide with the exact solution up to high values of the nominal pressures. New approximate solutions of 2D contact problems for a periodic system of parabolic asperities with single and double contact segments within a period are derived. It is also shown that the ratio of the contact zone size to the distance between asperities, at which the interaction effect becomes significant, only slightly depends on asperities shape.

Keywords: regular microgeometry, elastic contact, asperities interaction effect, localization principle, contact characteristics

INTRODUCTION

In general case, a surface topography is represented by a combination of deterministic and random functions (Whitehouse, 1994) determined by natural factors or technological treatment of the surface. Deterministic components are formed either as a result of imperfections in the operation of technological equipment or in stationary operating conditions (for example, the steady shape of a worn surface Goryacheva, 1997). In addition, a regular microgeometry on the surface can be created to control the operational properties of friction pairs, in particular their tribological characteristics. The geometric structure of surfaces has a great significance on the friction processes in an elastic contact under the condition of minimal wear. Prediction of the contact characteristics of surfaces with a given regular microgeometry, as well as the control of its optimal microgeometry are the urgent problems for micro- and nanotribology (Myshkin and Goryacheva, 2016).

The term “regular microgeometry” suggests that on the surface there is a periodic or a non-periodic system of asperities (or grooves) of a certain shape, mathematically described by a continuous or a piecewise function. The most common types of microgeometry used in

tribological applications include isotropic (created in the transverse and longitudinal directions) and anisotropic (created in one direction only) ones.

Many operational characteristics of joints, such as stiffness, thermal and electrical conductivity, tightness, adhesion strength, etc. are determined by characteristics of normal contact such as the contact pressure distribution, the real contact area and the additional compliance caused by the existence of microgeometry. For their correct calculation, it is necessary to formulate and to solve a periodic contact problem that takes into account the density of contact spots and the interaction of asperities. Such problems and methods for solving them have been extensively investigated. Most of the results obtained relate to the solution of the plane periodic problem of the linear elasticity theory, corresponding to the anisotropic (two-dimensional) geometric texture. The approaches for solving this class of problems are well developed. The main analytical ones are based on the methods using complex potentials (Westergaard, 1939; Kuznetsov, 1976; Krishtafovich et al., 1994; Manners, 1998; Soldatenkov, 2013; Xu and Jackson, 2018), the method of superposition of Flamant solutions (Schmueser and Comninou, 1979; Block and Keer, 2008; Tsukanov, 2018a,b), the dual series equations method (Dundurs et al., 1973; Carbone and Mangialardi, 2004), and also their modifications. Based on these methods, problems were also solved with more complex boundary conditions, for example, taking friction into consideration, or the influence of adhesion forces (e.g., see the review of Goryacheva and Martynyak, 2014). The two-dimensional (2D) periodic problems including ones, having several segments of integration within one period can be also effectively solved numerically, e.g., with the use of iterative methods (Chekina and Keer, 1999; Manners, 2003). Boundary element method (BEM) (Ciavarella et al., 2005) and finite element method (FEM) (Paggi and Reinoso, 2018) are also successfully applied for solving problems of that type. Unlike the problems for a single contact zone, the close-form solution of a 2D periodic contact problem commonly contains trigonometric functions and it can be cumbersome for engineering analysis, particular if the shape of asperities is not simple.

In contrast to the case of 2D problem, the three-dimensional (3D) problem corresponding to an isotropic microgeometry was solved only by using semi-analytical and numerical methods (e.g., BEM, FEM) due to the lack of a direct inverse of the corresponding integral equations. The non-periodic contact problems for a system of asperities, having various shape, location and height distribution were solved by semi-analytical iteration methods (Goryacheva, 1997; Shen et al., 2018). The significant influence of contact spots density, depending on the number of asperities and relative distance between them, is emphasized. The numerical methods for 3D periodic contact problems in linear elasticity are well developed now (Müser et al., 2017). They usually include the Fast Fourier transform (FFT) technique to meet the periodic boundary conditions and to reduce the calculation costs (Stanley and Kato, 1997; Yastrebov et al., 2015). A rather effective approximate approach for estimating the distribution of contact pressures on the real contact spots for the surfaces with regular microgeometry, taking into account the asperities interaction, is the use of the localization principle

(Goryacheva, 2006). It allows calculating the real contact areas and pressure distribution on them for periodic systems of asperities, having equal or different heights. This method was further developed to analyze the contact pressure distribution during the indentation of a multilevel periodic system of spherical asperities into a two-layer half-space and to analyze the influence of asperities density on the surface layer stress state and the additional compliance due to the existence of surface microgeometry (Goryacheva and Torskaya, 2019).

The main purpose of this study is the development of the approximate analytical method of solution for periodic contact problems, involving regular surface microgeometry on the basis of localization principle. This class of contact problems is widely used in modeling of artificial texture; however, the close-form solutions are rarely available, especially for 3D problems. In this paper, the unified approximate method to solve 3D and 2D contact problems for a periodic system of asperities of equal height is analytically developed. The accuracy of the method is estimated based on the known exact solutions of the corresponding contact problems. New approximate solutions of 2D contact problems for a periodic system of parabolic asperities with single and double contact segments within a period are obtained. Also, new features of the dependences of contact characteristics (pressure distribution, contact zone size) on load are revealed.

LOCALIZATION PRINCIPLE IN SOLVING THE DISCRETE CONTACT PROBLEMS

In general case, the problem of discrete contact of a nominally flat surface, having regular microgeometry, with an elastic half-space can be formulated as follows. There is a finite or an infinite system of punches with a given shape, loaded with a constant nominal pressure, which is penetrated into the elastic half-space. For solving the problem it is necessary to determine the contact pressure distribution on the real contact spots, taking into account the curvature of the deformed half-space surface due to the interaction of punches (asperities). The main point of the localization principle (Goryacheva, 2006) is that the real contact pressure on a single contact spot with a sufficient accuracy is determined by setting the exact contact conditions only on the fixed spot under consideration and adjacent to it. The influence of the remaining spots is replaced by the action of the nominal (averaged over the remaining part of the surface) pressure. With this approach, the normal displacements of the half-space surface under an arbitrary fixed asperity from the action of the remaining asperities are taken into account. They are approximately calculated from the distributed nominal pressure acting outside the certain region which depends on the contact density.

The general formulation of the linear elastic 3D periodic contact problem for a system of asperities (with equal or different heights) and its solution using the localization principle are given in Goryacheva (1998). In particular, the results of a numerical analysis of the integral relations obtained (Goryacheva, 1998) show that for a system of periodically arranged axisymmetric

asperities with equal height, the pressure distribution $p(r)$ on the real contact spot (for not very intimate contact) with sufficient accuracy is determined by the expression:

$$p(r) = \frac{E^*}{4\pi^2} \int_0^a \Delta f(\rho) H_1(r, \rho) d\rho + \frac{2}{\pi} \bar{N} P_s \arctan \left(\frac{\sqrt{a^2 - r^2}}{\sqrt{A^2 - a^2}} \right) \quad (1)$$

where $f(r)$ is the function describing the shape of the punch contact surface, $\Delta f(r)$ is the Laplacian of the function $f(r)$; \bar{N} is the average number of contact spots per unit area, P_s is the load acting on a single contact spot, a is the contact spot radius, $A = (\pi \bar{N})^{-1}$ is the radius of circle outside which the value of nominal pressure acting on the elastic half-space is taken into account (**Figure 1**), E^* is the reduced modulus of elasticity of the contacting bodies, determined by the following expression

$$\frac{1}{E^*} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2}, \quad (2)$$

where E_1 , ν_1 and E_2 , ν_2 are the Young's moduli and Poisson's ratios of the materials of contacting bodies, respectively.

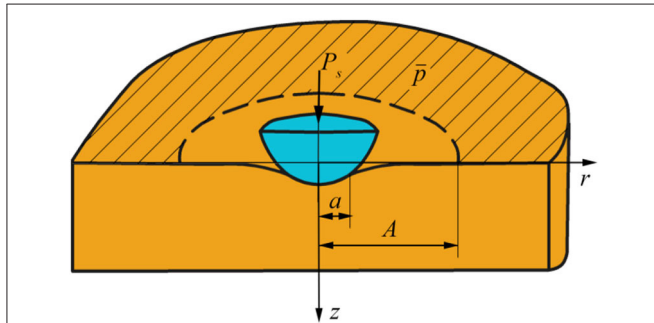


FIGURE 1 | Representation of contact interaction of a system of axisymmetric asperities with an elastic half-space according to the localization principle.

The kernel of the integral operator in Equation (1) is determined for the circular contact region in the case of bounded contact pressure as Goryacheva (1998):

$$H_1(r, \rho) = \int_0^{2\pi} \frac{2\rho}{\pi \sqrt{r^2 - 2r\rho \cos \theta + \rho^2}} \arctan \left(\frac{\sqrt{a^2 - r^2} \sqrt{a^2 - \rho^2}}{a \sqrt{r^2 - 2r\rho \cos \theta + \rho^2}} \right) d\theta. \quad (3)$$

The total load acting on a single contact spot is determined by the equilibrium equation

$$P_s = 2\pi \int_0^a p(r) r dr. \quad (4)$$

To increase the accuracy of contact characteristics calculation, especially at high contact density, it is necessary to solve the inhomogeneous integral equation of the second kind when determining the contact pressure (Goryacheva, 1998).

This approach was used (Goryacheva and Torskaya, 1995; Goryacheva, 1998) for calculation of contact pressures and internal stresses at indentation of a periodic system of equally high asperities into an elastic homogeneous and an inhomogeneous (coated) half-space. The calculation results show a significant effect of the asperities density on the distribution of contact pressures under a single asperity (**Figure 2A**) and on the dependence of radius of a single contact spot on the nominal pressure $\bar{p} = P/\bar{N}$ (**Figure 2B**).

In the case of indentation of periodic system of asperities with different heights into an elastic homogeneous or layered half-space the penetration of the highest asperities is calculated firstly. Then the curvature of the half-space boundary between asperities is calculated, and the nominal pressure at which the asperities of the next height level come into contact is determined. The described method for 3D problems with several height levels of asperities is suggested by Goryacheva (1997, 1998).

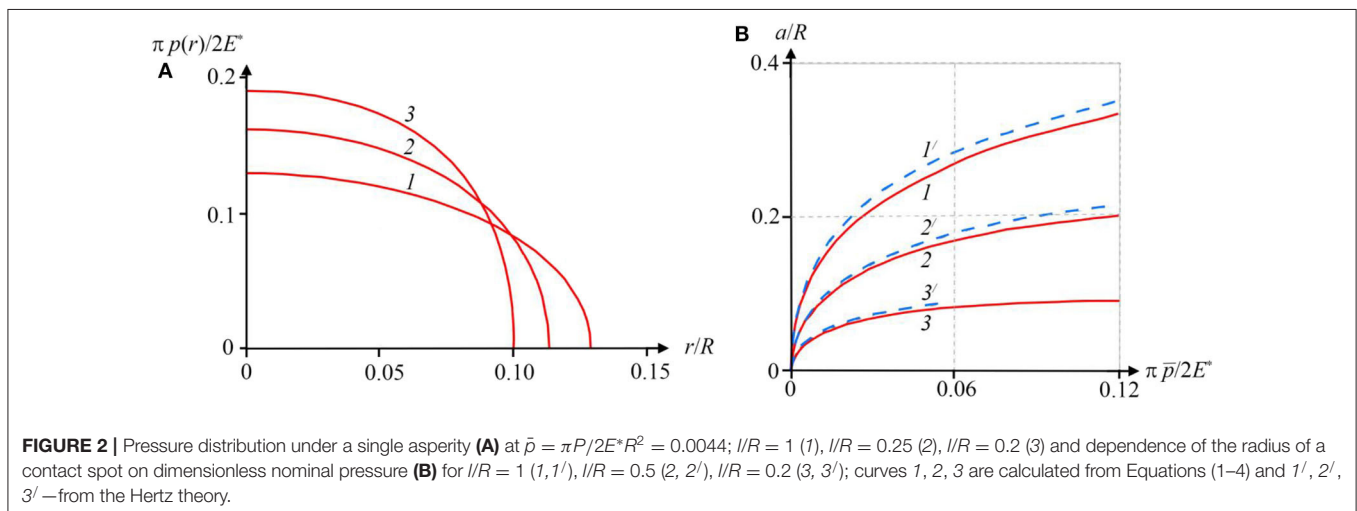


FIGURE 2 | Pressure distribution under a single asperity (**A**) at $\bar{p} = \pi P/2E^*R^2 = 0.0044$; $l/R = 1$ (1), $l/R = 0.25$ (2), $l/R = 0.2$ (3) and dependence of the radius of a contact spot on dimensionless nominal pressure (**B**) for $l/R = 1$ (1, 1'), $l/R = 0.5$ (2, 2'), $l/R = 0.2$ (3, 3'); curves 1, 2, 3 are calculated from Equations (1–4) and 1', 2', 3'—from the Hertz theory.

DETERMINATION OF THE CONTACT CHARACTERISTICS OF SURFACES WITH A MICRORELIEF IN THE CONDITIONS OF A PLANE ELASTICITY PROBLEM

In the general case, a plane-strain periodic contact problem without friction involves solving the following integral equation having a Hilbert kernel (Schtaierman, 1949; Barber, 2018):

$$\frac{E^*}{2} h'(x) = \frac{1}{2\pi} \int_{-a}^a p(\xi) \cot \frac{x-\xi}{2} d\xi, \quad (5)$$

here $h'(x)$ is the derivative of the initial gap function between surfaces, $p(x)$ is the contact pressure distribution; a is the half-width of the contact zone.

In the absence of a direct inversion of Equation (5) on an arbitrary segment $[-a, a]$ a closed form expression for determining the distribution of contact pressure is possible only for certain initial gap functions.

The method to solve the contact problems based on the localization principle, in the simplest formulation, involves the contact problem solution for a single asperity (determination of contact pressure distribution) taking into account the normal displacements of a half-plane boundary inside the considered single contact zone from action of the averaged pressure, applied outside the certain region (see section Localization Principle in Solving the Discrete Contact Problems).

The contact problem for a single contact segment is described by the integral equation with the Cauchy kernel (Muskhelishvili, 1953):

$$\frac{E^*}{2} h'(x) = \frac{1}{2\pi} \int_{-a}^a \frac{p(\xi)}{x-\xi} d\xi, \quad (6)$$

having for a symmetric initial gap function and bounded pressure on both ends of contact segment the following inversion (Muskhelishvili, 1953):

$$p(x) = \frac{E^*}{2\pi} \sqrt{a^2 - x^2} \int_{-a}^a h'(\xi) \frac{1}{\sqrt{a^2 - \xi^2}} \frac{1}{\xi - x} d\xi. \quad (7)$$

The initial gap function within the contact zone is determined by the following condition:

$$h(x) = \delta - (f(x) + f_2(x)), \quad (8)$$

where $f(x)$ is the asperity shape function, δ is the contact approach, $f_2(x)$ is the function, describing curvature of the half-plane boundary caused by the action of the remaining asperities except for the one under consideration.

According to the localization principle, the action of the remaining asperities is replaced by a uniform pressure acting outside a strip of width $2b$. The value of $2b$ is determined from the condition of equality of the mean pressure inside and outside of

this strip. The mean pressure in a plane periodic contact problem is determined as $\bar{p} = P_s/L$, where P_s is the total load on a single contact segment, L is the distance between the peaks of asperities (period); therefore $2b = L$. To determine the total load on a single contact zone P_s , the equilibrium equation is used

$$P_s = \int_{-a}^a p(x) dx. \quad (9)$$

The function $f_2(x)$ can be represented as a difference between displacements from a uniform load distributed over the entire half-plane and displacements from the same load inside a strip of width L (Johnson, 1985):

$$f_2(x) = -\frac{2}{\pi E^*} \frac{P_s}{L} \left(C - \int_{-a}^x \ln \left[\frac{L/2 + \xi}{L/2 - \xi} \right] d\xi \right). \quad (10)$$

Displacements from a uniform load distributed over the all half-plane are constant. This statement can be justified as follows. If one represents the uniform pressure distributed over a strip of arbitrary width $2b$ in the form of Fourier series with a period L as

$$p(x) = \bar{p} \frac{2b}{L} + \frac{2}{\pi} \bar{p} \sum_{n=1}^{\infty} \frac{1}{n} \sin \left(\frac{2\pi nb}{L} \right) \cos \left(\frac{2\pi nx}{L} \right), \quad (11)$$

then the derivative of vertical displacements of the half-plane boundary from the pressure distribution (11) is determined by the Hilbert transform (Srivastava and Lowengrub, 1970):

$$\bar{u}'_z(x) = H \{p(x)\} = \frac{4}{\pi} \bar{p} \sum_{n=1}^{\infty} \frac{1}{n} \sin \left(\frac{2\pi nb}{L} \right) \sin \left(\frac{2\pi nx}{L} \right). \quad (12)$$

Taking into account that $2b = L$, the right-hand side of Equation (12) becomes zero, and the half-plane displacements are constant. After differentiating Equation (10) and substitution the result in Equation (8), we obtain the expression for the derivative of the gap function inside the single contact zone:

$$\begin{aligned} h'(x) &= f'(x) + \frac{2}{\pi E^*} \frac{P_s}{L} \left(\ln \left[\frac{L/2 + x}{L/2 - x} \right] \right) \\ &= f'(x) + \frac{4}{\pi E^*} \frac{P_s}{L} \operatorname{artanh} \left(\frac{2x}{L} \right), \end{aligned} \quad (13)$$

Using Equation (7), one can determine the contact pressure taking into account the elastic interaction of asperities.

$$\begin{aligned} p(x) &= \frac{E^*}{2\pi} \sqrt{a^2 - x^2} \int_{-a}^a \left[f'(\xi) + \frac{4}{\pi E^*} \frac{P_s}{L} \operatorname{artanh} \left(\frac{2\xi}{L} \right) \right] \\ &\quad \frac{1}{\sqrt{a^2 - \xi^2}} \frac{1}{\xi - x} d\xi. \end{aligned} \quad (14)$$

The total load on the asperity, taking into account the symmetry of the function $f(x)$, is determined directly using Equation (13) (Schtaierman, 1949; Barber, 2018):

$$P = \frac{E^*}{2} \int_{-a}^a \frac{f'(\xi)\xi d\xi}{\sqrt{a^2 - \xi^2}} + \frac{2P_s}{\pi L} \int_{-a}^a \frac{\xi}{\sqrt{a^2 - \xi^2}} \operatorname{artanh}\left(\frac{2\xi}{L}\right) d\xi. \quad (15)$$

Calculation of the integral in the second term of Equation (15) with condition of $2a < L$ gives the following expression for the load:

$$P = \frac{E^*}{2} \int_{-a}^a \frac{f'(\xi)\xi d\xi}{\sqrt{a^2 - \xi^2}} + \frac{P_s}{L} \left(L - \sqrt{L^2 - 4a^2} \right). \quad (16)$$

Using Equation (16), it is possible to simplify the integration of the second term in square brackets in Equation (14). For this purpose the method based on the Abel transform of the function $\partial P/\partial a$ (Barber, 2018) was used. Then Equation (14) is reduced to the following form

$$p(x) = \frac{1}{\pi} \int_x^a \frac{P'_a(\xi) d\xi}{\sqrt{\xi^2 - x^2}} = \frac{E^*}{2\pi} \sqrt{a^2 - x^2} \int_{-a}^a \frac{f'(\xi) d\xi}{\sqrt{a^2 - \xi^2} (\xi - x)} + \frac{2P_s}{\pi L} \arctan\left(\frac{2\sqrt{a^2 - x^2}}{\sqrt{L^2 - 4a^2}}\right), \quad (17)$$

where $P'_a = \partial P/\partial a$.

It should be noted that in the 2D periodic problem the effect of elastic interaction on the contact pressure under the central asperity is characterized by a function similar to the 3D case for axisymmetric asperities arranged at the nodes of hexagonal lattice (see Equation 1).

Let us consider some examples of using the localization principle in 2D periodic contact problems for specific types of microgeometry, common in engineering applications.

Sinusoidal Profile

A two-dimensional profile described by the function $f(x) = \Delta(1 - \cos(2\pi x/L))$, where Δ , L are the amplitude and the period, is the simplest way to describe waviness or roughness of a surface formed along one direction (longitudinal or transverse). Expressions for determining the contact pressure distribution and the dependence of mean pressure \bar{p} on a contact zone half-width were first obtained by Westergaard (1939):

$$p(x) = \frac{\sqrt{2\pi} E^* \Delta}{L} \cos(\pi x/L) \sqrt{\cos(2\pi x/L) - \cos(2\pi a/L)}; \quad (18)$$

$$\bar{p} = p^* \sin^2(\pi a/L). \quad (19)$$

Here $p^* = \pi E^* \Delta/L$ is the pressure required to achieve complete contact between surfaces.

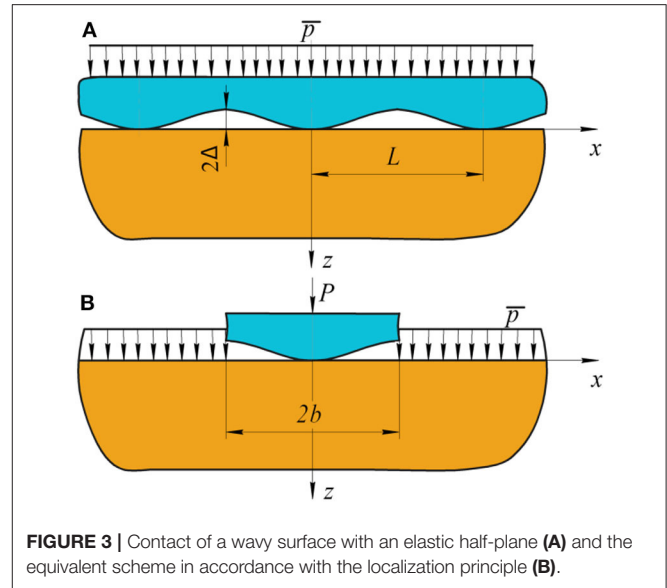


FIGURE 3 | Contact of a wavy surface with an elastic half-plane (A) and the equivalent scheme in accordance with the localization principle (B).

Consider the solution of this problem using the localization principle. The initial and the equivalent schemes of the problem are shown in Figure 3.

An analytical expression for contact pressure distribution under a single sinusoidal asperity, which is a solution of Equation (6), was obtained by Tsukanov (2018a) in the form of an infinite series of Chebyshev polynomials:

$$p_s(x) = \frac{2\pi \Delta E^*}{L} \sqrt{1 - (x/a)^2} \sum_{k=0}^{\infty} (-1)^k J_{2k+1}(2\pi a/L) U_{2k}(x/a), \quad (20)$$

where $U_i(x)$ is the Chebyshev polynomial of the second kind of degree i ; $J_k(x)$ is the Bessel function of the first kind of integer order k .

Equations for determining the total force, applied to a single asperity and the maximum pressure are obtained in closed form Tsukanov (2018a):

$$P_s = \frac{\pi^2 \Delta E^*}{L} a J_1(2\pi a/L); \quad (21)$$

$$p_{s\max} = p_s(x)|_{x=0} = \frac{\pi \Delta E^*}{L} \int_0^{2\pi a/L} J_0(t) dt. \quad (22)$$

Using Equations (16–17) and (20–22), the approximate expressions can be obtained to determine the contact characteristics of a wavy surface indenting into an elastic half-plane. The expressions for the mean and the maximum pressure have the following form

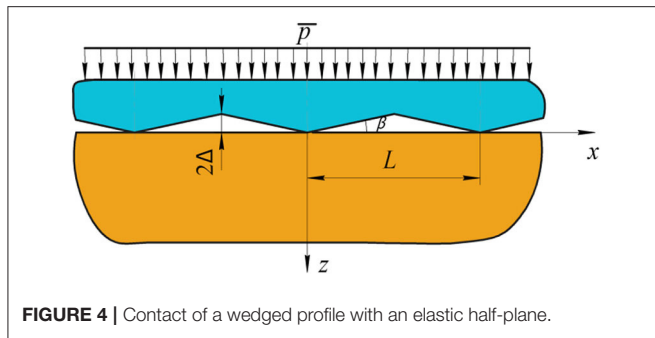


FIGURE 4 | Contact of a wedged profile with an elastic half-plane.

$$\bar{p} = \frac{\pi^2 \Delta E^*}{L^2} a J_1(2\pi a/L) \left(2 - \frac{\sqrt{L^2 - 4a^2}}{L} \right); \quad (23)$$

$$p_{\max} = \frac{\pi \Delta E^*}{L} \int_0^{2\pi a/L} J_0(t) dt + \frac{2\pi \Delta E^*}{L^2} a J_1(2\pi a/L) \left(\arctan \left(\frac{2a}{\sqrt{L^2 - 4a^2}} \right) \right). \quad (24)$$

Wedged Profile

This type of microrelief can be found on surfaces after very rough edge machining. The radius of curvature of asperities is much smaller than their height, and it can be considered as negligible. The scheme of the problem is shown in **Figure 4**. The equivalent scheme according to localization principle is similar to **Figure 3**.

For the possibility of applying the methods of the linear theory of elasticity, the angle $\beta = \arctan(4\Delta/L)$ should be small.

The exact solution of this problem was obtained by Block and Keer (2008):

$$p(x) = \frac{E^* \tan \beta}{\pi} \operatorname{arcosh} \left(\frac{\tan(\pi a/L)}{\tan(\pi |x|/L)} \right). \quad (25)$$

In accordance with the localization principle (see Equations 16, 17) and using well-known expressions for determining contact pressure and total load on a single blunt wedge (Johnson, 1985), the following approximate equations for determining the contact characteristics have been reduced:

$$p(x) = \frac{E^* \tan \beta}{\pi} \operatorname{arcosh} \left(\frac{a}{x} \right) + \frac{2E^* a \tan \beta}{\pi L} \arctan \left(\frac{2\sqrt{a^2 - x^2}}{\sqrt{L^2 - 4a^2}} \right); \quad (26)$$

$$\bar{p} = \frac{E^* a \tan \beta}{L} \left(2 - \frac{\sqrt{L^2 - 4a^2}}{L} \right). \quad (27)$$

Periodic System of Parabolic Asperities With a Single Contact Segment Within a Period

A common in practical applications type of texture (e.g., after laser surfacing) is a periodic system of cylindrical (parabolic)

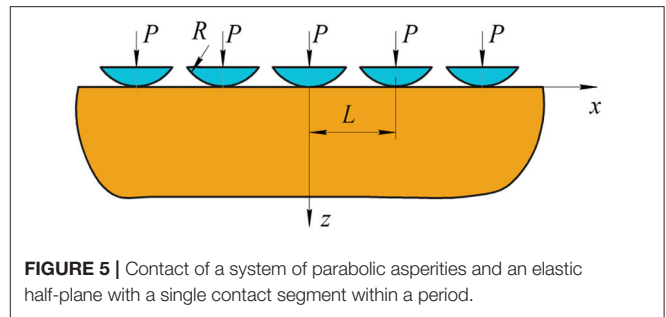


FIGURE 5 | Contact of a system of parabolic asperities and an elastic half-plane with a single contact segment within a period.

asperities in which the radius of curvature significantly exceeds their height. The scheme of the problem is shown in **Figure 5**.

This problem was previously considered by Kuznetsov (1978). An exact expression to calculate the contact pressure distribution in a closed form has not been derived, however the relation between the half-width a of a contact zone and the total load per one asperity P_s was found (Kuznetsov, 1978):

$$a = \frac{L}{\pi} \arccos \left(\exp \left(-\frac{2\pi P_s R}{L^2 E^*} \right) \right). \quad (28)$$

Using the well-known equations for determining the contact characteristics of a single smooth indenter (Johnson, 1985) and Equations (16, 17), the following expressions for the pressure distribution and the total load on a single asperity, taking into account elastic interaction are derived:

$$p(x) = \frac{E^*}{2R} \sqrt{a^2 - x^2} + \frac{E^* a^2}{2RL} \arctan \left(\frac{2\sqrt{a^2 - x^2}}{\sqrt{L^2 - 4a^2}} \right); \quad (29)$$

$$P_s = \frac{\pi E^* a^2}{4R} \left(2 - \frac{\sqrt{L^2 - 4a^2}}{L} \right). \quad (30)$$

Periodic System of Asperities With Two Contact Segments Within a Period

The more complicated problem arises, if there are multiple contact segments within one period. Without significant loss of generality let us consider the contact problem for a periodic non-uniform system of parabolic asperities and an elastic half-plane, with two interacting asperities within a single period. This type of contact problem has no exact solution, but can be approximately solved in a close-form using the localization principle.

The problem scheme is shown in **Figure 6**. Pairs of parabolic asperities form a periodic system with a period L . The shape of two interconnected asperities (see **Figure 6**) can be expressed by a biquadratic function:

$$f(x) = \frac{x^4}{8Rc^2} - \frac{x^2}{4R}. \quad (31)$$

Here R is a curvature radius of each asperity and $2c$ is a distance between asperity centers. The system of asperities is under the action of the nominal pressure \bar{p} .

For the two contact segments and asperities shape, defined by Equation (31), the contact pressure and the total load on one

contact segment are determined by the following expressions (Gladwell, 1980):

$$p_0(x) = \frac{E^* \sqrt{x^2 (x^2 - b_0^2) (a_0^2 - x^2)}}{4Rc^2}, \quad a_0 \leq |x| \leq b_0; \quad (32)$$

$$P_s = \frac{\pi E^* (b_0^2 - a_0^2)^2}{64Rc^2}, \quad (33)$$

where $2c = \sqrt{2(b_0^2 + a_0^2)}$.

The contact pressure in a general case of multizone 2D contact problem is determined by the following expression (Schtaierman, 1949; Muskhelishvili, 1953):

$$p(x) = \frac{E^*}{2\pi X(x)} \sum_{k=1}^n \int_{a_k}^{b_k} \frac{h'(\xi) X(\xi) d\xi}{\xi - x} + \frac{2iQ_{n-1}(x)}{X(x)}. \quad (34)$$

Here $h(x)$ is an initial gap function; n is a number of contact segments, $k = 1 \dots n$; a_k, b_k are the coordinates of the k -th contact segment; $X(x) = \sqrt{(x - a_1)(x - b_1) \dots (x - a_n)(x - b_n)}$; $Q_{n-1}(x) = D_0 x^{n-1} + D_1 x^{n-2} + \dots + D_{n-1}$; coefficients $D_0 \dots D_{n-1}$ are determined from the system of equations taking into account continuity and boundary conditions at the contact segment ends (Ghanati and Adibnazari, 2019).

According to the localization principle, to obtain a solution of a periodic problem we must consider the contact of the asperities at one period taking into account an additional curvature $f_2(x)$ of the half-plane boundary within the contact zones due to the influence of the other asperities. It follows from Equation (13), that the derivative of the initial gap function within each contact zone, can be expressed as:

$$h'(x) = f'(x) + f'_2(x) = \frac{x^3}{2Rc^2} - \frac{x}{2R} + \frac{4\bar{p}}{\pi E^*} \operatorname{artanh}\left(\frac{2x}{L}\right). \quad (35)$$

Taking into account the boundary conditions within one period and approximating the hyperbolic arctangent function in Equation (35) by the cubic polynomial under the assumption that

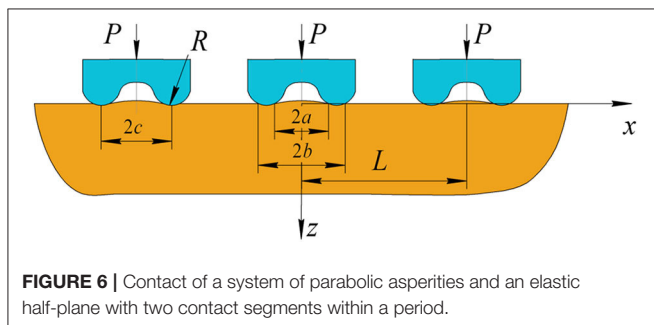


FIGURE 6 | Contact of a system of parabolic asperities and an elastic half-plane with two contact segments within a period.

$2c < L$, from Equation (34) we obtain the following expression for the contact pressure in periodic problem:

$$p(x) = \frac{E^* \sqrt{Z(x)}}{2\pi} \int_a^b \frac{h'(\xi) d\xi}{(\xi - x) \sqrt{Z(\xi)}}, \quad (36)$$

Here $Z(x) = (x^2 - b^2)(a^2 - x^2)$ (a and b are the ends of the contact zones in periodic contact problem) and $h'(x)$ is approximated by the following expression:

$$h'(x) = \left(\frac{1}{2Rc^2} + \frac{32\bar{p}}{3\pi E^* L^3} \right) x^3 + \left(\frac{8\bar{p}}{\pi E^* L} - \frac{1}{2R} \right) x. \quad (37)$$

Substituting Equations (35) and (37) in Equation (36) and following Gladwell (1980), we derive the following expression for the contact pressure in the periodic contact problem with two parabolic asperities within the period:

$$p(x) = E^* \left(\frac{1}{4Rc^2} + \frac{16\bar{p}}{3\pi E^* L^3} \right) \sqrt{x^2 (x^2 - b^2) (a^2 - x^2)}, \quad a \leq |x| \leq b, \quad (38)$$

where the ends of the contact zones follow the relation $b^2 + a^2 = 2c^2$.

A more general problem, involving multiple contact zones, can be treated similarly using numerical methods for solution of Equation (34).

For comparison of the results obtained in the periodic contact problem with the contact pressure distribution (32) for two asperities we assume that the distance between asperities $2c$ and the nominal (mean) pressure \bar{p} are the same; the value of the nominal pressure is calculated from the relation

$$\bar{p} = \frac{2P_s}{L}, \quad (39)$$

where P_s is given by Equation (33).

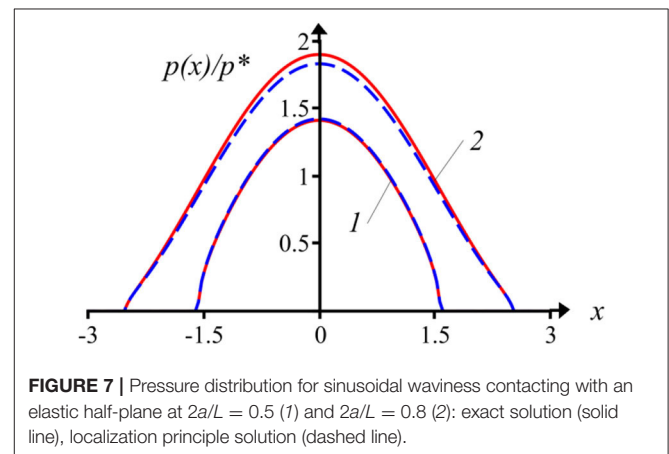
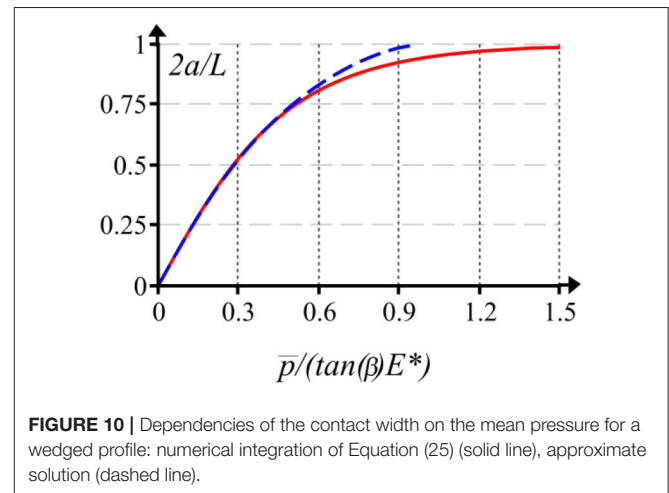
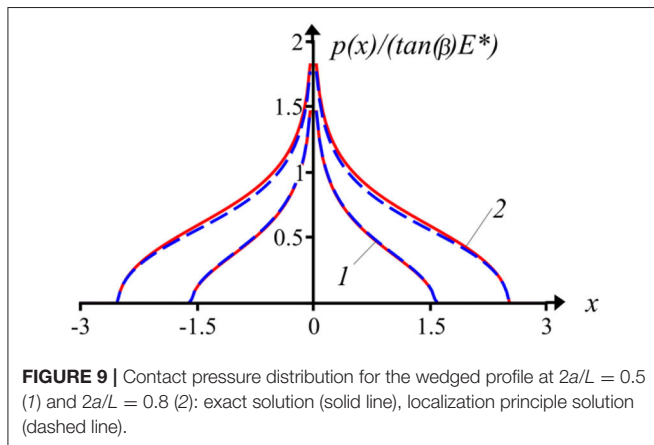
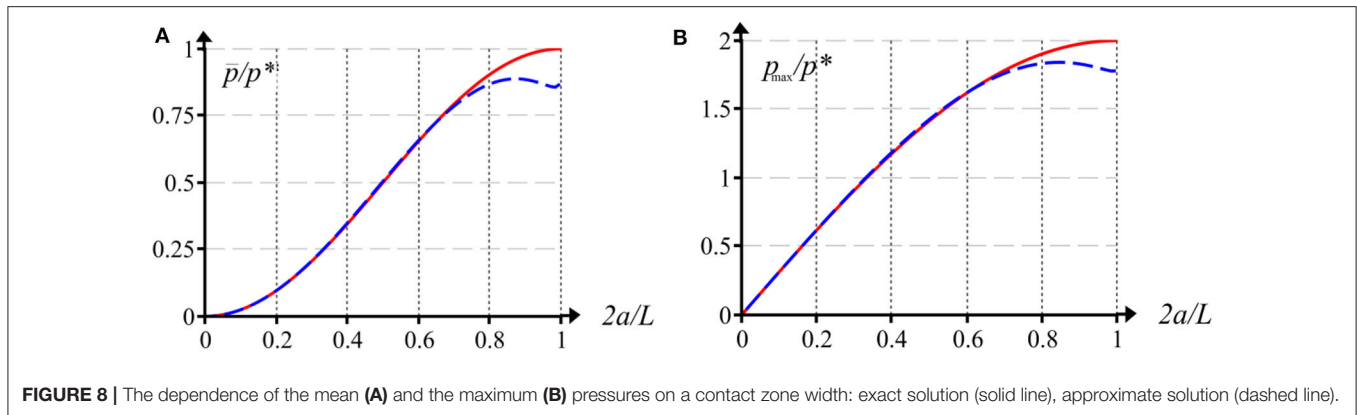


FIGURE 7 | Pressure distribution for sinusoidal waviness contacting with an elastic half-plane at $2a/L = 0.5$ (1) and $2a/L = 0.8$ (2): exact solution (solid line), localization principle solution (dashed line).



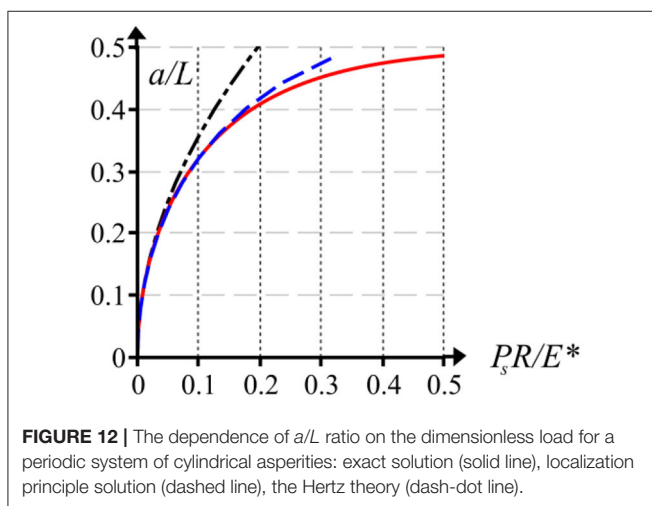
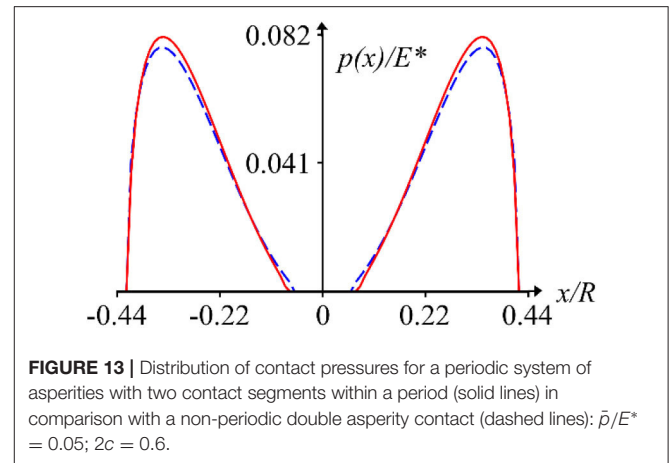
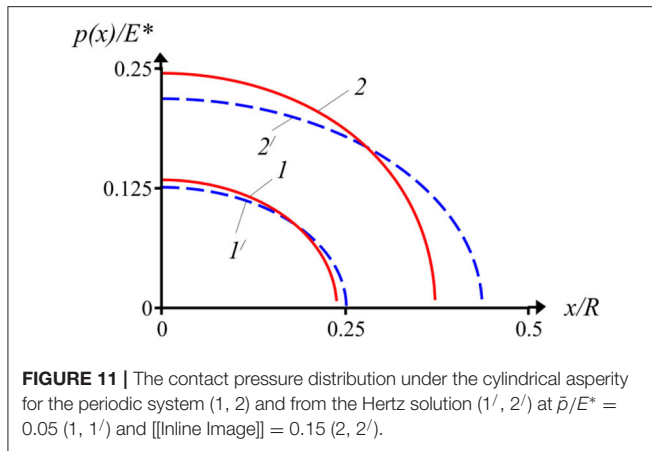
RESULTS AND DISCUSSION

The analytical solutions derived from the method of localization were used for the analysis of the contact characteristics in the periodic contact problems and for comparison with the available exact solutions. The dimensionless pressure distributions for the contact of a sinusoidal wavy surface and an elastic half-plane at two values of a dimensionless contact zone width are shown in **Figure 7**. The results indicate that the solution based on the localization principle allows predicting the distribution of contact pressures with sufficient accuracy up to high loads (high contact density, characterizing by the ratio $2a/L$). The comparison of the dependencies of the dimensionless mean and maximum contact pressures vs. contact zone width $2a$ for the exact and the approximate solutions is presented in **Figure 8**. The results indicate that a significant discrepancy between the exact and the approximate values of contact characteristics begins only at high contact density ($2a/L \approx 0.7$). Note, that for such high values of contact density the solution for almost complete contact can be applied (Johnson, 1985).

The distributions of contact pressure for a wedged profile indented into an elastic half-plane at two values of dimensionless contact zone width are shown in **Figure 9**. Comparison of the results calculated based on the exact solution and the localization principle makes it possible to conclude that the approximate

pressure distribution is close to the exact one even for the case with infinite peaks due to angle point in the wedged profile. The dependencies of the dimensionless contact zone width on the dimensionless mean pressure for the wedged profile, indenting into an elastic half-plane are shown in **Figure 10**. For the wedged profile, as well as for the wavy one, the discrepancy of the dependencies of the contact width on the mean contact pressure begins at $2a/L \approx 0.7$. At further increase of the applied pressure the approximate solution gives overestimated values of contact zone width.

The pressure distributions calculated for a periodic system of cylindrical (parabolic) asperities with use of Equation (29) in comparison with the Hertz theory (in a two-dimensional formulation) are shown in **Figure 11**. The asperity interaction effect is revealed in a decrease in the contact zone size and an increase in the peak pressure. With a further increase in load or asperities density, the pressure profile is significantly different from the Hertzian one. This result is in good agreement with the 3D case for spherical asperities (see **Figure 2**). The comparative graphs of a/L ratio on the dimensionless load for a periodic system of parabolic asperities with a single contact segment within a period are shown in **Figure 12**. The curve corresponding to the localization principle solution is close to the



exact solution up to the value $a/L \approx 0.35$ ($2a/L \approx 0.7$), as well as for other considered profiles. The discrepancy with the Hertzian curve corresponding to non-interacting asperities begins at $a/L \approx 0.125$. For the values of $a/L < 0.125$ the error of calculation using the Hertz theory is $< 3\%$. The last result is similar to the case of a 3D system of spherical asperities (see **Figure 2**). At the large values of load the approximate solution overestimates the contact zone size in comparison with the exact Equation (28), however the discrepancy is smaller than for a wedged profile.

Contact pressure distributions for a periodic system of parabolic asperities with two contact segments within a period ($L = 2.4c$) are shown in **Figure 13** in comparison with the non-periodic double asperity contact. The graphs show, that as well as for a single contact segment within a period (see **Figure 11**) the increase in the density of asperities leads to an increase in the peak pressure, and also to the reduction of contact width ($b-a$) at fixed nominal pressure. The considered case can be used in the analysis of a short-range and a long-range elastic interaction between asperities. The short-range elastic interaction, depending mainly on the distance between

two asperities, leads to asymmetry of the contact pressure distribution between the adjacent sides of the asperities. The long-range interaction is determined by the number and the shape of asperities within one period and the value of L . Effect of the long-range interaction is smaller than of the short-range; approximately it can be considered on the basis of localization principle. The maximum effect of the long-range elastic interaction is reached at merging of contact zones between the asperities ($a = 0$).

CONCLUSIONS

The unified approximate analytical method to solve 3D and 2D contact problems for a regular surface microgeometry penetrating the elastic half-space (half-plane) is developed. The equations derived for 3D and 2D cases, using the localization principle, show the identical structure for different dimensionality of the problem. The accuracy of the obtained method for two-dimensional problems was estimated for wavy and wedged profiles, by comparison with the exact solutions. New approximate solutions of 2D contact problems for a periodic system of parabolic asperities with single and double contact segments within a period are obtained.

The results obtained show that the application of the presented method in a plane contact problem for bodies with a periodic regular microrelief allows to calculate the contact characteristics with high accuracy up to large contact densities ($2a/L \approx 0.7$). Further increasing of the applied load or contact density leads to moderate overestimation of contact characteristics in comparison with the exact solutions. For all considered cases the increase in asperities density leads to an increase in the peak pressure, and also to the reduction of the contact half-width at a fixed load. Qualitatively, the form of contact pressure distribution is generally defined by the shape of asperities.

The advantage of the developed approach is the ability to separately consider the effects associated with the shape of

asperities and the relative distance between them. The method allows to simplify the calculation of contact characteristics for a complex-shaped regular texture, for which a straightforward analytical solution does not exist, including more general multizone contact problems. The approach developed can be also used for the solution of the 2D and 3D periodic contact problems with complicated boundary conditions (e.g., with adhesion of different nature at the interface Makhovskaya, 2003).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article and the supplementary material.

REFERENCES

- Barber, J. R. (2018). *Contact Mechanics*. Cham: Springer International Publishing.
- Block, J. M., and Keer, L. M. (2008). Periodic contact problems in plane elasticity. *J. Mech. Mater. Struct.* 3, 1207–1237. doi: 10.2140/jomms.2008.3.1207
- Carbone, G., and Mangialardi, L. (2004). Adhesion and friction of an elastic half-space in contact with a slightly wavy rigid surface. *J. Mech. Phys. Solids*. 52, 1267–1287. doi: 10.1016/j.jmps.2003.12.001
- Chekina, O. G., and Keer, L. M. (1999). A new approach to calculation of contact characteristics. *J. Tribol.* 121, 20–27. doi: 10.1115/1.2833804
- Ciavarella, M., Demelio, G., and Murolo, G. (2005). A numerical algorithm for the solution of two-dimensional rough contact problems. *J. Strain Anal. Eng.* 40, 463–476. doi: 10.1243/030932405X15936
- Dundurs, J., Tsai, K. C., and Keer, L. M. (1973). Contact between elastic bodies with wavy surfaces. *J. Elast.* 3, 109–115. doi: 10.1007/BF00045817
- Ghanati, P., and Adibnazari, S. (2019). A study on the extent of the contact and stick zones in multiple contacts. *Arch. Appl. Mech.* 89, 1825–1836. doi: 10.1007/s00419-019-01545-w
- Gladwell, G. M. L. (1980). *Contact Problems in the Classical Theory of Elasticity*. Sijthoff, Noordhoff.
- Goryacheva, I. (1997). *Contact Mechanics in Tribology*. Dordrecht: Springer.
- Goryacheva, I. G. (1998). Periodic contact problem for the elastic half-space. *J. Appl. Math. Mech.* 62, 1036–1044. doi: 10.1016/S0021-8928(98)00122-1
- Goryacheva, I. G. (2006). Mechanics of discrete contact. *Trib. Int.* 39, 381–386. doi: 10.1016/j.triboint.2005.04.020
- Goryacheva, I. G., and Martynyak, R. M. (2014). Contact problems for textured surfaces involving frictional effects. *P. I. Mech. Eng. J. J. Eng.* 228, 707–716. doi: 10.1177/1350650114528318
- Goryacheva, I. G., and Torskaya, E. V. (1995). Periodic contact problem for a system of punches and the elastic layer bonded to the elastic half-space. *J. Frict. Wear.* 17, 642–652.
- Goryacheva, I. G., and Torskaya, E. V. (2019). Contact of multi-level periodic system of indenters with coated elastic half-space. *FU Mech. Eng.* 17, 149–159. doi: 10.22190/FUME190511022G
- Johnson, K. L. (1985). *Contact Mechanics*. Cambridge: University Press.
- Krishtafovich, A. A., Martynyak, R. M., and Shvets, R. N. (1994). Contact between anisotropic half-plane and rigid body with regular microrelief. *J. Frict. Wear.* 15, 15–21.
- Kuznetsov, E. A. (1976). Periodic contact problem for half-plane allowing for forces of friction. *Sov. Appl. Mech.* 12, 1014–1019. doi: 10.1007/BF00885047
- Kuznetsov, Ye. A. (1978). The use of automorphic functions in the plane theory of elasticity. *Mech. Solids*. 6, 35–44.
- Makhovskaya, Y. Y. (2003). Discrete contact of elastic bodies in the presence of adhesion. *Mech. Solids*. 38, 39–51.
- Manners, W. (1998). Partial contact between elastic surfaces with periodic profiles. *Proc. R. Soc. Lond.* 454, 3203–3221. doi: 10.1098/rspa.1998.0298
- Manners, W. (2003). Methods for analysing partial contact between surfaces. *Int. J. Mech. Sci.* 45, 1181–1199. doi: 10.1016/S0020-7403(03)00143-7

AUTHOR CONTRIBUTIONS

IG, localization principle formulation with applications to 3D contact problems. IT, development of the localization method to 2D contact problems.

FUNDING

This research was funded by the Russian Foundation for Basic Research, grant 20-01-00400. The solution and analysis of contact characteristics for wavy surface falls under the framework of the grant from the Russian Foundation for Basic Research 19-08-00615.

- Müser, M. H., Dapp, W. B., and Bugnicourt, R. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65:118. doi: 10.1007/s11249-017-0900-2
- Muskhelishvili, N. I. (1953). *Some Basic Problems of the Mathematical Theory of Elasticity*. Holland: Noordhoff.
- Myshkin, N. K., and Goryacheva, I. G. (2016). Tribology: trends in the half-century development. *J. Frict. Wear.* 37, 513–516. doi: 10.3103/S106836661606009X
- Paggi, M., and Reinoso, J. (2018). A variational approach with embedded roughness for adhesive contact problems. *Mech. Adv. Mater. Struct.* 1–17. doi: 10.1080/15376494.2018.1525454
- Schmueser, D., and Comninou, M. (1979). The periodic array of interface cracks and their interaction. *Int. J. Solids Struct.* 15, 927–934. doi: 10.1016/0020-7683(79)90022-2
- Schtaierman, I. Y. (1949). *Contact Problem of Theory of Elasticity*. Moscow: Gostekhizdat.
- Shen, L., Quanzhou, Y., Qunyang, L., Xi-Qiao, F., and Huajian, G. (2018). Contact stiffness of regularly patterned multi-asperity interfaces. *J. Mech. Phys. Solids* 111, 277–289. doi: 10.1016/j.jmps.2017.10.019
- Soldatenkov, I. A. (2013). The periodic contact problem of the plane theory of elasticity. Taking friction, wear and adhesion into account. *J. Appl. Math. Mech.* 77, 245–255. doi: 10.1016/j.jappmath.2013.07.017
- Srivastava, K. N., and Lowengrub, M. (1970). Finite Hilbert transform technique for triple integral equations with trigonometric kernels. *Proc. Roy. Soc. Edn.* LXVIII, 309–321. doi: 10.1017/S0080454100008463
- Stanley, H. M., and Kato, T. (1997). An FFT-based method for rough surface contact. *J. Tribol.* 119, 481–485. doi: 10.1115/1.2833523
- Tsukanov, I. Y. (2018a). Partial contact of a rigid multisinusoidal wavy surface with an elastic half-plane. *Adv. Tribol.* 1–8. doi: 10.1155/2018/8431467
- Tsukanov, I. Y. (2018b). Periodical contact problem for a surface with two-level waviness. *Mech. Solids*. 53, 129–136. doi: 10.3103/S0025654418030160
- Westergaard, H. M. (1939). Bearing pressures and cracks. *J. Appl. Mech. T. ASME*. 6, 49–52.
- Whitehouse, D. J. (1994). *Handbook of Surface Metrology*. Bristol: Inst. of Physics.
- Xu, Y., and Jackson, R. L. (2018). Periodic contact problems in plane elasticity – the fracture mechanics approach. *J. Tribol.* 140, 404–415. doi: 10.1115/1.4036920
- Yastrebov, V. A., Anciaux, G., and Molinari, J. F. (2015). From infinitesimal to full contact between rough surfaces: evolution of the contact area. *Int. J. Solids Struct.* 52, 83–102. doi: 10.1016/j.ijsolstr.2014.09.019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Goryacheva and Tsukanov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Stress Tensor and Gradient of Hydrostatic Pressure in the Half-Space Beneath Axisymmetric Bodies in Normal and Tangential Contact

Fabian Forsbach*

Department of System Dynamics and Friction Physics, Institute of Mechanics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Liran Ma,
Tsinghua University, China
Vladislav Aleshin,
UMR8520 Institut d'Électronique, de
Microélectronique et de
Nanotechnologie (IEMN), France

*Correspondence:

Fabian Forsbach
fabian.forsbach@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 10 March 2020

Accepted: 13 May 2020

Published: 26 June 2020

Citation:

Forsbach F (2020) Stress Tensor and
Gradient of Hydrostatic Pressure in
the Half-Space Beneath Axisymmetric
Bodies in Normal and Tangential
Contact. *Front. Mech. Eng.* 6:39.
doi: 10.3389/fmech.2020.00039

The stress state in the volume of contacting bodies may essentially influence the material behavior. For evaluating various modes of inelastic behavior and/or failure, such as plastic deformation, crack initiation, and propagation or fatigue, the complete stress tensor beneath the contact interface may be of importance. For many geotechnical and biomechanical applications, the hydrostatic pressure gradient beneath the contact is of interest as well. However, most theories for normal and tangential contact provide only few stress components in the contact surface. In the present paper, we show that the full stress state in the half-space can be easily found for axisymmetric bodies. We provide expressions in form of one-dimensional integrals for all components of the stress tensor and the hydrostatic pressure gradient inside the half-space. In terms of numerical complexity, the proposed method can be advantageous to other elaborate methods.

Keywords: stress state, pressure gradient, normal contact, tangential contact, friction, axial symmetry, method of dimensionality reduction

INTRODUCTION

Since Huber's (1904) solution of the Hertzian contact, it is known that the von Mises equivalent stress and, thus, the endangered region for plastic failure, lies beneath the surface of the indented half-space. It was later shown that the same holds for the tangentially loaded Hertzian contact under sliding conditions for coefficients of frictions below 0.3 (Hamilton and Goodman, 1966) and for other indenter shapes (Ciavarella et al., 1998). However, many solutions for contact problems are limited to few components of surface stresses and, thus, do not allow for predictions with respect to plastic failure or more complicated failure mechanisms.

Another example, where the stress state inside the half-space or, more precisely, the hydrostatic pressure gradient is of interest, are natural joints such as knee or hip joints. When layers of cartilage are in contact, it was found experimentally that dynamic compression that may result from activities like walking and running promotes regenerative processes. In this loading scenario, a high rate of in- and outflow of interstitial fluid in the surface zone of the porous, fluid saturated medium is measured (Wong and Carter, 2003). It is believed that the fluid flow is responsible for transport of nutrients and waste and is, thus, crucial for cartilage health (Zhang et al., 2009). Hence, Popov (2019) proposed a growth law for cartilage based on the changing rate of the hydrostatic pressure gradient. Here, following Darcy's law for fluid flow in porous media, the hydrostatic pressure gradient is considered a measure for fluid flow.

The brute force method to obtain the stresses in the half space is to use superpositions of the point force solutions by Boussinesq (1885) and Cerruti (1882). However, the resulting numerical integrations are very expensive in terms of computational complexity. On a three-dimensional $n \times n \times n$ grid, the complexity of this method is ($\mathcal{O}n^5$). This is inconvenient, whenever the stresses must be computed repeatedly due to changing indenter shapes (in case of wear or growth) or due to variations in loading. Of course, more elaborate methods with lower complexity exist. An example with the complexity ($\mathcal{O}n^3 \log n$) is the FFT-based boundary element method (BEM) for the half-space (Pohrt and Li, 2014).

In the present paper, we employ the superposition idea first described by Mossakowski (1963) to derive expressions in form of one-dimensional integrals for the whole stress tensor in the half-space. Starting point for the frictionless normal contact problem is the complete analytical solution for the Hertzian contact by Huber (1904). For the tangential contact problem, we operate in the Cattaneo-Mindlin approximation and base our derivation on the Hertzian solution by Hamilton (1983). The resulting expressions may be used to numerically or analytically determine stresses and hydrostatic pressure gradients for arbitrary axisymmetric indenters and arbitrary loading histories. Employing the same method, Willert et al. (2020) derived simple analytic expressions for the stresses and the hydrostatic pressure gradient, but only for contact plane. In terms of complexity, we show that the proposed method can be advantageous, even compared to elaborate methods such as the FFT-based BEM.

The paper is organized as follows: sections Frictionless Normal Contact and Tangential Contact in the Cattaneo-Mindlin Approximation are concerned with the normal and tangential contact problem, respectively. In both sections, we first obtain the respective solution for the cylindrical flat punch from the Hertzian solution and then derive expression for arbitrary indenter shapes by superposition of flat punch solutions. In case of the tangential contact, we also discuss the states of partial and full slip and arbitrary loading histories. In the end of both sections, the von Mises stresses and hydrostatic pressure gradients are shown for some examples. The paper closes with a discussion and some conclusive remarks.

FRICITIONLESS NORMAL CONTACT

In this section, we derive expressions for the stress state and the pressure gradient in the half-space for the axisymmetric Boussinesq problem. As suggested by Mossakowski (1963), the solution for arbitrary axisymmetric profiles can be obtained by superimposing incremental flat punch solutions.

Conversely, this superposition idea can also be employed to obtain the flat punch solution from any known axisymmetric solution. Thus, we first derive the flat punch solution inside the half-space using the well-known Hertzian solution provided by Huber (1904) and later use this solution to derive expressions for arbitrary axisymmetric profiles.

Hence, the stress components for the indentation by a rigid cylindrical flat punch (superscript FP) with the radius a can be derived from the Hertzian solution (superscript H) using

$$\sigma_{ij}^{\text{FP}} = \delta \frac{\partial \sigma_{ij}^{\text{H}}}{\partial \delta} = \delta \frac{\partial \sigma_{ij}^{\text{H}}}{\partial a} \frac{da}{d\delta} = \delta \frac{\partial \sigma_{ij}^{\text{H}}}{\partial a} \frac{R}{2a}, \quad (1)$$

with the indentation depth δ and $\delta^{\text{H}} = a^2 / R$ in the Hertzian case. The obtained flat punch solution is presented in **Appendix A**.

Now, we consider the case of general axisymmetric profiles. Under the assumption of a convex smooth profile $f = f(r)$, the relation between penetration depth and contact radius is a unique function

$$\delta = g(a). \quad (2)$$

In the framework of the MDR (Popov and Heß, 2015), it is shown that this function can be derived from the profile $f(r)$ by the simple Abel integral transform,

$$g(x) = |x| \int_0^{|x|} \frac{f'(r)}{\sqrt{x^2 - r^2}} dr. \quad (3)$$

By again employing the superposition idea described above, we obtain expressions for the general axisymmetric profile (superscript AS) *via*

$$\sigma_{ij}^{\text{AS}} = \int d\sigma_{ij}^{\text{AS}} = \int_0^a \frac{\partial \sigma_{ij}^{\text{AS}}}{\partial \delta} \frac{d\delta}{d\tilde{a}} d\tilde{a} = \int_0^a \frac{\partial \sigma_{ij}^{\text{FP}}}{\partial \delta} \frac{d\delta}{d\tilde{a}} d\tilde{a}. \quad (4)$$

Thus, we interpret the indentation process as a series of incremental indentations of flat punches with increasing contact radii. The obtained expressions read

$$\begin{aligned} \sigma_{zz}^{\text{AS}}(r, z; a) &= -\frac{E^*}{\pi} \int_0^a \left(\frac{z}{\sqrt{u}} \right)^3 \left(3 - \frac{\tilde{a}u_a}{2u} - 2 \frac{\tilde{a}^2 z^2 + \tilde{a}u u_a}{u^2 + \tilde{a}^2 z^2} \right) \frac{\tilde{a}u g'(\tilde{a})}{u^2 + \tilde{a}^2 z^2} d\tilde{a}, \\ \sigma_{rr}^{\text{AS}}(r, z; a) &= \frac{1-2\nu}{2\pi r^2} F_N(a) - \sigma_{zz}^{\text{AS}}(r, z; a) + \frac{E^*}{\pi} \int_0^a \left\{ (1-2\nu) \frac{\tilde{a}^2}{r^2} \left(\frac{z}{\sqrt{u}} \right)^3 \left(\frac{\tilde{a}u_a}{2u} - 1 \right) \right. \\ &\quad \left. + \frac{z}{2\sqrt{u}} \left[\frac{2\tilde{a}u_a}{u} - 4 + (1-\nu) \frac{(\tilde{a}^2 - u)(\tilde{a}u_a - 2u)}{(u + \tilde{a}^2)^2} \right] \right\} \frac{g'(\tilde{a})}{\tilde{a}} d\tilde{a}, \\ \sigma_{\varphi\varphi}^{\text{AS}}(r, z; a) &= -\frac{E^*(1+\nu)}{\pi} \int_0^a \frac{\tilde{a}z}{u} \frac{2u - \tilde{a}u_a}{\tilde{a}^2 + u} \frac{g'(\tilde{a})}{\sqrt{u}} d\tilde{a} - \sigma_{rr}^{\text{AS}}(r, z; a) - \sigma_{zz}^{\text{AS}}(r, z; a), \\ \sigma_{rz}^{\text{AS}}(r, z; a) &= -\frac{E^*}{\pi} \frac{r}{z} \int_0^a \left(\frac{z}{\sqrt{u}} \right)^3 \frac{u}{u + \tilde{a}^2} \left(3 + \frac{\tilde{a}u_a}{2u} - 2 \frac{\tilde{a}^2 z^2 + \tilde{a}u u_a}{u^2 + \tilde{a}^2 z^2} \right. \\ &\quad \left. - \frac{2\tilde{a}^2 + \tilde{a}u_a}{u + \tilde{a}^2} \right) \frac{\tilde{a}u g'(\tilde{a})}{u^2 + \tilde{a}^2 z^2} d\tilde{a}, \end{aligned} \quad (5)$$

where u as well as its derivative u_a are shortcuts for

$$u = \frac{1}{2} (A + S), \quad u_a = \frac{\partial u}{\partial \tilde{a}} = -\tilde{a} \left(1 + \frac{A - 2z^2}{S} \right), \quad \text{with}$$

$$A = r^2 + z^2 - \tilde{a}^2 \text{ and } S = \sqrt{A^2 + 4\tilde{a}^2 z^2}. \quad (6)$$

In Equation (5), F_N is the total normal force,

$$F_N(a) = 2E^* \int_0^a (\delta - g(\tilde{a})) d\tilde{a}, \quad (7)$$

$E^* = E/(1 - \nu^2)$ is the effective Young's modulus and ν is the Poisson's ratio. In the contact plane ($z = 0$) and in the axis of contact ($r = 0$), Equation (5) simplify significantly. The resulting expressions are presented in **Appendices B** and **C**, respectively. Using the same procedure as described above, the expressions of the stress components at the surface were already presented by Willert et al. (2020).

It is elemental that the superposition idea also works with any linear operation on the stress components. As an example, the hydrostatic pressure $p_{hs} = (\sigma_{rr} + \sigma_{\varphi\varphi} + \sigma_{zz})/3$ and the components of the gradient of the hydrostatic pressure are obtained as well. Again, we first calculate the flat punch solutions from the Hertzian solution (presented in **Appendix A**) and then obtain expressions for the axisymmetric case:

$$\begin{aligned} p_{hs}^{AS}(r, z; a) &= -\frac{E^*(1 + \nu)}{3\pi} \int_0^a \frac{\tilde{a}z}{u} \frac{2u - \tilde{a}u_a}{\tilde{a}^2 + u} \frac{g'(\tilde{a})}{\sqrt{u}} d\tilde{a}, \\ \frac{dp_{hs}^{AS}}{dr}(r, z; a) &= \frac{E^*(1 + \nu)}{3\pi} \int_0^a \frac{z}{\sqrt{u}} \frac{\tilde{a}^2}{u + \tilde{a}^2} \left(u_{ar} + 3\frac{u_r}{\tilde{a}} \right. \\ &\quad \left. - \frac{3}{2} \frac{u_a u_r}{u} - \frac{2\tilde{a}u_r + u_a u_r}{u + \tilde{a}^2} \right) \frac{g'(\tilde{a})}{u} d\tilde{a}, \\ \frac{dp_{hs}^{AS}}{dz}(r, z; a) &= \frac{E^*(1 + \nu)}{3\pi} \int_0^a \frac{\tilde{a}}{\sqrt{u}} \frac{\tilde{a}^2}{u + \tilde{a}^2} \left[u_{az} \left(\frac{z}{\tilde{a}} + \frac{\tilde{a}z}{u} \right) - 2 \right. \\ &\quad \left. + \frac{u_a}{\tilde{a}} \left(1 + \frac{\tilde{a}^2}{u} \right) - \frac{2u}{\tilde{a}^2} + \frac{u_z z}{2u} \right. \\ &\quad \left. \left(2 + \frac{6u - 5\tilde{a}u_a}{\tilde{a}^2} - \frac{3\tilde{a}u_a}{u} \right) \right] \frac{g'(\tilde{a})}{u + \tilde{a}^2} d\tilde{a}, \quad (8) \end{aligned}$$

with the derivatives

$$\begin{aligned} u_r(r, z; \tilde{a}) &= \frac{\partial u}{\partial r} = r \left(1 + \frac{A}{S} \right), \\ u_z(r, z; \tilde{a}) &= \frac{\partial u}{\partial z} = z \left(1 + \frac{A + 2\tilde{a}^2}{S} \right), \\ u_{ar}(r, z; \tilde{a}) &= \frac{\partial^2 u}{\partial \tilde{a} \partial r} = -\frac{4\tilde{a}r z^2 (A + 2\tilde{a}^2)}{S^3}, \\ u_{az}(r, z; \tilde{a}) &= \frac{\partial^2 u}{\partial \tilde{a} \partial z} = \frac{4\tilde{a}r^2 z A}{S^3}. \quad (9) \end{aligned}$$

The corresponding expressions at the contact plane and at the axis of contact ($r = 0$) are, again, presented in **Appendices B** and **C**, respectively.

Examples for the Frictionless Normal Contact

The two already discussed indenter shapes, the paraboloid (Hertzian contact) and the cylindrical flat punch, as well as two more imperfect shapes, the paraboloid with a parabolic cap and the cylindrical flat punch with rounded edges, serve as examples for the frictionless normal contact in the following. The profiles and the relation between penetration depth δ and contact radius a are listed in Popov et al. (2019):

a) Paraboloid with radius of curvature R (see **Figure 1A**):

$$f(r) = \frac{r^2}{2R} \quad \text{and} \quad \delta = g(a) = \frac{a^2}{R}. \quad (10)$$

b) Cylindrical flat punch (see **Figure 1B**):

$$f(r) = \begin{cases} 0, & r \leq a \\ \infty, & r > a \end{cases}. \quad (11)$$

c) Paraboloid with a parabolic cap with radius of the cap b and radii of curvature R_1 and R_2 (see **Figure 1C**):

$$\begin{aligned} f(r) &= \begin{cases} \frac{r^2}{2R_1}, & r \leq b \\ \frac{r^2 - h^2}{2R_2}, & r > b \end{cases} \quad \text{and} \\ \delta = g(a) &= \begin{cases} \frac{a^2}{R_1}, & a \leq b \\ \frac{a^2}{R_1} + \frac{a^2}{R^*} \sqrt{a^2 - b^2}, & a > b, \end{cases} \quad (12) \end{aligned}$$

where $h^2 = b^2 \left(1 - \frac{R^2}{R_1} \right)$ and $R^* = \frac{R_1 R_2}{R_1 - R_2}$.

d) Cylindrical flat punch with rounded edges with radius of the blunt end b and radius of curvature R (see **Figure 1D**):

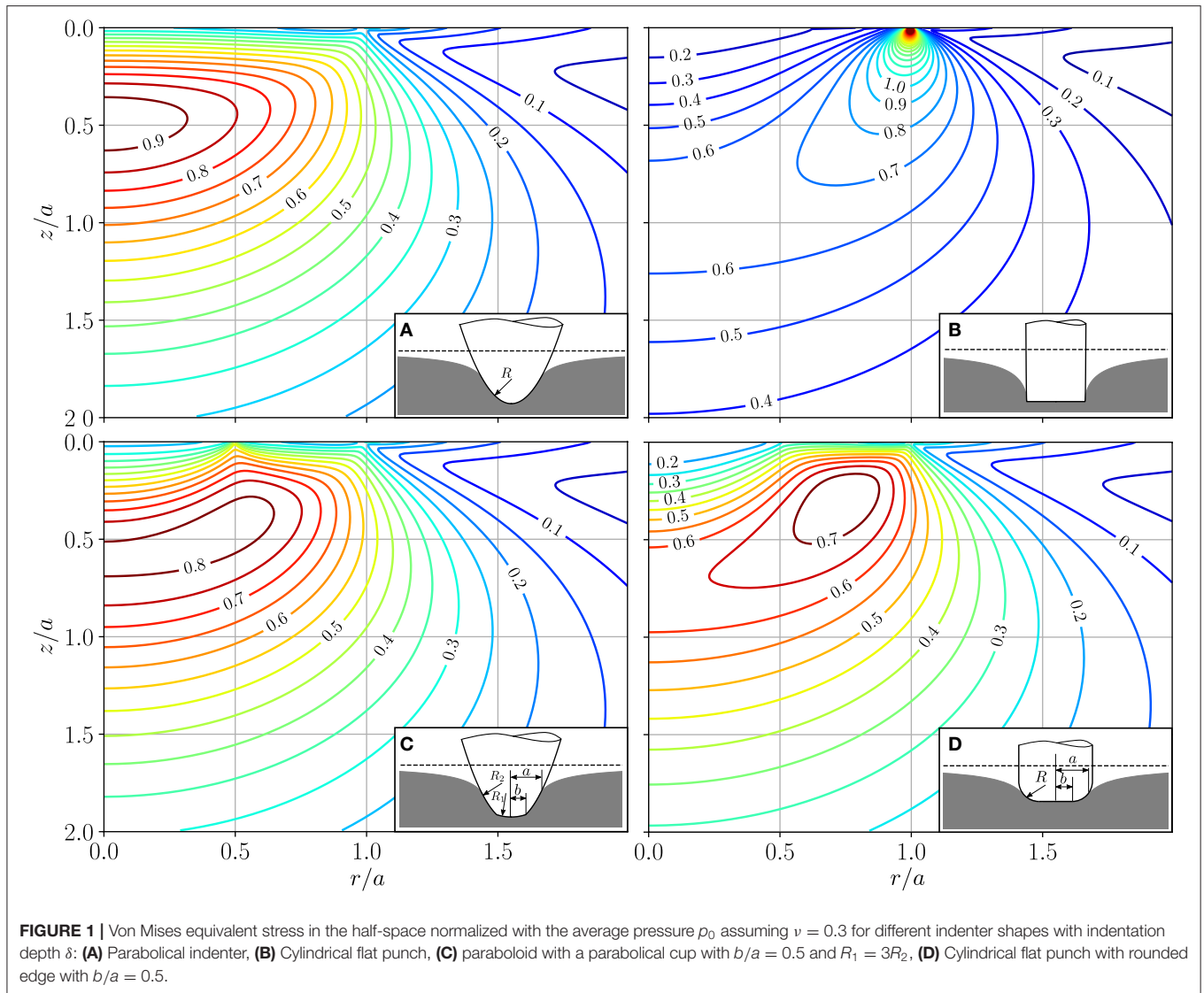
$$\begin{aligned} f(r) &= \begin{cases} 0, & r \leq b \\ \frac{(r-b)^2}{2R}, & r > b \end{cases} \quad \text{and} \\ \delta = g(a) &= \begin{cases} 0, & a \leq b \\ \frac{a}{R} \left[\sqrt{a^2 - b^2} - \text{barccos} \left(\frac{b}{a} \right) \right], & a > b \end{cases} \quad (13) \end{aligned}$$

With these relations, the integrals in Equations (5) and (8) and are evaluated numerically, with exception of the flat punch where stress components and pressure gradients are explicitly given in Equations (26) and (28) of **Appendix A**.

As a common measure for plastic failure, we calculate the von Mises equivalent stress

$$\sigma_{\text{mises}} = \sqrt{\frac{1}{2} \left\{ (\sigma_{rr} - \sigma_{\varphi\varphi})^2 + (\sigma_{\varphi\varphi} - \sigma_{zz})^2 + (\sigma_{zz} - \sigma_{rr})^2 + 6\sigma_{rz}^2 \right\}}. \quad (14)$$

The von Mises stress normalized with the average pressure in contact is plotted in **Figure 1**. Of course, the well-known distribution for the paraboloid with the maximum at $z/a = 0.5$ is simply the reproduction of the solution by Huber (1904). For the paraboloid with a parabolic cap with $b/a = 0.5$ and $R_1 = 3R_2$,



the maximum is widened and is located below the surface as well. Due to the small kink at $b/a = 0.5$ where the curvature changes, the distribution beneath is distorted toward the surface. The flat punch produces a stress singularity at the surface owing to the sharp edge. The lowest maximum with a normalized von Mises stress of 0.74 is obtained for the Cylindrical flat punch with rounded edges. It occurs under the rounded outer area of the indenter at a depth of $z/a \approx 0.3$.

In **Figure 2**, the absolute value of the hydrostatic pressure gradient,

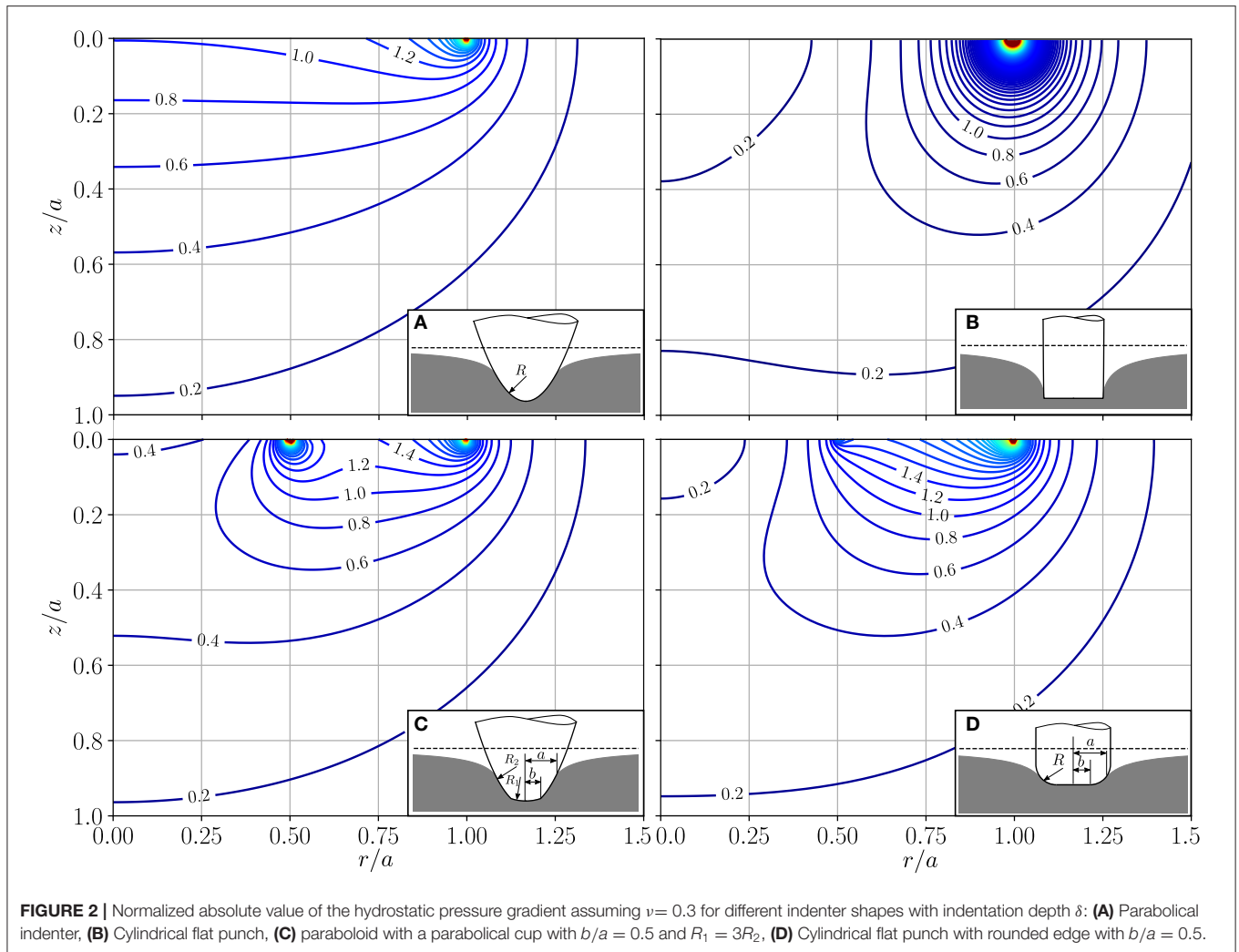
$$|\vec{\nabla} p| = \sqrt{\left(\frac{dp^{AS}}{dr}\right)^2 + \left(\frac{dp^{AS}}{dz}\right)^2}, \quad (15)$$

is plotted. It is normalized by the pressure gradient in the center of the contact plane produced by the equivalent

parabolical indenter with the same contact radius and indentation depth,

$$p'_0 = \frac{2E^*(1+\nu)\delta}{3a^2}. \quad (16)$$

Unlike the von Mises stress, the pressure gradient reaches its maximum at the surface in all cases. Maxima in form of singularities are found at the contact edge and at places of sudden geometrical change. Thus, the paraboloid with a parabolic cup and the flat punch with rounded edges exhibit additional singularities at $r = b$. Furthermore, the comparison of the pressure gradient under flat and curved section shows that, apart from the above-mentioned singularities, it is much larger for small radii of curvature.



TANGENTIAL CONTACT IN THE CATTANEO-MINDLIN APPROXIMATION

In this section, we consider the stress state and pressure gradient in the half-space due to tangential surface loads in the form

$$\sigma_{yz} = 0, \quad \sigma_{xz} \neq 0, \quad r < a. \quad (17)$$

In the Cattaneo-Mindlin approximation, the tangential contact can be reduced to superpositions of contact pressure distributions resulting from the frictionless normal indentation problem (Jäger, 1995). Thus, we can use the same procedure as for the frictionless normal contact. Further, we will discuss the cases of partial slip and complete slip, that is, if the stick condition $|\sigma_{xz}(r)| < \mu p(r)$ is violated.

For the Hertzian contact, the tangential surface loading of a globally sliding indenter is

$$\sigma_{xz}^H(r \leq a; a) = \mu \frac{2E^*}{\pi R} \sqrt{a^2 - r^2}, \quad (18)$$

with the coefficient of friction μ . The resulting stresses were given by Hamilton and Goodman (1966) as imaginary parts of a

complex function and, later, more conveniently, in explicit form by Hamilton (1983) (note, that in the equation for σ_{xx} it should be $x^2 z^2 / S$ instead of $x^2 z^2 / 3$).

With these expression, we, again, obtain the flat punch solutions *via* Equation (1). The surface loading for a sliding flat punch, for example, is

$$\sigma_{xz}^{\text{FP}}(r \leq a; a) = \frac{\mu E^* \delta}{\sqrt{a^2 - r^2}}. \quad (19)$$

A flat punch can either completely slip or completely stick (Popov et al., 2019). The solutions for the case of complete stick, are given by the substitution

$$\mu E^* \delta \rightarrow G^* u_{x,0}, \quad (20)$$

in the solutions for the sliding flat punch with the effective shear modulus $G^* = 4G / (2 - \nu)$ and the tangential rigid body displacement $u_{x,0}$. For the sake of brevity, the complete set of equations for the tangentially loaded flat punch is not explicitly presented here. It is more convenient to compute and store

the derivatives in Equation (1) using a symbolic engine such as *Wolfram Mathematica*.

For the axisymmetric sliding contact with a convex smooth profile, we, again, determine the stresses from a superposition of incremental sliding flat punch indentations with increasing radii. Hence, the stresses in the half-space, resulting only from the tangential surface loading, are given by Equation (4) and the computed flat punch solutions. It should be noted that for the full stress state, the stresses (5) caused by normal loading must be added.

For monotonic tangential loading, under constant normal force, partial slip will occur for convex profiles. Starting with the tangential loading, slip will propagate from the edge of contact. Following Jäger (1995, 1998) and Ciavarella (1998), the surface stresses σ_{xz} in the partial slip case can be written as a superposition

$$\sigma_{xz}^{AS}(r; a) = -\mu \left[\sigma_{zz}^{AS}(r; a) - \sigma_{zz}^{AS}(r; c) \right], \quad (21)$$

with the surface stress due to normal loading σ_{zz}^{AS} given in Equation of **Appendix B** and the radius c of the stick area, giving

$$\sigma_{xz}^{AS}(r; a) = \mu \frac{E^*}{\pi} \int_{\max(r, c)}^a \frac{g'(\tilde{a}) d\tilde{a}}{\sqrt{\tilde{a}^2 - r^2}}. \quad (22)$$

The radius of the stick area is determined by

$$G^* u_{x,0} = \mu E^* [\delta - g(c)] \quad \text{or} \quad \frac{F_x}{\mu F_N} = \frac{F_N(a) - F_N(c)}{F_N(a)}, \quad (23)$$

Popov et al. (2019), with F_N given in Equation (7). Hence, the stresses in the half-space for arbitrary profiles resulting from the tangential loading are given by a superposition of incremental sliding flat punch contacts with radii increasing from the stick radius to the contact radius,

$$\sigma_{ij}^{AS}(r, z; a) = \int_c^a \frac{\partial \sigma_{ij}^{FP}}{\partial \delta} \frac{dg}{d\tilde{a}} d\tilde{a}, \quad (24)$$

with ij in $\{xx, yy, zz, xy, yz, zx\}$ and the solutions for a sliding flat punch σ_{ij}^{FP} .

In the case of arbitrary loading histories, the solution is simply a finite number of superpositions in the form of Equation (21). Most conveniently, this can be modeled as one-dimensional tangential spring deflections $u_{x,1D}$ in the framework of MDR (see (Popov et al., 2019) for details), yielding

$$\sigma_{ij}^{AS}(r, z; a) = -\frac{G^*}{\mu E^*} \int_0^a \frac{\partial \sigma_{ij}^{FP}}{\partial \delta} \frac{d}{d\tilde{a}} [u_{x,1D}(\tilde{a})] d\tilde{a}. \quad (25)$$

Obviously, as for the normal contact, the integral Equations (24) and (25) also hold for the hydrostatic pressure and the components of the hydrostatic pressure gradient, if the stress components of the sliding flat punch σ_{ij}^{FP} are substituted with hydrostatic pressure p_{hs} and the components of the pressure gradient, respectively.

Examples for Monotonic Tangential Loading

As examples for the tangential contact, we discuss the parabolical indenter and the cylindrical flat punch with rounded edges (see section Examples for the Frictionless Normal Contact for details) under monotonic tangential loading. For a coefficient of friction of $\mu = 0.3$ and tangential rigid body displacement of $\delta/4$ and $\delta/2$, respectively, we obtain the cases of partial slip and complete slip for both indenter shapes.

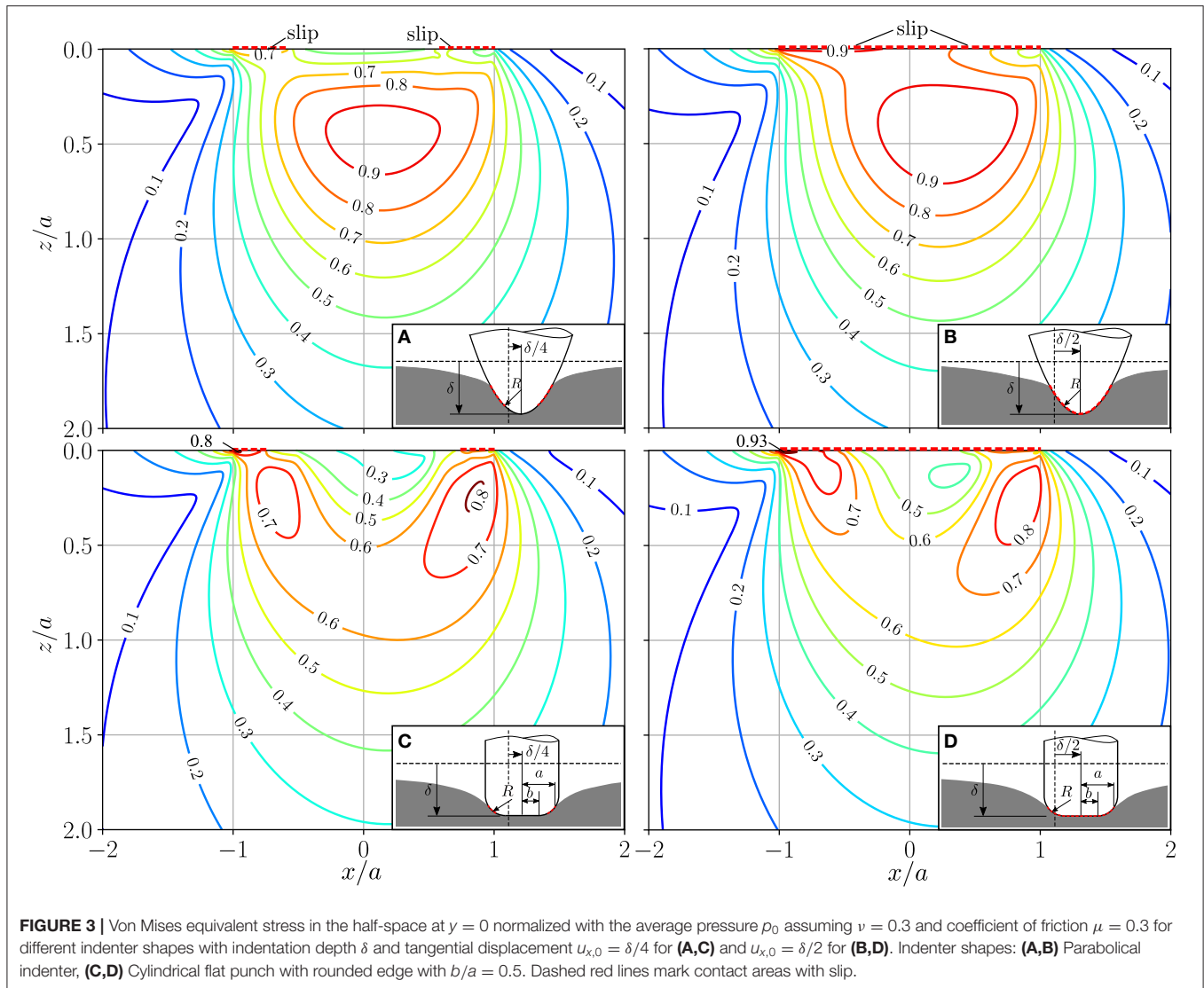
Figure 3 shows the von Mises equivalent stress for both indenters and loading scenarios in the x - z -plane. In all cases, the tangential surface loading leads to an additional local maximum at the trailing edge in the contact plane and the maximum beneath the surface is shifted toward the leading edge. For the parabolical indenter in the state of partial slip, the global maximum is beneath the surface as for the frictionless normal contact. In case of a complete slip, however, both maxima are of similar magnitude. This corresponds with the statement of Johnson (1985), that the maximum travels to the contact plane for $\mu > 0.3$ in the case of a sliding Hertzian contact. For the flat punch with rounded edges in the partial slip scenario, both maxima are of similar magnitude, while a distinct global maximum at the trailing edge in the contact plane is found in the case of complete slip.

Figure 4 shows the absolute value of the hydrostatic pressure gradient for the same indenter shapes and loading scenarios. In the case of complete slip, the gradient is increased at the leading edge and decreased at the trailing edge compared to the frictionless normal contact. However, the contribution of the tangential contact is, at least for $\mu = 0.3$, small in comparison to the normal contact. In the case of partial slip, the discontinuity due to the beginning slip area yields an additional minimum at the leading edge and an additional maximum at the trailing edge.

DISCUSSION

The evaluation of the integral kernels of the one-dimensional integrals (5) or (6) comes, of course, with some computational cost. It is thus advantageous to evaluate the analytic expressions $\partial \sigma_{ij}^{FP} / \partial \delta$ numerically on a grid over the half-space domain of interest and save the resulting matrices. These matrices can then be used to calculate the stress state for arbitrary axisymmetric indenters and loading histories.

For a grid of $n \times n$ points in the x - z plane, the one-dimensional integrals can be reduced to n matrix-vector multiplications of complexity $\mathcal{O}(nm)$, where m is the number of flat punch superpositions. Thus, the overall complexity simply scales with the number of points (in this case n^2) and flat punch superpositions, $\mathcal{O}(n^2 m)$. As it is not necessary to compute the stresses in more points than needed, the proposed formulation can be advantageous compared to other effective methods such as the FFT-based boundary element method (BEM) for the half-space (Pohrt and Li, 2014). For this example, the FFT-based BEM is of complexity $\mathcal{O}(n^3 \log n)$ because it requires evaluation of stresses on a three-dimensional $n \times n \times n$ grid. However, if stresses in $n \times n$ points in a x - y plane beneath the surface are sought for, the proposed

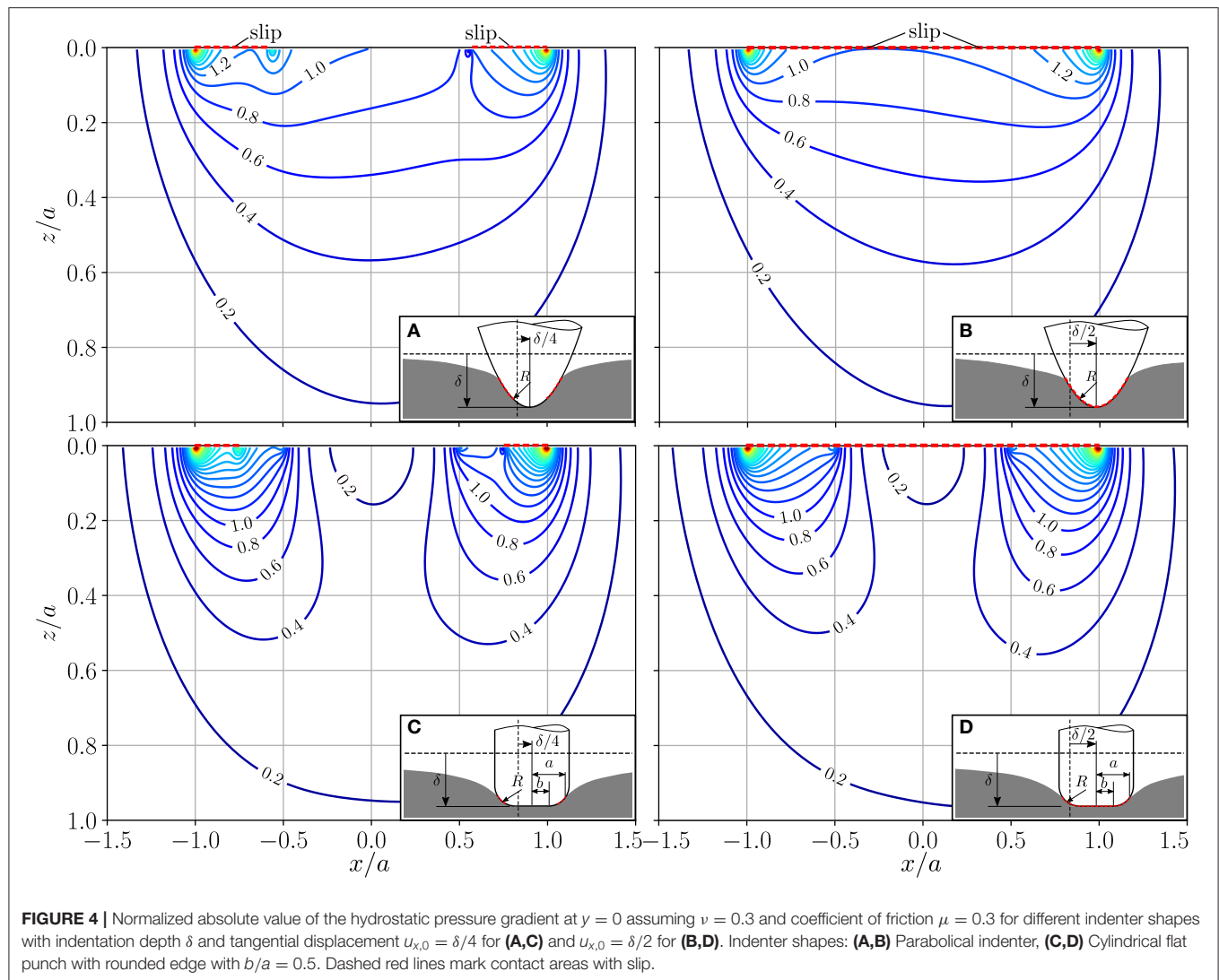


method still has $\mathcal{O}(n^2 m)$, whereas the FFT-based BEM is only of complexity $\mathcal{O}(n^2 \log \tilde{n})$.

We used stresses derived with the superposition idea to calculate the von Mises equivalent stresses for different indenter shapes and loadings. The examples showed that the endangered regions with respect to plastic failure often lie beneath the surface. Further, we provided expressions for direct calculation of the components of the hydrostatic pressure gradient. In fluid saturated media like cartilage, pressure gradients promote fluid flow which can be linked to tissue growth (Popov, 2019). The examples show that the absolute values of the pressure gradient are higher under small radii of curvature of the indenting body and singular at the surface, where discontinuities such as contact radius, stick radius, or sudden topography changes of the indenting body are located. It should be noted that the superposition idea is not bound

to homogenous half-spaces, but can, for example, also be used for layered, graded media like cartilage (Argatov et al., 2018).

The provided expressions for the stress state and pressure gradient in the half-space beneath an axisymmetric indenter in terms of the relation of indentation depth to contact radius $\delta = g(a)$ can, thus, be seen as a useful addition to the framework of the Method of Dimensionality Reduction (MDR) (Popov and Heß, 2015). Here, the same function $g(x)$ is interpreted as a one-dimensional profile pressed into a spring bedding. The MDR allows for rapid numerical simulations of profile changes due to wear or growth (in biological joints). However, the presented method of obtaining the full stress tensor can also be generalized with regard to other contact solutions for arbitrary loading histories such as the Method of Memory Diagrams (MMD) (Aleshin et al., 2015).



CONCLUSIONS

By exploiting the superposition idea by Mossakovski (1963) and Jäger (1998), we derived expressions in form of one-dimensional integrals for the full stress tensor and the components of the hydrostatic pressure gradient in the half-space beneath normally and tangentially loaded axisymmetric contacts. These expressions allow for efficient pointwise numerical evaluation for arbitrary indenter shapes and are suitable for contact simulations involving shape change due to wear or growth. With respect to plastic failure, we showed the importance of considering the whole half-space for selected indenter shapes and loadings. Further, we calculated hydrostatic pressure gradient distributions and discussed the relevance with respect to growth in biological contacts.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I am very grateful to Valentin Popov and Emanuel Willert for valuable discussions on the topic.

REFERENCES

- Aleshin, V., Bou Matar, O., and Van Den Abeele, K. (2015). Method of memory diagrams for mechanical frictional contacts subject to arbitrary 2D loading. *Int. J. Sol. Struct.* 60, 84–95. doi: 10.1016/j.ijsolstr.2015.02.016
- Argatov, I., Heß, M., and Popov, V. L. (2018). The extension of the method of dimensionality reduction to layered elastic media. *J. Appl. Math. Mech. ZAMM* 98, 622–634. doi: 10.1002/zamm.201700213
- Boussinesq, J. (1885). *Applications des potentiels à l'étude de l'équilibre et du mouvement des solides élastiques*. Paris: Gauthiers-Villars (in French).
- Cerruti, V. (1882). Ricerche intorno all' equilibrio de'corpi elastici isotropi. *Memorie Dell'accademia Nazionale Dei Lincei* 13, 81–122 (in Italian).
- Ciavarella, M. (1998). Tangential loading of general three-dimensional contacts. *J. Appl. Mech.* 65, 998–1003. doi: 10.1115/1.2791944
- Ciavarella, M., Hills, D. A., and Monno, G. (1998). The influence of rounded edges on indentation by a flat punch. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* 212, 319–327. doi: 10.1243/0954406981521259
- Hamilton, G. M. (1983). Explicit equations for the stresses beneath a sliding spherical contact. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* 197, 53–59. doi: 10.1243/PIME_PROC_1983_197_076_02
- Hamilton, G. M., and Goodman, L. E. (1966). The stress field created by a circular sliding contact. *J. Appl. Mech.* 33, 371–376. doi: 10.1115/1.3625051
- Huber, M. T. (1904). Zur theorie der berührung fester elastischer körper. *Ann. Phys.* 14, 153–163. doi: 10.1002/andp.19043190611
- Jäger, J. (1995). Axi-symmetric bodies of equal material in contact under torsion or shift. *Arch. Appl. Mech.* 65, 478–487. doi: 10.1007/BF00835661
- Jäger, J. (1998). A new principle in contact mechanics. *J. Tribol.* 120, 677–684. doi: 10.1115/1.2833765
- Johnson, K. L. (1985). *Contact Mechanics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139171731
- Mossakowski, V. I. (1963). Compression of elastic bodies under conditions of (axisymmetric case). *J. Appl. Math. Mech. PMM* 27, 630–643. doi: 10.1016/0021-8928(63)90150-3
- Pohrt, R., and Li, Q. (2014). Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17, 334–340. doi: 10.1134/S1029959914040109
- Popov, V. L. (2019). Active bio contact mechanics: concepts of active control of wear and growth of the cartilage in natural joints. *AIP Conf. Proc.* 2167:020285. doi: 10.1063/1.5132152
- Popov, V. L., and Heß, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin: Springer Verlag. doi: 10.1007/978-3-642-53876-6
- Popov, V. L., Heß, M., and Willert, E. (2019). *Handbook of Contact Mechanics: Exact Solutions of Axisymmetric Contact Problems*. Berlin: Springer Verlag. doi: 10.1007/978-3-662-58709-6
- Willert, E., Forsbach, F., and Popov, V. L. (2020). Stress tensor and gradient of hydrostatic pressure in the contact plane of axisymmetric bodies under normal and tangential loading. *ZAMM*. doi: 10.20944/preprints202003.0157.v1
- Wong, M., and Carter, D. R. (2003). Articular cartilage functional histomorphology and mechanobiology: a research perspective. *Bone* 33, 1–13. doi: 10.1016/S8756-3282(03)00083-8
- Zhang, L., Hu, J., and Athanasiou, K. A. (2009). The role of tissue engineering in articular cartilage repair and regeneration. *Crit. Rev. Biomed. Eng.* 37, 1–57. doi: 10.1615/critrevbiomedeng.v37.i1-2.10

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Forsbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Appendix A: Flat Punch Solution for the Frictionless Normal Contact Problem

The stress components of the cylindrical flat punch are obtained by Equation (1) with the Hertzian solution by Huber (1904):

$$\begin{aligned}\sigma_{zz}^{\text{FP}} &= -\frac{E^*\delta}{\pi a} \left(\frac{z}{\sqrt{u}} \right)^3 \frac{a^2 u}{u^2 + a^2 z^2} \left(3 - \frac{au_a}{2u} - 2 \frac{a^2 z^2 + auu_a}{u^2 + a^2 z^2} \right), \\ \sigma_{rr}^{\text{FP}} &= \frac{E^*\delta}{\pi a} \left\{ (1-2\nu) \frac{a^2}{r^2} \left[1 - \left(\frac{z}{\sqrt{u}} \right)^3 \left(1 - \frac{au_a}{2u} \right) \right] \right. \\ &\quad + \left(\frac{z}{\sqrt{u}} \right)^3 \frac{a^2 u}{u^2 + a^2 z^2} \left(3 - \frac{au_a}{2u} - 2 \frac{a^2 z^2 + auu_a}{u^2 + a^2 z^2} \right) \\ &\quad + \frac{z}{2\sqrt{u}} \left[\frac{2au_a}{u} - 4 + (1-\nu) \frac{(a^2 - u)(au_a - 2u)}{(u + a^2)^2} \right. \\ &\quad \left. \left. + (1+\nu) \frac{2u - au_a}{u + a^2} \right] \right\}, \\ \sigma_{\phi\phi}^{\text{FP}} &= -\frac{E^*\delta}{\pi a} \left\{ (1-2\nu) \frac{a^2}{r^2} \left[1 - \left(\frac{z}{\sqrt{u}} \right)^3 \left(1 - \frac{au_a}{2u} \right) \right] \right. \\ &\quad + \frac{z}{2\sqrt{u}} \left[\nu \left(4 - \frac{2au_a}{u} \right) + (1-\nu) \frac{(a^2 - u)(au_a - 2u)}{(u + a^2)^2} \right. \\ &\quad \left. \left. - (1+\nu) \frac{2u - au_a}{u + a^2} \right] \right\}, \\ \sigma_{rz}^{\text{FP}} &= -\frac{E^*\delta}{\pi a} \frac{r}{z} \left(\frac{z}{\sqrt{u}} \right)^3 \frac{a^2 u}{u^2 + a^2 z^2} \frac{u}{u + a^2} \\ &\quad \left(3 + \frac{au_a}{2u} - 2 \frac{a^2 z^2 + auu_a}{u^2 + a^2 z^2} - \frac{2a^2 + au_a}{u + a^2} \right),\end{aligned}\quad (26)$$

with the shortcuts

$$\begin{aligned}u(r, z; a) &= \frac{1}{2} (A + S), \\ u_a(r, z; a) &= \frac{\partial u}{\partial a} = -a \left(1 + \frac{A - 2z^2}{S} \right),\end{aligned}\quad (27)$$

where $A = r^2 + z^2 - a^2$ and $S = \sqrt{A^2 + 4a^2 z^2}$. With the components in Equation (26), the hydrostatic pressure $p_{hs} = (\sigma_{rr} + \sigma_{\phi\phi} + \sigma_{zz})/3$ and the components of the hydrostatic pressure gradient can be obtained as well:

$$\begin{aligned}p_{hs}^{\text{FP}} &= -\frac{E^*(1+\nu)\delta}{3\pi a} \frac{z}{\sqrt{u}} \frac{a^2}{u} \frac{2u - au_a}{a^2 + u}, \\ \frac{dp_{hs}^{\text{FP}}}{dr} &= \frac{E^*(1+\nu)\delta}{3\pi u} \frac{z}{\sqrt{u}} \frac{a^2}{u + a^2} \\ &\quad \left(u_{ar} + 3 \frac{u_r}{a} - \frac{3}{2} \frac{u_a u_r}{u} - \frac{2au_r + u_a u_r}{u + a^2} \right), \\ \frac{dp_{hs}^{\text{FP}}}{dz} &= \frac{E^*(1+\nu)\delta}{3\pi (u + a^2)} \frac{a}{\sqrt{u}} \frac{a^2}{u + a^2} \left[u_{az} \left(\frac{z}{a} + \frac{az}{u} \right) \right. \\ &\quad \left. - 2 + \frac{u_a}{a} \left(1 + \frac{a^2}{u} \right) - \frac{2u}{a^2} + \frac{u_z z}{2u} \right]\end{aligned}$$

$$\left(2 + \frac{6u - 5au_a}{a^2} - \frac{3a u_a}{u} \right) \Big], \quad (28)$$

with the derivatives

$$\begin{aligned}u_r(r, z; a) &= \frac{\partial u}{\partial r} = r \left(1 + \frac{A}{S} \right), \\ u_z(r, z; a) &= \frac{\partial u}{\partial z} = z \left(1 + \frac{A + 2a^2}{S} \right), \\ u_{ar}(r, z; a) &= \frac{\partial^2 u}{\partial a \partial r} = -\frac{4arz^2 (A + 2a^2)}{S^3}, \\ u_{az}(r, z; a) &= \frac{\partial^2 u}{\partial a \partial z} = \frac{4ar^2 z A}{S^3}.\end{aligned}\quad (29)$$

Appendix B: Axisymmetric Solution at the Surface for the Frictionless Normal Contact

At the surface ($z = 0$), the non-vanishing stress components in Equation (5) simplify to

$$\begin{aligned}\sigma_{zz}^{\text{AS}}(r; a) &= -\frac{E^*}{\pi} \int_r^a \frac{g'(\tilde{a}) d\tilde{a}}{\sqrt{\tilde{a}^2 - r^2}}, \\ \sigma_{rr}^{\text{AS}}(r; a) &= \sigma_{zz}^{\text{AS}}(r; a) + \frac{1-2\nu}{2\pi r^2} F_N(a) \\ &\quad - \frac{E^*(1-2\nu)}{\pi r^2} \int_r^a \frac{\sqrt{\tilde{a}^2 - r^2} g'(\tilde{a}) d\tilde{a}}{r}, \\ \sigma_{\phi\phi}^{\text{AS}}(r; a) &= (1+2\nu) \sigma_{zz}^{\text{AS}}(r; a) - \sigma_{rr}^{\text{AS}}(r; a).\end{aligned}\quad (30)$$

The hydrostatic pressure and the pressure gradient components in Equation (8) simplify to

$$\begin{aligned}p_{hs}^{\text{AS}}(r; a) &= \frac{2}{3} (1+\nu) \sigma_{zz}^{\text{AS}}(r; a), \\ \frac{dp_{hs}^{\text{AS}}}{dz}(r \leq a; a) &= \frac{2E^*(1+\nu)}{3\pi} \left\{ \frac{g'(0)}{r} + \int_0^r \frac{g''(\tilde{a})}{\sqrt{r^2 - \tilde{a}^2}} d\tilde{a} \right\}, \\ \frac{dp_{hs}^{\text{AS}}}{dz}(r > a; a) &= -\frac{2E^*(1+\nu)}{3\pi} \int_0^a \frac{\tilde{a} g'(\tilde{a})}{(r^2 - \tilde{a}^2)^{3/2}} d\tilde{a}, \\ \frac{dp_{hs}^{\text{AS}}}{dr}(r; a) &= \frac{2E^*(1+\nu)}{3\pi} \left\{ \frac{a g'(a)}{r\sqrt{a^2 - r^2}} - \int_r^a \frac{\tilde{a} g''(\tilde{a})}{r\sqrt{\tilde{a}^2 - r^2}} d\tilde{a} \right\},\end{aligned}\quad (31)$$

at the surface. All of the above equations except dp_{hs}^{AS}/dr were already derived by Willert et al. (2020).

Appendix C: Axisymmetric Solution in the Axis of Contact for the Frictionless Normal Contact

In the axis of contact ($r = 0$), the non-vanishing stress components in Equation (5) simplify to

$$\sigma_{zz}^{\text{AS}}(z; a) = -\frac{E^*}{\pi} \int_0^a \frac{\tilde{a}^3 + 3\tilde{a}z^2}{(\tilde{a}^2 + z^2)^2} g'(\tilde{a}) d\tilde{a},$$

$$\begin{aligned}\sigma_{rr}^{\text{AS}}(z; a) &= \sigma_{\varphi\varphi}^{\text{AS}}(z; a) \\ &= -\frac{E^*}{2\pi} \int_0^a \frac{(1+2\nu)\tilde{a}^3 - (1-2\nu)\tilde{a}z^2}{(\tilde{a}^2 + z^2)^2} g'(\tilde{a}) d\tilde{a}. \quad (32)\end{aligned}$$

The hydrostatic pressure and the non-vanishing pressure gradient component in Equation (5) simplify to

$$\begin{aligned}p_{hs}^{\text{AS}}(z; a) &= -\frac{2E^*(1+\nu)}{3\pi} \int_0^a \frac{\tilde{a} g'(\tilde{a})}{\tilde{a}^2 + z^2} d\tilde{a}, \\ \frac{dp_{hs}^{\text{AS}}}{dz}(z; a) &= \frac{4E^*(1+\nu)}{3\pi} \int_0^a \frac{\tilde{a} z g'(\tilde{a})}{(\tilde{a}^2 + z^2)^2} d\tilde{a}, \quad (33)\end{aligned}$$

in the axis of contact.



Ratio of Loss and Storage Moduli Determines Restitution Coefficient in Low-Velocity Viscoelastic Impacts

Emanuel Willert*

Department of System Dynamics and Friction Physics, Institute of Applied Mechanics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Irina Goryacheva,
Institute for Problems in Mechanics
(RAS), Russia

Reviewed by:

Ivan Argatov,
Technische Universität
Berlin, Germany
Andrey Dimaki,
Institute of Strength Physics and
Materials Science (ISPMS SB
RAS), Russia
Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine

*Correspondence:

Emanuel Willert
e.willert@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 04 November 2019

Accepted: 14 January 2020

Published: 25 February 2020

Citation:

Willert E (2020) Ratio of Loss and
Storage Moduli Determines Restitution
Coefficient in Low-Velocity
Viscoelastic Impacts.
Front. Mech. Eng. 6:3.
doi: 10.3389/fmech.2020.00003

Impact tests are an important tool to analyze dynamic material properties of viscoelastic media in technology and biology. In this context, rigorous contact mechanical models of the collision problem are necessary to adequately interpret data from impact experiments. It is shown here theoretically that the coefficient of restitution in these types of testing is mainly a function of one specific material property, namely, the ratio between the loss and storage moduli of the viscoelastic probe at the characteristic timescale of the impact. Explicit dependencies of the restitution coefficient on factors like impact velocity, impactor shape, general material rheology, and functional grading—beyond the fact that those may influence the impact duration and the dynamic modulus associated with it—are weak.

Keywords: impact testing, restitution, viscoelasticity, rheological models, graded materials

INTRODUCTION

Impact tests are considered as a fast and simple method to gain insights about the material behavior of viscoelastic media—for example, rubber or soft tissues like articular cartilage—under dynamic loading.

In the simplest case, only the rebound velocity (or some directly related measure, e.g., the rebound height) of a rigid impactor is determined. In technical contexts, this testing procedure is often referred to as the determination of the “rebound elasticity” of the material [see, e.g., the industrial standard DIN EN ISO, 8307 (2018)]. However, how the so-obtained coefficient of restitution (COR) generally relates to the viscoelastic rheology, or put bluntly, what kind of material property the “rebound elasticity” actually represents, remains unclear.

Biological tissues, especially in joints, are often subject to impact loading, for example, during sporting activities or accidents. Therefore, impact tests are also performed on these tissues to analyze material properties and to determine damage thresholds. For this purpose, Burgin and Aspdén (2007) constructed a drop tower, which was later developed further by Kang et al. (2017). In this regard, Burgin et al. (2014) pointed out that the material behavior of articular cartilage under impact loading is quite different from the one under slow loading conditions. Özcan et al. (2011) used the frequency response function in impact tests with a hammer to characterize frequency-dependent material properties of the human liver.

Contact mechanics models of viscoelastic impacts have a long history. In solving the impact problem, these models, obviously, heavily rely on the solutions of the contact problem. For linear materials, the latter ones are obtained based on the correspondence between linearly elastic and linearly viscoelastic boundary value problems established by Lee (1955). Although the correspondence principle strictly only holds, if the regions of boundary themselves do not change

in time (this is generally not the case for dynamic contact problems, when the contact area is not known *a priori*), Lee and Radok (1960) were able to show that the correspondence solution of the viscoelastic Hertzian problem is consistent with all boundary conditions, if the contact radius is a non-decreasing function of time. If the contact radius possesses a single maximum (this is the case for the impact problem), the solution needs to be adjusted, as was acknowledged by Lee and Radok in their paper. The corrected contact solution for the restitution phase (when the contact radius is decreasing in time) has been presented by Hunter (1960) (for the Hertzian case) and Graham (1965) and Ting (1966) (for arbitrary convex, axisymmetric profile geometries).

Hunter also applied his contact solution to analyze the respective contact-impact problem. Bassi (1978) was the first to examine the problem, to which extent impact characteristics (for the linear impact onto an incompressible Kelvin solid, see *Impact Onto an Incompressible Kelvin Solid*) allow for the determination of dynamic material properties. The linear impact problem for viscoelastic materials (i.e., the impact of a rigid flat cylindrical punch onto a viscoelastic medium) has been also studied by Butcher and Segalman (2000), Schwager and Pöschel (2007), and Argatov (2013). For the impact of a sphere onto a Kelvin solid (see *Impact of a Paraboloid*), Kuwabara and Kono (1987) proposed a model, later studied in detail by Ramírez et al. (1999), in which the simple form of the equation of motion during the compression phase (when the contact radius is increasing) is also used for the restitution. This, from a contact mechanics perspective, is slightly erroneous (as pointed out above) but allows for a mathematically simple treatment of the impact. A contact mechanically more rigorous solution of this problem was given by Willert et al. (2017). Selyutina et al. (2015) and Springhetti and Selyutina (2018) theoretically studied the impact response of articular cartilage on the basis of a quasi-linear Kelvin material model. The impact problem for a quasi-linear standard solid has been analyzed by Argatov et al. (2016).

In the present manuscript, it is analyzed in detail whether the primary result of a viscoelastic impact test, the COR, can be directly related to a specific material property and how this relation is in turn affected by the shape of the impactor (flat punch or spherical), the general material rheology, and functional grading—as biomaterials usually are graded media. For this purpose, the low-velocity (i.e., quasi-static) central normal impact of a rigid body with mass m and initial velocity v_0 onto an incompressible viscoelastic half-space is studied. First, in *Impact Onto an Incompressible Kelvin Solid*, the simplest viscoelastic material, the Kelvin solid with complete decoupling of elastic and viscous properties, is considered, and the known respective impact solutions are briefly summed up. To capture the influence of material rheology, a more general material model is considered in *Influence of Material Rheology*, first in a rigorous way and after that in an approximate sense, which demonstrates that the COR indeed is bound to a specific material property. *Influence of Viscoelastic Grading* is devoted to the effect of graded viscoelasticity.

IMPACT ONTO AN INCOMPRESSIBLE KELVIN SOLID

The arguably minimal model of a viscoelastic material is the Kelvin solid with shear modulus G_∞ and shear viscosity η . The complex dynamic modulus in this case simply reads as

$$\hat{G}(\omega) = G_\infty + i\eta\omega, \quad (1)$$

with the angular frequency of oscillation ω and the imaginary unit i . In the following, the known solutions for the impact of a cylindrical flat punch with radius a , and a sphere (approximated in the vicinity of the contact as a paraboloid) with radius R onto such a material, occupying a half-space, are briefly recapitulated.

Impact of a Cylindrical Flat Punch

In case of the impact of a rigid, cylindrical punch, the impact solution can be looked up in Butcher and Segalman (2000), Schwager and Pöschel (2007), Argatov (2013), and Willert (2020). It only depends on a nondimensional damping parameter

$$2D := \frac{\eta\omega_0}{G_\infty} = \frac{\text{Im}\{\hat{G}(\omega_0)\}}{\text{Re}\{\hat{G}(\omega_0)\}}, \quad \omega_0 := \sqrt{\frac{8aG_\infty}{m}}. \quad (2)$$

Note that ω_0 is the characteristic angular frequency of the elastic problem. “Im” and “Re” denote the imaginary and real parts of a complex quantity, respectively. Hence, D is given by the ratio of the loss and storage moduli at the characteristic frequency scale of the impact problem.

The COR for weak damping is given by Argatov (2013).

$$\varepsilon = \exp\left[-\frac{2D}{\sqrt{1-D^2}} \arctan\left(\frac{\sqrt{1-D^2}}{D}\right)\right], \quad D < 1, \quad (3)$$

whereas for strong damping (Willert, 2020)

$$\varepsilon = \exp\left[-\frac{2D}{\sqrt{D^2-1}} \operatorname{artanh}\left(\frac{\sqrt{D^2-1}}{D}\right)\right], \quad D > 1. \quad (4)$$

Impact of a Paraboloid

For a spherical impactor, contact mechanically rigorous modeling is slightly more complicated than in the previous case, because the contact radius is not prescribed by the radius of the flat punch and therefore increasing and decreasing during the impact. Hence, compression and restitution must be considered separately. A self-consistent model of the impact is most conveniently implemented within the framework of the “method of dimensionality reduction” (MDR; Popov and Heß, 2015; Popov et al., 2018). In the model by Kuwabara and Kono (1987), the simple form of the equation of motion for compression is also used for the restitution phase. It has been shown recently by the author in the monograph (Willert, 2020) that both models differ only slightly from each other. Also, both agree well with experimental results by Van Zeebroeck et al. (2003) on impacts with rubber and various biomaterials.

The impact duration and thus the characteristic frequency for the elastic problem were given by Hertz in his classical paper (Hertz, 1882). Hence, the nondimensional damping parameter, in analogy to Equation, reads (omitting numerical factors of the order of one)

$$2D = \eta \left(\frac{Rv_0}{m^2 G_\infty^3} \right)^{1/5}. \quad (5)$$

Figure 1 shows the COR as a function of the damping parameter for the flat punch and the spherical impactor. The curves are practically the same. Hence, the explicit dependence of the COR on the impactor shape is very weak. However, the shape obviously influences the damping parameter and therefore, implicitly, the impact solution. Especially note that the damping parameter for the spherical impactor depends on the impact velocity, so the COR is velocity dependent. This is not the case for the linear contact with a cylindrical punch, where the characteristic frequency is strictly a system property.

INFLUENCE OF MATERIAL RHEOLOGY

The fact that model predictions based on the Kelvin solid agree well with experimental data raises an interesting question: If, obviously, no real viscoelastic material exhibits such a trivial rheology, as suggested by the Kelvin model, how is it possible for the latter to accurately predict real impact behavior? Although the answer is quite simple, to demonstrate it, we have to briefly look into more general rheological models.

Rheological Models in Linear Viscoelasticity

The rheological behavior of “real” linearly viscoelastic materials is often captured in terms of a Prony series. In case of a generalized

Maxwell model, the time-dependent shear modulus reads

$$G(t) = \tilde{G}_\infty + \sum_{i=1}^N G_i \exp\left(-\frac{t}{\tau_i}\right), \quad (6)$$

with several different relaxation times τ_i (usually arranged in a logarithmic scale) and associated moduli G_i . If the problem of interest itself has an inherent timescale τ^* (for impact problems, this is the impact duration), it seems reasonable to assume that all relaxation processes faster than τ^* run infinitely fast and all slower than τ^* run infinitely slow (note that relaxation times in a Prony series are usually listed in a logarithmic scale). As under this assumption a Maxwell element of the Prony series with very slow relaxation degenerates to a spring, and one with very fast relaxation degenerates to a dashpot, we arrive at the following rheological model (see **Figure 2**): a Maxwell element in parallel with a Kelvin element. The modulus and viscosity of the Kelvin element are given by summing up the fast and slow relaxation processes,

$$\begin{aligned} G_\infty &:= \tilde{G}_\infty + \sum_{\tau_i \gg \tau^*} G_i, \\ \eta_0 &:= \sum_{\tau_i \ll \tau^*} G_i \tau_i. \end{aligned} \quad (7)$$

Hence, the complex dynamic modulus is

$$\hat{G}(\omega) = G_\infty + G^* \frac{(\omega\tau^*)^2}{1 + (\omega\tau^*)^2} + i \left[\eta_0 \omega + G^* \frac{\omega\tau^*}{1 + (\omega\tau^*)^2} \right]. \quad (8)$$

Now, why is the Kelvin solid such a convenient model? Because for harmonic loading with angular frequency ω , owing to Equation (1), any linearly viscoelastic material can be interpreted as a Kelvin solid, with the elastic modulus and viscosity given by the storage and loss modulus at ω ! Loading under impact is not exactly harmonic but can be approximated well enough by

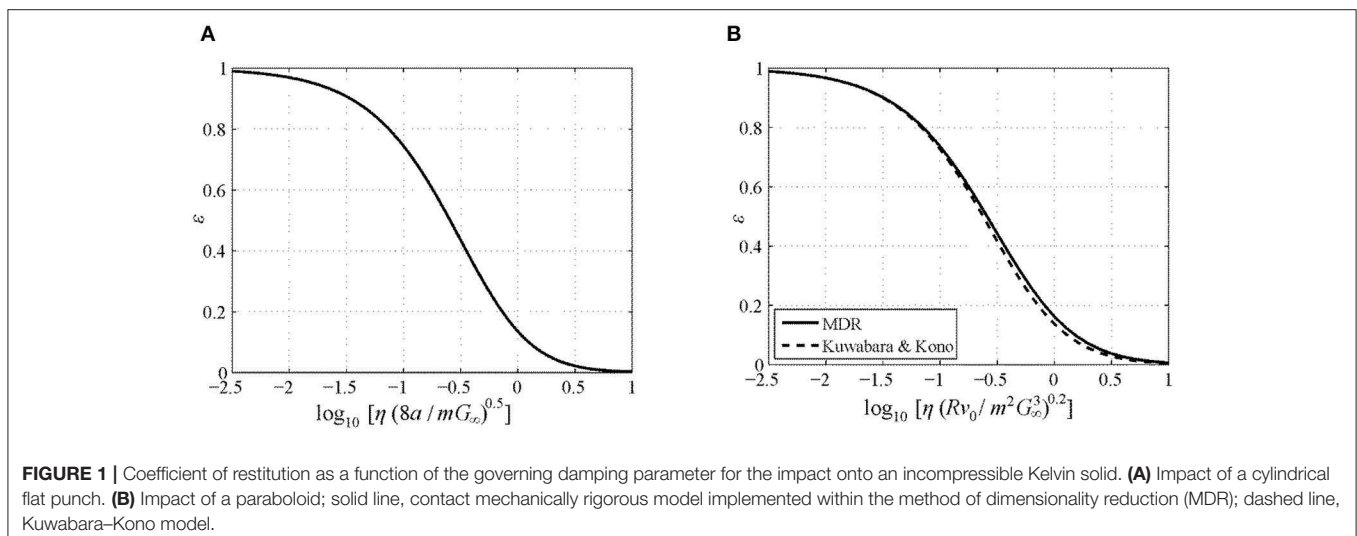
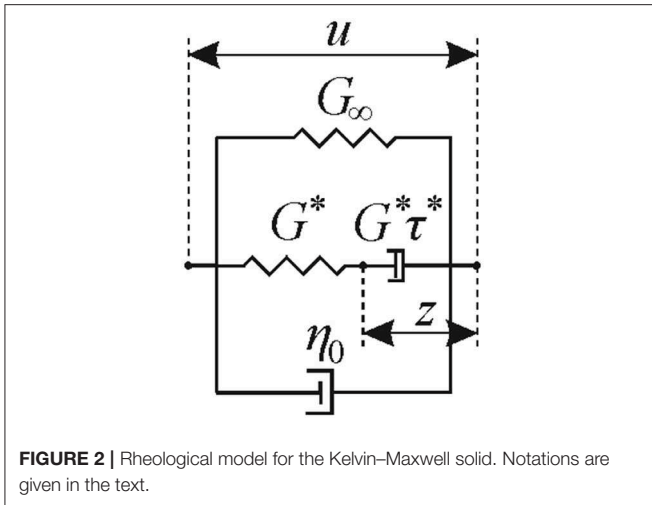


FIGURE 1 | Coefficient of restitution as a function of the governing damping parameter for the impact onto an incompressible Kelvin solid. **(A)** Impact of a cylindrical flat punch. **(B)** Impact of a paraboloid; solid line, contact mechanically rigorous model implemented within the method of dimensionality reduction (MDR); dashed line, Kuwabara-Kono model.



harmonic functions; see, for example, the work of Hunter (1957) on the Hertzian impact problem. Note that exactly sinusoidal indentation has been studied in detail by Argatov (2012) and that harmonic loading is also used in models of impact tests (Ruta and Szydło, 2005).

Solution of the Impact Problem for the Kelvin-Maxwell Solid

Let us analyze the consequences of this idea in case of the Kelvin-Maxwell solid introduced above. For simplicity, the impact of a rigid flat cylindrical punch with radius a and mass m is considered, as it has been shown before that the impactor shape is of secondary importance. The system of equations of motion for the outer and inner degrees of freedom of the rheological element shown in **Figure 2**, u and z , reads

$$\begin{aligned} 0 &= m\ddot{u} + 8\eta_0 a \dot{u} + 8G_\infty a u + 8G^* a (u - z), \\ 0 &= \tau^* \ddot{z} + \dot{z} - u. \end{aligned} \quad (9)$$

Written in proper nondimensional variables, the problem only depends on the following parameters:

$$\delta_0 := \eta_0 \sqrt{\frac{8a}{G_\infty m}}, \quad \delta^* := G^* \tau^* \sqrt{\frac{8a}{G_\infty m}}, \quad \gamma := \frac{G_\infty}{G^*}. \quad (10)$$

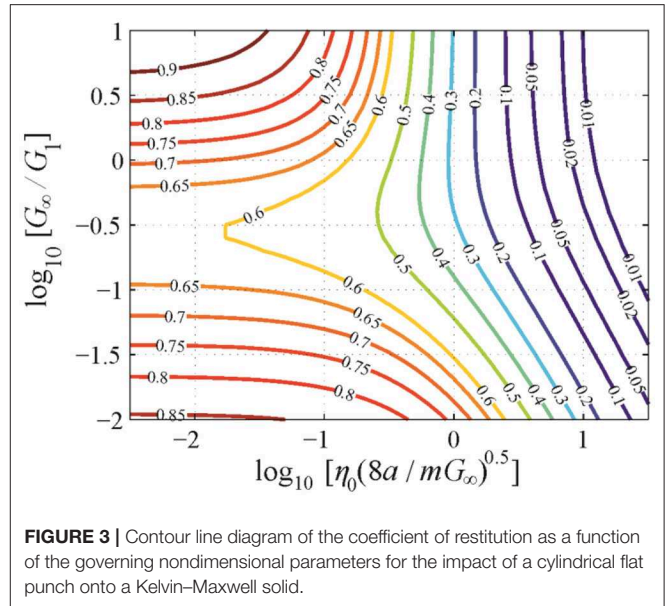
In **Figure 3** the COR—obtained by explicit solution of Equation —is shown in contour lines as a function of the governing parameters for

$$\gamma \delta^* \equiv 1 \quad \Leftrightarrow \quad \tau^* \equiv \sqrt{\frac{m}{8G_\infty a}} \quad (11)$$

in a logarithmic scale.

Approximate Solution via the Solution for the Kelvin Solid

These results can be reproduced in a very simple way, based on the idea laid out above: If, for harmonic loading, a viscoelastic



material can always be considered a Kelvin solid, and loading during impact is close to harmonic, it should be possible to obtain the impact solution for an arbitrary viscoelastic rheology based on the closed-form analytic solution for the Kelvin solid in Equations (3) and (4). For the Kelvin-Maxwell solid, Equation (2) suggests the following form for the damping parameter D :

$$2D = \frac{\text{Im} \left\{ \hat{G}(\omega_0) \right\}}{\text{Re} \left\{ \hat{G}(\omega_0) \right\}}, \quad \omega_0 \tau^* = 1, \quad (12)$$

which has been proposed recently by the author in the context of a three-element standard solid (Willert, 2020), that is, the ratio of loss and storage moduli at the characteristic frequency of the elastic problem. This form, however, only provides good results, if the impact duration does not deviate too strongly from the elastic case. A significant improvement can be achieved if the actual timescale of the viscoelastic problem is used, that is,

$$2D = \frac{\text{Im} \left\{ \hat{G}(\omega^*) \right\}}{\text{Re} \left\{ \hat{G}(\omega^*) \right\}}, \quad \omega^* = \frac{\pi}{T}, \quad (13)$$

where the impact duration T for the calculation has been taken from the exact solution of the problem (see *Solution of the Impact Problem for the Kelvin-Maxwell Solid*).

Figure 4 shows the coefficients of restitution obtained by this simple procedure—calculating the damping ratio according to Equation or and applying the analytical solution for the Kelvin solid—in the same fashion as in **Figure 3**. Usage of the modified damping parameter in Equation results in very good agreement with the exact solution. This strongly supports the idea that the general material behavior in viscoelastic impacts is of secondary importance; the main influencing factor is the ratio of loss

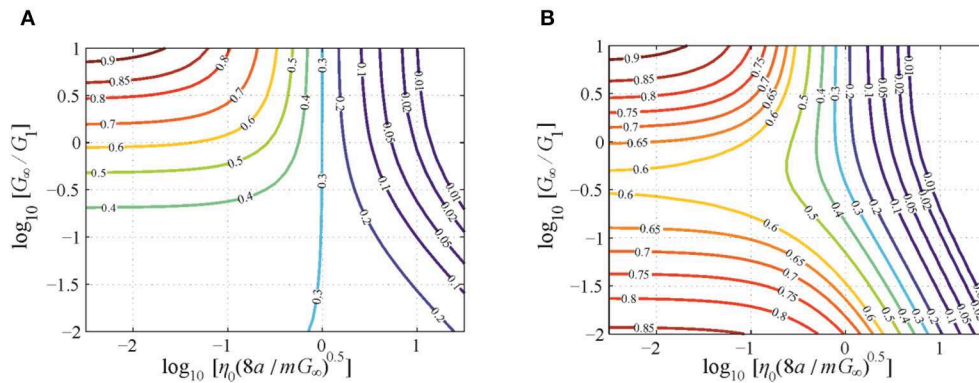


FIGURE 4 | Contour line diagrams of the coefficient of restitution as a function of the governing nondimensional parameters for the impact of a cylindrical flat punch onto a Kelvin–Maxwell solid; approximate solutions based on the solution for the Kelvin solid. **(A)** Based on Equation (12) for the damping parameter D . **(B)** Based on Equation (13) for the damping parameter.

and storage moduli at the timescale of the impact. Hence, this quantity can be determined in a stable way measuring only the COR.

INFLUENCE OF VISCOELASTIC GRADING

Besides viscoelasticity or poroelasticity, most biomaterials exhibit functional grading. Whether and how this influences the applicability of the idea described above shall be discussed in the following.

Correspondence Principle in Graded Viscoelasticity

All hitherto given viscoelastic contact (or contact-impact) solutions are based on the elastic-viscoelastic correspondence principle. Already Hashin (1965), as early as in 1965, considered the viscoelastic correspondence for heterogeneous (composite) materials. Only roughly 40 years later did Paulino and Jin (2001) and Mukherjee and Paulino (2003) show that the correspondence principle applies also to graded viscoelastic materials, if the spatial and temporal variations of the moduli are separable, that is, if every point exhibits the same stress relaxation behavior in time. Additionally, Jin (2006) pointed out that in case of two-phase graded composites, for the relaxation functions to be separable, the relaxation behavior in shear and dilatation must be the same.

Impact Onto a Power-Law Graded Kelvin Solid

So let us consider the impact of a rigid flat cylindrical punch onto a power-law graded Kelvin solid with the depth- and time-dependent shear and bulk moduli,

$$\begin{aligned} G(z, t) &= G_0 z^q [1 + \tau \delta(t)], \\ K(z, t) &= K_0 z^q [1 + \tau \delta(t)], \end{aligned} \quad (14)$$

with exponent q of the power-law grading, some characteristic time τ and the Dirac δ function. The static Poisson ratio is

$$\nu := \frac{3K_0 - 2G_0}{6K_0 + 2G_0} \approx 0.5, \quad (15)$$

because the material is considered to be incompressible. The elastic solution of the contact problem was given by Booker et al. (1985a),

$$F_{cl} = 2c_N(q) a^{1+q} d = k(q) d, \quad (16)$$

where c_N is a lengthy expression, which can be constructed from the fundamental solution in Booker et al. (1985b).

As the viscoelastic correspondence principle holds for this problem, the equation of motion reads

$$m\ddot{d} + k(q)\tau\dot{d} + k(q)d = 0, \quad (17)$$

and is thus formally the same as in the homogeneous case. Hence, also for graded viscoelastic materials, the COR is determined by the ratio of the loss and storage moduli at the timescale of the impact—at least, if the relaxation functions in shear and dilatation are separable and identical in form. Nonetheless, grading obviously affects the contact stiffness and therefore the impact duration.

DISCUSSION

As has been shown, the COR mainly depends on a ratio of moduli. To determine the values of the moduli themselves, one has to consider at least one dimensional property of the impact, for example, the impact duration or the maximum indentation depth. However, the relation between these quantities and the dynamic modulus seems to exhibit a slightly stronger explicit dependence on the material rheology than the restitution coefficient itself. For the flat cylindrical punch, a decent estimate for the absolute value of the dynamic modulus (showing a relative error of <10% for the Kelvin–Maxwell solid with $\gamma\delta^* \equiv 1$ and

not too small values of the restitution coefficient, $\varepsilon \geq 0.3$) is given by

$$\left| \hat{G}(\omega^*) \right| \approx \frac{\pi^2}{T^2} \frac{m}{8a}, \quad (18)$$

with the impact duration T .

All results presented are strictly valid only for incompressible media. Although many technological and biological materials can in good approximation be considered to be incompressible, there is, of course, no physical necessity to neglect dilatation. However, it has been shown recently by the author that the effect of compressibility in viscoelastic impacts is often very small (Willert, 2020). Moreover, the restriction to quasi-static processes, which requires the impact velocity to be much smaller than the smallest speed of wave propagation in the deformable material, poses a serious constraint of the calculations and considerations presented above, especially for very soft materials, as in this case wave propagation can be a significant source of energy dissipation (Hunter, 1957; Roylance, 1973).

To obtain a general linearly viscoelastic material model, which is nonetheless manageable in terms of the material parameters used, the analyzed Kelvin–Maxwell solid neglects the finite character of the relaxation processes that are either much faster or much slower than the impact itself. Thus, the results obtained can probably only serve as a “zero-order” approximation for “real” general linearly viscoelastic media (whose sets of material parameters, though, are usually unmanageable with respect to comprehensive analysis).

In sinusoidal indentation, the behavior is characterized by incomplete storage and loss moduli (Argatov, 2012). Their use may well be appropriate also for the analysis of the impact problem.

REFERENCES

- Argatov, I. I. (2012). Sinusoidally-driven flat-ended indentation of time-dependent materials: asymptotic models for low and high rate loading. *Mech. Mater.* 48, 56–70. doi: 10.1016/j.mechmat.2011.12.010
- Argatov, I. I. (2013). Mathematical modeling of linear viscoelastic impact: application to drop impact testing of articular cartilage. *Tribol. Int.* 63, 213–225. doi: 10.1016/j.triboint.2012.09.015
- Argatov, I. I., Selyutina, N. S., and Mishuris, G. S. (2016). Impact problem for the quasi-linear viscoelastic standard solid model. *J. Strain Anal. Eng. Design* 51, 294–303. doi: 10.1177/0309324715610027
- Bassi, A. C. (1978). Dynamic modulus of rubber by impact and rebound measurements. *Polymer Eng. Sci.* 18, 750–754. doi: 10.1002/pen.760181004
- Booker, J. R., Balaam, N. P., and Davis, E. H. (1985a). The behaviour of an elastic non-homogeneous half-space. Part II—circular and strip footings. *Int. J. Num. Anal. Methods Geomech.* 9, 369–381. doi: 10.1002/nag.1610090406
- Booker, J. R., Balaam, N. P., and Davis, E. H. (1985b). The behaviour of an elastic non-homogeneous half-space. Part I—line and point loads. *Int. J. Num. Anal. Methods Geomech.* 9, 353–367. doi: 10.1002/nag.1610090405
- Burgin, L. V., and Aspden, R. M. (2007). A drop tower for controlled impact testing of biological tissues. *Med. Eng. Phys.* 29, 525–530. doi: 10.1016/j.medengphy.2006.06.002
- Burgin, L. V., Edelsten, L., and Aspden, R. M. (2014). The mechanical and material properties of elderly human articular cartilage subject to impact and

CONCLUSIONS

The low-velocity impact of a rigid impactor onto a linearly viscoelastic flat has been studied within the framework of viscoelastic contact mechanics. It is found that the COR in this impact configuration is mainly a simple, unique, decreasing function of the ratio of the loss and storage moduli at the impact timescale. The explicit dependencies of the COR on the profile shape of the impactor, the general material rheology, and material grading (at least, if the relaxation behavior in shear and dilatation are the same)—beyond the fact that all these quantities may influence the impact timescale and the loss and storage moduli associated with it—are very weak. Therefore, as a simple and good approximation for the impact solution, the closed-form analytical solution for the impact of a flat cylindrical punch on a Kelvin solid can be used. Contact mechanically semi-rigorous approaches, like the Kuwabara–Kono model for parabolic impact, can also often be used. The obtained results will help to appropriately interpret impact test measurements on soft materials like rubber or articular cartilage.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

We acknowledge support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

- slow loading. *Med. Eng. Phys.* 36, 226–232. doi: 10.1016/j.medengphy.2013.11.002
- Butcher, E. A., and Segalman, D. J. (2000). Characterizing damping and restitution in compliant impacts via modified K-V and higher-order linear viscoelastic models. *J. Appl. Mech.* 67, 831–834. doi: 10.1115/1.1308578
- DIN EN ISO 8307 (2018). *Flexible Cellular Polymeric Materials - Determination of Resilience by Ball Rebound*.
- Graham, G. A. C. (1965). The contact problem in the linear theory of viscoelasticity. *Int. J. Eng. Sci.* 3, 27–46. doi: 10.1016/0020-7225(65)90018-2
- Hashin, Z. (1965). Viscoelastic behavior of heterogeneous media. *J. Appl. Mech.* 32, 630–636. doi: 10.1115/1.3627270
- Hertz, H. (1882). Über die Berührung fester elastischer Körper. *J. Reine Angew. Math.* 92, 156–171.
- Hunter, S. C. (1957). Energy absorbed by elastic waves during impact. *J. Mech. Phys. Solids* 5, 162–171. doi: 10.1016/0022-5096(57)90002-9
- Hunter, S. C. (1960). The Hertz problem for a rigid spherical indenter and a viscoelastic half-space. *J. Mech. Phys. Solids* 8, 219–234. doi: 10.1016/0022-5096(60)90028-4
- Jin, Z. H. (2006). Some notes on the linear viscoelasticity of functionally graded materials. *Math. Mech. Solids* 11, 216–224. doi: 10.1177/1081286504040401
- Kang, W., Chen, Y. C., Bagchi, A., and O’Shaughnessy, T. J. (2017). Characterization and detection of acceleration-induced cavitation in soft materials using a drop-tower-based integrated system. *Rev. Sci. Instruments* 88:125113. doi: 10.1063/1.5000512

- Kuwabara, G., and Kono, K. (1987). Restitution coefficient in a collision between two spheres. *Jap. J. Appl. Phys.* 26, 1230–1233. doi: 10.1143/JJAP.26.1230
- Lee, E. H. (1955). Stress analysis in visco-elastic bodies. *Q. Appl. Math.* 13, 183–190. doi: 10.1090/qam/69741
- Lee, E. H., and Radok, J. R. M. (1960). The contact problem for viscoelastic bodies. *J. Appl. Mech.* 27, 438–444. doi: 10.1115/1.3644020
- Mukherjee, S., and Paulino, G. H. (2003). The elastic-viscoelastic correspondence principle for functionally graded materials, revisited. *J. Appl. Mech.* 70, 359–363. doi: 10.1115/1.1533805
- Ozcan, M. U., Ocal, S., Basdogan, C., Dugusoy, G., and Tokat, Y. (2011). Characterization of frequency-dependent material properties of human liver and its pathologies using an impact hammer. *Med. Image Anal.* 15, 45–52. doi: 10.1016/j.media.2010.06.010
- Paulino, G. H., and Jin, Z. H. (2001). Correspondence principle in viscoelastic functionally graded materials. *J. Appl. Mech.* 68, 129–132. doi: 10.1115/1.1331286
- Popov, V. L., and Heß, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin; Heidelberg: Springer Verlag.
- Popov, V. L., Willert, E., and Heß, M. (2018). Method of dimensionality reduction in contact mechanics and friction: a user's handbook. iii. viscoelastic contacts. *Facta Univ. Ser. Mech. Eng.* 16, 99–113. doi: 10.22190/FUME180327013P
- Ramirez, R., Pöschel, T., Brilliantov, N. V., and Schwager, T. (1999). Coefficient of restitution of colliding viscoelastic spheres. *Phys. Rev. E* 60, 4465–4472. doi: 10.1103/PhysRevE.60.4465
- Roylance, D. (1973). Wave propagation in a viscoelastic fiber subjected to transverse impact. *J. Appl. Mech.* 40, 143–148. doi: 10.1115/1.3422914
- Ruta, P., and Szydló, A. (2005). Drop-weight test based identification of elastic half-space model parameters. *J. Sound Vibrat.* 282, 411–427. doi: 10.1016/j.jsv.2004.02.052
- Schwager, T., and Pöschel, T. (2007). Coefficient of restitution and linear-dashpot model revisited. *Granul. Matter* 9, 465–469. doi: 10.1007/s10035-007-0065-z
- Selyutina, N. S., Argatov, I. I., and Mishuris, G. S. (2015). On application of Fung's quasi-linear viscoelastic model to modeling of impact experiments for articular cartilage. *Mech. Res. Commun.* 67, 24–30. doi: 10.1016/j.mechrescom.2015.04.003
- Springhetti, R., and Selyutina, N. S. (2018). Viscoelastic modeling of articular cartilage under impact loading. *Meccanica* 53, 519–530. doi: 10.1007/s11012-017-0717-y
- Ting, T. C. T. (1966). The contact stresses between a rigid indenter and a viscoelastic half-space. *J. Appl. Mech.* 33, 845–854. doi: 10.1115/1.3625192
- Van Zeebroeck, M., Tijskens, E., VanLiedekerke, P., Deli, V., DeBaerdemaeker, J., Ramon, H., et al. (2003). Determination of the dynamical behaviour of biological materials during impact using a pendulum device. *J. Sound Vibrat.* 266, 465–480. doi: 10.1016/S0022-460X(03)00579-0
- Willert, E. (2020). *Stoßprobleme in Physik, Technik und Medizin: Grundlagen und Anwendungen*. Berlin: Springer Vieweg.
- Willert, E., Kusche, S., and Popov, V. L. (2017). The influence if viscoelasticity on velocity-dependent restitutions in the oblique impact of spheres. *Facta Univ. Ser. Mech. Eng.* 15, 269–284. doi: 10.22190/FUME170420006W

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Willert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



On Applications of Semi-Analytical Methods of Contact Mechanics

Vladislav V. Aleshin^{1,2*}

¹ Department of Tomsk, Univ. Lille, CNRS, Centrale Lille, ISEN, Univ. Valenciennes, UMR 8520—IEMN, LIA LICS/LEMAC, Lille, France, ² Scientific Department for Research in Materials and Technologies, Tomsk State University, Tomsk, Russia

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Ivan Argatov,
Technical University of
Berlin, Germany
Feodor M. Borodich,
Cardiff University, United Kingdom

*Correspondence:

Vladislav V. Aleshin
vladislav.aleshin@univ-lille.fr

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 09 March 2020

Accepted: 24 April 2020

Published: 05 June 2020

Citation:

Aleshin VV (2020) On Applications of
Semi-Analytical Methods of Contact
Mechanics. *Front. Mech. Eng.* 6:30.
doi: 10.3389/fmech.2020.00030

This paper is concerned with possible applications of semi-analytical methods of frictional contact mechanics. The semi-analytical solutions, such as the Method of Memory Diagrams, enable the load-displacement relationship for contact of two axisymmetric bodies with friction to be written as an analytical expression with parameters calculated via a numerical procedure. As a result, a complex history-dependent solution is obtained for an arbitrary loading history in a computationally efficient way. This fact allows one to calculate hysteretic responses to extremely complex loading histories, such as random vibrations. Another case when complex loading histories appear is in a harmonic excitation of a dynamic contact system in which inertia is taken into account. Both examples are considered here. The random excitation case can be used as a basis for modeling for wear in frictional contacts while the second one may be extended to describe coupled dynamic contact systems, stick-slip phenomena, or friction-induced instabilities.

Keywords: contact mechanics, Hertz-Mindlin, friction, semi-analytical contact solutions, method of memory diagrams

INTRODUCTION

This paper concerns the use of one semi-analytical method of frictional contact mechanics. The term “frictional contact mechanics” can comprise a large variety of problems. Here it is used in the following sense: (i) friction is a phenomenon that arises due to tangential interactions of bodies in contact, and (ii) frictional interaction is governed by the classical one-term law of friction (Coulomb friction law) originating in the works of Amontons. The second aspect is discussed in more detail from a historical point of view by Desplanques (2015) or Borodich and Savencu (2017). The first feature is essential as friction can also appear in a purely compressional loading case [see e.g., Borodich and Keer (2004)]. In any situation, friction results in a hysteretic response which depends not only on instantaneous values of drive parameters, but also on their history.

Further, methods belonging to the semi-analytical class (Dobry et al., 1991; Jäger, 2005; Aleshin et al., 2015, 2019; Popov and Heß, 2015; Popov et al., 2019) allow one to obtain a frictional contact response as an analytical solution that depends on parameters determined by an algorithm. For the price of accepting a certain number of simplifications (such as axisymmetric contact geometry, neglect for elastic dissimilarity of the bodies, and approximate fulfillment of the Coulomb friction law) extremely rapid and efficient calculation techniques are developed. The high computational performance makes these methods suitable for implementation in complex loading histories (e.g., random vibrations or acoustic waves).

In the fourth of the cited methods, the Method of Memory Diagrams (MMD), the contact characteristics such as local and global displacements, stress distributions, and contact loads are calculated via an internal memory fiction or diagram constantly updated following the excitation protocol. The memory function is defined on an adaptive grid whose points are dynamically adjusted in accordance to the applied excitation. In addition to the mechanical response, the friction-induced energy dissipation is analytically described. The MMD is sketched in section method of the memory diagrams of this paper preceded by section Hertz-Mindlin mechanics in which some basic solutions of contact mechanics are recalled.

In sections friction-induced energy dissipation in spherical contacts excited by random vibrations and section contact system with a finite mass dynamically excited by external forces, two examples of MMD applications are considered. The first example is concerned with a prestressed axisymmetric contact excited by random vibrations that represent random displacement time histories having fractal properties. In such a system, the mechanical and energetic responses depend on a very restrained number of parameters, such as the fractal dimension, the normal, and tangential displacement amplitudes normalized on the prestress value, and the ratio of the maximum and minimum frequencies of the excitation spectrum. In particular, it can be shown that at high amplitudes most of the energy is dissipated near the contact center while for low amplitudes the energy dissipation zone represents an annulus located inside of the mean contact circle. These results can be of use for determining parts of mechanical systems in which wear or thermal fatigue is most likely to appear.

The second example is a dynamic (i.e., having a mass and inertia) frictional contact system excited by a harmonic tangential force. Even the simplest case of a single contact with constant compression demonstrates rich dynamic effects. In particular, various time scales can appear under various combinations of a system's parameters. The responses can be categorized into several classes characterized by growing or complex oscillatory behavior.

In another case considered elsewhere (Aleshin et al., 2019), a solid material with a frictional crack is insonified by ultrasound. It can be shown that rough faces of cracks behave approximately as effective axisymmetric bodies (Jäger, 1995; Ciavarella, 1998) having the same normal reaction, thus suggesting the use of the MMD. The MMD-based contact model has been integrated with a finite-element environment (COMSOL). Doing so, a simulation tool called MMD-FEM has been elaborated for modeling acoustic responses of damaged materials, including nonlinear acoustic effects that are usually applied in modern nondestructive testing technologies.

HERTZ-MINDLIN MECHANICS

Similarly to other semi-analytical methods in contact mechanics, the MMD can be regarded as a direct generalization of the classical Cattaneo-Mindlin (Cattaneo, 1938; Mindlin and Deresiewicz, 1953) solution developed for elastic spheres in

contact loaded by a subsequent application of constant normal and tangential forces. As it was shown, the contact zone consists of stick and slip areas that represent a central circle, and outer annulus, respectively. In the stick zone, no relative tangential displacement between close points belonging to the opposite surfaces is possible. In the slip zone, the shear stress τ equals the normal stress σ times the friction coefficient μ , in accordance to the Coulomb friction law. At the same time, in that zone, the relative tangential displacement is a nonzero vector that must be directed as the local shear stress vector. The latter condition can be called the orientation aspect or property of the Coulomb friction law.

The most compact derivation (Jäger, 1995) is based on a superposition of the Bossiness solutions for rigid punches straining an elastic half-space in both normal and tangential senses. The smallest punch in the superposition coincides with the stick circle radius that guarantees the no-slip condition in the stick zone, and the largest one is the size of the contact zone itself. By a proper choice of “strengths” of the punches in the normal and tangential directions it is possible to satisfy the Coulomb condition $\tau = \mu \sigma$. However, the orientational property is satisfied only approximately. The issue is that for punches applied in the x -direction parallel to the half-space surface, the local vectors $\vec{\tau}$ are all directed along the same x -axis, while the tangential displacement vector has a non-zero in-plane y -component [Equation (28c) in Jäger (1995)]. Another simplification is related to the neglect of the second term in the second line of Equation (28b) in the cited paper. In addition, the Cattaneo-Mindlin approximation disregards dissimilarity phenomena (Munisamy et al., 1994) which, if present, can produce local tangential displacement for purely normal compression, since the Poisson effect can be of different magnitudes for non-equal spheres of different materials. However, despite some assumptions in the analysis, the Cattaneo-Mindlin solution remains a good approximation of frictional contact interaction of axisymmetric bodies largely used since 1950's. For equal bodies with the elastic constants E and ν having the contact geometry as in **Figure 1** (contact forces N and T , and displacements a and b , contact zone radius c , and stick zone radius s) the solution has the following form:

$$N = \frac{4E^*}{3R} c^3, \quad (1)$$

$$a = \frac{1}{R} c^2, \quad (2)$$

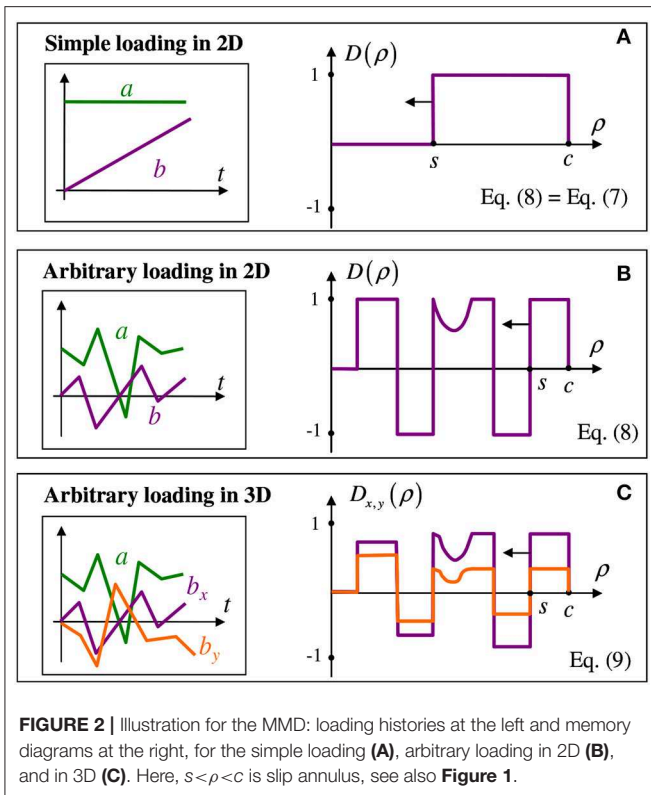
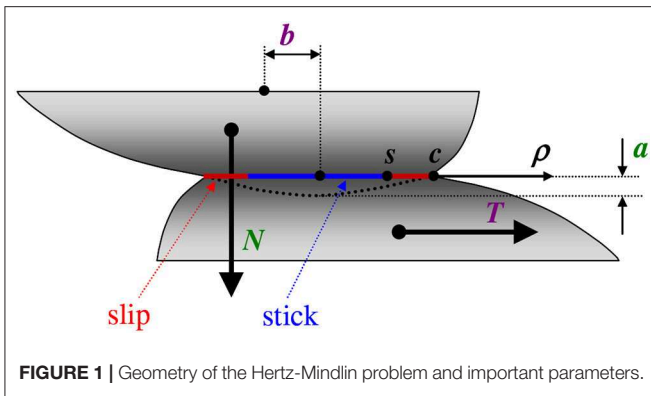
$$T = \frac{4\mu E^*}{3R} (c^3 - s^3), \quad (3)$$

$$b = \frac{\mu\theta}{R} (c^2 - s^2), \quad (4)$$

with E^* and θ defined as

$$E^* = \frac{E}{2(1-\nu^2)}, \quad (5)$$

$$\theta = \frac{2-\nu}{2(1-\nu)}. \quad (6)$$



In this solution, all geometric features of the contact system are taken into account through the dependences $N = N(c)$ and $a = a(c)$ in Equations (1, 2). The result can be rewritten using these functions as

$$\begin{cases} b = \theta \mu (a(c) - a(c=s)) \\ T = \mu (N(c) - N(c=s)) \end{cases} \quad (7)$$

where in the last terms of each equation the argument of functions $N(\cdot)$ and $a(\cdot)$ is the radius of the stick zone s . The result Equation (7) is frequently referred to as the reduced elastic friction principle (Jäger, 2005). The principle Equation (7), in contrast to Equations (1–4), is valid for any axisymmetric contact geometry, not necessarily spherical.

METHOD OF THE MEMORY DIAGRAMS

The MMD (Aleshin et al., 2015, 2019) develops the principle described in the previous section by applying it to more general loading histories which consist in arbitrarily changing oblique compressions in 2D or in 3D (the former means that the normal and tangential forces stay in one plane). The calculation is organized with the use of an auxiliary inter function $D(\rho)$, called a memory diagram, that encodes all memory information in the frictional system. In 2D, the solution reads

$$\begin{cases} b = \theta \mu \int_0^c D(\rho) \frac{da}{dc} \Big|_{c=\rho} d\rho \\ T = \mu \int_0^c D(\rho) \frac{dN}{dc} \Big|_{c=\rho} d\rho \end{cases} \quad (8)$$

In the previously considered “simple loading case” (Figure 2A) [i.e., when the tangential action is added after application of constant normal compression], the memory diagram has a simple rectangular shape that corresponds to the classical result Equation (7) after calculation of the integral in Equation (8).

An arbitrary loading history in 2D corresponds to a more complex shape of the memory diagram that can consist of positive and negative horizontal elements as well as from curvilinear sections (Figure 2B). The algorithm (Aleshin et al., 2015) keeps track of the evolution of the loading parameters and updates the diagram shape accordingly, in order to keep the balance equation (8). This formula does not require any additional assumptions in comparison to the reduced elastic friction principle Equation (7). Limitations related to this principle are discussed by Jäger (2005) and also mentioned in the paper (Aleshin et al., 2015) where the MMD is described in more detail.

If the loading parameters are allowed to arbitrarily vary in 3D, the system can be described via a vector counterpart of Equation (8) which reads,

$$\begin{cases} \vec{b} = \theta \mu \int_0^c \vec{D}(\rho) \frac{da}{dc} \Big|_{c=\rho} d\rho \\ \vec{T} = \mu \int_0^c \vec{D}(\rho) \frac{dN}{dc} \Big|_{c=\rho} d\rho \end{cases} \quad (9)$$

However, the analysis (Aleshin and Bou Matar, 2016) neglects the orientational aspect of the Coulomb friction law and therefore should be considered as an approximation.

Equations (8) or (9), together with the algorithm governing the memory diagram evolution, provide the possibility to calculate the hysteretic tangential load-displacement relationship through the known normal load-displacement relationship given by $N = N(c)$ and $a = a(c)$. The method works when the forces are considered as arguments and displacements are unknown or vice versa. At the same time, it is important to emphasize that the MMD introduced above is only valid for partial slip (i.e., when some stick zone remains around $\rho = 0$. If $|T|$ reaches μN or $|b|$ reaches $\theta \mu a$, the stick zone disappears). The force-driven system excited by a tangential force exceeding μN will experience accelerated movement which violates the current

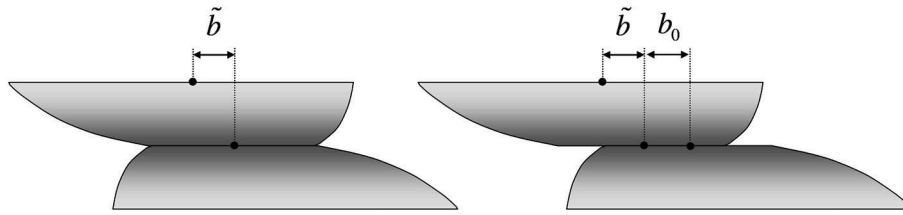


FIGURE 3 | Partial tangential displacements due to shearing of the bodies and due to the shift between contact centers.

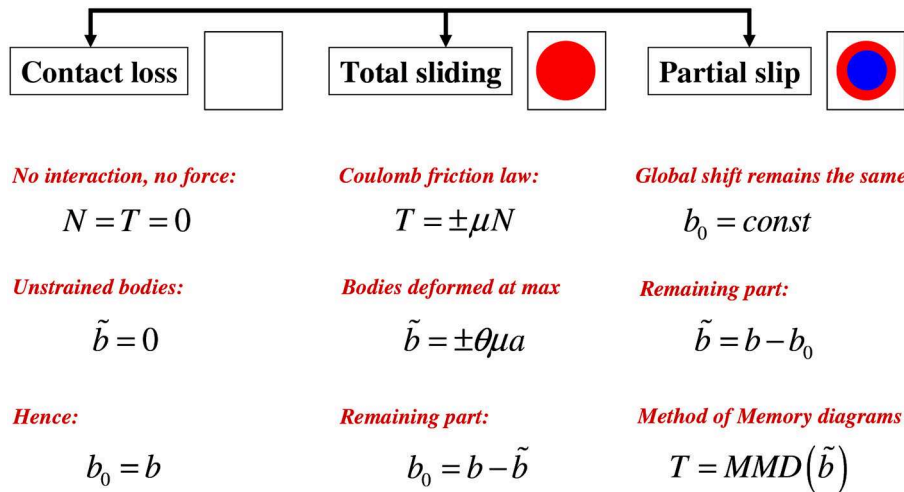


FIGURE 4 | Three contact states (contact loss, total sliding, and partial slip) and the corresponding solutions for T obtained via repartition $b = b_0 + \tilde{b}$ for loading in 2D (normal and tangential displacements always stay in one plane).

quasi-static character of description and generally complicates the problem. Fortunately, in the case where the system is driven by displacements, there exists a simple way to construct a quasi-static force-displacement relationship valid in all situations which may be encountered: partial slip, total sliding, and contact loss.

To do so, we introduce two displacement components of the total displacement

$$\vec{b} = \vec{b}_0 + \vec{\tilde{b}}. \quad (10)$$

as illustrated in **Figure 3**; $\vec{\tilde{b}}$ reflects deformation of one of the contacting bodies due to shearing while \vec{b}_0 is a tangential shift between the contact centers that develops due to total sliding. Since we already consider small displacements in comparison to all geometric features, the effects of the slight drop of the upper body because of the tangential mismatch or contact plane rotation are neglected.

For the 2D case, the algorithm that provides the unknown tangential force is shown in **Figure 4**. When the contact is lost, there is no contact interaction, and the bodies are unstrained (i.e., $N = T = 0$). When total sliding takes place, $T = \pm \mu N$ with the sign depending on the sliding direction. Finally, for partial slip the MMD algorithm has to be applied, which is symbolically expressed as $T = \text{MMD}(\vec{\tilde{b}})$. In each case, one of

the components, b_0 or \tilde{b} , is known directly, and the other one is immediately found since their sum equals the known argument. Numerically, the algorithm is applied to small increments Δb and Δa and updates previous values with small changes calculated at the current step which become previous values at the next step, etc.

In the 3D loading case, tangential displacement b , its components b_0 and \tilde{b} , and force T in **Figure 4** become vectors. In addition, the formulas for the total sliding case have to be further modified since sliding does not occur in the positive or negative direction as in 2D, but in a direction given by the unit vector $\vec{l} \uparrow \Delta \vec{b}_0$ where $\Delta \vec{b}_0$ is an infinitesimal slip vector. These vectors are also collinear with the tangential force, $\vec{l} \uparrow \vec{T}$, since slip is caused by \vec{T} (orientational aspect of the Coulomb friction law). From the previous considerations we also know that $|\vec{\tilde{b}}| = \theta \mu a$ [assume $s = 0$ in **Figure 2A** or in Equation (7)]. Then the repartition Equation (10) takes the form

$$\vec{b} = \vec{b}_{0p} + \Delta \vec{b}_0 + \vec{l} \theta \mu a \quad (11)$$

in which \vec{b}_{0p} is the known component \vec{b}_0 at the previous step, and the two last vectors are collinear. Finally, \vec{l} is obtained as a

unit vector collinear to

$$\vec{l} \uparrow \vec{b} - \vec{b}_{0p}, \quad (12)$$

and then $\vec{T} = \vec{l}\mu N$. The infinitesimal slip vector $\Delta\vec{b}_0$ becomes equal $\Delta\vec{b}_0 = \vec{b} - \vec{b}_{0p} - \vec{l}\theta\mu a$ which means that all components of the repartition Equation (11) at the current step are determined.

For brevity, the term MMD comprises the extension to the contact loss and total sliding cases (Figure 4), not only three partial slip situation in Equations (8) or (9). The formulation shown in Figures 2, 4 illustrated the efficiency of the method. Indeed, instead of considering detailed evolving distributions of local stresses and displacements, it is enough to introduce and update one inner memory function (two functions in 3D). Moreover, the function frequently contains constant segments thus the memorization of only the beginning and the end of each segment and not all intermediate points. The MMD algorithm is based on an adaptive grid whose points are created and deleted following the loading protocol instead of being predefined at fixed positions. As a result, the method is especially suitable to complex loading protocols such as random or acoustical excitation. At the same time, the contact geometry should remain relatively simple in order to be imitated by axisymmetric shapes.

In the next sections it is shown how the MMD solution to the contact problem can be used for an efficient description of frictional contacts excited by complex loading histories.

FRICTION-INDUCED ENERGY DISSIPATION IN SPHERICAL CONTACTS EXCITED BY RANDOM VIBRATIONS

The semi-analytic MMD formulation of the solution to the contact problem makes it possible to derive an expression (Truyaert et al., 2019) for the friction-induced energy dissipation in the incremental form valid for all three contact states, regardless the shape of the memory diagram:

$$\Delta W = 2\mu \left(|\Delta b| - \theta\mu\Delta a \right) \left[N(a) - N(q) + (q-a) \frac{dN}{da} \Big|_{a=q} \right]. \quad (13)$$

Here the normal load-displacement relationship $N(a)$ derived [e.g., in (Jäger, 1995)] for any axisymmetric contact geometry is used. Moreover, the same considerations (Truyaert et al., 2019) enable one to write the surface density of the energy loss defined by

$$\Delta W = \int_s^c \Delta \varpi(\rho) 2\pi\rho d\rho$$

in the form

$$\Delta \varpi(\rho) = 2\mu \left(|\Delta b| - \theta\mu\Delta a \right) \sigma(\rho) \left(1 - \frac{2}{\pi} \arcsin \left(\frac{s}{\rho} \right) \right), \quad (14)$$

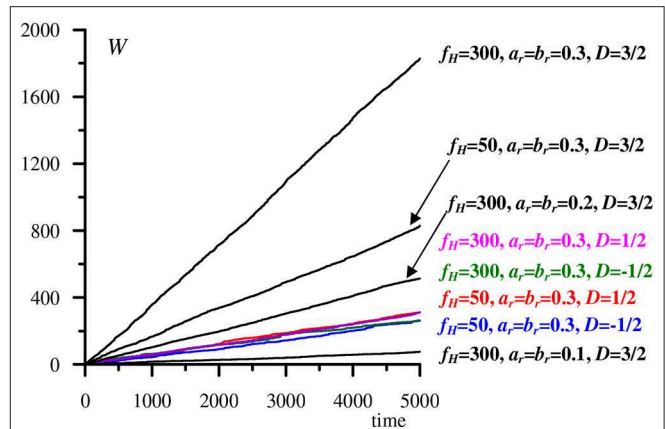


FIGURE 5 | Dissipated energy W as a function of time for different values of problems' parameters. The curves are approximately linear which means that the average dissipated power P is constant.

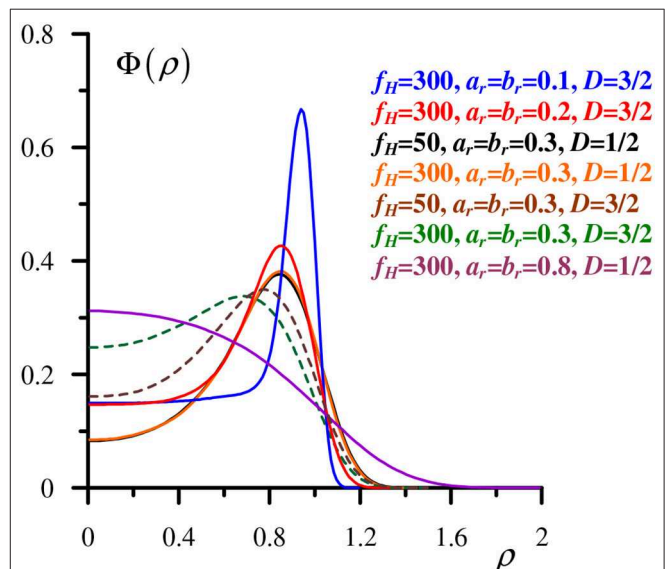


FIGURE 6 | Normalized form-factor of the surface density of the dissipated energy for different values of parameters.

in which the explicit knowledge of the normal stress distribution $\sigma(\rho)$ is required.

To reveal general tendencies in the frictional dissipation behavior, it is meaningful to consider a frictional system with a very restrained number of parameters. Aleshin and Papangelo (2020) suggested the use of a prestressed contact of two spheres excited by random normal and tangential displacements of equal rms having fractal time dependences in a certain frequency range. By the proper choice of normalization, it is possible to limit the number of parameters to three: rms amplitudes $a_r = b_r$ normalized on the prestress displacement, fractal dimension D , and higher cut-off frequency f_H normalized on the lower one. For

a fractal curve, the power spectral density is given by

$$S(f) \sim \begin{cases} f^{-5-2D} & \text{if } f_L \equiv 1 < f < f_H \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where D belongs to the interval $1 < D < 2$. It is straightforward to show (Aleshin and Papangelo, 2020) that extending the range for D to $0 < D < 1$ and $-1 < D < 0$ in Equation (15) will lead to fractal behavior for the velocity and acceleration time dependences. The fractal shape for $S(f)$ is selected since fractality of a random curve is related to (i) the Gaussian distribution of the random value which is in turn a consequence of a large number of factors that impact that value, as well as (ii) the power-law frequency dependence for $S(f)$ which means the absence of a characteristic frequency in the spectrum and reduces the number of free parameters by one. Certainly, in a real particular contact system, the statistical properties of the random excitation can differ.

As expected, average power P dissipated during a sizable interval of dimensionless time normalized on the inverse lower cut-off frequency is approximately constant (Figure 5). In our simulation, the number of time points equals $2^{19} = 524288$ and is high enough to neglect differences in two particular realizations which accumulate a lot of random features during a long observation time of 5,000 normalized units. The stiffness of curves in Figure 5 having the sense of the dissipated power depends on all parameters of the problem. Power P grows with increasing amplitudes $a_r = b_r$, with increasing frequency range f_H , and with increasing D . The last two effects are due to the fact that higher frequencies added by extending the range or amplifying existing HF components additionally generate loading-unloading cycles. Despite their small amplitude, they are frequent and therefore significantly contribute to the total energy loss.

In Figure 5 each curve is primarily characterized by a single parameter, its slope. More information is contained in the surface density of the dissipated power. To compare surface densities for different values of parameters, we calculate the form-factor $\Phi(\rho)$ introduced by the relationship

$$W = Pt \left\langle \int_s^c \Phi(\rho) 2\pi\rho d\rho \right\rangle \quad (16)$$

and normalized in order to have unit integral

$$\int_0^c \Phi(\rho) 2\pi\rho d\rho = 1$$

for each realization.

Figure 6 shows that form-factors can vary a lot depending on the problem's parameters. First of all, for small (such as 0.1 of the prestress displacement) amplitudes of vibrations, most of the energy is dissipated in a thin annulus located close to the average contact border. For moderate amplitudes, the annulus in which most of the dissipation takes place progresses inward and becomes smeared. Finally, for strong vibrations (of about of 0.8

and more), the maximum dissipation occurs in the contact center (i.e., the dissipation zone becomes circular, as the violet curve in Figure 6 indicates). Frequency content in the spectrum also influences the form-factor, but to a lesser degree. For instance, for low $D = 1/2$ most of the energy is contained in low frequencies, and increasing the upper frequency limit practically does not change anything (black and orange curves). At the same time, for high $D = 3/2$, extending the frequency range produces a minor effect (brown and green dashed curves). Generally, enhancing high frequency content shifts the annulus closer to the contact center, acting similarly to an amplitude increase, since higher frequencies make the total traveled path longer.

The fact that the form-factors determining the dissipated energy density portray the system's parameters in a finer way than the almost linearly growing total dissipation curves is additionally illustrated in Figure 7. By changing the frequency content of the vibrations spectrum with a simultaneous variation of the vibrations amplitudes, the slopes of two total dissipation curves can be matched. Indeed, in Figure 7A the red and blue curves are close, except that the blue one which represents the response on low-frequency vibrations can locally differ from the average inclination to a greater degree. At the same time, the surface densities of the dissipated energy are essentially distinct (Figure 7B). For small amplitudes, sharp peaks in surface densities curves are typically found. In practice, this means that a profiled joint subject to small but prolonged vibrations will experience wear in a thin annulus close to the contact border. For strong vibrations, wear should start near the contact center.

CONTACT SYSTEM WITH A FINITE MASS DYNAMICALLY EXCITED BY EXTERNAL FORCES

Another example of a problem that can be successfully solved with the use of the MMD, as well as by another method belonging to the semi-analytical class, is the dynamics of a simple system that consists of an axisymmetric body excited by a horizontal tangential force. In the considered case, the body is vertically prestressed on an elastic half-space and has certain mass m . Below it is shown that this geometrically and physically simple system has a rich dynamic behavior arising due to the presence of friction.

The equation of motion for such a body reads:

$$m \frac{d^2 b}{dt^2} = T_a \sin 2\pi f t - \mu N_0 T_{MMD} \left(\frac{b}{\theta \mu a_0} \right), \quad b(0) = \dot{b}(0) = 0, \quad (17)$$

where b is, as previously, the tangential displacement, T_a is the external tangential force amplitude, f is the harmonic frequency, N_0 is the constant vertical compression force, and a_0 is the normal displacement caused by that force. Here T_{MMD} is a dimensionless function of a dimensionless tangential displacement; T_{MMD} equals the friction force normalized on μN_0 .

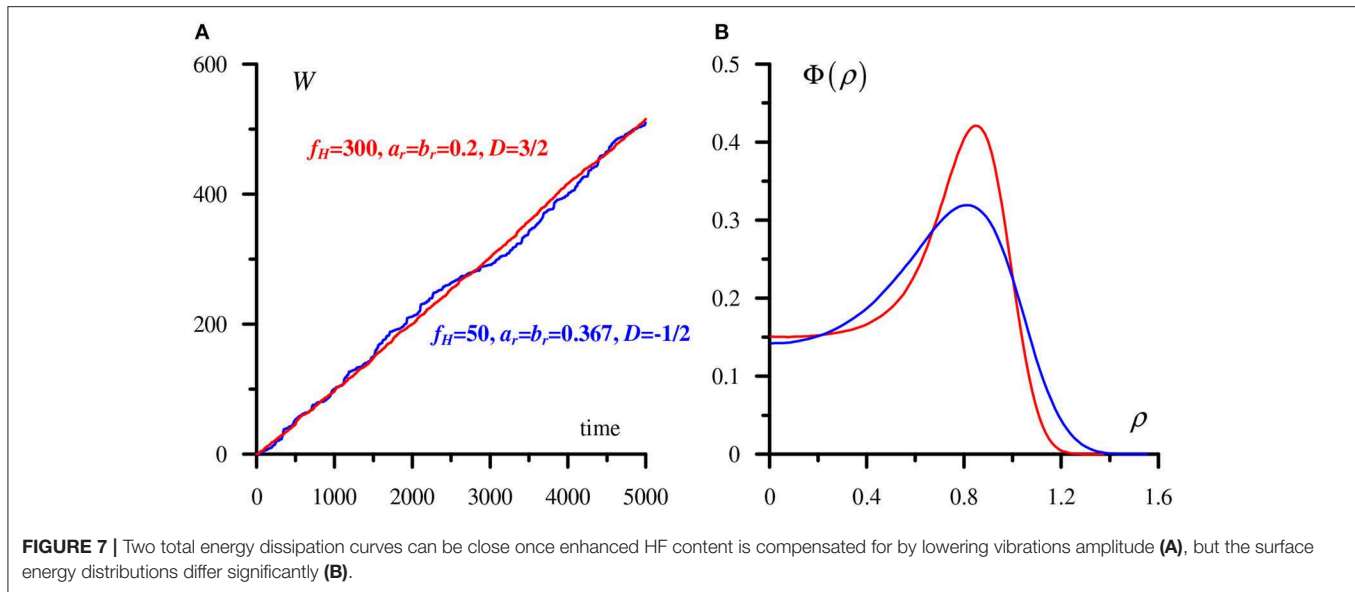


FIGURE 7 | Two total energy dissipation curves can be close once enhanced HF content is compensated for by lowering vibrations amplitude **(A)**, but the surface energy distributions differ significantly **(B)**.

In the considered case involving inertia, it is meaningful to distinguish between excitation and loading protocol/history. Here the former one is the sinusoidal term in Equation (17) while the latter one is the argument of T_{MMD} function. We will see that in the considered frictional system the loading protocol becomes rather complex even for a simple excitation signal (sinewave). In the previous quasi-static case (section friction-induced energy dissipation in spherical contacts excited by random vibrations) this difference did not exist, the same way as the equation of motion.

It is convenient to rewrite Equation (17) using dimensionless variables

$$b^* = \frac{b}{\theta \mu a_0}, \quad t^* = ft, \quad m^* = \frac{\theta a_0 f^2 m}{N_0}, \quad T_a^* = \frac{T_a}{\mu N_0} \quad (18)$$

and get an equation in which all variables and parameters are dimensionless:

$$m^* \frac{d^2 b^*}{dt^{*2}} = T_a^* \sin 2\pi t^* - T_{MMD}(b^*). \quad (19)$$

Actually, it has only two parameters: m^* that characterizes inertial properties for a given frequency, and T_a^* that corresponds to a relative strength of the external force compared to the friction force. In order to have comparable tangential responses for highly different parameters values, it is appropriate to introduce another dimensionless displacement

$$\bar{b}(t^*) = m^* b^* / T_a^*. \quad (20)$$

When the excitation amplitude is very high, friction is negligible, and the solution has a simple form

$$\bar{b}(t^*) = \frac{1}{2\pi} \left(t^* - \frac{1}{2\pi} \sin 2\pi t^* \right) \quad (21)$$

that does not depend on any parameters. Here the increasing term $\sim t^*$ appears due to the second boundary condition in

Equation (17) and may change once this boundary condition alters. For lower T_a^* friction becomes important, and the system has a whole range of different behaviors illustrated in **Figures 8, 9** plotted for heavy ($m^* = 100$, **Figure 8**) and light ($m^* = 0.01$, **Figure 9**) bodies. The curves represent the tangential displacements (at the left) and the corresponding velocities (at the right) for decreasing drive amplitudes $T_a^* = 10^4, 10^3, 10^2, 10, 1, 10^{-2}$ marked by various colors. Solution Equation (21) is plotted in black and is labeled $T_a^* \rightarrow \infty$.

The principal feature that shows up in **Figures 8, 9** is the presence of very different time scales. All curves always oscillate with the period of 1 that corresponds to the driving frequency. Besides the lowest scale of one, characteristic times of about 50 (**Figures 8B,C**), 20 [(**Figure 9C**) for $T_a^* = 10^{-2}$, in blue] or another can appear. Generally, observation time of 500 is sufficient to see the character of the dependence. The entire curves for velocities containing 500 oscillations are shown in gray for all drive amplitudes, whereas their fragments at the beginning and at the end of the observation time are plotted in colors corresponding to the particular amplitudes.

For heavy and light bodies excited by strong tangential forces of amplitudes $T_a^* = 10^4, 10^3$, displacement's behavior represents climbing saturating oscillations [it is expected that the curve for $T_a^* = 10^4$ will finally saturate as it does for $T_a^* = 10^3$, only for $T_a^* \rightarrow \infty$ no saturation is present, Equation (21)]. A progressive decrease in T_a^* leads to a quicker saturation at a lower level; finally, any climbing disappears. Indeed, blue and orange curves for heavy and light bodies are generally symmetric so the positive trend [linear term in Equation (21)] is absent. For small drive amplitudes, the behavior differs for heavy and light bodies. The heavy one demonstrates secondary oscillations (**Figures 8B,C**) while the orange curve in **Figure 9C** contains only the smallest oscillations of a constant level.

The rich behavior illustrated in **Figures 8, 9** is difficult to reproduce without using a semi-analytical method.

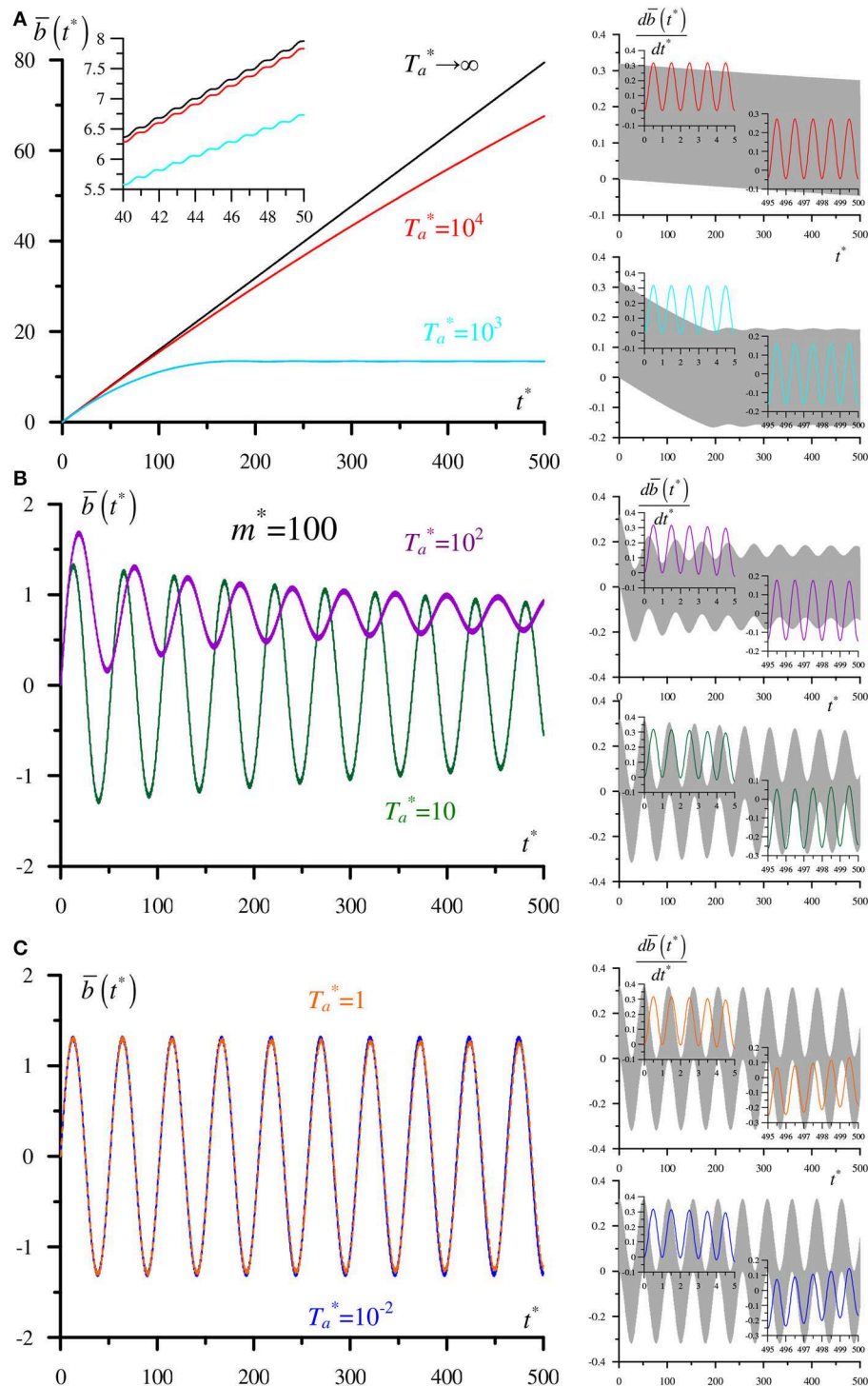


FIGURE 8 | Normalized tangential displacement (at the left) and corresponding velocity (at the right) for different excitation amplitudes T_a^* for $m^* = 100$: **(A)** $T_a^* \rightarrow \infty$, $T_a^* = 10^4$, and $T_a^* = 10^3$; **(B)** $T_a^* = 10^2$ and $T_a^* = 10$; **(C)** $T_a^* = 1$ and $T_a^* = 10^{-2}$.

For very high drive amplitudes when a partial slip is not essential, or for very low amplitudes when the contact behaves as a linear lossless spring, some asymptotic

analysis is possible. However, for the most important range of moderate amplitudes only numerical treatment is appropriate.

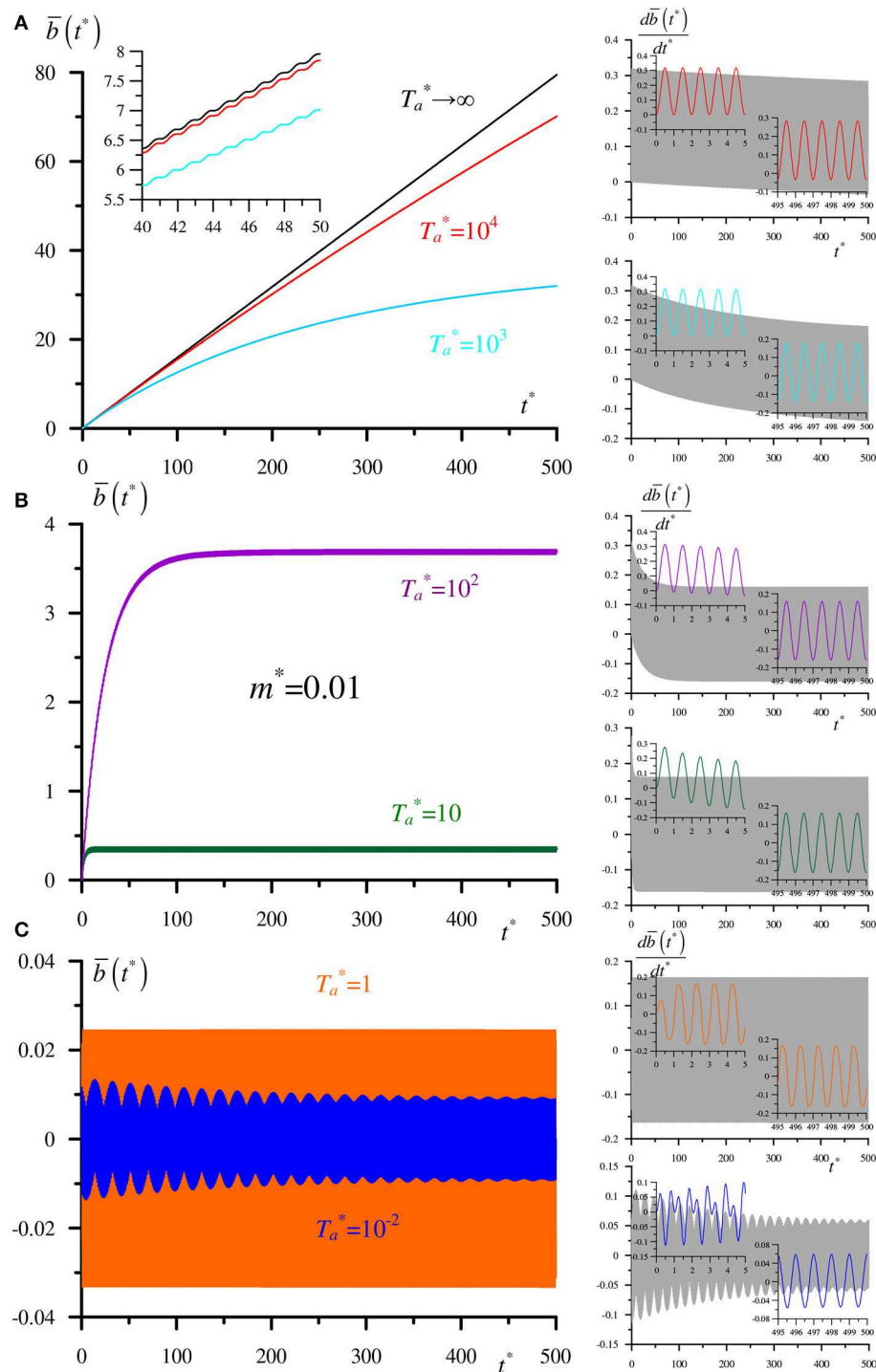


FIGURE 9 | Normalized tangential displacement (at the left) and corresponding velocity (at the right) for different excitation amplitudes T_a^* for $m^* = 0.01$: **(A)** $T_a^* \rightarrow \infty$, $T_a^* = 10^4$, and $T_a^* = 10^3$; **(B)** $T_a^* = 10^2$ and $T_a^* = 10$; **(C)** $T_a^* = 1$ and $T_a^* = 10^{-2}$.

CONCLUSIONS AND PERSPECTIVES

Semi-analytical methods in frictional contact mechanics enable the efficient calculation of a hysteric tangential force-displacement relationship of an axisymmetric contact system

for an arbitrary loading history. For instance, in the Method of Memory Diagrams (MMD), all history-dependent information is encoded in the internal function that is updated following the loading history in accordance with certain rules. Updating the memory diagram is computationally much cheaper than detailed

calculations of the local stress and displacement fields in the contact zone. In this paper, two simple examples are considered which help determine a class of practical problems for which the semi-analytical solutions can be of use.

The first example concerns the calculation of friction-induced energy losses in contact with two spheres excited by random vibrations. Knowledge of the mechanical response of a system makes it possible to describe its energetic response (i.e., the total energy dissipated during certain time interval together with a spatial distribution of this energy over the contact zone). The former dependence is close to a direct proportionality since the average dissipated power should be constant as long as the system is excited by a stationary random process. The latter one is more informative; in particular, most of the energy is dissipated in an annulus close to the average contact border for weak excitation amplitudes, while for stronger amplitudes the inner border of the annulus propagates inward so that eventually the annulus becomes a circle. An obvious goal of this kind of calculation is modeling for wear in contact systems. Indeed, adding a wear model to MMD simulations provides an opportunity to predict where and when wear is most likely to occur for known statistical properties of the random excitation. In that regard, it would be of interest to compare a final shape of the profile to known results (Argatov et al., 2018) obtained without a detailed analysis of energy dissipation or wear processes, but from the assumption that the final contact area coincides with the initial stick area for a harmonic tangential excitation. Besides, the cited paper, as well as the work by Chai and Argatov (2018), reports generalization on the Cattaneo-Mindlin theory for transversely isotropic elastic bodies that can be potentially incorporated into semi-analytical contact analysis.

The second example is related to a contact associated with a particular mass. Mass and inertial properties add dynamics to the contact system and give rise to a very rich behavior even for a simple harmonic excitation. In particular, for various combinations of two system's parameters (normalized mass and normalized excitation amplitude), a number of different time scales can be found in the tangential response. Besides the drive period, characteristic times tens or hundreds of times longer

than the drive period can be found. Depending on the mass and the drive amplitude, regimes of climbing saturated oscillations, decaying LF oscillations on top of weak HF ones, and others show up. The application field for numerical simulations of this kind can cover coupled frictional systems, stick-slip phenomena, friction-induced instabilities, or acoustic emission (squealing).

To summarize, three main conclusions can be formed:

- The MMD is especially suitable for modeling responses on complex loading protocols in frictional contact systems of simple geometries;
- The MMD allows one to model quasi-static mechanical and energetic responses on random vibrations and eventually to make a prediction on a configuration of wear zones;
- A contact system having a certain mass demonstrates rich dynamic behavior when excited even by a simple harmonic signal. Several classes of solutions have been identified that do not exist in the point mass case.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The author acknowledges funding of the French National Research Agency (ANR, grant number ANR-17-CE08-0013) as well as support from the Tomsk State University competitiveness improvement program.

ACKNOWLEDGMENTS

The author is grateful to colleagues from TU Berlin (V. Popov, L. Voll, I. Argatov) and from Politecnico Bari (M. Ciavarella and A. Papangelo) for fruitful discussions.

REFERENCES

- Aleshin, V. V., and Bou Matar, O. (2016). Solution to the frictional contact problem via the method of memory diagrams for general 3D loading histories. *Phys. Mesomech.* 19, 130–135. doi: 10.1134/S102995991602003X
- Aleshin, V. V., Bou Matar, O., and Van Den Abele, K. (2015). Method of memory diagrams for mechanical frictional contacts subject to arbitrary 2D loading. *Int. J. Solids Struct.* 60, 84–95. doi: 10.1016/j.ijsolstr.2015.02.016
- Aleshin, V. V., Delrue, S., Van Den Abele, V., and Bou Matar, O. (2019). "Nonlinear and hysteretic constitutive models for wave propagation in solid media with cracks and contacts," in *Nonlinear Ultrasonic and Vibro-Acoustical Techniques for Nondestructive Evaluation (Chapter 5)*, ed T. Kundu (Cham: Springer), 175–224. doi: 10.1007/978-3-319-94476-0_5
- Aleshin, V. V., and Papangelo, A. (2020). Friction-induced energy losses in mechanical contacts subject to random vibrations. *Int. J. Sol. Struct.* 190, 148–155. doi: 10.1016/j.ijsolstr.2019.10.026
- Argatov, I. I., Bae, J. W., and Chai, Y. S. (2018). The limiting shape of the transversely isotropic elastically similar solids in fretting. *Int. J. Appl. Mech.* 10:1850089. doi: 10.1142/S1758825118500898
- Borodich, F. M., and Keer, L. M. (2004). Contact problems and depth-sensing nanoindentation for frictionless and frictional boundary conditions. *Int. J. Solids Struct.* 41, 2479–2499. doi: 10.1016/j.ijsolstr.2003.12.012
- Borodich, F. M., and Savencu, O. (2017). "Hierarchical models of engineering rough surfaces and bio-inspired adhesives. Ch. 10," in *Bio-Inspired Adhesives*, eds L. Heepe, S. Gorb, and L. Xue (Berlin: Springer), 179–219. doi: 10.1007/978-3-319-59114-8_10
- Cattaneo, C. (1938). Sul contatto di due corpi elastici: distribuzione locale degli sforzi. *Rend. Accad. Naz. Lincei* 27, 342–348.
- Chai, Y. S., and Argatov, I. I. (2018). Local tangential contact of elastically similar, transversely isotropic elastic bodies. *Meccanica* 53, 3137–3143. doi: 10.1007/s11012-018-0870-y

- Ciavarella, M. (1998). The generalized Cattaneo partial slip plane contact problem. *I-theory. Int. J. Solids Struct.* 35, 2349–2362. doi: 10.1016/S0020-7683(97)00154-6
- Desplanques, Y. (2015). Amontons-coulomb friction laws, A review of the original manuscript. *SAE Int. J. Mater. Manufact.* 8, 98–103. doi: 10.4271/2014-01-2489
- Dobry, R., Ng, T., Petrakis, E., and Seridi, A. (1991). General model for contact law between two rough spheres. *J. Eng. Mech.* 117, 1365–1381. doi: 10.1061/(ASCE)0733-9399(1991)117:6(1365)
- Jäger, J. (1995). Axi-symmetric bodies of equal material in contact under torsion or shift. *Arch. Appl. Mech.* 65, 478–487. doi: 10.1007/BF00835661
- Jäger, J. (2005). *New Solutions in Contact Mechanics*. Southampton: WIT Press.
- Mindlin, R., and Deresiewicz, H. (1953). Elastic spheres in contact under varying oblique forces. *J. Appl. Mech.* 20, 327–344.
- Munisamy, R. L., Hills, D. A., and Nowell, D. (1994). Static axisymmetrical Hertzian contacts subject to shearing forces. *ASME J. Appl. Mech.* 61, 278–283. doi: 10.1115/1.2901441
- Popov, V., and Heß, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin; Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-53876-6
- Popov, V. L., Heß, M., and Willert, E. (2019). *Handbook of Contact Mechanics. Exact Solutions of Axisymmetric Contact Problems*. Berlin; Heidelberg: Springer-Verlag. doi: 10.1007/978-3-662-58709-6
- Truyaert, K., Aleshin, V. V., Van Den Abeele, K., and Delrue, S. (2019). Theoretical calculation of the instantaneous friction-induced energy losses in arbitrarily excited axisymmetric mechanical contact systems. *Int. J. Sol. Struct.* 158, 268–276. doi: 10.1016/j.ijsolstr.2018.09.014

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Aleshin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



FFT-Based Methods for Computational Contact Mechanics

Q. Jane Wang^{1*}, Linlin Sun^{1,2}, Xin Zhang¹, Shuangbiao Liu^{1*} and Dong Zhu³

¹ Mechanical Engineering and Center for Surface Engineering and Tribology, Northwestern University, Evanston, IL, United States, ² School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an, China, ³ Tri-Tech Solutions, Mt. Prospect, IL, United States

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Aydar Akchurin,
SKF, Netherlands
Qiang Li,
Technical University of
Berlin, Germany

*Correspondence:

Q. Jane Wang
qwang@northwestern.edu
Shuangbiao Liu
shuangbiaoli@yahoo.com

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 29 May 2020

Accepted: 30 June 2020

Published: 28 August 2020

Citation:

Wang Q, Sun L, Zhang X, Liu S and
Zhu D (2020) FFT-Based Methods for
Computational Contact Mechanics.
Front. Mech. Eng. 6:61.
doi: 10.3389/fmech.2020.00061

Computational contact mechanics seeks for numerical solutions to contact area, pressure, deformation, and stresses, as well as flash temperature, in response to the interaction of two bodies. The materials of the bodies may be homogeneous or inhomogeneous, isotropic or anisotropic, layered or functionally graded, elastic, elastoplastic, or viscoelastic, and the physical interactions may be subjected to a single field or multiple fields. The contact geometry can be cylindrical, point (circular or elliptical), or nominally flat-to-flat. With reasonable simplifications, the mathematical nature of the relationship between a surface excitation and a body response for an elastic contact problem is either in the form of a convolution or correlation, making it possible to formulate and solve the contact problem by means of an efficient Fourier-transform algorithm. Green's function inside such a convolution or correlation form is the fundamental solution to an elementary problem, and if explicitly available, it can be integrated over a region, or an element, to obtain influence coefficients (ICs). Either the problem itself or Green's functions/ICs can be transformed into a space-related frequency domain, via a Fourier transform algorithm, to formulate a frequency-domain solution for contact problems. This approach converts the original tedious integration operation into multiplication accompanied by Fourier and inverse Fourier transforms, and thus a great computational efficiency is achieved. The conversion between ICs and frequency-response functions facilitate the solutions to problems with no explicit space-domain Green's function. This paper summarizes different algorithms involving the fast Fourier transform (FFT), developed for different contact problems, error control, as well as solutions to the problems involving different contact geometries, different types of materials, and different physical issues. The related works suggest that (i) a proper FFT algorithm should be used for each of the cylindrical, point, and nominally flat-flat contact problems, and then (ii) the FFT-based algorithms are accurate and efficient. In most cases, the ICs from the 0-order shape function can be applied to achieve satisfactory accuracy and efficiency if (i) is guaranteed.

Keywords: contact of materials, fast fourier transform, FFT algorithms, contact pressure, contact stress, tribology

INTRODUCTION

Contact of materials is a common engineering phenomenon, and the solution to a contact problem, in terms of pressure, deformations, and stresses, as well as flash temperature, is usually among first steps in the design and analysis of an engineering system or a functional device. A contact problem is solved first for the information of the contacting interface, such as contact pressure, surface

interaction, contact area, and interfacial friction, followed by a boundary-value solution process for the stresses in each contacting body. At least two sets of convolutions, each for a group of excitations and Green's functions, are involved, i.e., displacements and stresses, in response to surface tractions, and the former are calculated first (Conry and Seireg, 1971; Kalker and van Randen, 1972; Kalker, 1986; Polonsky and Keer, 1999). When solving the contact of bodies involving an inclusion-containing material, the mathematical correlation between an eigenstrain and a Green's function appears (Liu et al., 2012). In most cases, the convolution nature makes it possible to formulate and solve a contact problem by means of an efficient fast Fourier transform (FFT) algorithm (Ju and Farris, 1996; Stanley and Kato, 1997; Ai and Sawamiphakdi, 1999; Hu et al., 1999; Polonsky and Keer, 1999, 2000; Liu et al., 2000; Liu and Wang, 2002). We will discuss and summarize the theories, algorithms, and numerical methods of FFT-based contact modeling approaches.

If the contact body can be treated as a mathematically “semi-infinite” medium, or a half-space material, analytical influence coefficients (ICs), or frequency-response functions (FRFs), may be derived for certain problems. Normally such a simplification can be made if deformations are small and the radii of curvature of contact bodies are sufficiently larger than the effective contact size of these two bodies. In the case of the macroscopic contact of two spheres, or two cylinders, of the same size and material, if the contact radius, or half width, a , is 10% of the radius of the contact body, the maximum error from using the z -direction deformation to replace the radial deformation at the contact edge is within 0.5%. However, stresses need more attention. Two contact problems, (1) two steel spheres of the same radius and (2) their equivalent sphere and a half space, are analyzed using the finite-element method (FEM) without considering plasticity. **Figure 1** shows maximum relative errors of the von Mises stress along the central axis, due to ignoring curvature, as a function of a/R_e . The error is about 7% when a/R_e reaching 10%, and this situation is beyond the “small elastic deformation” assumption. In the elastic range, using steels as an example, the maximum error of the von Mises stress is <3% if the half-space solution is pursued. Similar small errors due to the use of the half-space assumption in the elastic range have also been reported by Londhe et al. (2018), in the comparison of the results from the FEM and Hertz formulas for different types of contacts.

In a contact simulation, the computational complexity of evaluating a convolution via direct summation of the products of ICs and surface traction is on the order of $O(N^2)$, where N is the mesh number. If N is large and the convolution has to be repeatedly calculated in an iteration process, the computation burden is very heavy. The works by Ju and Farris (1996), Nogi and Kato (1997), Stanley and Kato (1997), Ai and Sawamiphakdi (1999), Hu et al. (1999), Polonsky and Keer (2000), and Liu et al. (2000) are in a chain of studies to apply the FFT to evaluate the convolutions for elastic deformation and stresses efficiently in the field of contact mechanics and tribology. Several papers have reviewed such efforts of solving tribology problems via the FFT and analyzed the sources of errors (Liu et al., 2000; Wang et al., 2003; Liu S. et al., 2007; Wang and Zhu, 2019). Although most of the effort is on solving non-conformal contact problems, the

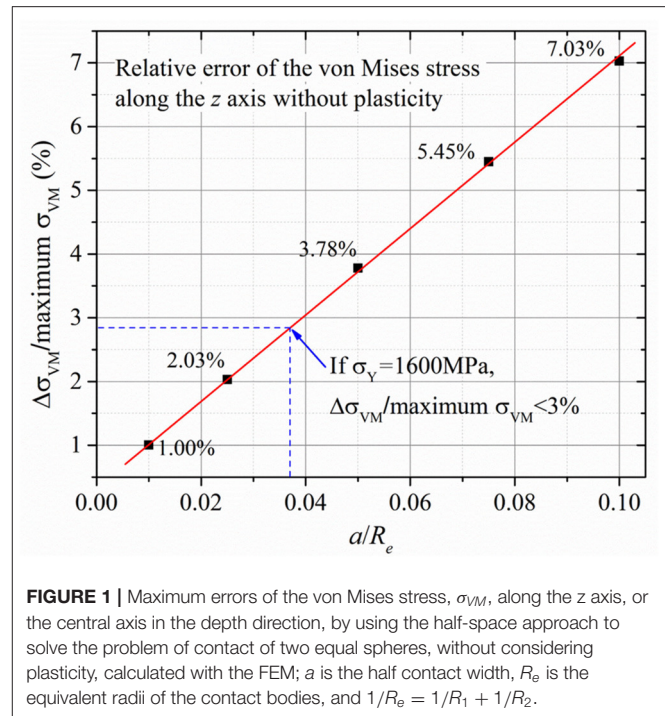


FIGURE 1 | Maximum errors of the von Mises stress, σ_{VM} , along the z axis, or the central axis in the depth direction, by using the half-space approach to solve the problem of contact of two equal spheres, without considering plasticity, calculated with the FEM; a is the half contact width, R_e is the equivalent radii of the contact bodies, and $1/R_e = 1/R_1 + 1/R_2$.

FFT methods suit for certain conformal-contact problems as well if they involve a convolution and have ICs obtained analytically or numerically. Actually, the circular nature of a cylindrical structure fits the circular convolution theorem perfectly. Liu and Chen (2012) and Liu (2013) reported an FFT-based conformal-contact model for two-dimensional (2D) problems. Wang and Jin (2004) conducted the fluid-film lubrication analysis for artificial joints, which requires the determination of the elastic deformation of the bearing surface of both the acetabular cup and the femoral head. They used FFT along with the spherical distance and numerical ICs from an FEM calculation.

The FFT methods greatly help reduce the computation burden. For example, for a three-dimensional (3D) point-contact problems, the FFT operation is on the order of $O(12N \log_2 [4N])$, to be discussed in detail later. Its ratio to the operation needed for calculating the convolution is $12N \log_2 (4N)/N^2 = 12 \log_2 (4N)/N = 0.00024$, if N is 1024×1024 . This is a significant saving of computational time. The key issues to be addressed are (1) how should the Fourier transform (FT) method be properly used to solve contact problems? (2) Can one solution algorithm be used for all problems?

Theoretically, the Fourier transform can be used for infinite-domain and the Fourier series for periodic problems, but most contact problems do not satisfy these conditions. For example, a point-contact problem has its pressure only on a small region of contacting surfaces. If the FFT is directly used to solve such a problem, the results near the borders have notable errors. In order to reduce the periodicity error, Ju and Farris (1996) substantially extended the domain, Ai and Sawamiphakdi (1999) decomposed the total pressure into a smooth portion and a zero-mean fluctuating portion, and Polonsky and Keer

(2000) developed a hybrid algorithm by adding a special correction procedure. Each of these brings in a certain accuracy improvement while introducing new complications. We have investigated the theories of contact mechanics and signal analyses, and realized that proper convolution theorems should be considered in solving different contact problems of non-conformal and conformal configurations (**Figure 2**) (Liu et al., 2000; Liu and Wang, 2002; Liu S. et al., 2007; Liu and Hua, 2009; Liu and Chen, 2012), as summarized by Wang and Zhu (2019, Chapter 4).

In this paper, we will discuss the basic issues of the FFT methods for contact analyses from the convolution theorems and the tree of the Fourier-transform algorithms for solving different contact problems, such as (1) the algorithm of discrete-convolution and fast-Fourier-transform (DC-FFT), with double domain extension in each dimension, for non-periodic problems, and the discrete-convolution and fast-Fourier-transform algorithm (DC-FFT) without domain extension for journal bearing problems, (2) the algorithm of continuous-convolution and Fourier-transform (CC-FT) for

periodic (or infinite) contact problems, (3) the algorithm of discrete convolution with duplicated padding and FFT (DCD-FFT), that of discrete-continuous convolutions and FFT (DC-CC-FFT), and that of the discrete convolution with IC summation and FFT (DCS-FFT) for 3D line-contact problems that are periodic (or infinite) in one direction but non-periodic in the other direction, (4) the algorithm of discrete-correlation and FFT (DCR-FFT) for inclusion problems, (5) the FRF-IC conversion method, as well as the applications of them to solve the contact problems involving layered materials, anisotropic elastic materials, and viscoelastic polymers, or those subjected to multifields, and (6) a non-uniform DCS-FFT method, recently developed by Sun et al. (2020), for solve large-scale contact problems. Most of the contents in this paper are based on the works by Liu et al. (2000, 2002), Liu and Wang (2002), Boucly et al. (2005), Chen et al. (2008), Liu and Hua (2009), Yu et al. (2014, 2016), Zhang X. et al. (2017, 2018), Zhang and Wang (2019), Sun (2020), Sun et al. (2020), Zhang et al. (2020a,b), and Chapter 4 in the book by Wang and Zhu (2019). **Table 1** summarizes the FFT-based approaches and their applications.

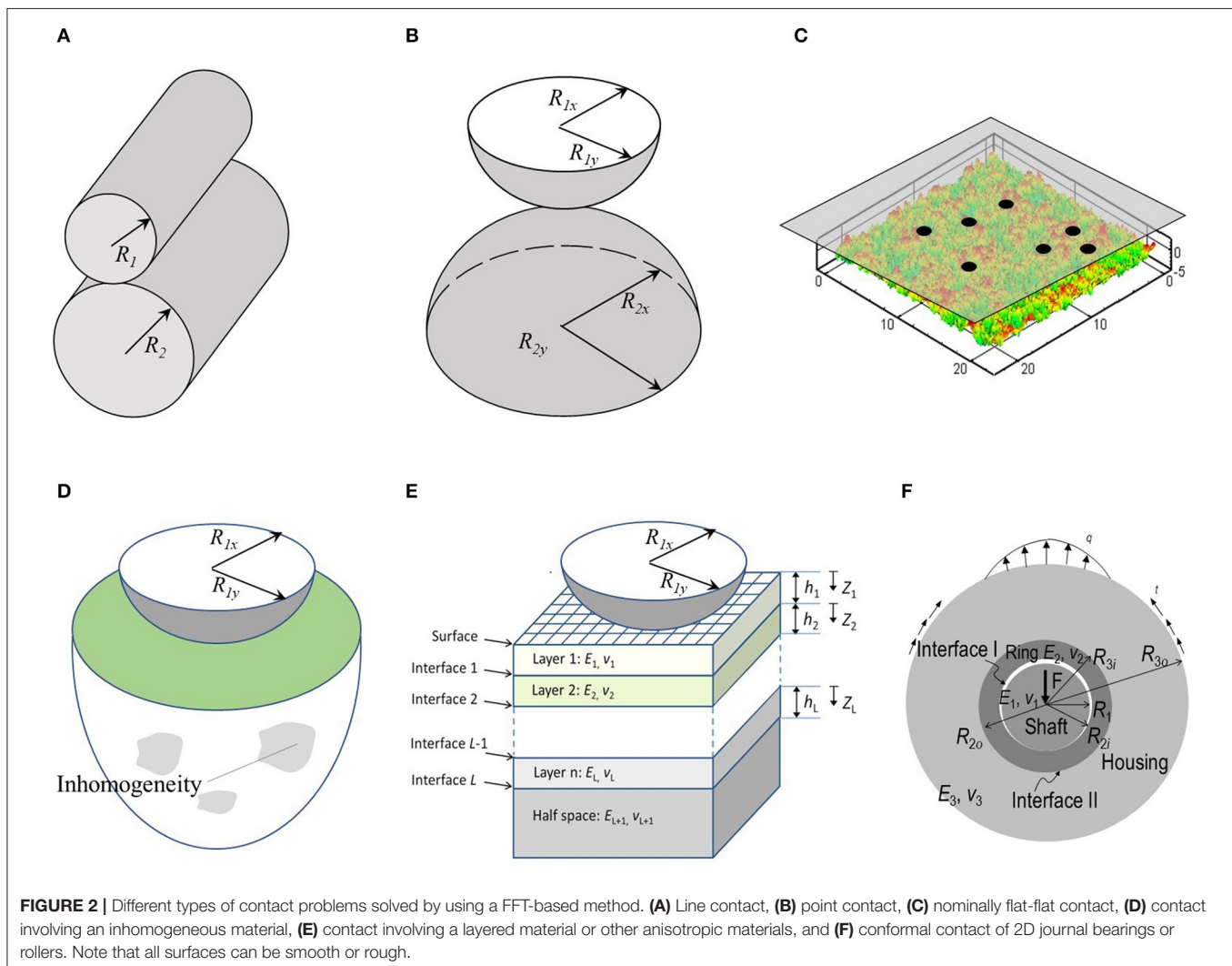


TABLE 1 | FFT methods for contact analyses.

Name and references	Algorithm and method	Problem to solve
DC-FFT (Liu et al., 2000; Liu, 2001) (Liu and Chen, 2012; Liu, 2013)	Discrete-convolution and FFT	Point-contact problems Cylindrical contact problems, counterformal and conformal
CC-FT (Ju and Farris, 1996; Liu, 2001; Liu et al., 2002)	Continuous-convolution and FFT	Nominally flat-flat contact problems
DCSS-FFT (Sun, 2020)		
DCD-FFT (Chen et al., 2008)	Discrete convolution with duplicated padding and FFT	3D line-contact problems
DC-CC-FFT (Liu et al., 2006)	Discrete-convolution, continuous-convolution and FFT	
DCS-FFT (Liu and Hua, 2009; Sun et al., 2020)	Discrete convolution with IC summation and FFT	
DCR-FFT (Liu and Wang, 2005)	Discrete-convolution-correlation and fast-Fourier-transform	Materials with residual strains, inclusions and/or inhomogeneities, contact plasticity problems
IC conversion (Liu et al., 2000; Liu, 2001; Liu and Wang, 2002; Liu S. et al., 2007)	FRFs known, which can be transformed to ICs, followed by the DC-FFT algorithm or other proper ones	Layered, viscoelastic, transversely isotropic materials, coupled-stress problems, multifield contact problems
Non-uniform DCS-FFT (Sun et al., 2020)	Implementation of DCS-FFT, or others, to segments of different mesh densities	Large-scale problems, 3D line-contact problem considering defect, crown, and edge effects
SFFT (Wang and Jin, 2004)	Discrete-convolution and FFT	Sphere-cup contact problems

It should be mentioned that the pressure and contact area that satisfy the Kuhn-Tucker type complementary conditions can be solved with different methods, and the conjugate gradient method (CGM) (Polonsky and Keer, 1999; Jin et al., 2013) is currently the widely accepted one.

In the following, the elastic field means the distributions of stresses and displacements, and the target domain means the physical domain, on which a physical contact problem is defined. The algorithms and methods will be explained mainly through deformation calculations; details of the FFT-based computations of stresses, flash temperature, and other physical fields can be found in the reports by Liu (2001), Liu and Wang (2002), Chen et al. (2008), Zhang X. et al. (2018), and Wang and Zhu (2019), as well as those mentioned in the previous paragraph.

CONVOLUTION, FREQUENCY RESPONSE FUNCTION, AND INFLUENCE COEFFICIENTS

Convolution and ICs

Let's use the pressure-displacement relationships, such as the Flamant and Boussinesq equations (Johnson, 1987), as examples. Here, an excitation at ξ , or (ξ, η) , and a response at x , or (x, y) , are related to each other through a Green's function defined with the distance between the two, which is either $|x - \xi|$ or $R^I = \sqrt{(x - \xi)^2 + (y - \eta)^2}$.

The surface normal displacement of a cylinder in a line contact, $u_z(x)$, due to pressure $p(x)$ on surface region S_x is

$$u_z(x) = -\frac{4}{\pi E'} \int_{S_x} \ln |x - \xi| p(\xi) d\xi = C \int_{S_x} G(x - \xi) p(\xi) d\xi \quad (1)$$

where $C = -4/(\pi E')$, E' is the effective Young's modulus and the corresponding Green's function is $C \ln |x|$. The integral kernel, G , is defined as

$$G(x) = \ln |x| \quad (2)$$

In numerical modeling, the equation above can be discretized and re-written as the summation of the products of influence coefficients $D(k, i)$, or $D_{i,j}$, and nodal pressures p_i .

$$u_z(x_k) = C \sum_{i=1}^{N_p} D(k, i) p_i = C \sum_{i=1}^{N_p} D_{k,i} p_i, \quad (k = 1, 2, 3, \dots, N_d) \quad (3)$$

where N_p , N_d are the total numbers of nodes for pressure and deformation, respectively.

A shape function, Y_s , may be used to distribute pressure, or other excitations, around a nodal point, and the commonly used shape functions are 0-order, 1st-order, and 2nd-order polynomials. Detailed expressions and use of these shape functions can be found in the book by Wang and Zhu (2019).

The ICs are from the elementary integration of Green's function and shape function Y_s , implying unit nodal pressure, or from G and shape function Y_s without including coefficient C . The latter is used here. In general,

$$D(k, i) = \int_{x_l=x_i-\Delta_2-x_k}^{x_u=x_i+\Delta_1-x_k} G(x_k - \xi_i) Y_s(\xi_i) d\xi = \{f(x_u) - f(x_l)\} \quad (4)$$

where f is the integration result, and x_u and x_l are the upper and lower boundaries of the element integration at x_i , with Δ_1 and Δ_2 marking the element size.

If the zero-order shape function is used, the influence coefficient expression becomes

$$D(k, i) = (x_{i+1/2} - x_k) (\ln |x_{i+1/2} - x_k|) - (x_{i-1/2} - x_k) (\ln |x_{i-1/2} - x_k|) + (x_{i-1/2} - x_{i+1/2}) \quad (5)$$

Note that here the influence coefficients, D , depend only on the geometric factors of the grid. When a uniform grid is used with the constant mesh spacing $x_{i+1} - x_i = x_i - x_{i-1} = 2\Delta$, the above becomes

$$D(k, i) = (x_i + \Delta - x_k) (\ln |x_i + \Delta - x_k|) - (x_i - \Delta - x_k) (\ln |x_i - \Delta - x_k|) - 2\Delta \quad (6)$$

For point-contact problems, surface normal displacement $u_z(x, y)$ due to pressure $p(x, y)$ on surface area Ω is,

$$u_z(x, y) = \frac{2}{\pi E'} \iint_{\Omega} \frac{p(\xi, \eta)}{\sqrt{(x - \xi)^2 + (y - \eta)^2}} d\xi d\eta = C \iint_{\Omega} \frac{p(\xi, \eta)}{\sqrt{(x - \xi)^2 + (y - \eta)^2}} d\xi d\eta \quad (7)$$

where $C = \frac{2}{\pi E'}$, G is from Green's function, defined as $G = \frac{1}{\sqrt{x^2 + y^2}}$.

Likewise, the equation above can be re-written as the summation of the products of influence coefficients $D_{ij}^{k,l}$ and nodal pressures p_{ij} . The ICs are from the elementary integration of Green's function and a shape function, Y_s , or G and shape function Y_s without including the material properties, and the latter is used here.

$$u_z(x_k, y_l) = C \sum_{i=1}^{N_{px}} \sum_{j=1}^{N_{py}} D_{ij}^{k,l} p_{ij} \quad (k = 1, 2, \dots, N_{dx}, l = 1, 2, \dots, N_{dy}) \quad (8)$$

where N_{dx} , N_{dy} are the total numbers of nodes for deformation in the x, y directions, and N_{px} , N_{py} are the total numbers of nodes for pressure in the x, y directions, respectively.

$$D_{ij}^{k,l} = \int_{x_l}^{x_u} \int_{y_l}^{y_u} G(x_k, y_l, \xi_i, \eta_j) Y_s(\xi_i, \eta_j) d\xi d\eta \quad (9)$$

$$\text{Or } D = \sum_{\text{Elemental-contribution}} \{f(x_u, y_u) + f(x_l, y_l) - f(x_u, y_l) - f(x_l, y_u)\} \quad (10)$$

where f is the integration result, and (x_u, y_u) (x_l, y_l) are the upper and lower boundaries of the element integration.

If the zero-order shape function is used, the influence coefficient expression is (Love, 1929),

$$D_{ij}^{k,l} = (x_k - x_i + a) \ln \left[\frac{(y_l - y_j + b) + \sqrt{(y_l - y_j + b)^2 + (x_k - x_i + a)^2}}{(y_l - y_j - b) + \sqrt{(y_l - y_j - b)^2 + (x_k - x_i + a)^2}} \right] + (y_l - y_j + b) \ln \left[\frac{(x_k - x_i + a) + \sqrt{(y_l - y_j + b)^2 + (x_k - x_i + a)^2}}{(x_k - x_i - a) + \sqrt{(y_l - y_j + b)^2 + (x_k - x_i + a)^2}} \right] + (x_k - x_i - a) \ln \left[\frac{(y_l - y_j - b) + \sqrt{(y_l - y_j - b)^2 + (x_k - x_i - a)^2}}{(y_l - y_j + b) + \sqrt{(y_l - y_j + b)^2 + (x_k - x_i - a)^2}} \right] + (y_l - y_j - b) \ln \left[\frac{(x_k - x_i - a) + \sqrt{(y_l - y_j - b)^2 + (x_k - x_i - a)^2}}{(x_k - x_i + a) + \sqrt{(y_l - y_j - b)^2 + (x_k - x_i + a)^2}} \right] \quad (11)$$

Replacing $x_{k-i} = x_k - x_i$ and $y_{l-j} = y_l - y_j$ leads to the following.

$$D_{ij}^{k,l} = (x_{k-i} + a) \ln \left[\frac{(y_{l-j} + b) + \sqrt{(y_{l-j} + b)^2 + (x_{k-i} + a)^2}}{(y_{l-j} - b) + \sqrt{(y_{l-j} - b)^2 + (x_{k-i} + a)^2}} \right] + (y_{l-j} + b) \ln \left[\frac{(x_{k-i} + a) + \sqrt{(y_{l-j} + b)^2 + (x_{k-i} + a)^2}}{(x_{k-i} - a) + \sqrt{(y_{l-j} + b)^2 + (x_{k-i} + a)^2}} \right] + (x_{k-i} - a) \ln \left[\frac{(y_{l-j} - b) + \sqrt{(y_{l-j} - b)^2 + (x_{k-i} - a)^2}}{(y_{l-j} + b) + \sqrt{(y_{l-j} + b)^2 + (x_{k-i} - a)^2}} \right] + (y_{l-j} - b) \ln \left[\frac{(x_{k-i} - a) + \sqrt{(y_{l-j} - b)^2 + (x_{k-i} - a)^2}}{(x_{k-i} + a) + \sqrt{(y_{l-j} - b)^2 + (x_{k-i} + a)^2}} \right] \quad (12)$$

where a and b are the half length of the rectangular integration element.

Frequency Response Functions and ICs

The Fourier transform can be applied to the IC matrix, D , Equation (4), to obtain,

$$\tilde{D} = \tilde{G} \cdot \tilde{Y}_s \quad (13)$$

$$\text{and } \tilde{G} = \tilde{D} / \tilde{Y}_s \quad (14)$$

where \tilde{G} is the frequency response function excluding the elastic parameter, C . These two equations show how to obtain one from the other.

If discrete Fourier transform \hat{D} has already been obtained from a set of known ICs via the FFT, \tilde{G} can be solved from $\tilde{G} = \tilde{D} / \tilde{Y}_s$ once Fourier series coefficients \tilde{D} can be obtained from \hat{D} . Based on the sampling theorem, the one-dimensional (1D) relationship between the FT (\sim) and discrete Fourier transform (DFT) (\wedge) of the ICs, sampled with mesh interval 2Δ , can be obtained as

$$\hat{D}_i = \frac{1}{2\Delta} \sum_{r=-\infty}^{\infty} \tilde{D} \left(\frac{2\pi i}{n \cdot 2\Delta} - \frac{2\pi r}{2\Delta} \right) = \frac{1}{2\Delta} \sum_{r=-\infty}^{\infty} \tilde{D} \left(\frac{\pi i}{n\Delta} - \frac{\pi r}{\Delta} \right) \quad (i = 0, 1, 2, \dots, n-1) \quad (15)$$

An aliasing control parameter, AL , can be introduced instead of summation for r from $-\infty$ to ∞ , in order to satisfy a required accuracy while saving computation time.

$$\hat{D}_i = \frac{1}{2\Delta} \sum_{r=-AL}^{AL} \tilde{D}\left(\frac{2\pi i}{n \cdot 2\Delta} - \frac{2\pi r}{2\Delta}\right) = \frac{1}{2\Delta} \sum_{r=-AL}^{AL} \tilde{D}\left(\frac{\pi i}{n\Delta} - \frac{\pi r}{\Delta}\right) \quad (i = 0, 1, 2, \dots, n-1) \quad (16)$$

The above can be simplified if the sampling frequency is sufficiently high, or the datum interval is sufficiently small.

$$\hat{D}_i \approx \frac{1}{2\Delta} \tilde{D}\left(\frac{2\pi i}{n \cdot 2\Delta}\right) = \frac{1}{2\Delta} \tilde{D}\left(\frac{\pi i}{n\Delta}\right) \quad (i = 0, 1, 2, \dots, n-1) \quad (17)$$

The above becomes exact if and only if $m_s > 2m_{\max}$ (Morrison, 1994) with $\pm m_{\max}$ as the band limit (or m_{\max} as the highest frequency component), beyond which there is no Fourier transform results (Walker, 1996). Then, only the term at $r = 0$ is needed.

Likewise, the 2D relationship between the FT (\sim) and DFT (\wedge) of the ICs is

$$\hat{D}_{i,j} = \frac{1}{\Delta_x \Delta_y} \sum_{r_x=-AL}^{r_x=AL} \sum_{r_y=-AL}^{r_y=AL} \tilde{D}\left(\frac{2\pi}{n_x \Delta_x} i - \frac{2\pi}{\Delta_y} r_y, \frac{2\pi}{n_y \Delta_y} j - \frac{2\pi}{\Delta_y} r_x\right) \quad (i = 0, 1, 2, \dots, n_x - 1, \quad j = 0, 1, 2, \dots, n_y - 1) \quad (18)$$

where Δ_x and Δ_y are the mesh intervals in the x and y directions, and $\Delta_x = \Delta_y = 2\Delta$ if the meshes are uniform in both directions.

This means two operations. (1) Upon knowing \tilde{D} from \hat{D} via Equations (17) or (18) with a properly chosen discretization interval, $2\Delta < 2\pi/\omega_{\max}$, we can solve \tilde{G} using the equation below.

$$\begin{aligned} \tilde{G}_i &= \tilde{G}\left(\frac{2\pi i}{n \cdot 2\Delta}\right) = \tilde{D}\left(\frac{2\pi i}{n \cdot 2\Delta}\right) / \tilde{Y}\left(\frac{2\pi i}{n \cdot 2\Delta}\right) \\ &\approx 2\Delta \cdot \hat{D}_i / \tilde{Y}_s\left(\frac{2\pi i}{n \cdot 2\Delta}\right) \\ &= 2\Delta \cdot \hat{D}_i / \tilde{Y}_s\left(\frac{\pi i}{n\Delta}\right) \quad (i = 0, 1, 2, \dots, n-1) \end{aligned} \quad (19)$$

(2) Upon knowing \tilde{G} , we can get \hat{D} from \tilde{D} using Equations (13) and (17), or (18). In many cases where the solutions to frequency response functions are more convenient than those to Green's functions, and this operation can be utilized to convert the FRFs to the discrete Fourier transformed ICs.

CONVOLUTION THEOREMS

Equations (1) and (7) are both convolutions, which can be solved efficiently via Fourier transform followed by inverse Fourier transform. However, because the pressure may be in a discrete form, i.e., rough-surface contact pressure, and its application

domain may be in different sizes, i.e., finite or infinite, accurate solutions to these equations, and others in the same nature, require the use of different convolution theorems. The following explains these theorems with 1D datum series for convenience.

Continuous Linear Convolution

If a set of continuous functions of t , f and G , follows the convolution in Equation (20), resulting in O , then the Fourier transform of the convolution results, $\tilde{O}(\omega)$, satisfies Equation (21), which convert the integration in Equation (20) to multiplication of continuous Fourier transforms of f and G .

$$O(t) = \int_{-\infty}^{\infty} G(t-\tau)f(\tau)d\tau \equiv G(t) * f(t) \quad (20)$$

$$\tilde{O}(\omega) = \tilde{G}(\omega)\tilde{f}(\omega) \quad (21)$$

where O is accounted as the response of the continuous linear convolution, and symbol “ $*$ ” is the convolution operator; $\tilde{O}(\omega)$, $\tilde{G}(\omega)$, $\tilde{f}(\omega)$ are Fourier transformed results of O , G , and f , with ω for frequency corresponding to the domain of variable t . The tilde (\sim), means a 1D Fourier transform. For contact problems, f can be the excitation force, such as pressure, if it is a continuous function defined in the entire domain, and G is Green's function, both are function of space variable x . Then, frequency ω is corresponding to the space domain.

Periodic Convolution

If $D_p(i)$ and $f_p(i)$ are periodic functions with period N , the product of their discrete Fourier series (DFS) coefficients, $\tilde{D}_p(m)$ and $\tilde{f}_p(m)$, with m for the frequency, is \tilde{O}_p that can be expressed as (Oppenheim et al., 1999),

$$\tilde{O}_p(m) = \tilde{D}_p(m)\tilde{f}_p(m) \quad (22)$$

The periodic sequence, O_p , with the same period, N , is the periodic convolution analyzed in one period, N , shown below.

$$(O_p)_j = \sum_{r=0}^{N-1} (D_p)_{j-r} (f_p)_r \quad 0 \leq j \leq N-1 \quad (23)$$

For contact problems, $D_p(i)$ and $f_p(i)$ are $D_p(x_i)$ and $f_p(x_i)$. Here, the subscript can be negative.

Cyclic (Circular) Convolution

Based on Oppenheim et al. (1999), if $D_c(i)$ and $f_c(i)$ are sequences of finite length N , and their DFT results are $\hat{D}_c(m)$ and $\hat{f}_c(m)$, with m for the frequency, then their term-by-term product is \hat{O}_c , expressed as

$$\hat{O}_c(m) = \hat{D}_c(m)\hat{f}_c(m) \quad (24)$$

The finite sequence O_c is actually the circular convolution analyzed in length N , shown below.

$$(O_c)_j = \sum_{r=0}^{N-1} (D_c)_{j-r} (f_c)_r \quad 0 \leq j \leq N-1 \quad (25)$$

$$\text{Or } (O_c)_j = \sum_{r=0}^{N-1} (D_c)_{j-r+NH(r-j)} (f_c)_r \quad 0 \leq j \leq N-1 \quad (26)$$

where $H(r-j)$ is the Heaviside unit step function, which is 1 when $r-j$ is positive, or 0 otherwise. The term $NH(r-j)$ contributes to the subscript numbering only when the step function is not zero to avoid negative subscript. Here, subscript $-r$ reverses the sequence of the D series and subscript j shifts it in a circular fashion. More details has been given by Liu et al. (2000).

This means that the Fourier transform operation in Equation (24) is valid only when the convolution of $D_c(i)$ and $f_c(i)$ is in the form given in Equations (25) or (26). Although Equations (23) and (25) are for different events, they are expressed in the same form and lead to the same results in one period (Oppenheim et al., 1999).

The cyclic (circular) convolution is for series of finite lengths, and the circular fashion of the D series make it suitable for problems with either periodic features (e.g., nominally flat contact with a special IC treatment) or a circular configuration (e.g., cylindrical and journal bearing problems). Liu and Chen (2012) and Liu (2013) reported an FFT-based conformal-contact model for 2D problems with two concentric cylindrical interfaces, **Figure 1F**, for which we can directly apply the cyclic convolution theorem and 1D FFT operations to obtain the shaft deformation. However, for problems without the periodic features, such as counterformal line/point contact problems, special measures are needed to make the D series in such a needed circular fashion so that the cyclic convolution can be properly performed. The ICs and the pressure series can be properly handled based on the characteristics of the problems, e.g., Liu S. et al. (2000, 2007), Liu and Wang (2002), and Chen et al. (2008), so that they fit the need for the circular-convolution analyses. Because the FFT is a collection of algorithms for fast execution of DFT, the cyclic convolution of the two datum series (Equation 26) should also satisfy Equation (24) when the DFT is replaced by the FFT.

FFT ALGORITHMS FOR CONTACT MECHANICS

Cyclic Convolution and the DC-FFT Algorithm for Non-periodic Contact Problems

Consider the general line-contact displacement problem shown in Equation (3), subjected to a Hertzian pressure, or any localized pressure in a certain distribution, where influence coefficient $D(k, i)$ means u_z/C at node k caused by a unit pressure at node i , on a uniform grid of mesh spacing 2Δ . This is a problem of the convolution of two series of finite lengths; it is not infinite, nor periodic. Therefore, the cyclic convolution theorem,

Equation (26), should be applied in order to solve it with the Fourier transform method, for which the IC matrix has to be a cyclic matrix. This section explains how such a matrix is constructed from the original IC matrix via wrap-around order, and how this problem is solved properly and efficiently via the FFT. The wrap-around order requires one-to-one extension of the target domain on which the physical problem is defined, and the pressure on the extended domain should be set to zero (zero padding).

It should be mentioned that if such a problem were solved with the continuous convolution theorem, Equation (22), via FRF and Fourier transform of pressure in the finite target domain, a noticeable error would appear at the borders because it periodizes the problem mathematically. Ju and Farris (1996) depressed this error by five times domain extension. The analysis by Liu et al. (2000) indicated that a complete error removal would require 16 times domain extension, as if the problem were infinite. Error analyses will be discussed in a later section.

Because influence coefficient $D(k, i)$ only depends on the grid geometry, or, more specifically, the distance between points k and i , for a given uniform grid, it relies solely on $|k-i|$ no matter what k or i is. We can define X_{k-i} as the non-dimensional distance (normalized by a characteristic length, a) from X_k to X_i , or $X_{k-i} = X_k$ to X_i , then the IC component can be expressed as D_{k-i} . Subscript $k-i$ marks each element in the IC matrix. Obviously here for Equation (3), $X_{k-i} = -X_{i-k}$, and $D_{k-i} = D_{i-k}$, or $D_j = D_{-j}$. Using $\bar{X}_{k-i} = X_{k-i}/\bar{\Delta}$, $\bar{\Delta} = \Delta/a$, Equation (3) can be expressed as follows in a non-dimensional form,

$$\frac{u_z(X_k)}{ap_h C} = \sum_i \bar{\Delta} [(1 - \bar{X}_{k-i}) \ln |1 - \bar{X}_{k-i}| + (1 + \bar{X}_{k-i}) \ln |1 + \bar{X}_{k-i}|] + [2(\ln \bar{\Delta} + \ln a - 1)] P(X_i) \quad (27)$$

For example, if a and p_h are the Hertzian contact half width and maximum pressure, $P(X) = \sqrt{1 - X^2}$, and the Hertzian pressure distribution is then $p = P p_h$, and if the problem is digitized with $N_p = N_d = N = 5$ nodes, $k-i = [-4, 4]$, the non-dimensional matrix-form displacement for Equation (3), U , becomes,

$$\frac{-u_z(X_k)}{ap_h C} = \begin{bmatrix} D_0 & D_{-1} & D_{-2} & D_{-3} & D_{-4} \\ D_1 & D_0 & D_{-1} & D_{-2} & D_{-3} \\ D_2 & D_1 & D_0 & D_{-1} & D_{-2} \\ D_3 & D_2 & D_1 & D_0 & D_{-1} \\ D_4 & D_3 & D_2 & D_1 & D_0 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \\ P_2 \\ P_3 \\ P_4 \end{bmatrix} = \begin{bmatrix} U_0 \\ U_1 \\ U_2 \\ U_3 \\ U_4 \end{bmatrix} = U \quad (28)$$

The solution requires 5×5 , or $N \times N$, multiplication operations.

The IC matrix above is a Toeplitz matrix, or a diagonal-constant matrix. This matrix has to be converted to a cyclic one in order to utilize the cyclic convolution theorem (Liu and Wang, 2002). This can be done with the operation of the wrap-around order (Bracewell, 1978; Brigham, 1988; Press et al., 1992) by adding the reversed first column without the first element, which is $[D_4, D_3, D_2, D_1]$, to the end of the first row, $[D_0, D_{-1}, D_{-2}, D_{-3}, D_{-4}]$. Then, the extended first row becomes

$$[D_0 \ D_{-1} \ D_{-2} \ D_{-3} \ D_{-4} \ D_4 \ D_3 \ D_2 \ D_1] \quad (29)$$

The total number of the extended series is $N_c = 2N - 1 = 9$. By using the Heaviside step function notation, $j - r + N_c H(r - j)$, with $H(r - j) = 1$ for $r - j > 0$, Equation (29) becomes Equation (30). Series D_c in Equation (26) can be written as $[< D >]_{N_c = 2N - 1}$ to show the cyclic nature by $< >$ and the size information by the subscript. We can also use $[D]_{2N - 1}$ to express this extended wrap-around matrix. A short vertical bar is used to separate the original and extended terms for clarity.

$$\begin{aligned} [< D >]_{N_c = 2N - 1} &= [D_0 \ D_8 \ D_7 \ D_6 \ D_5 \ | \ D_4 \ D_3 \ D_2 \ D_1] \\ &= [D | EX]_{2N - 1} \end{aligned} \quad (30)$$

where EX means the extended components.

The nodal pressure vector is also extended by zero padding as follows

$$[P]_{N_c = 2N - 1} = [P_0 \ P_1 \ P_2 \ P_3 \ P_4 \ | \ 0 \ 0 \ 0 \ 0] = [P | EX]_{2N - 1} \quad (31)$$

Then Equation (3) becomes

$$\begin{aligned} \frac{-u_z(X_k)}{ap_h C} &= \sum_{i=0}^{N-1} (D_c)_{k-i+N_c H(i-k)} p_i \Leftrightarrow [< D >]_{N_c = 2N - 1} * [P]_{N_c = 2N - 1} \\ &= \text{IDFT} \left(\left[[< \hat{D} >]_{N_c = 2N - 1} \circ [\hat{P}]_{N_c = 2N - 1} \right] \right) \\ &= \text{IFFT} \left(\left[[< \hat{D} >]_{N_c = 2N - 1} \circ [\hat{P}]_{N_c = 2N - 1} \right] \right) \end{aligned} \quad (32)$$

where “ \circ ” means the operation of term-by-term type complex multiplication. Note that the equal sign with arrows indicates that the vector at the right-hand side contains exactly that at the left-hand side, but the former has extra useless terms in the extension. This is the expression for the discrete convolution (cyclic convolution) and fast Fourier transform (DC-FFT) algorithm, named by Liu et al. (2000), for the deformation calculation. Similar expressions can be obtained for stress calculations (Liu, 2001; Liu and Wang, 2002). Because the FFT operation of an N -number series is in the order of $N \log_2 N$, the operation of Equation (32) is in the order of $3N_c \log_2 N_c$, much smaller than that of the direct summation (DS) operation, which is $N \times N$, especially when N is large, as shown earlier in Introduction. Wang and Zhu (2019) offer a detailed numerical example, which shows that the direct summation, Equation (28), and the DFT-IDFT operation of the DC-FFT algorithm, Equation (32), lead to the same results in the analyzed accuracy.

In order to show the cyclic nature of this operation, the fully extended matrix, or the cyclic matrix, is completely constructed from $[< D >]_{N_c = 2N - 1}$ by circulating the last element in one row

to the first position in the next row, given below.

$$\begin{aligned} &\begin{bmatrix} D_0 & D_{-1} & D_{-2} & D_{-3} & D_{-4} & | & D_4 & D_3 & D_2 & D_1 \\ D_1 & D_0 & D_{-1} & D_{-2} & D_{-3} & | & D_{-4} & D_4 & D_3 & D_2 \\ D_2 & D_1 & D_0 & D_{-1} & D_{-2} & | & D_{-3} & D_{-4} & D_4 & D_3 \\ D_3 & D_2 & D_1 & D_0 & D_{-1} & | & D_{-2} & D_{-3} & D_{-4} & D_4 \\ D_4 & D_3 & D_2 & D_1 & D_0 & | & D_{-1} & D_{-2} & D_{-3} & D_{-4} \\ \hline & & & & & & EX & & & EX \end{bmatrix} \\ &= \begin{bmatrix} D_0 & D_8 & D_7 & D_6 & D_5 & | & D_4 & D_3 & D_2 & D_1 \\ D_1 & D_0 & D_8 & D_7 & D_6 & | & D_5 & D_4 & D_3 & D_2 \\ D_2 & D_1 & D_0 & D_8 & D_7 & | & D_6 & D_5 & D_4 & D_3 \\ D_3 & D_2 & D_1 & D_0 & D_8 & | & D_7 & D_6 & D_5 & D_4 \\ D_4 & D_3 & D_2 & D_1 & D_0 & | & D_8 & D_7 & D_6 & D_5 \\ \hline & & & & & & EX & & & EX \end{bmatrix} \\ &= \begin{bmatrix} IC & | & EX \\ \hline EX & | & EX \end{bmatrix} \end{aligned} \quad (33)$$

Only the top portion of the matrix is written out because they are related to the physical target domain of the original problem in the matrix operation. When using the DC-FFT algorithm, only the first row is needed, and only the first N rows of the IFFT results are the needed solutions while the others should be discarded.

IC wrap around order and pressure zero padding, shown in **Figure 3**, are two important operations for the DC-FFT algorithm to utilize the cyclic convolution and solve contact mechanics problems. Different implementation variations can be made; however, these two operations are necessary. The wrap-around order for deformation calculation can be done by shifting the negative side or flip those between $1 \sim N-1$. Caution should be paid for stress analyses because the shear-stress ICs are anti-symmetric. Sometimes, ICs also need zero padding, which should be done at node N where the IC is the smallest.

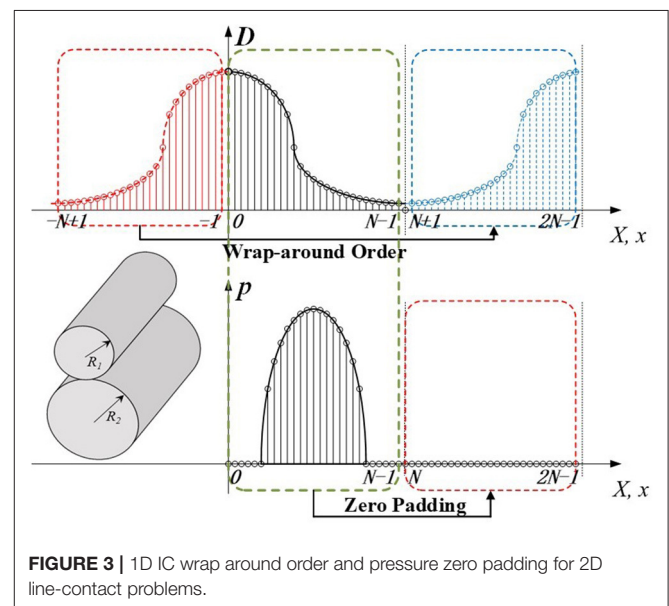


FIGURE 3 | 1D IC wrap around order and pressure zero padding for 2D line-contact problems.

Analogous to the line-contact problem discussed above, the IC matrix for a 3D point-contact problem can be constructed by extending the original physical domain in the two lateral directions, as shown in **Figure 4A**. Equation (9) becomes the following for a 3D contact problem, where a is a reference length, which could be the Hertzian radius of a spherical contact, and p_h a reference pressure, which could be the maximum Hertzian pressure.

$$U = \frac{-u_z(X_k, X_l)}{ap_h C} \Leftrightarrow [< D >]_{(2N_x)(2N_y)} * [P]_{(2N_x)(2N_y)} \\ = \text{IFFT} \left(\left[< \hat{D} > \right]_{(2N_x)(2N_y)} \left[\hat{P} \right]_{(2N_x)(2N_y)} \right) \quad (34)$$

where $N_{px} = N_{dx} = N_x$ and $N_{py} = N_{dy} = N_y$ are the numbers of nodes in the x and y directions, respectively.

The extended circular convolution IC matrix is

$$[< D >]_{(2N_x)(2N_y)} = \left[\left\langle \frac{D}{EX} \middle| \frac{EX}{EX} \right\rangle \right]_{(2N_x)(2N_y)} \quad (35)$$

where D on the right-hand side means the ICs corresponding to the physical domain.

It should be emphasized that two domains are involved, which are (1) the target domain, or the physical domain where the contact problem is defined, and (2) the extended domain, or the computation domain. At the end of the calculation, only the data for the target domain should be retained for the results. It is worth mentioning that the total lengths of P and D here are the minimum requirements; longer series than these should work, simply at a higher cost of computational efficiency. With these in mind, we can construct certain variations to implement this extension in order to apply the circular convolution for different situations, which are not discussed here.

The solution process involves IC matrix calculation, IC wrap-around order, pressure zero padding, and the FFT-IFFT. Its procedure is detailed as follow (Liu et al., 2000).

- 1) Calculate the influence coefficient matrix, $[D]_{2N_x \times 2N_y}$, from $-N_x$ to $N_x - 1$ in the x direction and $-N_y$ to $N_y - 1$ in the y direction;
- 2) Apply the wrap-around order and zero padding to convert the IC matrix into a cyclic matrix, $[< D >]_{2N_x \times 2N_y}$ in the calculation domain, marked as $[1 : 2N_x, 1 : 2N_y]$ (or $[0 : 2N_x - 1, 0 : 2N_y - 1]$), in the x and y directions, as shown in **Figure 4B**, and then apply the two-dimensional FFT to obtain the Fourier transformed IC matrix, $[< \hat{D} >]$;
- 3) Input pressure, conduct zero padding to convert pressure $[P]_{N_x \times N_y}$ into $[P]_{2N_x \times 2N_y}$ and then apply the two-dimensional FFT to get $[< \hat{P} >]$;
- 4) Obtain a temporary frequency series using element-by-element product of the two, i.e., $[< \hat{D} >] \circ [< \hat{P} >]$, where “ \circ ” means the operation of element complex multiplication;
- 5) Conduct two-dimensional IFFT $([< \hat{D} >] \circ [< \hat{P} >])$ to obtain the surface deformation and keep the result data within the original physical domain.

The error analyses by Liu et al. (2000) and Wang et al. (2003) have convinced that (1) the DC-FFT algorithm generates no additional inaccuracy beyond the discretization error, (2) its accuracy for solving elastic contact problems is nearly independent of the computation domain size accept for the necessary extension, and (3) the DC-FFT algorithm is the fastest among the commonly used contact analysis methods. **Figure 4B** presents a series of calculation results for the contacts of two honed rough surfaces subjected to several normal loads. The composite root-mean square (RMS) roughness is $R_q = 0.5$ micron. The contact ellipticity is $K = 2.0$, radii of curvature of the contact bodies are $R_x = 19.05$ mm, $R_y = 54.165$ mm, the equivalent elastic modulus is $E' = 226.4$ GPa, and the maximum Hertzian pressure is 2.72 GPa at the load of 7,680 N. No plasticity is considered in this set of analyses. The complementary conditions for contact modeling are given in the **Appendix**.

Continuous Convolution and Fourier Transform (CC-FT) for Nominally Flat-Flat Contact Problems

The contact of nominally flat but rough surfaces is a problem with an infinite domain, and it can also be considered as a periodic contact problem, i.e., the contact characteristics in a representative finite region repeat periodically in lateral directions. Since the FRF of Green's function exists and the periodic pressure distribution can be made into Fourier series, the continuous convolution theorem is applicable.

The continuous convolution and Fourier transform (CC-FT) algorithm has been suggested by Liu et al. (2000) and Liu S. et al. (2007) for this type of problems, which is so named because in the theoretical nature, the continuous convolution theorem is applied. If the FT in Equation (21) and the final IFT are replaced by the discrete Fourier transform and the inverse discrete Fourier transform (IDFT), which is actually that the FT divided by mesh interval is replaced by the DFT and the IFT multiplied by the mesh interval is replaced by the IDFT, the equation below can be executed directly. Here, the FFT and IFFT are applicable to execute the DFT and IDFT efficiently. The CC-FT algorithm is built upon the frequency response functions (FRFs) \tilde{G} and FFT of pressure, \hat{p} ,

$$U = \text{IDFT}[\tilde{G} \circ \hat{p}] \quad (36)$$

The FRFs are singular at the coordinate origin, which can be processed with the Gauss quadrature integration method. This CC-FT method is actually what Ju and Farris (1996) and others used before 2000 for non-periodic contact problems where it involved periodic errors. With the CC-FT algorithm, the solution can be computed only in a representative target domain as if this were one period of the rough surface laterally. Therefore, this method is highly efficient, as well as accurate, in computation.

FFT Algorithms for 3D Line-Contact Problems

The 3D line-contact problems involve a limited domain size in one of the lateral directions but a significantly long length in the other. Such a problem can be simplified as a 2D plane-strain issue

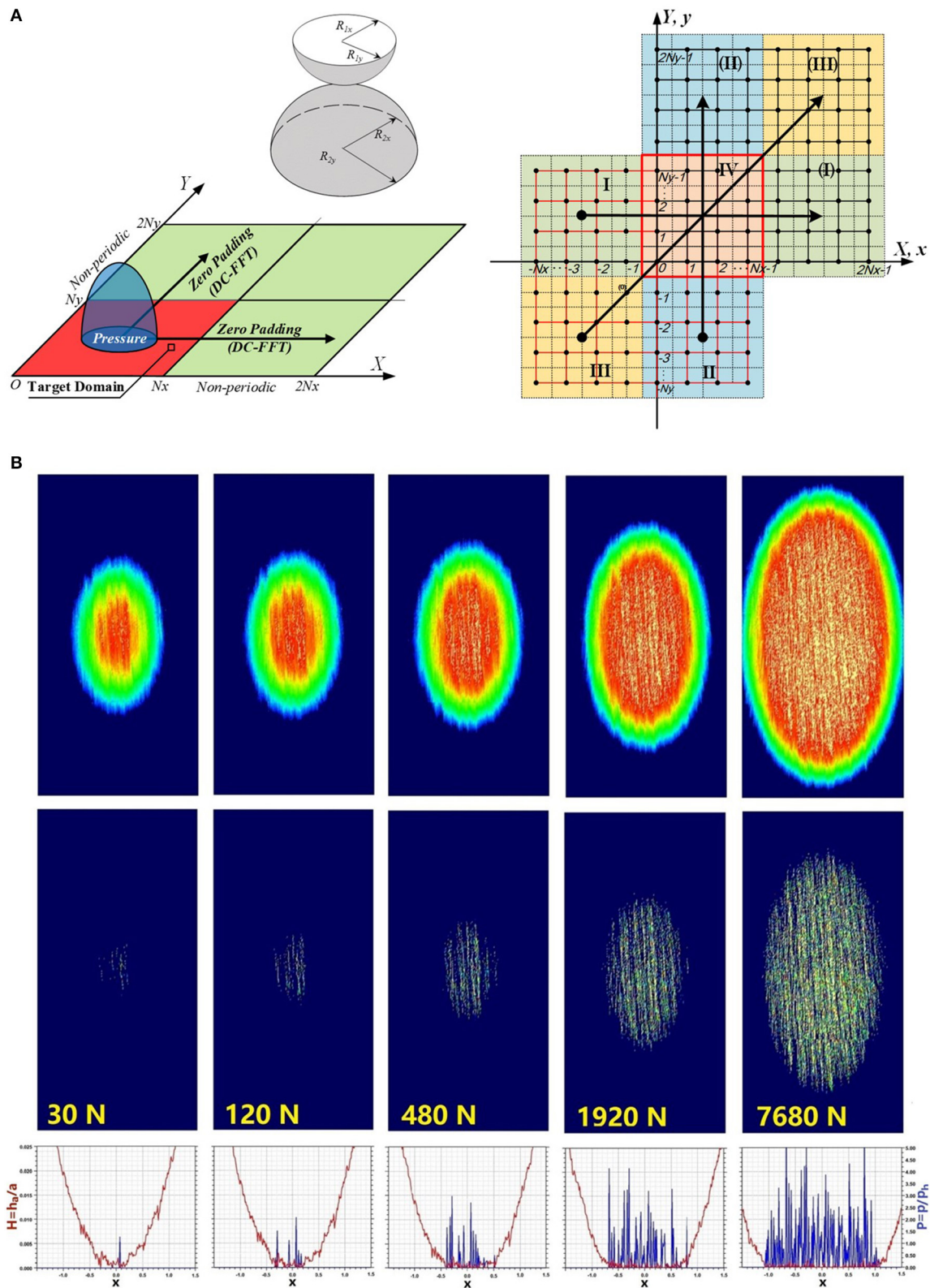


FIGURE 4 | 2D Wrap-around order in both x and y directions and zero padding for 3D point-contact problems **(A)**, and elliptical contacts of two rough surfaces under different normal loads **(B)**. In **(A)**, the arrows show the directions of the IC wrap-around order. In **(B)**, the light-colored patterns inside the blue (middle) and red (top) contours show asperity contact pressure and area.

if the contacting surfaces are ideally smooth and the materials homogeneous. To explore more detail, we are facing 3D problem with mixed issues. Three FFT-based approaches have been developed, to be discussed below, which are the algorithms of mixed discrete-continuous convolution with duplicated padding and FFT (DCD-FFT) (Chen et al., 2008), the hybrid discrete convolution, continuous convolution and FFT (DC-CC-FFT) (Liu and Hua, 2009), and the discrete-periodic convolution with IC summation and FFT (DCS-FFT) (Liu and Hua, 2009; Sun et al., 2020) to consider the effect of roughness and material inhomogeneity. They are mathematically and numerically the same in the finite direction (x), but different in the other dimension (y).

DCD-FFT Algorithm

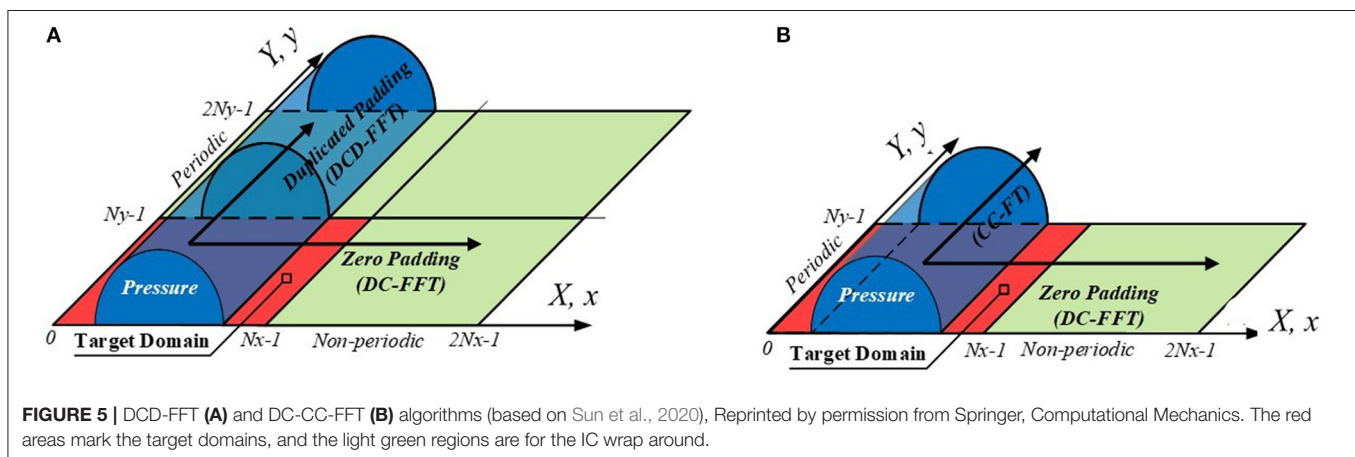
The DCD-FFT algorithm (**Figure 5A**) modifies the DC-FFT algorithm with a treatment in the y direction, simply by duplicating the pressures in the target domain to the extended region in the y direction, called duplicated padding (Chen et al., 2008). The length of the extended region should be at least the same as that of the target domain. This method is not as accurate as the DC-FFT one for non-periodic problems, especially when the extended domain in y is not sufficiently long, because it ignores the deformation influences from the region not included in the calculation, and the IC truncation error plays a role. Therefore, this method is mentioned here for a reference use only. However, the DCD-FFT method can effectively solve the normal deformation of finite-cylinder surfaces (a quarter-space problem), if the cylinders are not too short, with pressure duplication in the extended domain (Liu et al., 2020).

DC-CC-FFT Algorithm

The infinite extension of the problem in the y direction qualifies the direct use of the CC-FT method. Thus, the DC-CC-FFT algorithm combines the features of the discrete convolution theorem in the x direction, and the continuous convolution theorem in the y direction, shown in **Figure 5B**, involving hybrid FRFs and ICs (named ICs-FRF). Because there are two ways to obtain the FRFs for the y -direction solution, two variations of the DC-CC-FFT algorithm can be constructed.

DC-CC-FFT with IC-conversion. A simple way, with known ICs, described by Sun et al. (2020), to build the DC-CC-FFT algorithm is to process the 2D Fourier transform of the ICs sequentially in the y and the x directions to obtain ICs-FRF, where the FRF is from the IC conversion. After the y -direction FFT, the FRFs are calculated from Equations (14) and (17). Then the wrap-around order in the x direction is conducted, followed by the 1D Fourier transform in x . The calculation can be performed in the following steps with a slightly different procedure for the wrap around order.

- 1) Input the pressure, $[P]_{N_x \times N_y}$;
- 2) Extend the pressure $[P]_{N_x \times N_y}$ into $[P]_{2N_x \times N_y}$ with zero-padding in the x direction only;
- 3) Transform $[P]_{2N_x \times N_y}$ to $[\hat{P}]_{2N_x \times N_y}$ by applying 2D FFT, note here, the domain has been enlarged;
- 4) Calculate the IC matrix $[D]_{2N_x \times N_y}$;
- 5) Apply 1D FFT to $[D]_{2N_x \times N_y}$ in the y direction to get $[\hat{D}]_{2N_x \times N_y}$;
- 6) Calculate the ICs-FRF, $[\tilde{D}]_{2N_x \times N_y}$ by using Equations (14) and (17) in the y direction;
- 7) Treat $[\tilde{D}]_{2N_x \times N_y}$ with the wrap-around order only in the x direction and get $[\tilde{D}]_{2N_x \times N_y}$ which is only term flip because the domain has been enlarged in 4). Here, $[\tilde{D}]_{2N_x \times N_y}$ is the ICs after the wrap around order in the x direction but FRF in the y direction, and the latter is converted from the ICs done in the previous step;
- 8) Apply 1D FFT to $[\tilde{D}]_{2N_x \times N_y}$ in the x direction, and the resultant series is denoted as $[\hat{\tilde{D}}]_{2N_x \times N_y}$;
- 9) Obtain a temporary frequency series by element-by-element production between $[\hat{P}]_{2N_x \times N_y}$ and $[\hat{\tilde{D}}]_{2N_x \times N_y}$, which can be expressed as $[\hat{P} \circ \hat{\tilde{D}}]_{2N_x \times N_y}$, where “ \circ ” means complex multiplication in an element-by-element manner;



- 10) Apply a 2D IFFT to $\left[\langle \hat{\hat{D}} \rangle \right] \circ \left[\langle \hat{\hat{P}} \rangle \right]$;
- 11) Obtain the surface displacement by keeping values within the original target domain, and the resultant series are shown as $IFFT \left\{ \left[\langle \hat{\hat{D}} \rangle \right] \circ \left[\langle \hat{\hat{P}} \rangle \right] \right\}$.

The advantage of the algorithm of DC-CC-FFT with IC-conversion is that it does not require known FRFs if ICs can be calculated more easily. It is more accurate than the DCD-FFT method but still involves some IC truncation and conversion errors. The IC truncation error can be greatly reduced by using the IC summation method, to be discussed in section DCS-FFT Algorithm.

Precise DC-CC-FFT algorithm, presented by Liu and Hua (2009), is to conduct the analytical Fourier transform of Green's function in the length (y) direction, and calculate the ICs with respect to x , to obtain the hybrid ICs-FRFs, and then conduct the wrap-around order in x and the one-dimensional FFT of the new ICs-FRFs. This approach is theoretically accurate.

Both DC-CC-FFT algorithms only require extending the domain in the x direction twice the size of the target domain,

while the domain size in y is unchanged, which can be as small as possible, as long as it is sufficient represent the necessary features of surfaces and materials.

DCS-FFT Algorithm

Liu and Hua (2009) suggested another approach to deal with 3D line-contact problems, which is to consider the elasticity effect of the entire domain, or a sufficiently large domain, by IC summation, and Sun et al. (2020) have extended this method to solve the contact problems involving inhomogeneous materials. **Figure 6A** illustrates this idea. We can include M segments of equal length on each side of the target domain in the length direction, while making the target domain size as small as possible, as long as the main contact features are included. N_x in the x direction and N_y in the y direction may be different.

The ICs in each segment are the same, periodically shifted from those in the target domain. The influences of the pressures in the other segments on the deformation within the target domain can be included via the IC summation in the target domain, given below, with the superscripts for the segment numbers and the subscripts for the y direction coordinates in the

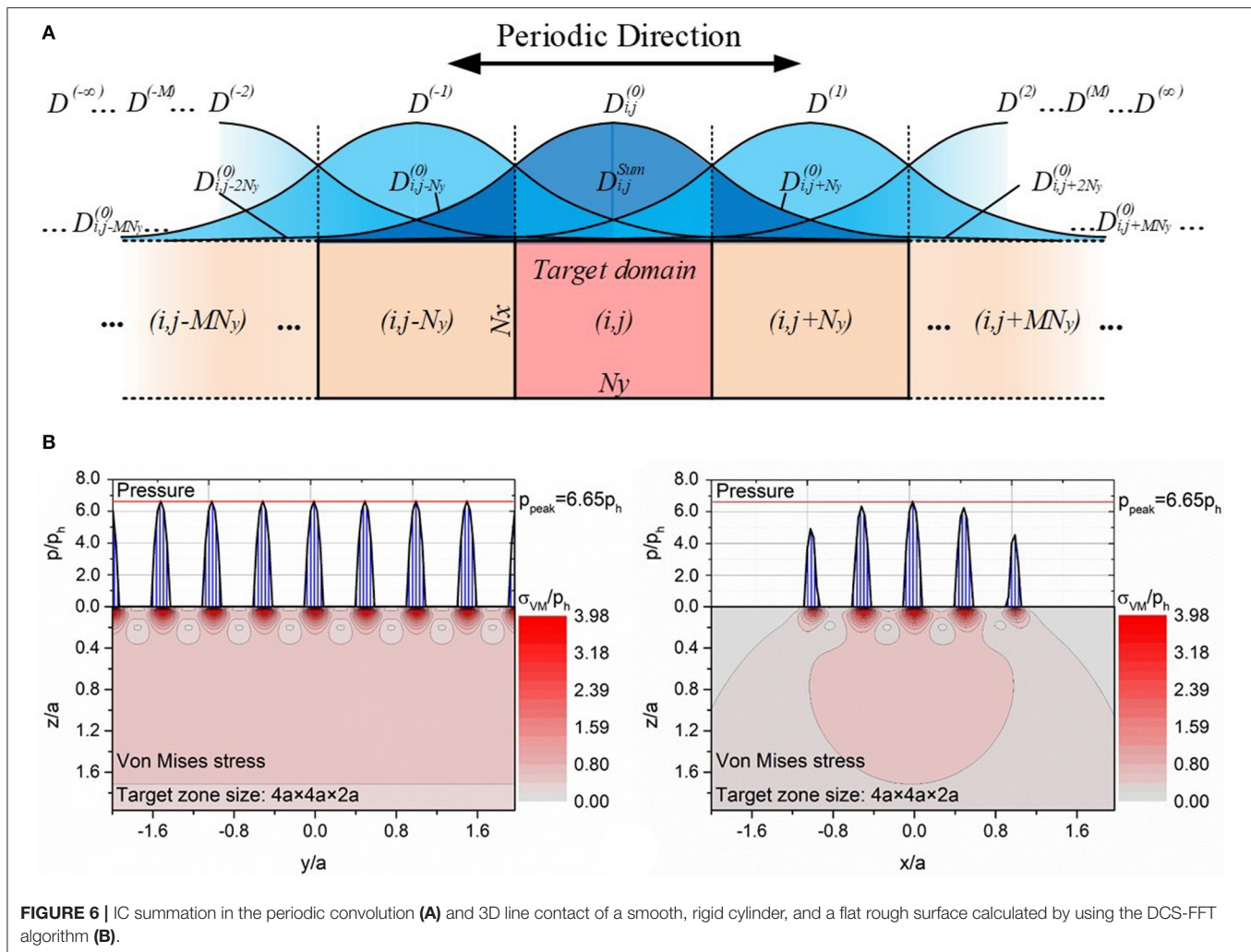


FIGURE 6 | IC summation in the periodic convolution (A) and 3D line contact of a smooth, rigid cylinder, and a flat rough surface calculated by using the DCS-FFT algorithm (B).

target domain.

$$D_{ij}^{Sum} = \sum_{r=-M}^M \left(D_{ij-rN_y}^{(r)} \right), \quad 0 \leq i \leq N_x - 1, 0 \leq j \leq N_y - 1 \quad (37)$$

Because the ICs in each segment are the same, only one copy of the IC, marked with superscript (0), is necessary, shown below with proper shifts for other copies.

$$D_{ij}^{Sum} = \sum_{r=-M}^M \left(D_{ij-rN_y}^{(0)} \right), \quad 0 \leq i \leq N_x - 1, 0 \leq j \leq N_y - 1 \quad (38)$$

Using D_{ij}^{Sum} in the DC-FFT algorithm should result in more accurate solutions to 3D line-contact problems than those by using the DCD-FFT algorithm. Of course D_{ij}^{Sum} can be used in the algorithm of DC-CC-FFT with IC-conversion discussed in the previous section. The solution procedure of the DCS-FFT algorithm can utilize the DC-FFT framework, which may involve the following steps:

- 1) Calculate the summed IC matrix, D_{ij}^{Sum} , from the ICs calculated in the region from $-N_x$ to N_x-1 in the x direction and from $-MN_y$ to MN_y-1 in the y direction, and construct $[D]_{N_x \times N_y}^{Sum}$;
- 2) Apply the wrap-around order, as well as zero padding if needed, to transfer the IC matrix into a cyclic matrix, $[< D >]_{2N_x \times 2N_y}^{Sum}$, similar to that in the DC-FFT algorithm, and then employ the 2D FFT to obtain the Fourier transformed IC matrix, $[< \hat{D} >]^{Sum}$;
- 3) Input pressure, conduct zero padding in the x direction and duplicated padding in the y direction to convert pressure $[P]_{N_x \times N_y}$ into $[P]_{2N_x \times 2N_y}$, and apply the 2D FFT to get the Fourier transformed pressure matrix, $[< \hat{P} >]$;
- 4) Obtain a temporary frequency series from the element-by-element complex product of the two, as $[< \hat{D} >]^{Sum} \circ [< \hat{P} >]$;
- 5) Obtain the surface deformation data from $IFFT \left\{ [< \hat{D} >]^{Sum} \circ [< \hat{P} >] \right\}$ and keep the resultant data within the original physical target domain.

Figure 6B presents the 3D cylindrical contact of a rigid infinite-length cylinder and an elastic half space material ($E = 200$ GPa, $\nu = 0.3$) with a sinusoidal rough surface, analyzed with the DCS-FFT algorithm. The 3D roughness is periodic in both the x and y directions, and the load is treated periodically in the y direction. With the DCS-FFT method, no edge effect appears at the borders of the target domain, which means that the elasticity effect of neighboring duplicated domains has been properly taken into account.

Compared with the DCD-FFT algorithm, the DCS-FFT method does not require a large target domain because of the

IC summation. As mentioned before, N_y can be very small, e.g., as small as one period of the pressure variation in the y direction, or as short as the length of a representative rough surface area and/or material inhomogeneity region. Among all the three algorithms for 3D line-contact problems, the DCS-FFT and the accurate DC-CC-FFT algorithms are recommended, and the former may be more preferred because it uses the same DC-FFT solution framework, convenient for programming, especially when the DC-FFT software is available as a set of the open-source codes (<http://othello.mech.northwestern.edu/qwang/OpenSourceCodeDCFFT/DC-FFTWeb.htm>). It should be mentioned that the DCS-FFT algorithm can also be used to construct a mechanism to solve the nominally flat-flat contact problems, named DCSS-FFT, which further modifies the ICs with 2D IC summations in both x and y directions (Sun, 2020).

DCR-FFT Algorithm

In comparison to the contact issues involving an excitation and response defined in a convolution, the relationship between eigenstrains and the components of the elastic field in a material involves both the convolution and the correlation (Liu and Wang, 2005; Liu et al., 2012), and the correlation theorem should be implemented as well. Here, eigenstrain is a generic term for various non-elastic strains, defined by Mura (1993), including thermal strain, plastic strain, fit-induced strain, phase transformation induced strain, and residual strain in a general sense. Many inhomogeneity problems can be solved via the Equivalent-eigenstrain method (EIM) (Eshelby, 1957), and correlations of functions are involved in the mathematical expressions of the eigenstrain-induced field.

Correlation Theorem

The Fourier transform of the correlation of two real datum series is the multiplication of the complex conjugate of Fourier transform of one function and the Fourier transform of the other.

$$\text{Correlation} \int_{-\infty}^{\infty} g(t + \tau) f(t) dt = R_{Gf}(\tau) \quad (39)$$

with $R_{Gf}(\tau) = R_{fG}(-\tau)$, and the Fourier transform is,

$$\tilde{R}_{Gf}(m) = \tilde{g}(m) \overline{\tilde{f}(m)} \quad (40)$$

where \tilde{f} is the complex conjugate of \tilde{f} .

DCR-FFT Algorithm

The DCR-FFT algorithm is an analogy to the DC-FFT algorithm, to be proven below, and it can be combined with the DC-FFT algorithm for a hybrid convolution-correlation operation to solve the inhomogeneity problem illustrated in **Figure 1D**. Details are given by Liu and Wang (2005), and applications can be found in the works by Liu et al. (2012), Wang et al. (2013a,b), Zhou et al. (2016), and Zhang M. Q. et al. (2018). A set of open-source codes can be downloaded from <http://othello.mech.northwestern.edu/qwang/OpenSourceCodeEigenstrainFiledHalfSpace/2012Web.htm>.

Let's use deformation as an example. Equation (41) (Liu et al., 2012) shows the link between eigenstrain $[e]$ in domain Ω and

surface deformation u_i in the $i = x, y$, or z direction, via a group of Green's functions, \mathbf{U}_i^s .

$$u_i = \frac{-1}{2\pi} \int_{\Omega} \mathbf{U}_i^s[e] d\mathbf{x}' \quad (41)$$

Similar to the expression of convolution related to potential function $R^I = \sqrt{(x-\xi)^2 + (y-\eta)^2}$ for the solution to contact elasticity, Equation (41) involves $R^I = \sqrt{(x-\xi)^2 + (y-\eta)^2 + (z-\zeta)^2}$ and $R = \sqrt{(x-\xi)^2 + (y-\eta)^2 + (z+\zeta)^2}$. Here, R has a plus sign inside the third term, which means a 1D cross correlation with $[e]$ on z for the solution.

The discrete form of the 1D cross correlation of influence coefficient D and $[e]$ in Equation (41) is given below

$$u_z = \sum_{i=1}^{N_e} D_{i+k} e_i \quad (42)$$

where N_e is the total number of nodes, and IC components D_{k+i} involves R .

More generally, for two series of complex numbers, f and g ,

$$R_{hg} = \int_{-\infty}^{\infty} \overline{f(t)} g(t+\tau) dt \quad (43)$$

where $\overline{f(t)}$ is the complex conjugate of $f(t)$. Equation (42) is equivalent to the following correlation,

$$R_{hg} = \int_{-\infty}^{\infty} \overline{f(t-\tau)} g(t) dt \text{ and } \tilde{R}_{hg} = \overline{\tilde{f}(t-\tau)} \tilde{g}(t) \quad (44)$$

$$\text{Or } R_{hg[n]} \equiv \sum_{m=-\infty}^{\infty} \overline{f[m-n]} g[m] \text{ and } \hat{R}_{hg} = \overline{\hat{f}(t-\tau)} \hat{g}(t) \quad (44')$$

This means that the convolution theorems should also be true to Equation (44) or (44'), no matter the series are real or complex. The numerical calculation of correlation can be done either by direct use of the correlation theorem (Equations 39–40) or the convolution given in Equation (44) (or Equation 44') with proper treatment of the two series, and the DC-FFT algorithm directly works for the latter. Caution should be given in conducting the Fourier transform.

By analogy, the DC-CC-FFT and CC-FT algorithms can all be extended to include the correlation operation for the analyses of the field due to eigenstrains.

Figure 7 shows a case of a cylinder in contact with the rough surface of an inhomogeneous half-space material (Sun et al., 2020), solved with the DCS-FFT algorithm. The representative piece of a ground surface is given in **Figure 7A** in a mesh of 128×128 , and $R_q = 1.18 \mu\text{m}$. A virtual ground rough surface subjected to the cylindrical contact is formed through periodically extending this patch along the y direction. The target

domain has the dimensions of $l_x \times l_y \times l_z$, and the length are all set to be $l_x = l_y = \frac{8a}{3}$, $l_z = \frac{4a}{3}$. All the cuboidal inhomogeneities are identical, with $a_x = a_y = a_z = 0.3a$, and they are distributed in the subsurface in the $x = 0$ plane in the depth of $z_d = 0.45a$. **Figure 7B** shows the 3D pressure distribution mapped on the contact surface, where sporadic pressure peaks can be found. The contours of the pressure and the von Mises stress distributions in the XOZ and $YOZOZ$ cross sections are given in **Figures 7C,D**. The 3D features of the pressure and stress are captured while their length-direction periodicities are also retained.

IC Conversion From FRFs

Figure 1E shows the contact involving a multi-layered material. The core analytical solutions to this type of problems are obtained, ignoring the body forces, by solving the governing differential equations in the frequency domain through the Fourier transform. The Fourier-domain solutions become the frequency response functions (FRFs) if surface tractions are unit valued. The conversion through Equations (13)–(18) leads to influence coefficients for the DC-FFT algorithm (Liu and Wang, 2002; Yu et al., 2014, 2016). Note that here, the direct inverse Fourier transform of the frequency-domain solution may result in inaccuracy when solving a non-periodic non-infinite contact problem (Liu et al., 2000).

The IC conversion method can be used to other contact cases as well. Our recent studies on the contacts involving magnetoelastoelectric or viscoelastic materials are all through the path of FRFs-ICs conversion and then the DC-FFT algorithm. Two such examples are given in **Figures 8, 9**.

Figure 8A shows the contact of two magnetoelastoelectric materials subjected to interfacial contact pressure, surface electric and magnetic charges. For this type of multifield material systems, the Fourier-domain solutions can be obtained by solving the coupled mechanical governing equations and the Maxwell equations in the frequency domain through the Fourier transform. The Fourier-domain solutions become the FRFs if the mechanical tractions, surface electric and magnetic charges are unit valued. The inverse Fourier transform through Equations (13) and (18) leads to influence coefficients to construct the DC-FFT algorithm. **Figure 8B** shows a case of an ellipsoid in contact with the surface of a magnetoelastoelectric material. The major and minor radii of the ellipsoid are $R_x = 300 \text{ mm}$ and $R_y = 200 \text{ mm}$, and the normal force is $P_z = 500 \text{ N}$. More details can be found in the work by Zhang X. et al. (2017, 2018, 2019).

Viscoelastic contact problems have drawn a great deal of attention (Goryacheva and Sadeghi, 1995; Chen et al., 2011; Putignano et al., 2015; Stepanov and Torskaya, 2018). It is also convenient to solve certain viscoelastic contact problems in the frequency domain. Here, two types of frequencies are involved, one with respect to time and the other to space. **Figure 9A** shows the sliding contact of a layer-substrate system, which actually represents four cases for each contacting bodies, (1) a viscoelastic layer on an elastic substrate, as shown, (2) a viscoelastic half space with the layer thickness $= \infty$, (3) an elastic half space with the layer thickness $= 0$, and (4) an elastic layer on a viscoelastic substrate by exchanging material properties under certain conditions, all solvable with the same model (Zhang et al.,

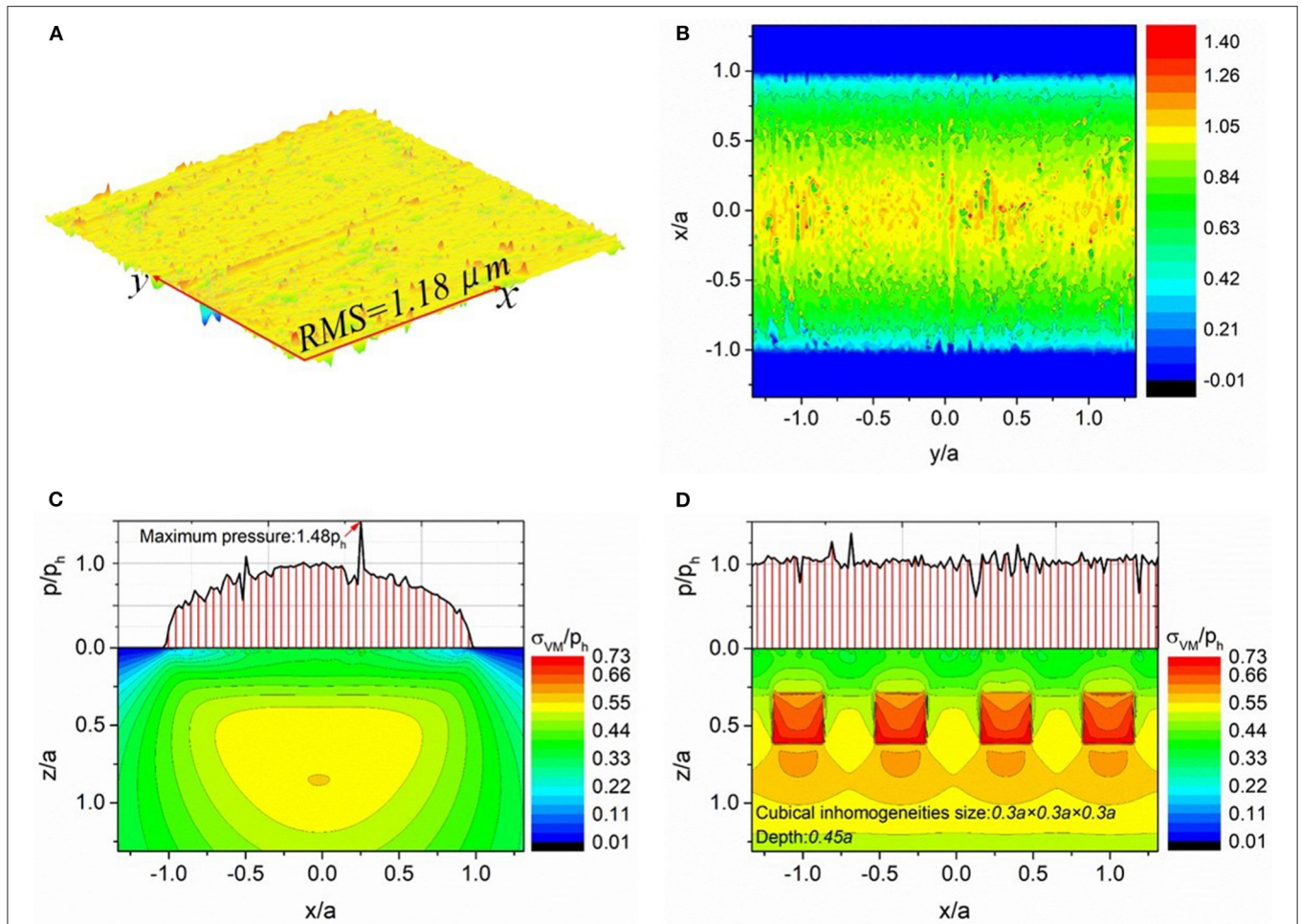


FIGURE 7 | Contact of a cylinder with inhomogeneous half-space and a rough surface **(A)** and inhomogeneities, 3D pressure distribution on the contact surface **(B)**, pressure and von Mises stress distributions in the XOZ plane **(C)**, and pressure and von Mises stress distribution in the YOZ plane **(D)**. The square inhomogeneities are shown in **(D)**. Note that **(C)** cuts through the center of the domain, where no inhomogeneity appears.

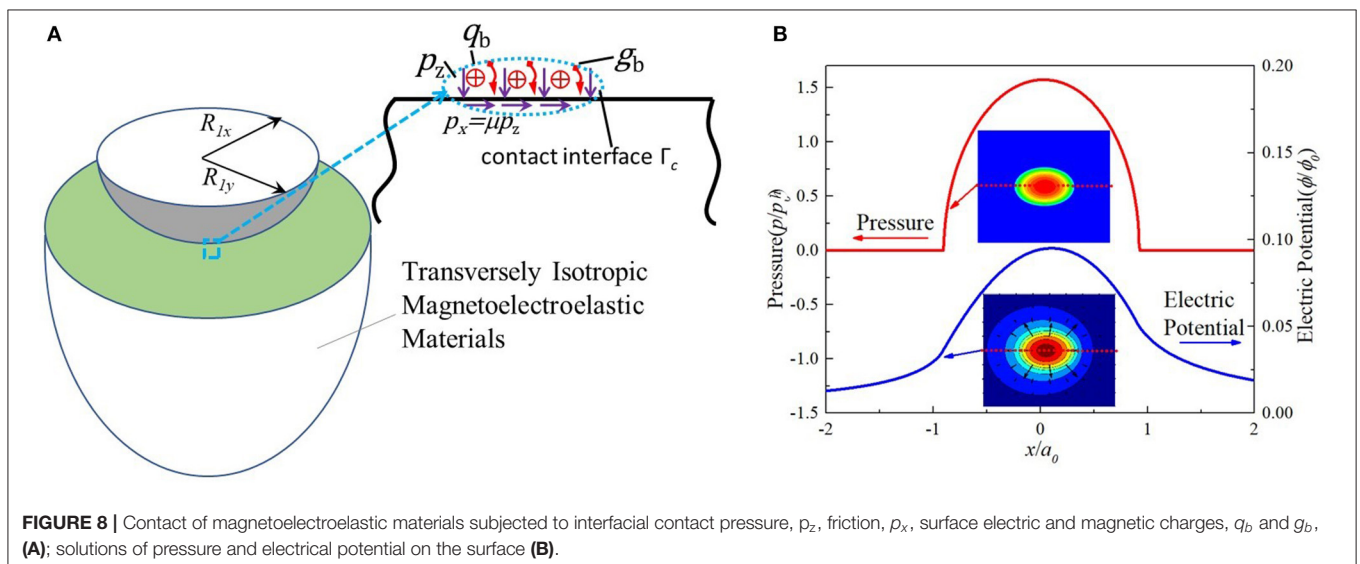
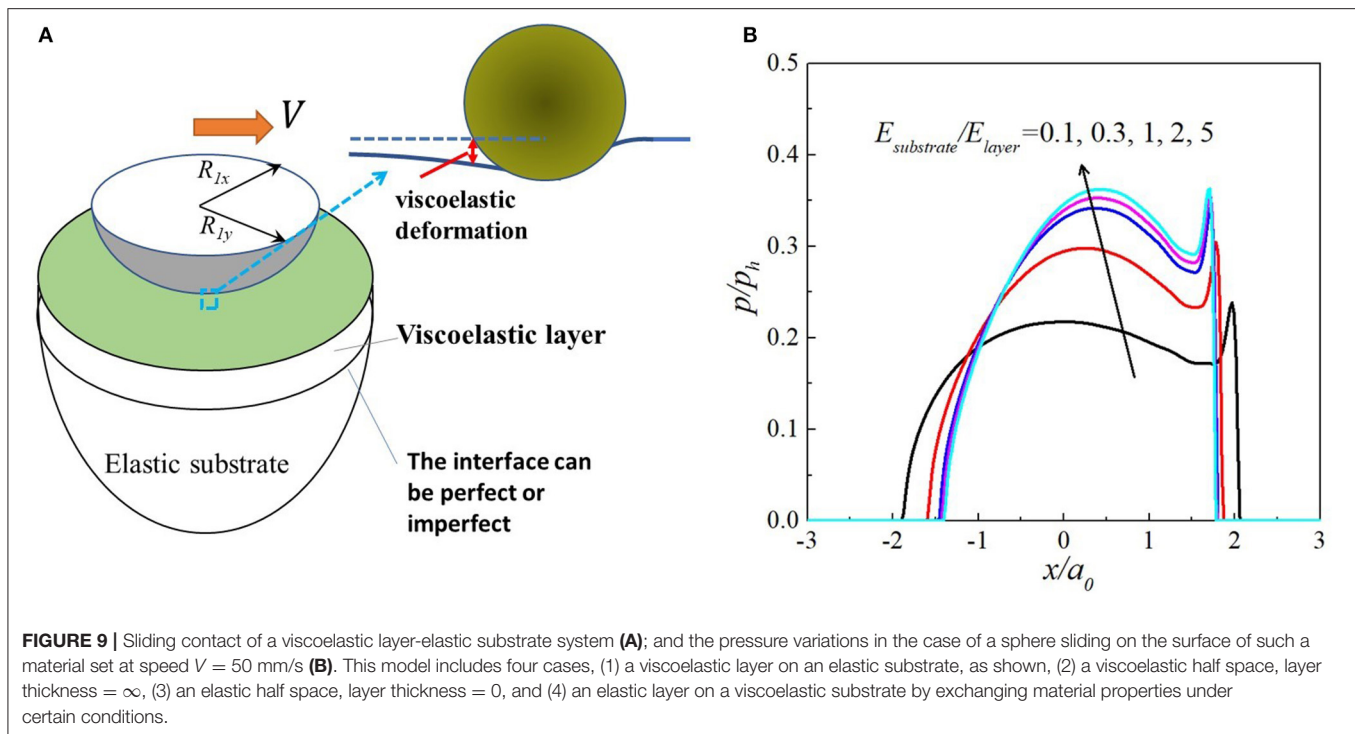


FIGURE 8 | Contact of magneto-electroelastic materials subjected to interfacial contact pressure, p_z , friction, p_x , surface electric and magnetic charges, q_b and g_b , **(A)**; solutions of pressure and electrical potential on the surface **(B)**.



2020a,b). The FRFs can be readily obtained from the elastic FRFs of elastic layered materials, by replacing the elastic modulus in elastic FRFs with the viscoelastic complex modulus. The inverse Fourier transform of the viscoelastic FRFs through Equations (13) and (18) leads to the viscoelastic influence coefficients to construct the DC-FFT algorithm in the viscoelastic contact framework. **Figure 9B** shows a case of a sphere sliding on the surface of a viscoelastic layer-elastic substrate system at a contact speed $V = 50$ mm/s. The layer thickness is 1 mm, the sphere radius is 10 mm, and the normal load is 1.48 N. Note that in **Figure 9A**, the interface between the layer and the substrate can be imperfect, and the spatial domain ICs have been solved by Wang et al. (2017a,b) and Zhang and Wang (2020). More studies related to viscoelastic contact solutions of layered materials with perfect or imperfect interfaces subjected to steady-state and transient conditions can be found in the work by Zhang et al. (2020a,b).

FFT With Non-uniform Mesh

In engineering systems, such as gears, journal bearings, and manufacturing tools, the contact regions can be large but the computation scale in a numerical simulation is limited. It is always a challenge to balance efficiency and accuracy. Take the line contact in a gear or a roller bearing as an example; the middle portion of the contact zone can be analyzed accurately with a coarse mesh, but the edges have to be modeled with a fine mesh in order to describe the drastic pressure variations there. The FEM deals with this type of issues with non-uniform meshes; however, the FFT-based methods built so far,

although efficient in a single mesh, are largely confined by the requirement of uniform grids. Sun (2020) has developed a method to extend the FFT-based algorithms to meshes of different densities, in which a finer mesh system is used in specific regions involving high pressure peaks while a coarser one is set in other regions under relatively smooth pressures. **Figure 10A** illustrates this non-uniform-mesh idea for a roller contact problem, where one side of the regions under the edge effects is meshed denser, while the other side is meshed with a coarse grid. In this particular example, the density of the fine grid is three times that of the coarse grid. The solution on the coarse-grid domain may be run for the entire region, depending on the size of the problem, but that on the fine-mesh domain is pursued just in a region somewhat larger than what it is designated. Extra data are discarded, and the joint deformations from the two meshes in both designated regions are used to evaluate the gap. This process involves overlapped calculations; however, for problems like the roller contact shown here, the larger the physical domain is, the more the saving of the computation time.

Figure 10B shows the comparison of the pressure calculated with a uniform mesh system and non-uniform meshes. The density of the uniform mesh is the same as that of the fine mesh of the non-uniform meshes. The pressure obtained on the effective zone of the fine mesh from the non-uniform mesh solution well matches that on the uniform fine mesh system. When the contact region is large and the pressure distribution is strongly non-uniform, the FFT method with non-uniform meshes offers a more efficient and flexible way to detail the

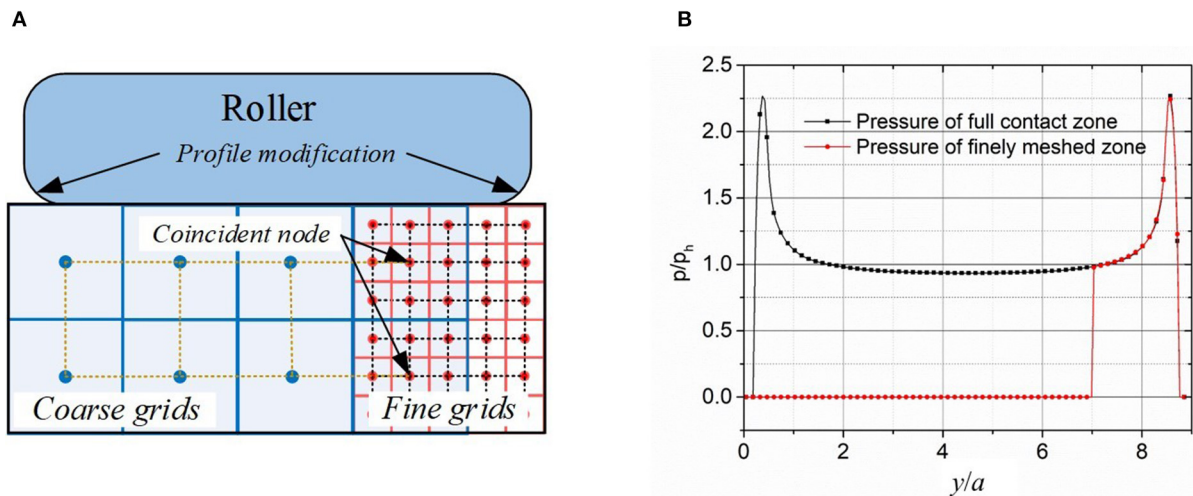


FIGURE 10 | Non-uniform mesh system used in the contact of a bearing roller with an end profile modification **(A)**, and a calculation result **(B)**.

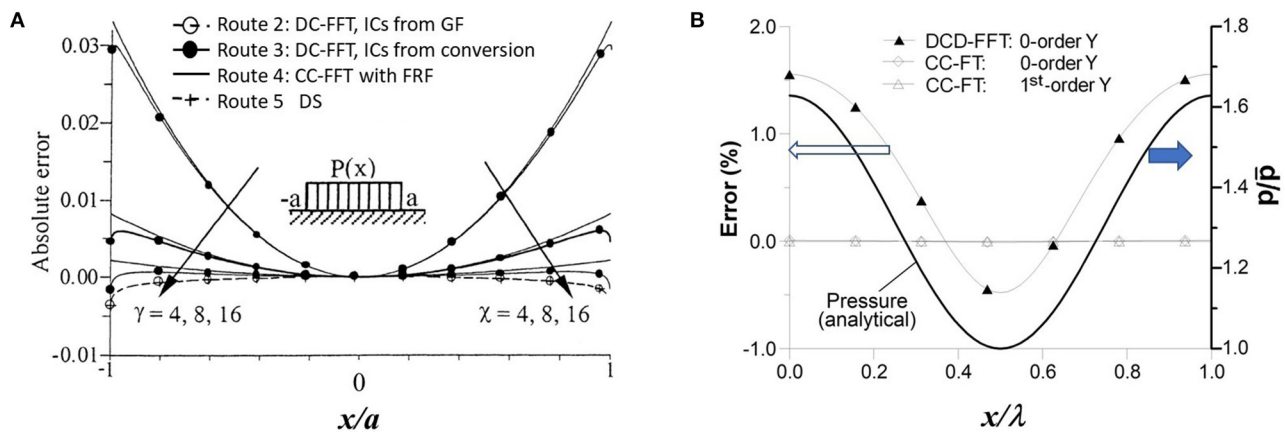


FIGURE 11 | Error comparisons. **(A)** DC-FFT and CC-FT algorithms in solving a finite-domain problem (Liu et al., 2000), Reprinted by permission from Elsevier, *Wear*, γ means the discretization density in the frequency domain and χ the ratio for spatial calculation domain extension, **(B)** CC-FT and DCD-FFT algorithms in solving the contact of a 3D sinusoidal wavy surface and a flat (Liu S. et al., 2007), Reprinted by permission from Elsevier, *Tribology International*; this is an infinite-domain, or periodic, problem. \bar{p} means the average pressure.

regions of special concerns, such as edges, interfaces, dents, and defects.

DISCUSSION AND CONCLUSIONS

Accuracy Comparisons

Figure 11A is for the results of a finite-domain problem with pressure acting on a region of $2a$ in length (Liu et al., 2000); it compares the errors from the DC-FFT algorithm structured in different routes with respect to the solution from the direct summation (DS) method. The continuous convolution and Fourier transform (CC-FT) algorithm is used to solve the same finite-domain problem, and its result is also compared. In this figure, the solution methods are named routes, which are Route 2:

DC-FFT with ICs from Green's function, Route 3: DC-FFT with ICs from conversion of frequency response functions, Route 4: CC-FT with frequency response functions, Route 5: the classic DS method with ICs from Green's function. Route 1 is not included here, which uses the FEM to calculate the ICs. The DC-FFT solution method (Route 2) appears to be accurate; its results overlap with those by the DS method. This means that the solution is not affected by the calculation domain size. Route 3, however, is different, depending on the discretization density in the frequency domain. The discrete influence coefficient from the frequency response function is the key for error control, the frequency domain sampling intervals, Δ_m and Δ_n , should satisfy $N_m \geq 2N_x$ and $N_n \geq 2N_y$, and Equations (16), or (17), or (18) should be implemented, followed by a proper wrap-around order (Liu and Wang, 2002).

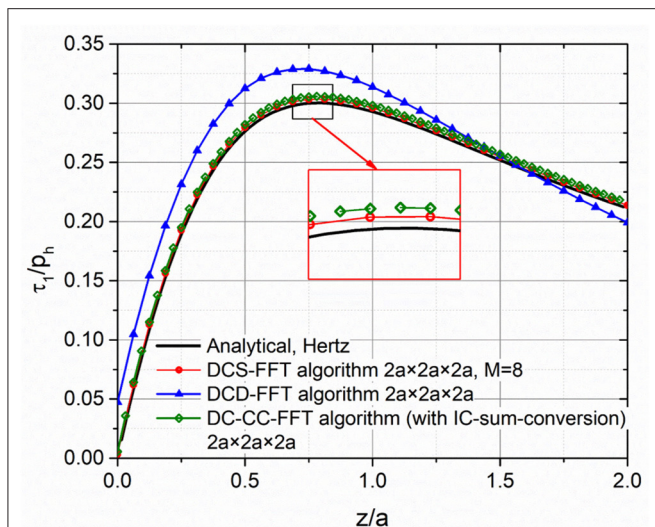


FIGURE 12 | Comparison of the algorithms of DCS-FFT, DCD-FFT and DC-CC-FFT with IC-sum conversion. The physical domains are marked as $x \times y \times z$, and $M = 8$ means the number of segments used in the IC preparation.

Figure 11B shows the behavior of the CC-FT algorithm in solving the periodic problem of a 3D sinusoidal wavy surface in contact with a flat; it compares the numerical solution with the corresponding analytical results (Liu S. et al., 2007). In addition, the DCD-FFT algorithm is also evaluated. It is evident that the CC-FT algorithm yields the solutions of high accuracy that is within machine precision. The CC-FT algorithm with either the zero-order or the first-order shape function yields nearly the same high accuracy. This indicates that a higher order shape function may be unnecessary if the algorithm is formulated properly.

The CC-FT algorithm is superb in analyzing the contact of nominally flat but actually rough surfaces.

Figure 12 compares the 3D cylindrical-contact algorithms of DCS-FFT, DCD-FFT, and DC-CC-FFT with IC-sum-conversion in calculating the maximum shear stress τ_1 , given below:

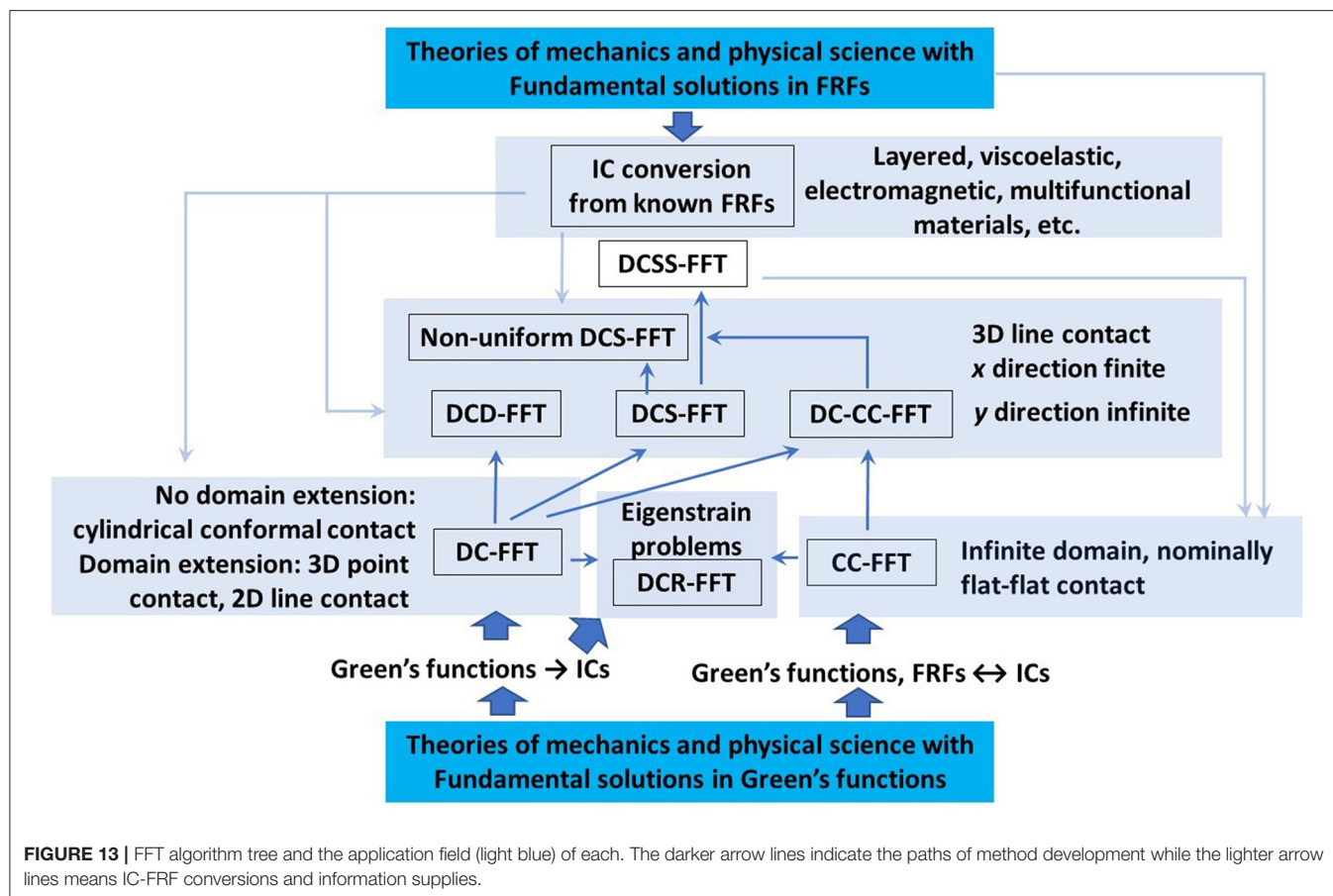
$$\tau_1 = \frac{1}{2} \sqrt{(\sigma_{11} - \sigma_{33})^2 + 4\sigma_{13}^2} \quad (45)$$

The identical target domain ($2a \times 2a \times 2a$) is used in all the three FFT-based methods. A large gap is shown between the result curves for the analytical and the DCD-FFT results, caused by the IC truncation error. The utilization of the DC-CC-FFT algorithm (IC-summation-conversion) greatly reduces the truncation error; however, some difference is still visible. The advantage of this DC-CC-FFT algorithm with the summarized ICs is the reduced computational burden because no duplicated-padding and wrap-around order are needed in the length direction. The DCS-FFT algorithm shows its accuracy and efficiency in dealing with this type of contact problems without knowing analytical FRFs. Theoretically the precise DC-CC-FT algorithm is more accurate (Liu et al., 2009), which is similar to the CC-FT shown in **Figure 11B**.

Range of Applications of the FFT Approaches

The FFT-based approaches are efficient. Many publications have shown the applications of these FFT algorithms on the contact analyses of elastic fields, plastic transition and yield, flash temperature, thermal stress, partial slip, and contact electrical and magnetic fields, dealing with science and mechanics issues for various systems from traditional mechanical components to emerging sensors and batteries (Zhang X. et al., 2019; Zhang et al., 2020c). The FFT algorithms discussed above are advantageous in solving the problems that are mathematically described in convolutions, correlations, combined convolutions and correlations, and in frequency response functions as well. The last one makes the FFT-based approaches more widely applicable and powerful because many governing differential equations can be more easily solved in the frequency domain through the Fourier transform, such as those for functionally graded materials (Ke and Wang, 2006), thermoelastically graded materials (Choi and Paulino, 2008), thermally graded materials (Zhang H. B. et al., 2018), materials with coupled stresses (Wang et al., 2020), and magnetoelectroelastic materials (Zhou and Lee, 2013). In addition, because the surface deformation analysis is an intermittent process of the elastohydrodynamic lubrication (EHL) calculation and related modeling, the FFT-based solutions are also building blocks in the models of mixed EHL in general (Liu et al., 2006, 2009), 3D line-contact EHL (Ren et al., 2009), finite-roller EHL (Zhang H. B. et al., 2017 with IC-overlapping DC-FFT; Liu et al., 2020 with DCD-FFT), coating EHL (Liu Y. et al., 2007 for single coatings, Wang et al. (2015) for multilayered coatings), plasto-EHL (Ren et al., 2010), EHL of inhomogeneous materials (Wang et al., 2014), wear in EHL (Zhu et al., 2007), EHL of transversely isotropic materials (Wang and Zhang, 2019), EHL of artificial joints (Wang and Jin, 2004), and EHL of 2D bearings (Liu and Chen, 2012). Moreover, adhesive contact problems can be solved with the FFT-based methods as well (Pohrt and Popov, 2015; Popov et al., 2017; Rey et al., 2017).

It is important to note that different contact types require different convolution theorems and thus different FFT algorithms. As summarized by Wang and Zhu (2019), contact with nominally flat surfaces should be tackled with the continuous convolution and Fourier transform (CC-FT) algorithm. Point-contact problems, either circular or elliptical, are non-periodic, and they should be formulated with the cyclic convolution and solved with the discrete convolution and FFT (DC-FFT) method utilizing zero padding and wrap-around order. Three-dimensional cylindrical (line) contact problems involve infinite domain extension in one direction but a finite domain width in the other, for which the discrete-continuous convolutions and FFT (DC-CC-FFT) and the discrete convolution with IC summation and FFT (DCS-FFT) methods should be used. The DCS-FFT algorithm is recommended if ICs are already known. Both the CC-FT and DCS-FFT algorithms are capable of solving periodic problems, and if the latter is used, the IC summation in the other direction is also needed, which makes it the DCSS-FFT algorithm mentioned in section DCS-FFT Algorithm. **Figure 13** summarizes the FFT algorithm



tree and the application field of each. The roots of this tree are the essential analytical solutions to mechanics and physical problems in the forms of Green's functions, ICs, or FRFs.

Recently, the fast Fourier transform method has been incorporated with the boundary-element method for solving problems with an arbitrary columnar geometry, not confined by the half-space assumption (Benad, 2018, 2019). This further extends the breadth of FFT applications. For the plane-strain problems associated with a columnar geometry formed by extruding a 2D shape in the length direction, e.g., the one solved by Benad (2018, 2019), the FFT solutions have a lower computational complexity, $O(n^3 \log n^{1.5})$, than the inversion of a standard BEM matrix, $O(n^4)$, where $n \times n$ is the total number of surface nodes. These problems are, mathematically, in the same nature as that in **Figure 2F**, automatically meeting the DC-FFT requirement with no need of the domain extension, as indicated in the first row of **Table 1** and **Figure 13**. Likewise, the FFT method should also be directly applicable to plane-stress problems of a disk-like geometry of any shape.

Limitations and Future Developments

The FFT-based methods mentioned here well fit the solutions of many engineering problems, not limited to mechanical contacts, as long as their model formulations contain convolution and/or correlation, or are solvable in the frequency domain, subjected to the assumptions of small deformation (linear or piecewise linear). So far, for counterformal contacts, the characteristic body

dimension, such as radius, should be much larger than that of the contact area; for conformal contacts, bodies involved should have axisymmetry or columnar geometries. Generally, these FFT-based solution approaches are confined by uniform meshes. Although section FFT With Non-uniform Mesh has briefly discussed the use of non-uniform meshes, more work is needed to make the non-uniform-mesh FFT algorithms more flexible and more efficient for contact analyses. Large deformation and the effect of body forces are also among the challenges to further developments of FFT-based methods for contact mechanics.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, open-source codes are available online in <http://othello.mech.northwestern.edu/qwang/OpenSourceCodeDCFFT/DC-FFTWeb.htm>; <http://othello.mech.northwestern.edu/qwang/OpenSourceCodeEigenstrainFiledHalfSpace/2012Web.htm>. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

QW leads the overall writing, LS is on some of the methods and results in Introduction, FFT Algorithms for 3D Line-Contact Problems, DCR-FFT Algorithm, FFT With Non-uniform Mesh, XZ on IC Conversion From FRFs, SL on

technical and algorithm review, and DZ on a part of Frequency Response Functions and ICs and rough surface contact results. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to express sincere gratitude to co-workers at the Center for Surface Engineering and Tribology, Northwestern University, Evanston, IL, USA, for years of

research on theoretical derivations of fundamental solutions, ICs, and FRFs, and developments and implementations of the FFT-based methods, especially to Drs. W. W. Chen, X. Q. Jin, W.-S. Kim, D. L. Li, Y. C. Liu, Z. Liu, A. Martini, N. Ren, M. J. Rodgers, X. J. Shi, Z. J. Wang, C. J. Yu, M. Q. Zhang, K. Zhou, and Q. H. Zhou, as well as Mr. H. Yu, Chongqing University, China, for their valuable contributions. The authors would also like to thank Professor L. M. Keer, Northwestern University, for long-term collaboration and research discussion, and the journal editor and reviewers for their valuable suggestions on manuscript revision.

REFERENCES

- Ai, X., and Sawamiphakdi, C. (1999). Solving elastic contact between rough surfaces as an unconstrained strain energy minimization by using CGM and FFT techniques. *ASME J. Tribol.* 121, 639–647. doi: 10.1115/1.2834117
- Benad, J. (2018). Efficient calculation of the BEM integrals on arbitrary shapes with the FFT. *Facta Univ. Ser. Mech. Eng.* 16, 405–417. doi: 10.22190/FUME180912034B
- Benad, J. (2019). Numerical methods for the simulation of deformations and stresses in turbine blade fir-tree connections. *Facta Univ. Ser. Mech. Eng.* 17, 1–15. doi: 10.22190/FUME190103008B
- Boucly, V., Nelias, D., Liu, S. B., Wang, Q., and Keer, L. M. (2005). Contact analyses for bodies with frictional heating and plastic behavior. *ASME J. Tribol.* 127, 355–364. doi: 10.1115/1.1843851
- Bracewell, R. N. (1978). *The Fourier Transform and Its Applications*. New York, NY: McGraw Hill.
- Brigham, E. O. (1988). *The Fast Fourier Transform and Its Applications*. Prentice Hall, NJ: Englewood Cliffs.
- Chen, W. W., Liu, S. B., and Wang, Q. J. (2008). Fast fourier transform based numerical methods for elasto-plastic contacts of nominally flat surfaces. *ASME J. Appl. Mech.* 75:011022. doi: 10.1115/1.2755158
- Chen, W. W., Wang, Q., Zhang, H., and Luo, X. (2011). Semi-analytical viscoelastic contact modeling of polymer-based materials. *J. Tribol.* 133:041404. doi: 10.1115/1.4004928
- Choi, H., and Paulino, G. (2008). Thermoelastic contact mechanics for a flat punch sliding over a graded coating/substrate system with frictional heat generation. *J. Mech. Phys. Solids* 56, 1673–1692. doi: 10.1016/j.jmps.2007.07.011
- Conry, T. F., and Seireg, A. (1971). A mathematical programming method for design of elastic bodies in contact. *ASME J. Appl. Mech.* 38, 387–392. doi: 10.1115/1.3408787
- Eshelby, J. D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc. A* 241, 376–396. doi: 10.1098/rspa.1957.0133
- Goryacheva, I. G., and Sadeghi, F. (1995). Contact characteristics of a rolling/sliding cylinder and a viscoelastic layer bonded to an elastic substrate. *Wear* 184, 125–132. doi: 10.1016/0043-1648(94)06561-6
- Hu, Y. Z., Barber, G. C., and Zhu, D. (1999). Numerical analysis for the elastic contact of real rough surfaces. *Tribol. Trans.* 42, 443–452. doi: 10.1080/10402009908982240
- Jin, X., Keer, L. M., and Wang, Q., and Chez, E. (2013). “Conjugate gradient method for contact analysis,” in *Encyclopedia of Tribology*, eds Q. Wang and Y. W. Chung (New York, NY; Heidelberg Dordrecht, London: Springer), 467.
- Johnson, K. L. (1987). *Contact Mechanics*. London: Cambridge University Press.
- Ju, Y. Q., and Farris, T. N. (1996). Spectral analysis of two-dimensional contact problems. *ASME J. Tribol.* 118, 320–328. doi: 10.1115/1.2831303
- Kalker, J. J. (1986). Numerical calculation of the elastic field in a half-space. *Commun. Appl. Num. Methods* 2, 401–410. doi: 10.1002/cnm.1630020412
- Kalker, J. J., and van Randen, Y. (1972). A minimum principle for frictionless elastic contact with application to non-hertzian half space contact problems. *J. Eng. Math.* 6, 193–206. doi: 10.1007/BF01535102
- Ke, L., and Wang, Y. (2006). Two-dimensional contact mechanics of functionally graded materials with arbitrary spatial variations of material properties. *Int. J. Solids Struct.* 43, 5779–5798. doi: 10.1016/j.ijsolstr.2005.06.081
- Liu, S. (2013). Numerical simulation of conformal contacts involving both interference and clearance. *Tribol. Transac.* 56, 867–878. doi: 10.1080/10402004.2013.806686
- Liu, S., and Chen, W. (2012). Two-dimensional numerical analyses of double conforming contacts with effect of curvature. *Int. J. Solids Struct.* 49, 1365–1374. doi: 10.1016/j.ijsolstr.2012.02.019
- Liu, S., Chen, W. W., Hua, D., and Wang, Q. (2007). Tribological modeling: application of fast fourier transform. *Tribol. Int.* 40, 1284–1293. doi: 10.1016/j.triboint.2007.02.004
- Liu, S., and Hua, D. Y. (2009). Three-dimensional semiperiodic line contact-periodic in contact length direction. *J. Tribol.* 131:021408. doi: 10.1115/1.3084237
- Liu, S., and Wang, Q. (2002). Studying contact stress fields caused by surface tractions with a discrete convolution and fast fourier transform algorithm. *ASME J. Tribol.* 124, 36–45. doi: 10.1115/1.1401017
- Liu, S., Wang, Q., and Liu, G. (2000). A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear* 243, 101–111. doi: 10.1016/S0043-1648(00)00427-0
- Liu, S. B. (2001). *Thermomechanical contacts of rough surfaces* [Ph. D. thesis]. Northwestern University, Evanston, IL, United States.
- Liu, S. B., Jin, X. Q., Wang, Z. J., Keer, L. M., and Wang, Q. (2012). Analytical solution for elastic fields caused by eigenstrains in a half-space and numerical implementation based on FFT. *Int. J. Plasticity* 35, 135–154. doi: 10.1016/j.iijplas.2012.03.002
- Liu, S. B., and Wang, Q. (2005). Elastic fields due to eigenstrains in a half-space. *ASME J. Appl. Mech.* 72, 871–878. doi: 10.1115/1.2047598
- Liu, S. B., Wang, Q., Rodgers, M. J., Keer, L. M., and Cheng, H. S. (2002). Temperature distributions and thermoelastic displacements in moving bodies. *Comput. Model. Eng. Sci.* 3, 465–481. doi: 10.3970/cmcs.2002.003.465
- Liu, Y., Chen, W., Liu, S., Zhu, D., and Wang, Q. (2007). An elastohydrodynamic lubrication model for coated surfaces in point contacts. *J. Tribol.* 129, 509–516. doi: 10.1115/1.2736433
- Liu, Y., Wang, Q., Hu, Y., Wang, W., and Zhu, D. (2006). Effects of differential schemes and mesh density on EHL film thickness in point contacts. *ASME J. Tribol.* 128, 641–653. doi: 10.1115/1.2194916
- Liu, Y., Wang, Q., Zhu, D., Wang, W., and Hu, Y. (2009). Effects of differential scheme and viscosity model on rough-surface point-contact isothermal EHL. *J. Tribol.* 131:044501. doi: 10.1115/1.2842245
- Liu, Z., Gu, T., Pickens, D., Nishino, T., and Wang, Q. (2020). Model assisted housing profile design for improved apex seal lubrication using a finite-length roller EHL model. *Rev. J. Tribol.*
- Londhe, N. D., Arakere, N. K., and Subhash, G. (2018). Extended hertz theory of contact mechanics for case-hardened steels with implications for bearing fatigue life. *ASME J. Tribol.* 140:021401. doi: 10.1115/1.4037359
- Love, A. E. H. (1929). Stress produced in a semi-infinite solid by pressure on part of the boundary. *Philod. Trans. R. Soc. A* 228, 337–420. doi: 10.1098/rsta.1929.0009
- Morrison, N. (1994). *Introduction to Fourier Analysis*. New York, NY: John Wiley and Sons.
- Mura, T. (1993). *Micromechanics of Defects in Solids*, 2nd Edn. Dordrecht: Kluwer Academic Publishers.
- Nogi, T., and Kato, T. (1997). Influence of a hard surface layer on the limit of elastic contact-part i: analysis using a real surface model. *ASME J. Tribol.* 119, 493–500. doi: 10.1115/1.2833525

- Oppenheim, A., Schaffer, R., and Buck, J. (1999). *Discrete-Time Signal Processing*, 2nd Edn, Pearson, Upper Saddle River, NJ.
- Pohrt, R., and Popov, V. L. (2015). Adhesive contact simulation of elastic solids using local mesh-dependent detachment criterion in boundary elements method. *Facta Univ. Ser. Mech. Eng.* 13, 3–10.
- Polonsky, I. A., and Keer, L. M. (1999). A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. *Wear* 231, 206–219. doi: 10.1016/S0043-1648(99)00113-1
- Polonsky, I. A., and Keer, L. M. (2000). Fast methods for solving rough contact problems: a comparative study. *ASME J. Tribol.* 122, 36–41. doi: 10.1115/1.555326
- Popov, V. L., Pohrt, R., and Li, Q. (2017). Strength of adhesive contacts: influence of contact geometry and material gradients. *Friction* 5, 308–325. doi: 10.1007/s40544-017-0177-3
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in Fortran 77-The Art of Scientific Computing*, 2nd Edn. Cambridge: Cambridge University Press.
- Putignano, C., Carbone, G., and Dini, D. (2015). Mechanics of rough contacts in elastic and viscoelastic thin layers. *Int. J. Solids Struct.* 69, 507–517. doi: 10.1016/j.ijsolstr.2015.04.034
- Ren, N., Zhu, D., Chen, W. W., Liu, Y., and Wang, Q. (2009). A three-dimensional deterministic model for rough surface line-contact EHL. *J. Tribol.* 131:011501. doi: 10.1115/1.2991291
- Ren, N., Zhu, D., Chen, W. W., and Wang, Q. (2010). Plasto-Elastohydrodynamic Lubrication (PEHL) in point contact. *J. Tribol.* 132:031501. doi: 10.1115/1.4001813
- Rey, V., Anciaux, G., and Molinari, J.-F. (2017). Normal adhesive contact on rough surfaces: efficient algorithm for FFT-based BEM resolution. *Comput. Mech.* 60, 69–81. doi: 10.1007/s00466-017-1392-5
- Stanley, H. M., and Kato, T. (1997). An FFT-based method for roughness surface contact. *ASME J. Tribol.* 119, 481–485. doi: 10.1115/1.2833523
- Stepanov, F., and Torskaya, E. (2018). Modeling of sliding of a smooth indenter over a viscoelastic layer coupled with a rigid base. *Mech. Solids* 53, 60–67. doi: 10.3103/S0025654418010077
- Sun, L. (2020). *Advanced FFT Algorithms for Contact of Materials with Inhomogeneities and Their Applications*. [Ph.D. thesis]. Northwestern Polytechnical University, Xian, China.
- Sun, L., Wang, Q., Zhang, M., Zhao, N., Keer, L. M., Liu, S. B., et al. (2020). Discrete convolution and FFT method with summation of influence coefficients (DCS-FFT) for three-dimensional contact of inhomogeneous materials. *Comput. Mech.* 65, 1509–1529. doi: 10.1007/s00466-020-01832-2
- Walker, J. S. (1996). *Fast Fourier Transforms*, 2nd Edn. Boca Raton, FL: CRC Press.
- Wang, F. C., and Jin, Z. M. (2004). Prediction of elastic deformation of acetabular cups and femoral heads for lubrication analysis of artificial hip joints. *Proc. Inst. Mech. Eng. Pt. J. J. Eng. Tribol.* 218, 201–209. doi: 10.1243/1350650041323331
- Wang, Q., and Zhu, D. (2019). *Interfacial Mechanics, Theories and Methods for Contact and Lubrication*. Boca Raton, FL; London, New York, NY: CRC Press.
- Wang, W.-Z., Wang, H., Liu, Y.-C., Hu, Y.-Z., and Zhu, D. (2003). A comparative study of the methods for calculation of surface elastic deformation. *Proc. Inst. Mech. Eng. Pt. J. J. Eng. Tribol.* 217, 145–152. doi: 10.1243/13506500360603570
- Wang, Y., Zhang, X., Shen, H., Liu, J., and Zhang, B. (2020). Couple stress-based 3D contact of elastic films. *Int. J. Solids Struct.* 191–192, 449–463. doi: 10.1016/j.ijsolstr.2020.01.005
- Wang, Z., Jin, X., Keer, L. M., and Wang, Q. (2013b). Novel model for partial-slip contact involving a material with inhomogeneity. *J. Tribol.* 135:041401. doi: 10.1115/1.4024548
- Wang, Z., Yu, H., and Wang, Q. (2017a). Layer-substrate system with imperfect bonding interface: coupled dislocation-like and force-like conditions. *Int. J. Solids Struct.* 122–123, 91–109. doi: 10.1016/j.ijsolstr.2017.06.004
- Wang, Z., Yu, H., and Wang, Q. (2017b). Layer-substrate system with an imperfectly bonded interface: spring-like condition. *Int. J. Mech. Sci.* 134, 315–335. doi: 10.1016/j.ijmecsci.2017.10.028
- Wang, Z. J., Jin, X. Q., Zhou, Q. H., Ai, X. L., Keer, L. M., and Wang, Q. (2013a). An efficient numerical method with a parallel computational strategy for solving arbitrarily shaped inclusions in elasto-plastic contact problems. *ASME J. Tribol.* 135:031401. doi: 10.1115/1.4023948
- Wang, Z. J., Yu, C., and Wang, Q. (2015). Model for elastohydrodynamic lubrication of multilayered materials. *J. Tribol.* 137:011501. doi: 10.1115/1.4028408
- Wang, Z. J., and Zhang, Y. (2019). An efficient numerical model of elastohydrodynamic lubrication for transversely isotropic materials. *ASME J. Tribol.* 141:091501. doi: 10.1115/1.4043902
- Wang, Z. J., Zhu, D., and Wang, Q. (2014). Elastohydrodynamic lubrication of inhomogeneous materials using the equivalent inclusion method. *J. Tribol.* 136:021501. doi: 10.1115/1.4025939
- Yu, C. J., Wang, Z. J., Liu, G., Keer, L. M., and Wang, Q. (2016). Maximum von mises stress and its location in trilayer materials in contact. *ASME J. Tribol.* 138:041402. doi: 10.1115/1.4032888
- Yu, C. J., Wang, Z. J., and Wang, Q. (2014). Analytical frequency response functions for contact of multilayered materials. *Mech. Mater.* 76, 102–120. doi: 10.1016/j.mechmat.2014.06.006
- Zhang, H. B., Wang, W. Z., Zhang, S. G., and Zhao, Z. Q. (2017). Elastohydrodynamic lubrication analysis of finite line contact problem with consideration of two free end surfaces. *J. Tribol.* 139:031501. doi: 10.1115/1.4034248
- Zhang, H. B., Wang, W. Z., Zhang, S. G., and Zhao, Z. Q. (2018). Semi-analytic solution of three-dimensional temperature distribution multilayered materials based on explicit frequency response functions. *Int. J. Heat Mass Transfer* 118, 208–222. doi: 10.1016/j.ijheatmasstransfer.2017.10.118
- Zhang, M. Q., Zhao, N., Wang, Z. J., and Wang, Q. (2018). Efficient numerical method with a dual-grid scheme for contact of inhomogeneous materials and its applications. *Comput. Mech.* 62, 991–1007. doi: 10.1007/s00466-018-1543-3
- Zhang, X., and Wang, Q. (2019). A SAM-FFT based model for 3d steady-state elastodynamic frictional contacts. *Int. J. Solids Struct.* 170, 53–67. doi: 10.1016/j.ijsolstr.2019.04.028
- Zhang, X., and Wang, Q. (2020). “Thermoelastic Contacts of Layered Materials with Interface Imperfections,” *Int. J. Mech. Sci.* 186:105904. doi: 10.1016/j.ijmecsci.2020.105904
- Zhang, X., Wang, Q., Harrison, K., Roberts, S., and Harris, S. J. (2019). Rethinking how external pressure can suppress dendrites in lithium metal batteries. *J. Electrochem. Soc.* 166, A3639–A3652. doi: 10.1149/2.0701914jes
- Zhang, X., Wang, Q., Harrison, K. L., Roberts, S. A., and Harris, S. J. (2020c). Pressure-driven interface evolution in solid state lithium metal batteries. *Cell Rep. Phys. Sci.* (2020) 1:100012. doi: 10.1016/j.xcrp.2019.100012
- Zhang, X., Wang, Q., and He, T. (2020a). Transient and steady-state viscoelastic contact response of layer-substrate systems with interfacial imperfections. *J. Mech. Phys. Solids* (2020).
- Zhang, X., Wang, Q., He, T., Liu, Y., Li, Z., Jin, H. J., et al. (2020b). Fully coupled modeling of thermo-viscoelastic contacts of layered materials. *Mech. Mater.*
- Zhang, X., Wang, Z. J., Shen, H. M., and Wang, Q. (2017). Frictional contact involving a multiferroic thin film subjected to surface magneto-electroelastic effects. *Int. J. Mech. Sci.* 131–132, 633–648. doi: 10.1016/j.ijmecsci.2017.07.039
- Zhang, X., Wang, Z. J., Shen, H. M., and Wang, Q. (2018). An efficient model for the frictional contact between two multiferroic bodies. *Int. J. Solids Struct.* 130–131, 133–152. doi: 10.1016/j.ijsolstr.2017.10.004
- Zhou, Q. H., Jin, X. Q., Wang, Z. J., Yang, Y., Wang, J. X., Keer, L. M., et al. (2016). A mesh differential refinement scheme for solving elastic fields of half-space inclusion problems. *Tribol. Int.* 93A, 124–136. doi: 10.1016/j.triboint.2015.09.009
- Zhou, Y., and Lee, K. (2013). Theory of sliding contact for multiferroic materials indented by a rigid punch. *Int. J. Mech. Sci.* 66, 156–67. doi: 10.1016/j.ijmecsci.2012.11.004
- Zhu, D., Martini, A., Wang, W., Hu, Y., Lisowsky, B., and Wang, Q. (2007). Simulation of sliding wear in mixed lubrication. *J. Tribol.* 129, 544–552. doi: 10.1115/1.2736439

Conflict of Interest: DZ is associated with Tri-Tech Solutions.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Wang, Sun, Zhang, Liu and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Contact Conditions

The contact of two bodies, (1) and (2), should satisfy the complementary “gap-G” and “flux-F” conditions, shown below, with 1, 2, 3 for x, y, z .

$$\begin{aligned} G_3(x_1, x_2) &> 0; & F_{33}^* &= 0 & \text{if } (x_1, x_2) \notin A_c \\ G_3(x_1, x_2) &= 0; & F_{33}^* &> 0 & \text{if } (x_1, x_2) \in A_c \end{aligned} \quad (\text{A1})$$

where $F_{33}^* = -p$ is the contact pressure, and

$$G_3 = \bar{B}^{(1)}(x_1, x_2) + u_3^{(1)}(x_1, x_2) + \bar{B}^{(2)}(x_1, x_2) + u_3^{(2)}(x_1, x_2) - \bar{g}$$

$$\text{with } g(x_1, x_2) = \left(\bar{B}^{(1)}(x_1, x_2) + u_3^{(1)}(x_1, x_2) \right) + \left(\bar{B}^{(2)}(x_1, x_2) + u_3^{(2)}(x_1, x_2) \right), \bar{g} = \frac{1}{N_c} \left[\sum_{(i,j) \in A_c} g(x_1, x_2) \right], \text{ where } u_3$$

is u_z given in Equation (3) or (10), $\bar{B}^{(i)}$ is for the geometry of body i , and N_c is the total nodes in the contact area A_c .

The contact area adjustment can be made through the following.

$$\begin{aligned} \text{If } G_3(x_1, x_2) < 0, & \text{ make } G_3(x_1, x_2) = 0 \text{ and} \\ & \text{add point } (x_1, x_2) \text{ to } A_c \end{aligned} \quad (\text{A2})$$

The overall the load balance is

$$\int_{A_c} F_{33}^* dA = \bar{P}_M \quad (\text{A3})$$

where \bar{P}_M is the normal load. The contact area and pressure distributions can be solved by using the conjugate gradient method (CGM) (Polonsky and Keer, 1999).

In addition, the contact should also satisfy the interfacial conditions for fluxes.

$$\begin{aligned} F_{3J}^* &= 0 & \text{if } (x_1, x_2) \notin A_c \\ F_{3J}^{*(1)}(x_1, x_2) &= -F_{3J}^{*(2)}(x_1, x_2) & \text{if } (x_1, x_2) \in A_c \quad J = 1, 2, 3 \end{aligned} \quad (\text{A4})$$

where $F_{3J}^{*(i)}(x_1, x_2) = \{\bar{t}_{M1}, \bar{t}_{M2}, -p\}$, and \bar{t}_{Mi} means the surface tangential tractions in the x_1 and x_2 directions. These conditions are also applicable to other multifield contact problems (Wang and Zhu, 2019).



Calculation of the BEM Integrals on a Variable Grid With the FFT

Justus Benad*

Department of System Dynamics and Friction Physics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Evgeny Viktorovich Shilko,
Institute of Strength Physics and
Materials Science (ISPMs SB
RAS), Russia
Shingo Ozaki,
Yokohama National University, Japan

*Correspondence:

Justus Benad
mail@jbenad.com

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 13 March 2020

Accepted: 30 April 2020

Published: 27 May 2020

Citation:

Benad J (2020) Calculation of the
BEM Integrals on a Variable Grid With
the FFT. *Front. Mech. Eng.* 6:32.
doi: 10.3389/fmech.2020.00032

In this study, an exemplary application of the *pFFT* is shown for the 2D Navier equation for a linear elastic continuum. Using this example, it is illustrated how the *pFFT* might be extended in order to decrease the computational complexity of the method. In the standard *pFFT* approach, all panel influences which are not calculated directly are obtained using a single regular grid. In the present study, a variable grid is suggested to obtain these influences. It is outlined how it is possible to apply the FFT on each level of this variable grid by rearranging segments of the shape boundary. A brief example is presented which indicates feasibility of the concept.

Keywords: boundary element method, aerodynamic panel method, *pFFT*, FFT, variable grid

INTRODUCTION

The Boundary Element Method (BEM) is used in countless engineering disciplines. Among them is the field of Contact Mechanics where the approach has been applied with great success in recent years (Putignano et al., 2012; Müser et al., 2017; Popov et al., 2017; Li et al., 2018; Paggi and Hills, 2020). Another, yet much older example for a very successful application of the Boundary Element Method is the Aerodynamic Panel Method. Conceived by Hess and Smith at *Douglas Aircraft* in the early 1960s (for details, see Smith, 1990), the Aerodynamic Panel Method was among the first numerical implementations of the Boundary Element Method when electronic computers became available (for details, see Cheng and Cheng, 2005). Today, more than 60 years later, the method is still an essential tool in the aircraft industry for initial design studies (Anderson, 2017; Raymer, 2018). An example is the investigation of entirely new airplane designs, such as the *Flying V* (Benad, 2015; Faggiano et al., 2017; Palermo and Vos, 2020; Rubio Pascual and Vos, 2020). In the very first conceptual design stage of this project, panel methods were used to assess the aircrafts aerodynamics. While they describe only potential flow (for details, see Katz and Plotkin, 2001), the two main advantages of such methods are their ease of use and their computational speed. From a numerical perspective, only the boundary of the domain needs to be discretized which leads to simple meshing. When no measures are taken to accelerate the calculation, the computational complexity of a problem with N surface discretization points is $O(N^3)$ when direct solvers are used, and $O(N^2)$ using iterative solvers. While an $O(N^2)$ complexity is by current standards considered as expensive, it is still practical for initial design studies with a low number of surface panels. However, if a high number of discretization elements is required, or large parameter studies shall be conducted,

one should consider accelerating the panel method. Several approaches exist for the reduction of the computational complexity of the Boundary Element Method (Phillips and White, 1997; Masters and Ye, 2004; Benedetti et al., 2008; Lim et al., 2008). An approach which has recently gained a lot of attention and has been very successful in the area of Contact Mechanics is to accelerate the calculation of the boundary integrals using the Fast Fourier Transformation (FFT) (Putignano et al., 2012; Pohrt and Li, 2014; Popov, 2017). It is natural perhaps, that this particular success should occur in the area of Contact Mechanics as one often considers a half-space surface in this field of research. A half-space surface aligns perfectly with a two-dimensional FFT-grid. Therefore, the calculation of the boundary integrals over the half-space can easily be performed with the FFT which reduces the computational complexity to $O(N \log N^{0.5})$. For arbitrary shapes however, this process is not as simple because the surface of an arbitrary shape does not align with a plane FFT-grid. In 1997, Phillips and White developed a method for utilizing the Fast Fourier Transformation even for the calculation of boundary integrals on arbitrary shapes (Phillips, 1997). As it is illustrated in Benad (2018), the boundary integrals can be regarded as convolutions over a space which is one dimension higher than that of the boundary of an arbitrary shape. However, there are two main challenges when this realization is put into practice: First, even a three-dimensional regular FFT-grid which fully encloses the surface of an arbitrary shape never cuts through this surface precisely at its discretization points. Instead, the nodes of the regular grid lie either outside of the arbitrary shape, or inside the shape, but in general never directly on the surface, as it is the case for the half-space in the area of Contact Mechanics. Second, the computational complexity rises, when compared to the half-space. While the half-space fully aligns with a two-dimensional FFT-grid, an arbitrary shape requires a fully enclosing FFT-grid, that is, a three-dimensional convolution. This results in a complexity of $O(N^{1.5} \log N^{0.75})$, which is lower than the $O(N^2)$ complexity of the standard boundary element procedure, but higher than the desired $O(N \log N^{0.5})$ complexity. Phillips and White addressed both these issues in 1997 with their precorrected FFT method, which has often been referred to as *pFFT*. Therein, nearby panel interactions are calculated directly as in the standard boundary element procedure, and the far field

is computed using the three-dimensional FFT-grid which fully encloses the surface. This FFT-grid may be much coarser than the surface mesh since higher order interpolation techniques are used to project the boundary values of the surface mesh points onto the nodes of the FFT-grid. This approach further reduces the $O(N^{1.5} \log N^{0.75})$ complexity. At best, it is an $O(N \log N)$ algorithm (Phillips and White, 1997).

It is remarkable how different the fields of application are, in which the *pFFT* approach can be used. When it was introduced by Phillips and White, it was applied in the field of electrostatic analysis. But of course, the method can also be regarded as a more general FFT-based BEM in the area of Contact Mechanics which enables not only the calculation displacements and stresses of a half-space with the FFT, but of any arbitrary shape: (Masters and Ye, 2004) used the *pFFT* approach for the calculation of displacements and stresses in linear elastic solids of arbitrary shape with the Navier equation achieving $O(N \log N)$ complexity. This makes the *pFFT* approach very appealing in the area of Contact Mechanics since the Boundary Element Method has proved to be a very robust tool in this area (Paggi and Hills, 2020), but is in its efficient practical application still somewhat restricted to the half-space. The *pFFT* may provide an opportunity to accelerate Boundary Element calculations of contact problems with a more difficult geometry. An exemplary application may be turbine blade fir-tree connections, a contact problem where the half-space theory is pushed to its limits (Benad, 2019). Naturally, the *pFFT* approach is not restricted to geometries where the domain of interest is the inner region which is enclosed by the boundary. Instead, it is also possible to consider the outer region. In 2006, Willis, Peraire and White utilized the *pFFT* to accelerate the Aerodynamic Panel Method achieving $O(N \log N)$ complexity (Willis et al., 2006). The study investigated complicated shapes, and even unsteady applications with varying geometries such as flapping wings.

It seems that many fields of research may benefit from fast Boundary Element solvers. In the present study, an exemplary application of the *pFFT* will be shown for the 2D Navier equation. Using this example, it is illustrated how the *pFFT* might be extended to allow the application of the FFT with multiple discretization levels. This may lead to a further reduction in computational complexity to obtain the boundary integrals.

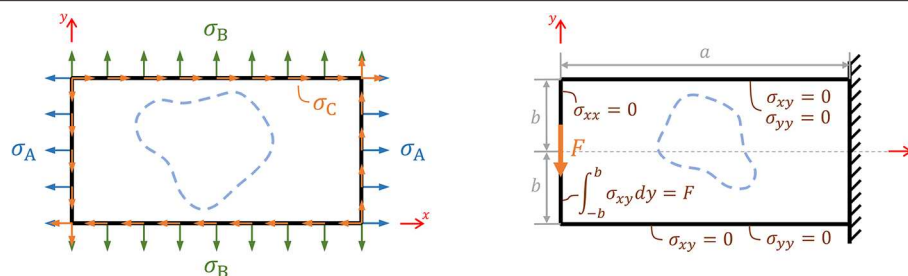
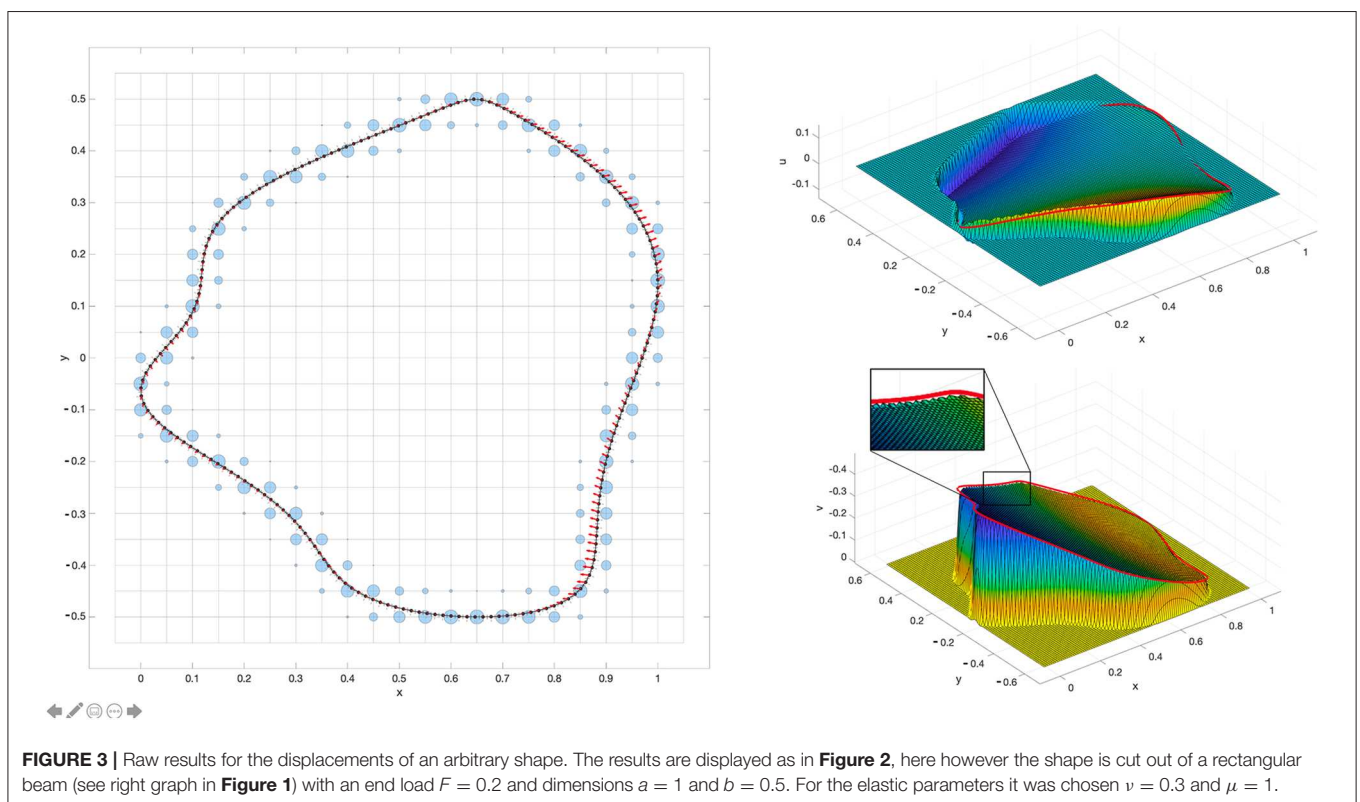
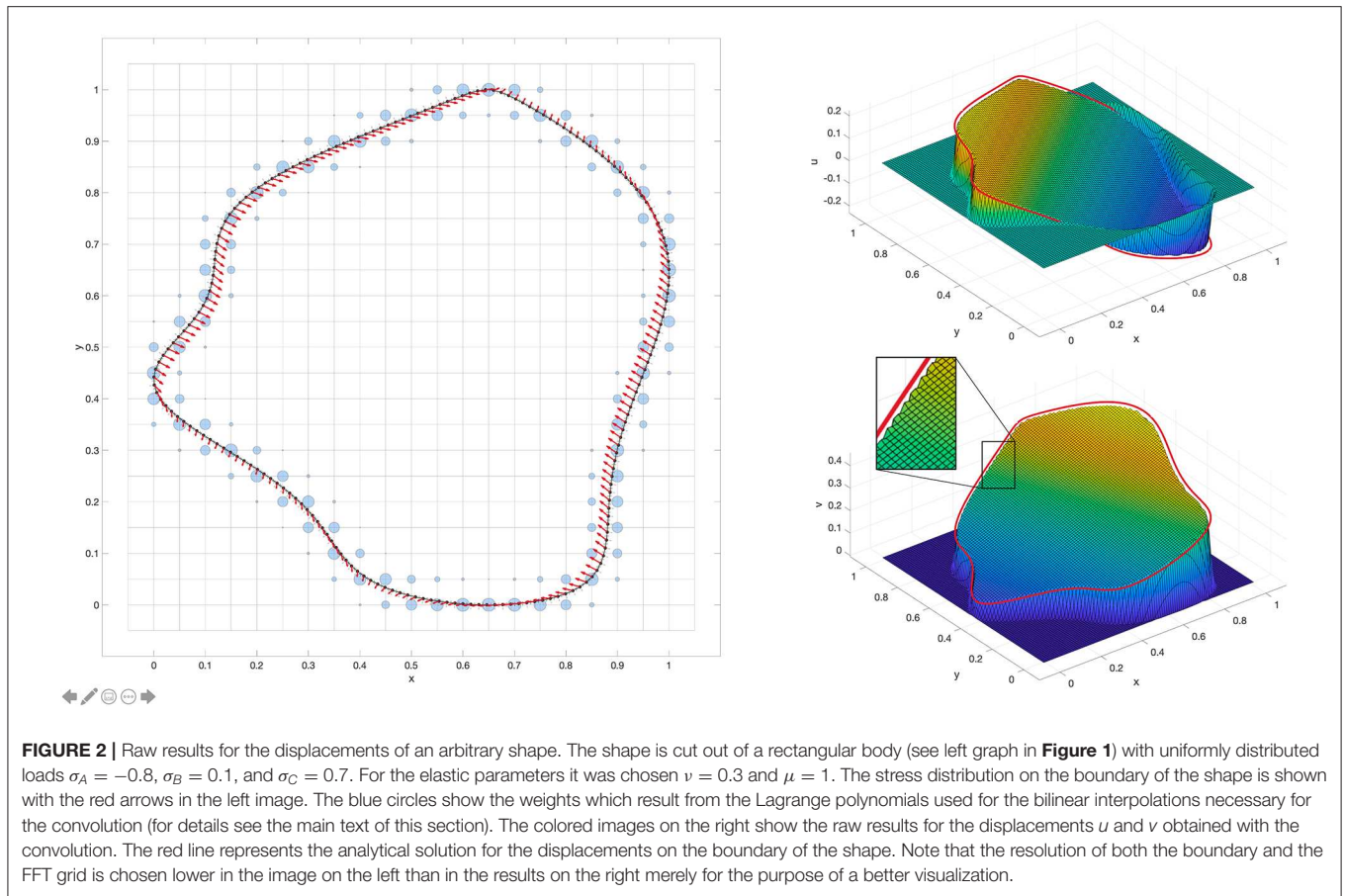


FIGURE 1 | Left: A rectangular body with uniformly distributed loads. Right: A rectangular beam with an end load.



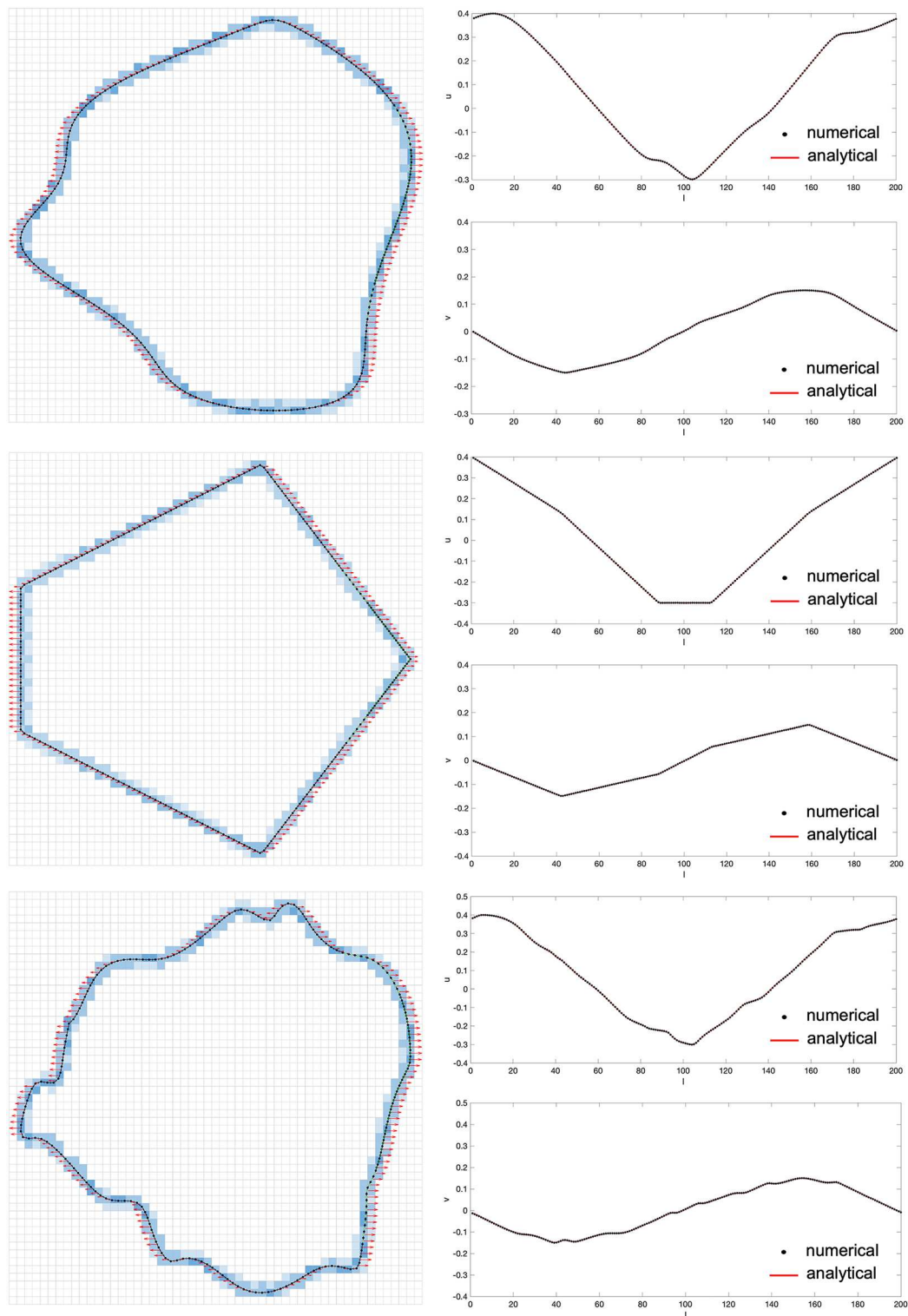


FIGURE 4 | Results for the boundary displacements of three different cutouts from a rectangular body (see left graph in **Figure 1**) with a uniformly distributed load $\sigma_A = 1$, $\sigma_B = 0$, and $\sigma_C = 0$. For the elastic parameters it was chosen $\nu = 0.3$ and $\mu = 1$. The given stress distribution on the boundary is displayed with red arrows (left images), and the numerical results for the displacements on the boundary are shown in the diagrams on the right with black dots plotted over the analytical solution. I denotes the number of the equally spaced boundary nodes.

GOVERNING EQUATION AND ANALYTICAL SAMPLE SOLUTIONS

Navier's equation is:

$$\frac{1}{1-2\nu} u_{j,ji} + u_{i,jj} + \frac{1}{\mu} b_i = 0 \quad (1)$$

where $i, j \in \{1, 2, 3\}$, u_i are the displacements, ν is the Poisson's ratio, μ is the shear modulus, and b_i is the force density field (Hahn, 1985). Note that we choose to denote partial derivatives with $\frac{\partial}{\partial x_i}(\dots) = (\dots)_{,i}$. For simplicity, we will in this work only consider the case of *plane displacements* where the Navier equation simply remains as in (1), but with $i, j \in \{1, 2\}$ (Irgens, 2008). This case can be transformed to the case of *plane stress* by replacing ν with $\frac{\nu}{(1+\nu)}$ and leaving the value for μ unchanged (Irgens, 2008), see also (Galin, 2008) and (Kelly, 2013). Before we proceed, let us introduce analytical solutions for (1) which can later be used to validate the numerical results.

Uniformly Distributed Loads on Rectangular Body

A solution of (1) is

$$\begin{aligned} u &= \frac{1-\nu}{2\mu} \sigma_A x - \frac{\nu}{2\mu} \sigma_B x + \frac{1}{2\mu} \sigma_C y, \\ v &= \frac{1-\nu}{2\mu} \sigma_B y - \frac{\nu}{2\mu} \sigma_A y + \frac{1}{2\mu} \sigma_C x. \end{aligned} \quad (2)$$

(Note that we choose to denote $u_{i=1} = u$, $u_{i=2} = v$, $x_{i=1} = x$ and $x_{i=2} = y$.) For plane displacements, the corresponding stress field is simply:

$$\sigma_{xx} = \sigma_A, \quad \sigma_{yy} = \sigma_B, \quad \sigma_{xy} = \sigma_C, \quad (3)$$

which belongs, for example, to a rectangular body with uniformly distributed loads on the sides, see the left image in **Figure 1**. This simple problem, as well as arbitrarily shaped cut-outs (see the blue dashed line), can be used for validation purposes.

Bending of a Rectangular Beam by an End Load

We now turn to another exact solution of (1), which is often used to approximate the solution for a system as shown in the right image in **Figure 1**. A stress field for this problem with boundary conditions imposed for the stresses in the weak form reads (see Barber, 2004):

$$\sigma_{xx} = \frac{3Fxy}{2b^3}, \quad \sigma_{yy} = 0, \quad \sigma_{xy} = \frac{3F(b^2 - y^2)}{4b^3}. \quad (4)$$

With the material law of a linearly elastic isotropic solid (Gross et al., 2014), one obtains:

$$\varepsilon_{xx} = \frac{1-\nu}{2\mu} \frac{3Fxy}{2b^3}, \quad \varepsilon_{yy} = -\frac{\nu}{2\mu} \frac{3Fxy}{2b^3}, \quad \varepsilon_{xy} = \frac{1}{2\mu} \frac{3F(b^2 - y^2)}{4b^3} \quad (5)$$

for the strains which belong to (4) for the case of plane displacements. Integration yields:

$$\begin{aligned} u &= \frac{Fy}{8\mu b} \left(6 + (1-\nu) \frac{3x^2}{b^2} + (\nu-2) \frac{y^2}{b^2} \right) - C_1 y + C_3 \\ v &= \frac{Fx}{8\mu b} \left(-(1-\nu) \frac{x^2}{b^2} - \nu \frac{3y^2}{b^2} \right) + C_1 x + C_2. \end{aligned} \quad (6)$$

Setting $u(x=a, y=0) = 0$ yields $C_3 = 0$. With $u(x=a, y=b) = 0$ one further obtains $C_1 = \frac{F}{8\mu b} \left(4 + (1-\nu) \frac{3a^2}{b^2} + \nu \right)$ and $v(x=a, y=0) = 0$ yields $C_2 = -\frac{Fa}{8\mu b} \left(4 + (1-\nu) \frac{2a^2}{b^2} + \nu \right)$.

The result (6) for the displacements is an *exact* solution of (1) and may thus be used together with the corresponding stress field (4) as an analytical reference for comparison with the numerical results.

Note that at the boundary $x = a$ the strains ε_{yy} in (5) are not zero as we would expect them to be for the beam which is displayed in the right image of **Figure 1**. Imposing the boundary conditions merely for the stresses in the weak form (see Barber, 2004) only gives an approximate solution for this set-up. The solution is fully accurate for a beam which is supported not by a solid wall, but a parabolic traction.

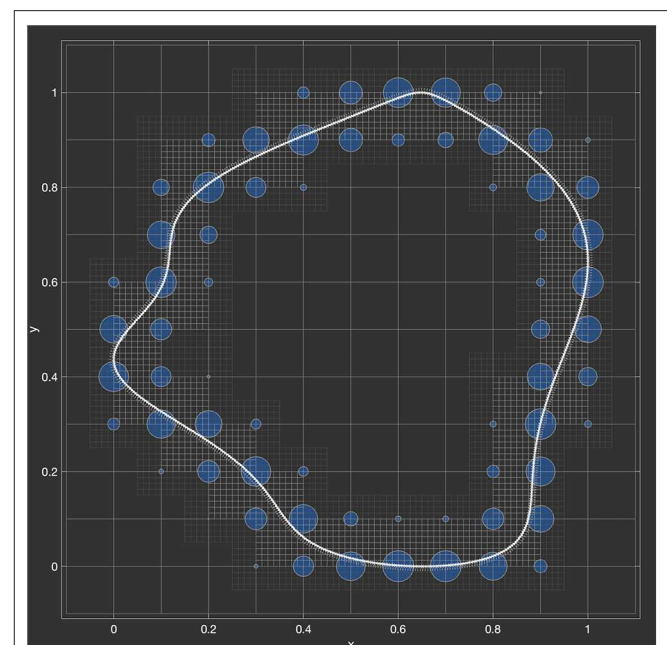


FIGURE 5 | An arbitrary shape fully enclosed with a grid which is refined in close proximity to the boundary of the shape. The blue dots represent the weights of the Lagrange polynomials used for the bilinear interpolation of the boundary values with the coarse grid.

BOUNDARY INTEGRAL FORMULATION

For the governing equation (1), Somigliana's identity (Gaul and Fiedler, 2013) relates the displacements $u_j(\underline{x})$ and stress vector $t_j(\underline{x}) = \sigma_{jk}(\underline{x}) n_k(\underline{x})$ on the boundary S (outward normal vector n_k) to the displacement $u_i(\underline{x}_0)$ at a particular inner point \underline{x}_0 . For $b_i = 0$, it is:

$$u_i(\underline{x}_0) = \int_S u_{ij}^*(\underline{x}, \underline{x}_0) t_j(\underline{x}) dS - \int_S t_{ij}^*(\underline{x}, \underline{x}_0) u_j(\underline{x}) dS \quad (7)$$

For the case of plane displacements which we consider in this study, u_{ij}^* and t_{ij}^* read:

$$u_{ij}^* = \frac{-(3-4\nu) \ln(r) \delta_{ij} + (x_i - x_{0i})(x_j - x_{0j})/r^2}{8\pi\mu(1-\nu)} \quad (8)$$

and

$$t_{ij}^* = -\frac{1}{4\pi(1-\nu)r} \left(\left((1-2\nu) \delta_{ij} + 2 \frac{(x_i - x_{0i})(x_j - x_{0j})}{r^2} \right) \frac{x_k - x_{0k}}{r} n_k \right) + \left((1-2\nu) \left(\frac{x_j - x_{0j}}{r} n_i - \frac{x_i - x_{0i}}{r} n_j \right) \right) \quad (9)$$

where $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$. Note that we choose to denote $x_{i=1}$ as x and $x_{i=2}$ as y . Relations (8) and (9) can be obtained from the *fundamental solution* of (1). For details see (Benad, 2019).

CALCULATION OF THE BOUNDARY INTEGRALS WITH THE FFT ON A SINGLE GRID

We now perform exemplary calculations of the boundary integrals (7) with the FFT on a single regular grid. Therein, we follow the procedure described in Phillips and White (1997). Additional information on the numerical implementation can be found in Benad (2018). The numerical results are then compared with the analytical sample solutions which were introduced in Section Governing Equation and Analytical Sample solutions.

Let us first examine the raw data which is obtained with the method right after the application of the convolutions. **Figures 2, 3** show this data for an arbitrary cut out from the rectangular body with uniformly distributed loads, and for an arbitrary cut out from the rectangular beam with an end load. The analytical solutions for these problems given with (2–4, 6) set the boundary values on the chosen shape. These are then distributed to the



FIGURE 6 | Segments of the arbitrary shape displayed in **Figure 5** aligned one after the other. Each segment (fine bright grid) is chosen with some additional space around it (fine dark grid) so as to assure that the segments do not influence each other upon convolution with the kernels.

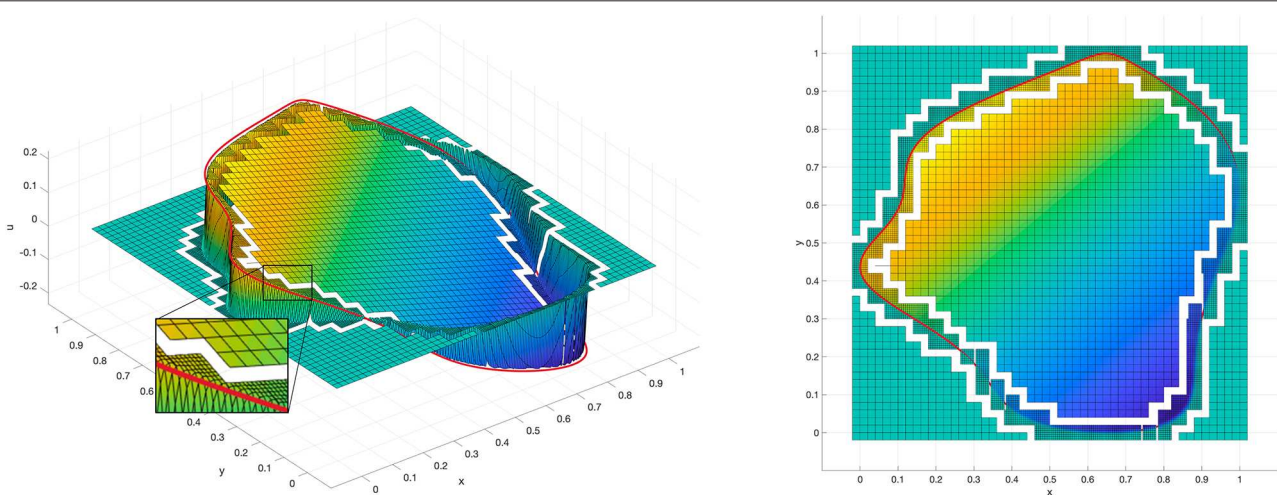


FIGURE 7 | Exemplary raw results for the displacement component u obtained on a grid with two discretization levels for the same problem as it is displayed in **Figure 2** with a single regular grid. The results on the fine grid level close to the boundary (the analytical solution is displayed with the red line) appear similar to the results which were obtained with the significantly more expensive technique used in **Figure 2**. The white borders between the two discretization levels are displayed for the purpose of a better visualization of the two discretization levels.

FFT grid via bilinear interpolation. The corresponding weights of the Lagrange polynomials are shown with the blue dots in **Figures 2, 3**. The right side of (7) can then be evaluated with a two dimensional convolution which is performed with the FFT. This operation then returns the left side of (7), that is the values $u_i(\underline{x}_0)$ at all points of the regular two-dimensional grid. This raw data is plotted in the colored diagrams on the right in **Figures 2, 3**. Therein, the analytical solution for u and v directly on the boundary is shown with the red line. While the raw results are highly accurate at a distance to the boundary, they begin to oscillate in close proximity to the boundary. For this reason, nearby panel interactions are evaluated directly in the *pFFT* technique.

So far we have used the analytical values of both the tractions and displacements on the boundary in order to obtain the FFT raw results. Let us now consider a problem where we only set the tractions on the boundary and seek a numerical solution for the displacements on the boundary. **Figure 4** shows the results of such a procedure for different cutouts from the uniformly loaded rectangular body. A conjugate gradient method with 10 iterations is used to solve for the displacements on the boundary. Therein, the projection of the FFT results onto the actual boundary is performed via bilinear interpolation. Furthermore, nearby panel interactions were calculated directly in a radius of 7 panels in order to eliminate oscillations. It can be observed that the method converges well to the analytical results for all three exemplary shapes.

VARIABLE FFT GRID

As we have observed in the previous section, the raw FFT data is highly accurate at a distance to the boundary but begins to oscillate in close proximity to the boundary. Consequently, the small number of nearby panel interactions is calculated directly in the *pFFT* technique and the larger number of all remaining panel interactions is computed with the FFT. Let us stress the point of having these two levels of accuracy. Then the question arises, why one does not use a hierarchy with even more levels of accuracy. Examining **Figures 2–4**, it seems obvious that the large number of unused discretization points of the FFT grid where no interpolation weights are placed are a waste of computational power. There are no numerical values at these points except zeros, and we have no interest in the results at these points after the convolution was applied. Philips and White addressed this by using an FFT grid which is coarser than the actual discretization of the boundary. This is enabled by using higher order interpolation techniques. A different or even additional approach might be a refinement of the grid at the boundary, see **Figure 5**. The main question which arises with a variable grid is how to apply the FFT which requires a regular grid. In the following, we illustrate that an $O(N \log N)$ algorithm seems technically feasible with this approach. Consider cutting the arbitrary shape shown in **Figure 5** into segments and aligning them one after the other as shown in **Figure 6**. Repositioning the segments in such a way creates an almost “one-dimensional” arrangement when compared to the original two-dimensional

shape. It is this reduction in dimension which we seek in order to lower the computational complexity. This new line of segments is longer than the circumference of the original shape because for each segment, we include some additional space around it (see **Figure 6** and compare with **Figure 5**). This way, it is ensured that while we convolute the kernels over the line of segments, the influence of the segments among each other is excluded. Exemplary raw results from this technique are shown in **Figure 7**. First, the convolutions are obtained on the rough grid level. From these results, we then subtract the influence obtained with the convolutions on the line of segments using also the rough grid level. Then, we add the influence of the convolutions over the line of segments using the fine grid level to our results using bilinear interpolation. Qualitatively, it can be observed that the results on the fine grid close to the boundary appear similar to the results which were obtained with the significantly more expensive technique used in **Figure 2**. This brief example can be regarded as a first sketch to indicate feasibility of the concept.

CONCLUSION

In this work, the *pFFT* was applied to solve the 2D Navier equation on arbitrary two-dimensional shapes. Furthermore, it was illustrated how the *pFFT* technique might be extended in order to decrease its computational complexity. In the standard *pFFT* approach, all panel influences which are not calculated directly are obtained on a single regular grid. In the present study, a variable grid was suggested to obtain these influences. It was outlined how it is possible to apply the FFT on each level of this variable grid by rearranging segments of the shape. A brief example was presented which indicates feasibility of the concept. However, more studies will be needed to research this concept further. Open questions include how to choose the grid size and number of levels of the variable grid, how high the order of the interpolation techniques used for projection onto the variable grid should be, how best to realize the interpolation between the grid levels, how the variable grid will influence iterative the solvers, and finally how the computational complexity and accuracy of the method compare to the standard *pFFT* technique and other methods designed to accelerate the Boundary Element Method.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author would like to thank V. L. Popov and M. Heß for valuable discussions on the topic and critical comments.

REFERENCES

- Anderson, J. (2017). *Fundamentals of Aerodynamics*. New York, NY: McGraw-Hill.
- Barber, J. (2004). *Elasticity*. Ann Arbor, MI: Kluwer Academic Publishers. doi: 10.1007/0-306-48395-5
- Benad, J. (2015). The Flying V - A new aircraft configuration for commercial passenger transport. *Deutscher Luft- und Raumfahrtkongress 2015* (Rostock). 101.
- Benad, J. (2018). Efficient calculation of the BEM integrals on arbitrary shapes with the FFT. *Facta Universitatis, Series Mechan. Eng.* 16, 405–417. doi: 10.22190/FUME180912034B
- Benad, J. (2019). Numerical methods for the simulation of deformations and stresses in turbine blade fir-tree connections. *Facta Universitatis, Series Mechan. Eng.* 17, 1–15. doi: 10.22190/FUME190103008B
- Benedetti, L., Aliabadi, M., and Davi, G. (2008). A fast 3D dual boundary element method based on hierarchical matrices. *Int. J. Solids Struct.* 45, 2355–2376. doi: 10.1016/j.ijsolstr.2007.11.018
- Cheng, A., and Cheng, D. (2005). Heritage and early history of the boundary element method. *Eng. Anal. Bound. Elem.* 29, 268–302. doi: 10.1016/j.enganabound.2004.12.001
- Faggiano, F., Vos, R., Baan, M., and Van Dijk, R. (2017). Aerodynamic design of a flying v aircraft. 17th AIAA aviation technology. *Integr. Operat. Conf.* 6:3589. doi: 10.2514/6.2017-3589
- Galin, L. (2008). “Plane elasticity theory,” in *Contact Problems*, ed G. Gladwell (Dordrecht: Springer). doi: 10.1007/978-1-4020-9043-1_2
- Gaul, L., and Fiedler, C. (2013). *Methode der Randelemente in Statik und Dynamik*. Berlin: Springer. doi: 10.1007/978-3-8348-2537-7
- Gross, D., Hauger, W., and Wriggers, P. (2014). *Technische Mechanik 4*. Berlin: Springer. doi: 10.1007/978-3-642-41000-0
- Hahn, H. (1985). *Elastizitätstheorie*. Wiesbaden: Vieweg Teubner Verlag. doi: 10.1007/978-3-663-09894-2
- Irgens, F. (Ed.). (2008). “Theory of elasticity,” in *Continuum Mechanics* (Berlin: Springer). doi: 10.1007/978-3-540-74298-2
- Katz, J., and Plotkin, A. (2001). *Low-Speed Aerodynamics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511810329
- Kelly, P. (2013). “Linear Elasticity,” in *An introduction to Solid Mechanics (Lecture Notes)* (University of Auckland). Available online at: http://homepages.engineering.auckland.ac.nz/~pkel015/SolidMechanicsBooks/How_to_reference.htm
- Li, Q., Voll, L., Starcevic, J., and Popov, V. L. (2018). Heterogeneity of material structure determines the stationary surface topography and friction. *Sci. Rep.* 8:32545. doi: 10.1038/s41598-018-32545-5
- Lim, K., He, X., and Lim, S. (2008). Fast Fourier transform on multipoles (FFTM) algorithm for Laplace equation with direct and indirect boundary element method. *Comput. Mechan.* 41, 313–323. doi: 10.1007/s00466-007-0187-5
- Masters, N., and Ye, W. (2004). Fast BEM solution for coupled electrostatic and linear elastic problems. *NSTI Nanotech* 2, 426–429. doi: 10.1016/j.enganabound.2004.02.001
- Müser, M., Dapp, W., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T., et al. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65:118. doi: 10.1007/s11249-017-0900-2
- Paggi, M., and Hills, D. (2020). *Modeling and Simulation of Tribological Problems in Technology*. Udine: Springer. doi: 10.1007/978-3-030-20377-1
- Palermo, M., and Vos, R. (2020). Experimental aerodynamic analysis of a 4.6%-scale flying-v subsonic transport. *AIAA Scitech 2020 Forum*, 2228. doi: 10.2514/6.2020-2228
- Phillips, J. (1997). *Rapid solution of potential integral equations in complicated 3-dimensional geometries*. (Doctoral thesis). Massachusetts Institute of Technology, Cambridge, MA, United States.
- Phillips, J., and White, K. (1997). A precorrected FFT-method for electrostatic analysis of complicated 3-D structures. *IEEE Trans. Comput Aided Des. Integr. Circ. Syst.* 16, 1059–1072. doi: 10.1109/43.662670
- Pohrt, R., and Li, Q. (2014). Complete Boundary Element Formulation for Normal and Tangential Contact Problems. *Phys. Mesomechan.* 17, 334–340. doi: 10.1134/S1029959914040109
- Popov, V. L. (2017). *Contact Mechanics and Friction*. Berlin: Springer. doi: 10.1007/978-3-662-53081-8
- Popov, V. L., Pohrt, R., and Li, Q. (2017). Strength of adhesive contacts: Influence of contact geometry and material gradients. *Friction* 5, 308–325. doi: 10.1007/s40544-017-0177-3
- Putignano, C., Afferrante, L., Carbone, G., and Demelio, G. (2012). A new efficient numerical method for contact mechanics of rough surfaces. *Int. J. Solids Struct.* 49, 338–343. doi: 10.1016/j.ijsolstr.2011.10.009
- Raymer, D. (2018). *Aircraft Design: A Conceptual Approach*. Reston, VA: American Institute of Aeronautics and Astronautics, Inc. doi: 10.2514/4.104909
- Rubio Pascual, B., and Vos, R. (2020). The effect of engine location on the aerodynamic efficiency of a flying-v aircraft. *AIAA Scitech 2020 Forum* 1954. doi: 10.2514/6.2020-1954
- Smith, A. (1990). The panel method: Its original development. *Appl. Comput. Aerodynam.* 125, 3–17. doi: 10.2514/5.9781600865985.0003.0017
- Willis, D., Peraire, J., and White, J. (2006). A combined pFFT-multipole tree code, unsteady panel method with vortex particle wakes. *Int. J. Num. Methods fluids* 53, 1399–1422. doi: 10.1002/fld.1240

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Benad. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Non-adhesive Contacts With Different Surface Tension Inside and Outside the Contact Area

Qiang Li* and Valentin L. Popov*

Department of System Dynamics and Friction Physics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Alessandro Ruggiero,
University of Salerno, Italy

Reviewed by:

Qian Jane Wang,
Northwestern University, United States
Aydar Akchurin,
SKF, Netherlands
Feodor M. Borodich,
Cardiff University, United Kingdom

*Correspondence:

Qiang Li
qiang.li@tu-berlin.de
Valentin L. Popov
v.popov@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 12 June 2020

Accepted: 02 July 2020

Published: 02 September 2020

Citation:

Li Q and Popov VL (2020)
Non-adhesive Contacts With Different
Surface Tension Inside and Outside
the Contact Area.
Front. Mech. Eng. 6:63.
doi: 10.3389/fmech.2020.00063

In the past decade, the influence of surface tension on contact properties has attracted much attention, especially in the context of contact of very soft materials (such as gels) or contacts at the nanoscale. However, in the most current studies it is assumed that the tension of the surface inside and outside the contact area is the same. In practical terms, this means that the object considered is an elastic body “coated” with a tensed membrane. In real contacts, there is no reason why the surface tensions of the “free interface” and the “contact interface” should be equal. On the contrary, especially in contacts of soft bodies with hard solid indenters, one can anticipate that they are completely different. In the present article, we consider an elastic contact taking into account different surface tensions inside and outside the contact area. However, the considered contacts are still “non-adhesive.” This means that the three surface energies in play (two surface energies of both bodies outside the contact and the interface energy in the contact region) fulfill the criterion that the work of separation vanishes. A numerical model based on the Fast Fourier transform–assisted boundary element method is implemented and is illustrated with a few examples.

Keywords: boundary element method, contact angle, contact mechanics, elastocapillary length, surface tension

INTRODUCTION

It is known that surface tension governs the contact behavior of liquids (Brown, 1974) but it is hardly seen in contacts of solid bodies at the macro- and mesoscales. It is clear that it should start to play an essential role if we consider a continuous transition from an elastic solid to a very soft body such as a gel, soft rubber, and biological tissue (Style et al., 2017). In biological studies, it is found that surface tension may determine the tissue growth and the kinetics of the regeneration of organs (Ehrig et al., 2019). Surface tension may become important even in stiff contacts if their size is very small. In particular, it may influence the function of microelectromechanical systems (MEMS) (Syms et al., 2003).

There exists a class of contact problems in which the surface energy has been considered already for almost 50 years; these are adhesive contacts. While non-adhesive contacts of isotropic elastic bodies are completely characterized by the elastic modulus E and the Poisson's ratio ν of bulk material, in adhesive contacts the surface energy plays a key role, as has been shown in the classic theory by Johnson, Kendall, and Roberts (JKR) in 1971 (Johnson et al., 1971). However, when talking about the “surface energy,” we have always to specify what surface energy is meant. At the contact boundary, three interfaces are coming together. The “surface energy” used in the JKR theory is better characterized as “specific work of separation,” which we will also call “work of adhesion,” w .

This quantity was introduced already by Dupre in 1869 (Dupré, 1869); it can be expressed as

$$w = \gamma_1 + \gamma_2 - \gamma_{12} \quad (1)$$

where γ_1 and γ_2 are specific free surface energies of two bodies and γ_{12} interfacial surface energy inside the contact area (Figure 1).

From a general theoretical point of view, the JKR theory is incomplete: It makes a step by including surface energy into consideration but at the same time assumes that the surface energies of bodies outside the contact area are zero. However, there are no theoretical reasons why the surface energy of the interface should be so dominant compared to the surface energies of the surfaces outside the contact. In a general case, all three surface energies are different and may have a comparable order of magnitude.

As academic theoretical models, one can consider various relations of the three surface energies and corresponding limiting cases.

- (I) If an elastic body is in contact with a rigid body and the specific surface energy of the elastic body outside the contact area can be neglected compared with the interface energy in the contact area, then we have an adhesive contact with specific work of separation $w = \gamma_2 - \gamma_{12}$. It is this case that is considered in the well-known work by Johnson et al. (1971). In reality, this case can be realized only in a contact of a high surface energy solid with low surface energy polymer but is hardly realizable in a contact of two materials with comparable surface energies.
- (II) If the surface energy of the elastic body outside the contact area is finite, $\gamma_1 \neq 0$, but the work of adhesion is zero,

$$\gamma_1 + \gamma_2 - \gamma_{12} = 0, \quad (2)$$

then we have a *non-adhesive* contact with surface tension.

- (IIa) The latter case usually has been further simplified with the assumption that the surface energies inside and outside contact are equal:

$$\gamma_1 = \gamma_{12}, \quad (3)$$

which automatically leads to the conclusion that $\gamma_2 = 0$. The corresponding theory was first developed for elastic foundations (Filonenko-Borodich, 1940; Kerr, 1964) and generalized to three-dimensional half-space contacts in Hajji (1978). It attracted much attention in recent years. However, this case cannot be realized physically, as there exist no substances with zero surface energy. In particular, in the cases that motivated consideration of this case—contacts of a hard indenter with a soft solid—one can assume that γ_2 not only does not vanish but is also the largest of the three relevant surface energies [(Popov, 2017), Chapter 3]. This popular case can practically be realized only as a composite consisting of an elastic body coated with a tensed membrane.

- (IIb) If the condition of the absence of adhesion is fulfilled, but the validity of the usual (and physical completely unrealistic) condition is not assumed, then we have a “general non-adhesive contact with surface tension,” which is characterized by two independent surface energies. It is this case that will be considered in the present article.
- (III) Finally, in the general case both work of adhesion and surface tension of the free surface are finite. This leads to a general adhesive contact with surface tension. The latter attracted much interest in the past two decades in the context of indentation of soft matter (e.g., gels) (Carrillo and Dobrynin, 2012; Style et al., 2013; Cao et al., 2014).

It is interesting to note that it took almost 50 years to make the next necessary (and, as a matter of fact, obvious) step after the JKR theory. On the other hand, the appearance of the JKR theory, which is nothing but a direct implementation of the Griffith crack theory, also took 50 years (Popova and Popov, 2018). For the sake of historical truth, let us, however, mention that a solution equivalent to that of JKR was obtained by Sperling already in 1964 in his unpublished doctoral thesis (Sperling, 1964; Borodich, 2014).

In the present article, we will briefly recapitulate the studies of case IIa and then consider a more general (and more complicated) case IIb. As already stated above, case IIa can be physically realized only as an elastic body coated with a stressed membrane. The fundamental solution determining vertical displacement of the surface under the action of a

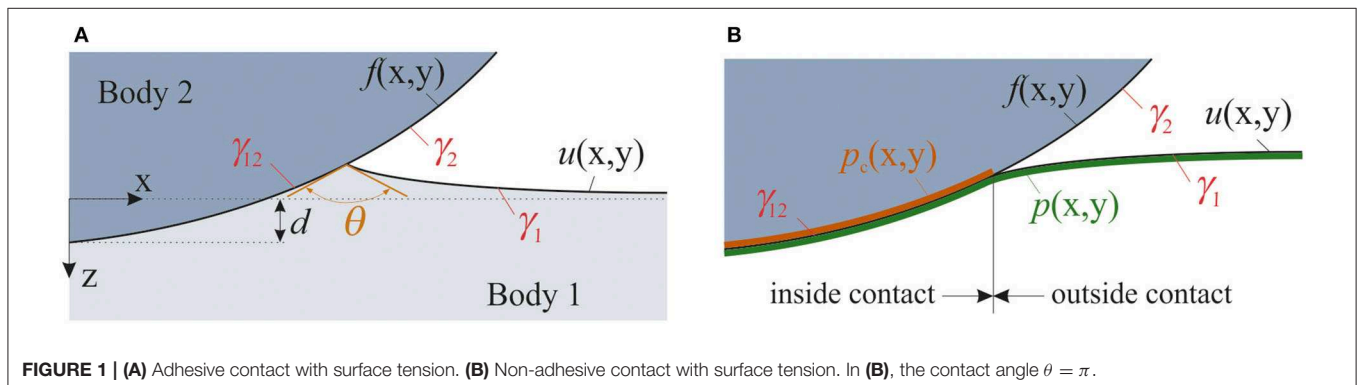


FIGURE 1 | (A) Adhesive contact with surface tension. (B) Non-adhesive contact with surface tension. In (B), the contact angle $\theta = \pi$.

concentrated force was given for this case by Hajji (1978). In this contact problem a characteristic length appears, γ_{12}/E^* , which is called “elastocapillary length,” determines the influence of surface tension: the surface tension effect is significant if the value of γ_{12}/E^* is comparable with the contact radius (Long and Chen, 2017). This leads to a “size-dependent” behavior in indentation testing (Style et al., 2013).

In the past few years, adhesive and non-adhesive contacts with surface tension have been studied intensively, for example, contact of homogeneous elastic half space including spherical contact (Long and Wang, 2013; Hui et al., 2015), conical contact (Long and Chen, 2017), and two-dimensional cylindrical contact (Liu et al., 2015), all based on the Hajji’s fundamental solution, as well as complicated inhomogeneous coatings based on the semianalytical modeling of the surface effect (Zhang et al., 2018).

The present article is structured as follows: We start with a consideration and discussion of the boundary conditions in the general case of arbitrary values of specific surface energies of all three interfaces. We then recapitulate the results of the simplest “membrane model” using a Fast Fourier transform (FFT)-based boundary element method (BEM) implementation. Finally, we generalize the FFT-BEM to the case of non-adhesive contact with different surface energies inside and outside the contact and illustrate this case with a simulation example.

BOUNDARY CONDITIONS TAKING INTO ACCOUNT SURFACE TENSION

Analysis of a contact between a rigid body and an elastic half space taking into account surface tension, carried out in Karpitschka et al. (2016) and Popov (2020), shows that the Young’s law determining the contact angle of liquids is also valid for solids. This leads to the following boundary conditions:

$$u(x, y) = d - f(x, y), \text{ inside the contact area} \quad (4)$$

$$p(x, y) = \gamma_1 \Delta u(x, y), \text{ outside the contact area} \quad (5)$$

$$\gamma_1 \cos \theta = \gamma_2 - \gamma_{12}, \text{ at the contact boundary} \quad (6)$$

where $f(x, y)$ is the profile of rigid indenter and $u(x, y)$ the normal displacement of the surface of elastic half space, d is indentation depth, θ is contact angle, Δ is Laplace operator, and $p(x, y)$ is the normal pressure at the surface immediately below the (infinitely thin) surface layer. Note that for definitions of the surface profile we used other direction of the z -axis (out of the elastic half-space) than for displacement and other quantities (into the half-space). It is also important to note that Equation (5) is written in approximation of small slopes of the surface of elastic body, which, however, is the necessary condition also for application of the superposition principle used throughout the paper. Equation (4) states that the surfaces of the indenter and the elastic half-space coincide in the contact area. Equation (5) determines the elastic stress at the surface under the tensioned surface layer. Equation (6) is the equilibrium condition of the contact boundary under the action of three surface tensions. Similarly to the contact angle in liquid contacts, θ is a thermodynamic property of the system. It does not depend on the shape of the body and deformation of the surface.

Using Equation (1) for the work of adhesion, w , Equation (6) can be rewritten as

$$w = \gamma_1 (1 + \cos \theta). \quad (7)$$

This equation is known as Young–Dupre equation (Young, 1805; Dupré, 1869). For non-adhesive contacts with surface tension it has $w = 0$. From Equation (7), it follows that $\cos \theta = -1$ and

$$\theta = \pi. \quad (8)$$

This equation means that at the boundary the slope of the surface profile is equal to the slope of the elastic half-space. Thus, in this specific case the condition (6) can be reformulated as a continuity of the surface slope at the contact boundary (Figure 1B). In this article, we focus on this non-adhesive contact with surface tension and its numerical simulation.

NUMERICAL SIMULATION OF NON-ADHESIVE CONTACT WITH SURFACE TENSION

A numerical solution to non-adhesive contact with surface tension for the case IIa ($\gamma_1 = \gamma_{12} = \gamma$) is based on Hajji’s equation for the vertical surface displacement u at the position (x, y) caused by the pressure distribution $p_c(x', y')$:

$$u(x, y) = \iint_A p_c(x', y') G(x - x', y - y') dx' dy', \quad (9)$$

with

$$G(x, y) = \frac{1}{4\gamma} \left[H_0 \left(\frac{\sqrt{x^2 + y^2}}{2\gamma/E^*} \right) - Y_0 \left(\frac{\sqrt{x^2 + y^2}}{2\gamma/E^*} \right) \right], \quad (10)$$

where H_0 and Y_0 are Struve and Bessel functions of the second kind of zero order (Hajji, 1978). It should be stressed that the pressure $p_c(x, y)$ is the pressure acting on the surface of the upper tensed layer and is *not equal* to the pressure $p(x, y)$ in Equation (5), which is the pressure *under* the tensed layer. Qualitatively speaking, $p_c(x, y)$ is the “external pressure” acting on the elastic body coated with membrane, while $p(x, y)$ is the pressure in the elastic body immediately under the membrane.

For vanishing surface tension the first term approximations to (9) and (10) become the classic Boussinesq’s equation

$$u(x, y) = \frac{1}{\pi E^*} \iint_A \frac{p(x', y')}{\sqrt{(x - x')^2 + (y - y')^2}} dx' dy'. \quad (11)$$

In this case, $p = p_c$.

As already stated, according to the theory of surface elasticity (Gurtin et al., 1998), the contact with surface tension can be considered as a contact with an elastic body covered by an (infinitely) thin membrane. The equilibrium and constitutive equations in the bulk of the elastic body still follow the theory of elasticity. Therefore, we can find a solution to contact with

surface tension by use of the fundamental solution for elastic half-space without surface tension. Using the definitions of pressure $p(x, y)$ and $p_c(x, y)$, as shown in **Figure 1B**, the equilibrium condition of the interface in contact the contact region is

$$p_c(x, y) = p(x, y) - \gamma_{12} \Delta u(x, y), \text{ inside the contact.} \quad (12)$$

The other two boundary conditions (4) and (5) should of course also be fulfilled. The deformation $u(x, y)$ is generated by pressure $p(x, y)$, which can be obtained from the Boussinesq's equation.

Realization in the Boundary Element Method

The BEM was developed in the past 30 years and represents an effective numerical tool for solving various contact problems including homogeneous material and layered systems, adhesive and non-adhesive contact, rough contact, and thermal contact (Nogi and Kato, 1997; Liu et al., 2000; Campa  a and M  ser, 2006). Some BEM formulations are based on integral equations of the form of (9) or (11). In a discretized form they can be written as

$$u^{ij} = K^{ij i' j'} p_c^{i' j'}, \quad (13)$$

where u^{ij} is the displacement of surface element at position (i, j) in two-dimensional discretization, $p_c^{i' j'}$ is stress acting on an element located on (i', j') , and $K^{ij i' j'}$ is the influence coefficient and its value is calculated analytically or numerically depending on the form of the fundamental solution. The matrix of influence coefficients can be found, e.g., in Pohrt and Li (2014). The convolution nature of integral equations of the form (9) allows formulating and solving a contact problem using the FFT. Using this technique increases computing efficiency by replacing the integration by multiplication and therefore reducing the complexity from N^4 to $N^2 \log N^2$ (for a system with discretization of $N \times N$ elements). For example, the algorithm of evaluating Equation (13) can be represented as

$$\mathbf{u} = \text{IFFT} [\text{FFT}(\mathbf{K}) \cdot \text{FFT}(\mathbf{p}_c)], \quad (14)$$

where \mathbf{p}_c is the pressure vector matrix, \mathbf{u} is the matrix of discretized displacements, and \mathbf{K} is the matrix of the influence coefficient. Note that for the non-periodic contact in a finite domain, the techniques of zero padding and wraparound order of matrix of pressure and influence coefficient in a doubled domain should be used to execute cyclic convolution. Thus, the \mathbf{p}_c and \mathbf{K} should be expanded to dimension $2N \times 2N$ when FFT is performed. The displacement \mathbf{u} is then extracted from the obtained displacement with the same dimension $2N \times 2N$ after the IFFT. The details of theories of the linear and cyclic convolutions and their applications to periodic and non-periodic contacts as well as numerical procedures can be found in Liu et al. (2000) and Ju and Farris (1996).

In Equation (14) the fundamental solution (given by the matrix \mathbf{K}) should be first Fourier transformed. Alternatively, one can directly use the fundamental equation in Fourier space,

which in most cases can be calculated very easily. For example, for an elastic half-space it is equal to

$$K = \frac{2}{E^* q}, \quad (15)$$

where $q = \sqrt{q_x^2 + q_y^2}$. This gives the relation of the Fourier components of pressure and displacement

$$u(q_x, q_y) = \frac{2}{E^*} \frac{p(q_x, q_y)}{q}. \quad (16)$$

Displacement in the coordinate space is then calculated as

$$\mathbf{u} = \text{IFFT} \left[\frac{2}{E^* q} \cdot \text{FFT}(\mathbf{p}) \right]. \quad (17)$$

The general form of the fundamental equation in Fourier space can be written as

$$u(q) = C(q) \cdot p_c(q), \quad (18)$$

where the Fourier transform of the kernel of convolution, $C(q)$, depends on a particular system. The only general prerequisite for this simple form is the lateral homogeneity of the system. In the direction perpendicular to the surface, it can be heterogeneous (e.g., a layered system or functionally graded material). $C(q)$ is of course also a function of the effective elastic modulus in the case of a homogeneous material and coating system.

Equations (12) and (5) for non-adhesive contact with surface tension can be written in Fourier space as follows:

$$p_c(q) = p(q) + \gamma_{12} \cdot q^2 \cdot u(q), \text{ in contact,} \quad (19)$$

$$p(q) + \gamma_1 \cdot q^2 \cdot u(q) = 0, \text{ out of contact.} \quad (20)$$

We first consider the simple case with equal surface energy inside and outside the contact area, $\gamma_1 = \gamma_{12} = \gamma$. Substitution of (18) into (19) gives the relation between pressure p_c and deformation u

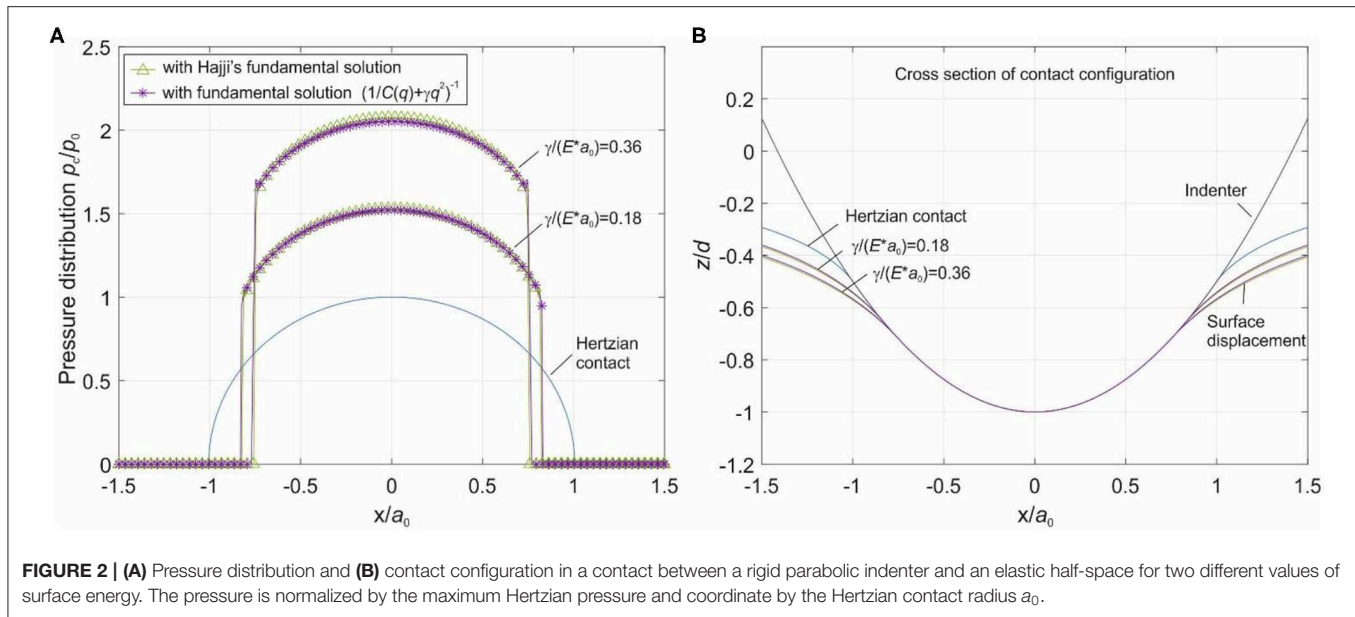
$$u(q_x, q_y) = \frac{p_c(q_x, q_y)}{1/C(q) + \gamma \cdot q^2}. \quad (21)$$

For the displacement field in coordinate space, it follows

$$\mathbf{u} = \text{IFFT} \left[\frac{1}{1/C(q) + \gamma \cdot q^2} \cdot \text{FFT}(\mathbf{p}_c) \right]. \quad (22)$$

Thus, any available BEM program can be used for simulating this system covered with a tensed membrane just by adding the term " γq^2 " in the Green's function. Note that this equation connects the surface displacement with the external pressure. Therefore, the boundary condition and numerical procedure can be dealt in the same way as in all existing BEM: the pressure outside the contact is zero:

$$p_c(x, y) = 0, \text{ outside of the contact,} \quad (23)$$



while the surface displacement inside the contact area meets the geometry condition. Of course, the zero padding and wraparound order are still needed for the non-periodic contacts as discussed in Liu et al. (2000). For an isotropic homogeneous material, Equation (22) becomes

$$\mathbf{u} = \text{IFFT} \left[\frac{1}{E^* q/2 + \gamma \cdot q^2} \cdot \text{FFT}(\mathbf{p}) \right]. \quad (24)$$

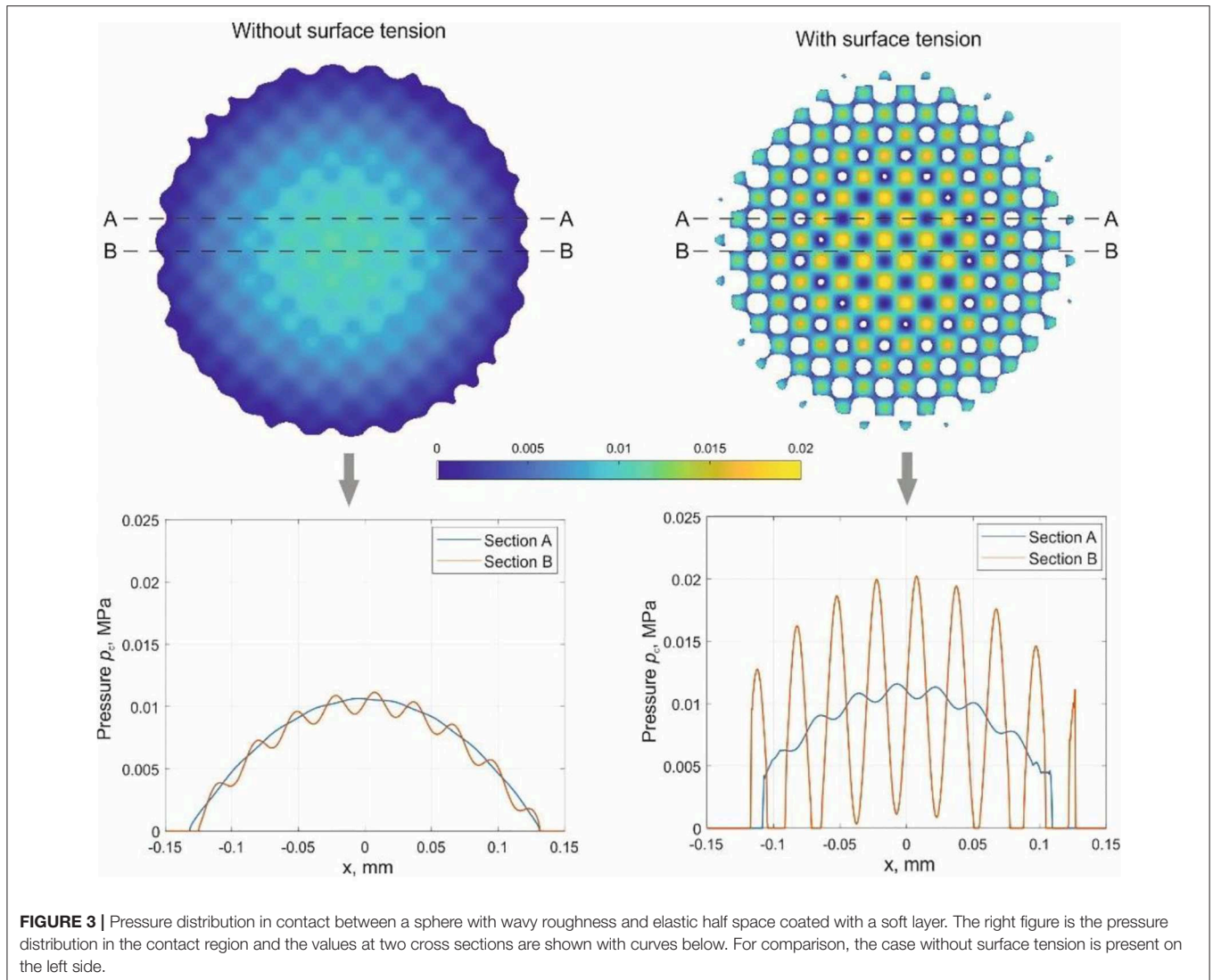
Case Study 1: Isotropic Homogeneous Elastic Half-Space Coated With a Tensed Membrane

Let us discuss numerical simulation of a “Hertzian” contact with surface tension: a contact of a rigid parabolic indenter with elastic half-space taking into account surface tension, which we first consider equal inside and outside the contact. The result obtained using the Hajji’s fundamental solution was given in Long and Wang (2013). Here we would like to compare this solution with the above-described solution based on the fundamental solution $(1/C(q) + \gamma \cdot q^2)^{-1}$ in the Fourier space. Parameters used in the simulation are as follows: $E^* = 0.011$ MPa; sphere radius $R = 1$ mm; indentation depth $d = 0.01$ mm; and two different specific surface energies, $\gamma = 0.2$ N/m and $\gamma = 0.4$ N/m, were considered, corresponding to the values $\gamma/(E^*a_0) = 0.18$ and 0.36 of the parameter $\gamma/(E^*a_0)$ (ratio of the elastocapillary length and the Hertzian contact radius a_0). **Figure 2** shows numerical solution obtained with discretization 512×512 points, where the triangles in **Figure 2A** are pressure distributions calculated based on the Hajji’s fundamental equation, and stars are results based on the alternative approach using the fundamental solution $(1/C(q) + \gamma \cdot q^2)^{-1}$. The calculating time by using this fundamental solution is 12 times smaller than that with the Hajji’s fundamental solution. The contact behavior found in Long

and Wang (2013) practically coincides with that found with the above-described procedure. In particular, the pressure at the contact boundary has a “jump” in the case with surface tension (**Figure 2A**), and the surface tension leads to a reduction of contact area (**Figures 2A,B**).

Case Study 2: Layered System

As explained above, any existing BEM formulation for a contact without surface tension can be trivially extended to include surface tension by a small change in the corresponding fundamental solution in Fourier space, according to Equation (21) or (24). Let us illustrate this with an example of an elastic layer with surface tension. We proceed from the BEM formulation without surface tension described in Li et al. (2019). Adding the term “ γq^2 ” in the fundamental solution, we include the surface tension. Here we show a case of soft layer bounded on the elastic half space. The layer had elastic modulus $E_1^* = 0.011$ MPa, specific surface energy $\gamma = 0.2$ N/m, and thickness $h_0 = 0.03$ mm. The elastic modulus of the underlying half-space is $E_2^* = 10E_1^*$. The profile of indenter is a sphere with radius $R = 1$ mm superposed with a two-dimensional waviness with wavelength $\lambda = 0.03$ mm and a small amplitude of $h = 0.0005$ mm. The indentation depth was $d = 0.01$ mm. The simulation results without and with surface tension are shown in **Figure 3**. The color map presents the pressure distribution in the contact region. Two cross sections of this map are selected to show the details. Under the above conditions, the contact area in the system without surface tension is compact (left figure), while “switching on” the surface tension makes it much more heterogeneous and even not simply connected (right figure). This leads to much more intensive oscillations of pressure. Note that the pressure shown in **Figure 3** is the pressure immediately under the surface of the indenter. This pressure is relevant for estimating the possible damage of the indenter.



Different Surface Energies Inside and Outside of the Contact Area

When surface energies inside and outside the contact are different, $\gamma_1 \neq \gamma_{12}$, the FFT of Equations (19) and (20) cannot be directly carried out for the whole region. However, we can rewrite Equation (12) in the following form:

$$p_1(x, y) = p(x, y) - \gamma_1 \Delta u(x, y), \text{ in the whole area,} \quad (25)$$

$$p(x, y) = p_1(x, y) + (\gamma_1 - \gamma_{12}) \Delta u(x, y), \quad (26)$$

in the contact area,

where $p_1(x, y)$ is an auxiliary function. In Fourier space, Equation (25) takes the form

$$p_1(q) = p(q) + \gamma_1 \cdot q^2 \cdot u(q) = [1/C(q) + \gamma \cdot q^2] \cdot u(q) \quad (27)$$

The computation algorithm is the following. Initially we assume that the contact area is the geometrical intersection A and the displacement of the surface of the elastic half-space in this area

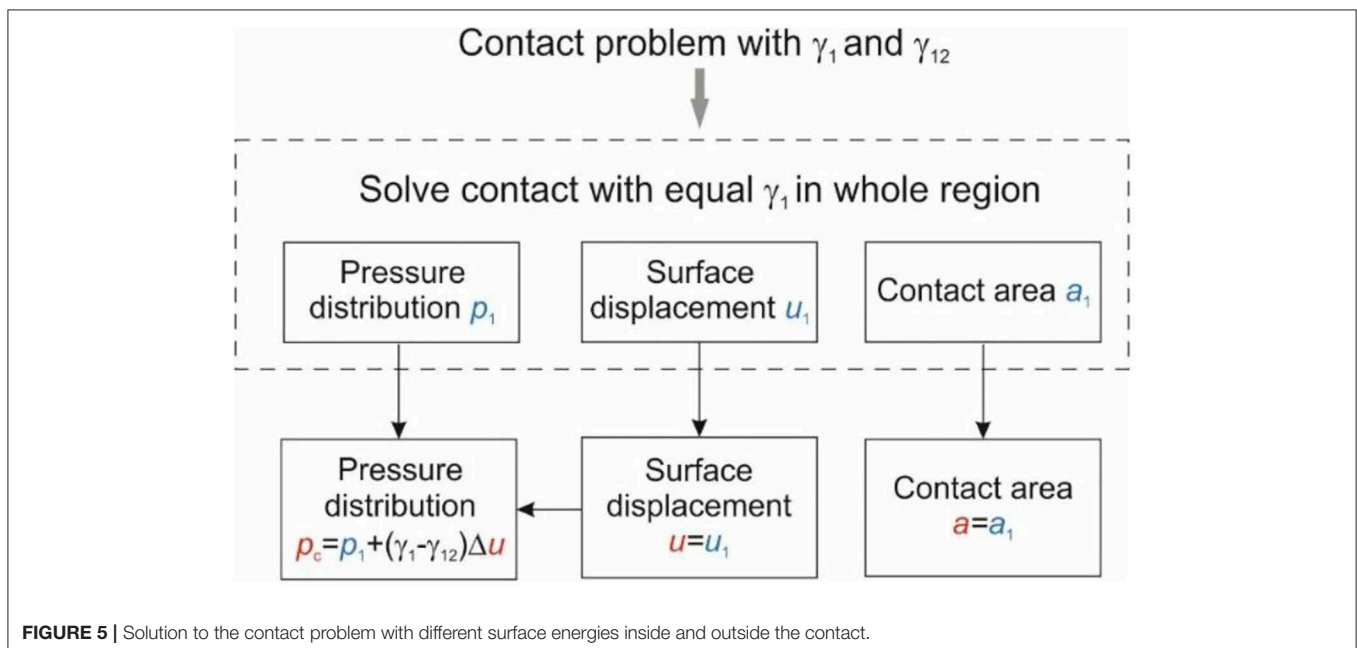
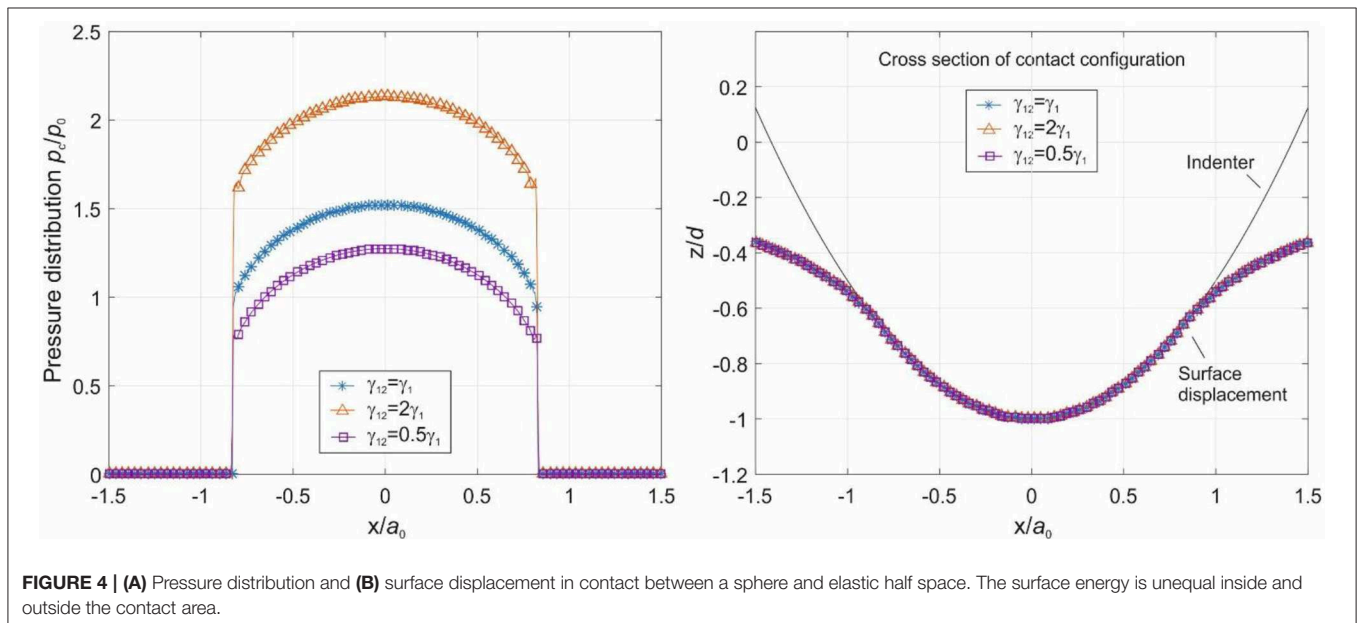
coincides with the profile of indenter $u_A(x, y)$; the corresponding FFT (u_A) is calculated. Using this FFT (u_A) one can obtain the auxiliary stress function $p_1(q)$ as well as $p_1(x, y)$ appearing in (25). After that, the deformation $u(x, y)$ can be calculated within the contact area. With $p_1(x, y)$ and $u(x, y)$ as well as the curvature of the surface, $\Delta u(x, y)$, one gets the pressure distribution $p(x, y)$ from Equation (26). After that, the usual correction and iteration procedure starts: the elements that have negative pressure or geometrical penetration out of contact are marked as “detached” so that a new contact area, A' , is generated. With this new contact area, the above iteration is repeated until both pressure and geometry conditions for all elements in contact area are met.

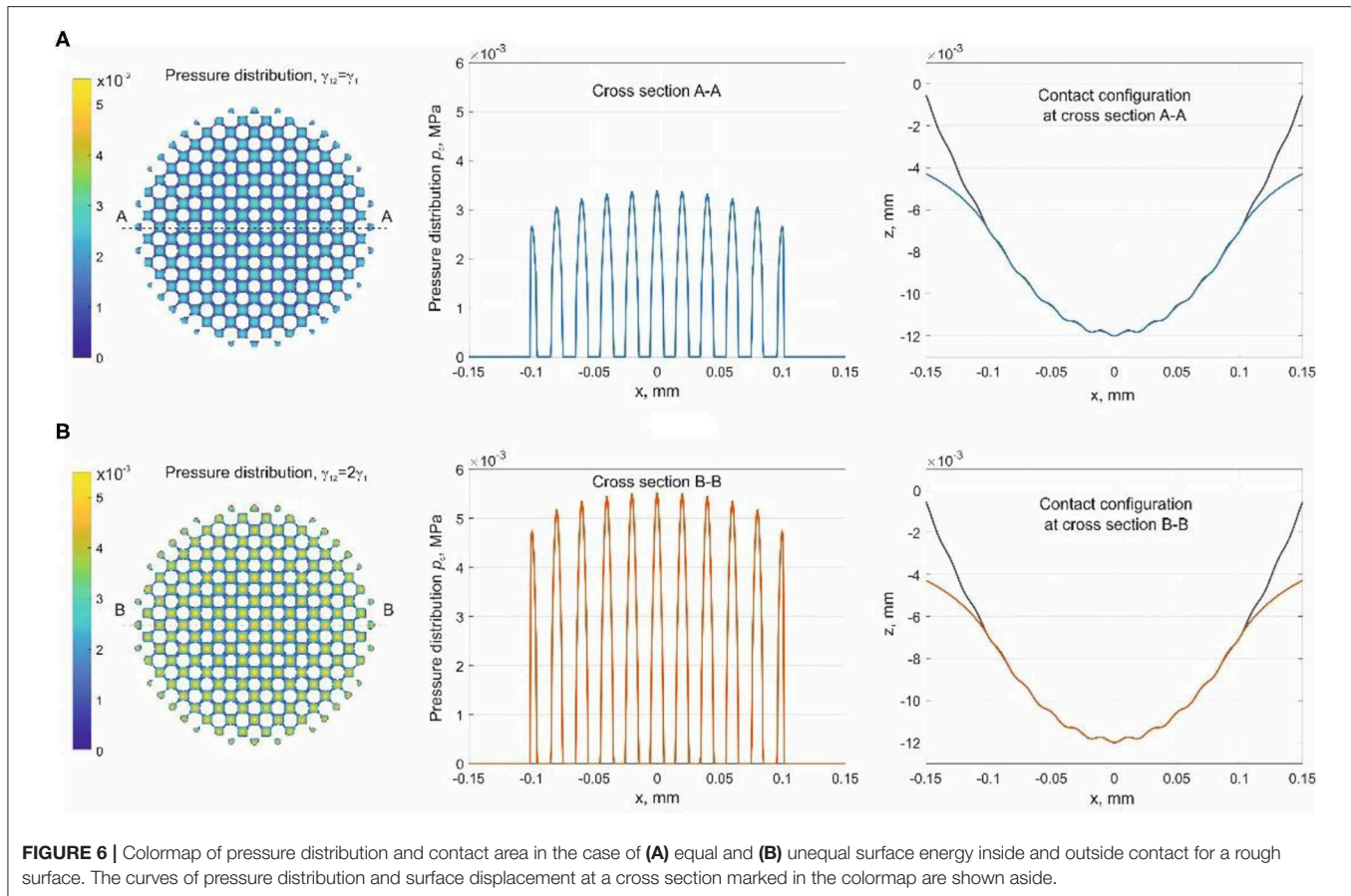
Example 1 We give an example with the same parameters as studied in the case of **Figure 2**, with the only difference that the surface energy outside contact area, γ_1 , is not equal to that inside, γ_{12} . Surface energy in the noncontact region was equal to $\gamma_1 = 0.2$ N/m. Two cases were studied: one with smaller surface energy inside contact $\gamma_{12} = 0.1$ N/m and the other one with larger $\gamma_{12} = 0.4$ N/m. The obtained pressure distribution

and surface displacement are shown in **Figure 4**. The results with equal surface energy are also shown with stars for a comparison. One can see that when the surface tension inside the contact is larger, the loading needed for generating the same indentation depth is also higher due to the surface tension effect (curve with triangles). Interestingly the contact area and the surface displacement are the same in all three cases. It is clear that the surface displacement in the contact coincides with the profile of the indenter shifted by the indentation depth. Displacement outside the contact is determined by the pressure p and surface tension γ_1 , where p is numerically calculated by Boussinesq's equation with the given surface deformation inside the contact.

So surface displacement will be the same for unequal surface energy inside the contact if the contact area is unchanged. The difference of surface energy changes only the pressure p_c with a constant value in the case of sphere contact because of the constant mean curvature.

Figure 4 shows that the shape of the surface of elastic body does not depend on the surface tension inside the contact area. *A posteriori*, this seems to be a trivial conclusion. However, it is trivial only under the assumption of non-adhesive contact, which guarantees the unchanged condition at the boundary of the contact area, which does not depend on the interface surface energy density. The only quantity that is influenced by the





specific surface energy of the contact interface is the pressure distribution. This means that the computation procedure in the case of two different surface energies can be simplified as shown in Figure 5.

1. In the first step, the contact problem is solved with a constant specific surface energy equal to the specific surface energy outside the contact area.
2. In the second step, the pressure distribution inside the contact area is corrected by the term $\delta p(x, y) = (\gamma_1 - \gamma_{12}) \Delta u(x, y)$.

Example 2 Consider a wavy surface similar to that shown in Figure 3 with the radius of curvature, $R = 1$ mm; wavelength, $\lambda = 0.02$ mm; and amplitude of waviness, $h = 0.0001$ mm. The elastic modulus of the half space is assumed $E_1^* = 0.013$ MPa and the indentation depth $d = 0.012$ mm. In the following two cases are considered: (1) equal specific surface energies $\gamma_1 = \gamma_{12} = 0.1$ N/m and (2) two different specific surface energies, $\gamma_1 = 0.1$ N/m and $\gamma_{12} = 0.2$ N/m, inside and outside the contact correspondingly. Numerical results are shown in Figure 6. The contact area and surface displacement are the same in both cases.

CONCLUSION

Non-adhesive contact with surface tension is usually solved using the fundamental solution of Hajji. This fundamental solution has a complicated form in the coordinate space but can be

derived in an extremely simple way in the Fourier space. More than that, this derivation has a universal character and can be applied for more complicated situations, as, e.g., for layered systems or functionally graded materials. With an example of Hertzian contact this method was validated by comparison with the fundamental solution in the coordinate space, and it was then applied to the rough contact of a coated system.

We argue that the approach based on the fundamental solution can be used only in the case when the surface tension inside and outside of the contact area are equal. There are no physical reasons why it should be the case. Therefore, in the general case, the approach based on the use of the fundamental solution and superposition principle does not work. We show how this problem can be reduced to the simpler contact problem with constant surface tension inside and outside the contact area.

Of course, the case of non-adhesive contact with different surface energies inside and outside the contact is a little bit an academic exercise. In a general “dry” contact of two solids, the condition of non-adhesive contact normally will not be fulfilled. However, the contact can be made non-adhesive by introducing an intermediate liquid with dielectric constant equal to that of one of the bodies. According to the theory of van der Waals’ interactions by Dzyaloshinskii et al. (1961), this leads to suppression of van der Waals force (and thus of the separation energy), while the surface tensions of both bodies remain generally non-zero.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

VP designed the concept of the article, formulated the analytical theory, and formulated the algorithm of contact simulation in the case of different surface energies inside and outside of the contact

area. QL implemented the algorithm into a BEM program and executed numerical simulations. All authors contributed equally to the analysis of simulation results as well as editing and review of the manuscript.

FUNDING

This work has been conducted under financial support from DFG (Grant number PO 810/55-1).

REFERENCES

- Borodich, F. M. (2014). The Hertz-Type and adhesive contact problems for depth-sensing indentation. *Adv. Appl. Mech.* 47, 225–366. doi: 10.1016/B978-0-12-800130-1.00003-5
- Brown, R. C. (1974). The surface tension of liquids. *Contemp. Phys.* 15, 301–327. doi: 10.1080/00107517408210795
- Campaná, C., and Müser, M. H. (2006). Practical green's function approach to the simulation of elastic semi-infinite solids. *Phys. Rev. B* 74, 1–15. doi: 10.1103/PhysRevB.74.075420
- Cao, Z., Stevens, M. J., and Dobrynin, A. V. (2014). Elastocapillarity: adhesion and wetting in soft polymeric systems. *Macromolecules* 47, 6515–6521. doi: 10.1021/ma5013978
- Carrillo, J.-M. Y., and Dobrynin, A. V. (2012). Contact mechanics of nanoparticles. *Langmuir* 28, 10881–10890. doi: 10.1021/la301657c
- Dupré, A. (1869). *Théorie Mécanique de la Chaleur*. Paris: Gauthier-Villars.
- Dzyaloshinskii, I. E., Lifshitz, E. M., and Pitaevskii, L. P. (1961). General theory of van der Waals' forces. *Sov. Phys. Usp.* 4, 153–176. doi: 10.1070/PU1961v004n02ABEH003330
- Ehrig, S., Schamberger, B., Bidan, C. M., West, A., Jacobi, C., Lam, K., et al. (2019). Surface tension determines tissue shape and growth kinetics. *Sci. Adv.* 5:eav9394. doi: 10.1126/sciadv.aav9394
- Filonenko-Borodich, M. M. (1940). Some approximate theories of the elastic foundation. *Uchenye Zapiski Moskovskogo Gosudarstvennogo Universiteta. Mekhanika* 46, 3–18.
- Gurtin, M. E., Weissmüller, J., and Larché, F. (1998). A general theory of curved deformable interfaces in solids at equilibrium. *Philos. Mag. A Phys. Condens. Matter. Struct. Defects Mech. Prop.* 78, 1093–1109. doi: 10.1080/01418619808239977
- Hajji, M. A. (1978). Indentation of a membrane on an elastic half space. *J. Appl. Mech. Trans. ASME* 45, 320–324. doi: 10.1115/1.3424295
- Hui, C. Y., Liu, T., Salez, T., Raphael, E., and Jagota, A. (2015). Indentation of a rigid sphere into an elastic substrate with surface tension and adhesion. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 471:20140727. doi: 10.1098/rspa.2014.0727
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. R. Soc. London A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Ju, Y., and Farris, T. N. (1996). Spectral analysis of two-dimensional contact problems. *J. Tribol.* 118, 320–328. doi: 10.1115/1.2831303
- Karpitschka, S., Van Wijngaarden, L., and Snoeijer, J. H. (2016). Surface tension regularizes the crack singularity of adhesion. *Soft Matter* 12, 4463–4471. doi: 10.1039/C5SM03079J
- Kerr, A. D. (1964). Elastic and viscoelastic foundation models. *J. Appl. Mech.* 31, 491–498. doi: 10.1115/1.3629667
- Li, Q., Pohrt, R., Lyashenko, I. A., and Popov, V. L. (2019). Boundary element method for nonadhesive and adhesive contacts of a coated elastic half-space. *Proc. Inst. Mech. Eng. J* 234, 73–83. doi: 10.1177/1350650119854250
- Liu, S., Wang, Q., and Liu, G. (2000). A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear* 243, 101–111. doi: 10.1016/S0043-1648(00)00427-0
- Liu, T., Jagota, A., and Hui, C. Y. (2015). Adhesive contact of a rigid circular cylinder to a soft elastic substrate-the role of surface tension. *Soft Matter* 11, 3844–3851. doi: 10.1039/C5SM00008D
- Long, J., and Chen, W. (2017). Effects of surface tension on the nanoindentation with a conical indenter. *Acta Mech.* 228, 3533–3542. doi: 10.1007/s00707-017-1901-9
- Long, J. M., and Wang, G. F. (2013). Effects of surface tension on axisymmetric Hertzian contact problem. *Mech. Mater.* 56, 65–70. doi: 10.1016/j.mechmat.2012.09.003
- Nogi, T., and Kato, T. (1997). Influence of a hard surface layer on the limit of elastic contact: analysis using a real surface model. *Japanese J. Tribol.* 42, 270–271. doi: 10.1115/1.2833525
- Pohrt, R., and Li, Q. (2014). Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17, 334–340. doi: 10.1134/S1029959914040109
- Popov, V. L. (2017). *Contact Mechanics and Friction. Physical Principles and Applications*. 2nd Edn. Berlin: Springer.
- Popov, V. L. (2020). Adhesive and non-adhesive contact of a rigid indenter and a thin elastic layer with surface tension. *arXiv preprint arXiv:2002.01874*.
- Popova, E., and Popov, V. L. (2018). Note on the history of contact mechanics and friction: interplay of electrostatics, theory of gravitation and elasticity from Coulomb to Johnson-Kendall-Roberts theory of adhesion. *Phys. Mesomech.* 21, 1–5. doi: 10.1134/S1029959918010010
- Sperling, G. (1964). *Eine Theorie der Haftung von Feststoffteilchen an Festen Körpern*. Karlsruhe: Unpublished doctoral dissertation, Technische Hochschule Karlsruhe.
- Style, R. W., Hyland, C., Boltyanskiy, R., Wettlaufer, J. S., and Dufresne, E. R. (2013). Surface tension and contact with soft elastic solids. *Nat. Commun.* 4, 2728–2733. doi: 10.1038/ncomms3728
- Style, R. W., Jagota, A., Hui, C.-Y., and Dufresne, E. R. (2017). Elastocapillarity: surface tension and the mechanics of soft solids. *Ann. Rev. Condensed Matter. Phys.* 8, 99–118. doi: 10.1146/annurev-conmatphys-031016-025326
- Syms, R. R. A., Yeatman, E. M., Bright, V. M., and Whitesides, G. M. (2003). Surface tension-powered self-assembly of microstructures - the state-of-the-art. *J. Microelectromech. Syst.* 12, 387–417. doi: 10.1109/JMEMS.2003.811724
- Young, T. (1805). III. An essay on the cohesion of fluids. *Philoso. Trans. R. Soc. London.* 95, 65–87. doi: 10.1098/rstl.1805.0005
- Zhang, X., Wang, Q. J., Wang, Y., Wang, Z., Shen, H., and Liu, J. (2018). Contact involving a functionally graded elastic thin film and considering surface effects. *Int. J. Solids Struct.* 150, 184–196. doi: 10.1016/j.ijsolstr.2018.06.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Li and Popov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Boundary Element Calculations for Normal Contact of Soft Materials With Tensed Surface Membrane

Weike Yuan^{1,2*} and Gangfeng Wang^{1*}

¹ Department of Engineering Mechanics, SVL, Xi'an Jiaotong University, Xi'an, China, ² Institute of Mechanics, Technical University of Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Irina Goryacheva,
Institute for Problems in Mechanics
(RAS), Russia

Reviewed by:

Xi-Qiao Feng,
Tsinghua University, China
Martin H. Müser,
Saarland University, Germany

*Correspondence:

Weike Yuan
w.yuan@campus.tu-berlin.de
Gangfeng Wang
wanggf@xjtu.edu.cn

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 26 March 2020

Accepted: 19 June 2020

Published: 29 July 2020

Citation:

Yuan W and Wang G (2020) Boundary
Element Calculations for Normal
Contact of Soft Materials With Tensed
Surface Membrane.
Front. Mech. Eng. 6:57.
doi: 10.3389/fmech.2020.00057

This work considers the non-adhesive frictionless contact problem of soft materials with surface being tensed by equi-biaxial tension. The boundary element method (BEM) based on Fast Fourier Transform and conjugate gradient algorithm is extended to deal with this problem. By comparing with existing analytical solutions for the axisymmetric contact between a rigid parabolic indenter and an elastic half space, our numerical simulations are validated having great accuracy. Moreover, the developed BEM algorithm is applied on the calculations of elastic responses of a soft substrate indented by a smooth indenter with general quadric profile and a rough indenter with self-affine fractal surface, respectively. Some essential contact behaviors resulted from the presence of membrane tension are presented and discussed.

Keywords: boundary element method, membrane/surface tension, contact mechanics, elliptical indenter, rough surface

INTRODUCTION

Many physiological systems can be modeled as the layer-foundation structure, and generally, the mechanical properties of surface layer differ from those of bulk interior. Early in 1978, Hajji (1978) investigated the indentation of lung under uniform pressure by considering the sample as an isotropic elastic half space ideally adhered with a tensed membrane. It is assumed that the surface membrane thickness is ultimately small so that its bending rigidity can be neglected. Besides, for small deformation, the membrane tension is assumed keeping constant. With the same assumptions, Kim and Gouldstone (2008) addressed the axisymmetric spherical indentation of elastic solid with strain-independent membrane tension. Moreover, the surface layer of some soft tissues and single cells can be regarded as pre-tensed “plate” or “shell” with finite thickness, which can also resist bending deformation (Zamir and Taber, 2004; Zhang and Zhang, 2009). Recently, Argatov and Sabina (2012) treated the surface layer as a reinforced membrane under generalized plane stress state. They hypothesized that no pre-tension exists in the membrane, and the flexural stiffness is negligible compared with the in-plane tensile stiffness. Such modeling method was employed in the deformation analysis of anisotropic articular cartilage (Argatov and Mishuris, 2016).

It is interesting to note that the above models originally developed for the biological materials are analogous to the well-known surface elasticity theory of Gurtin and Murdoch (1975), which characterizes the material surface with surface tension and surface elasticity. As Hajji stated (Hajji, 1978), the mathematical expressions are basically consistent, although they have completely different physical natures. The constant membrane tension in Hajji (1978) and Kim and Gouldstone (2008) is corresponding to the residual surface tension, and the superficial layer modeled in Argatov and Sabina (2012) and Argatov and Mishuris (2016) can be considered as a solid surface with surface elasticity. Recently, problems of surface loadings and contacts have been widely studied based on the surface elasticity theory (He and Lim, 2006; Wang and Feng, 2007; Zhou and Gao, 2013; Long et al., 2017; Li et al., 2020). It is found that the deformation behavior would be distinctly different from classical models when the size of loading is comparable or even smaller than a critical length, usually at nano/microscale. However, these analytical works are usually limited to simple cases under a given surface pressure or for the symmetric situations.

Physically, the surface elasticity theory is based on the concept of solid surface energy (Gurtin and Murdoch, 1975). For the general contact between two solid bodies, the surface energy of each free surface (γ_1, γ_2) and contacting interface (γ_{12}) should be taken into account simultaneously. Two simplified cases can be recognized according to values of these surface energies. When the energy difference $w = \gamma_1 + \gamma_2 - \gamma_{12}$ (called work of adhesion) is appreciable and the effect of surface energy outside the contact can be neglected, adhesive contact model with specific work of adhesion would be appropriate (Johnson et al., 1971). On the other hand, if the energy difference equals to zero ($w = 0$) but the surface energy of one contact body is much larger than the other (i.e., $\gamma_1 \gg \gamma_2$ and $\gamma_1 \approx \gamma_{12}$), the contact problem can be addressed in the framework of surface elasticity theory (He and Lim, 2006; Wang and Feng, 2007; Zhou and Gao, 2013; Long et al., 2017; Li et al., 2020). In this work, we consider the latter circumstance with constant surface tension or, equivalently, the non-adhesive contact of solids with tensed surface membrane.

The Fast Fourier Transform (FFT) based methods have been widely employed to solve the elastic contact problems for its great advantage on reducing the computation cost. For the efficient Green's function molecular dynamics (GFMD) (Campaña and Müser, 2006; Prodanov et al., 2014), the FFT technique plays an important role in the determination of substrate elastic response to external surface forces. In addition, the boundary element method (BEM) that exploits FFT to accelerate the calculation of displacements induced by given surface forces has been also proved fast and robust for the analysis of various contact problems (Nogi and Kato, 1997; Liu et al., 2000; Polonsky and Keer, 2000; Pohrt and Popov, 2012, 2015; Pohrt and Li, 2014; Rey et al., 2017; Bugnicourt et al., 2018). The basic principle of the FFT-based BEM is to evaluate the linear convolution of fundamental solution and pressure distribution based on the convolution theorem. For the classical elastic problem with half space approximation, the simple Boussinesq's fundamental solution or its form in Fourier space is commonly used in the

BEM for normal contact problems. However, when the surface of elastic half space is equi-biaxially tensed, things will be different.

This paper aims to extend the mature FFT-based BEM for the calculation of normal contact of material with membrane/surface tension. Several calculation examples are implemented to demonstrate the validity and usefulness of this method. This method provides a potential numerical approach for the study of contact behavior of some soft structures, which is of particular interest for the measurement of mechanical properties of biological tissues (Zhang et al., 2014).

NUMERICAL METHOD

We consider an ideally smooth elastic substrate. The surface is tensed by a constant, strain-independent tension τ_0 , and the bulk material is homogeneous having elastic modulus E and Poisson's ratio ν . The origin of the Cartesian coordinate system (o -xyz) is located at the surface of half space, and the z -axis is perpendicular to the surface, as shown in **Figure 1A**.

With some specified external pressure applied on the surface, the boundary value problem can be solved in the framework of classical linear elastic theory, provided that the unconventional boundary condition associated with surface tension is employed (Hajji, 1978; He and Lim, 2006; Wang and Feng, 2007; Zhou and Gao, 2013). For the axisymmetric situation under a uniform normal pressure p_0 within a circular region of radius a , the boundary condition reads,

$$\begin{cases} \tau_0 \left(\frac{d^2 u}{dr^2} + \frac{1}{r} \frac{du}{dr} \right) + \sigma_{zz} + p_0 H(a-r) = 0 \\ \sigma_{zr} = 0 \end{cases} \quad (1)$$

where $u(r)$ is the vertical displacement on the surface, σ_{zz} and σ_{zr} are the bulk normal stress and shear stress at $z = 0$, respectively, and H is the Heaviside step function.

By using Hankel integral transform and substituting the boundary condition into the general solutions, the vertical displacement caused by the uniform pressure p_0 can be derived (Hajji, 1978). Moreover, by implementing a limiting process in which the radius a decreases to zero with the resultant force held constant unity $\pi a^2 p_0 = 1$, the fundamental solution that is the vertical displacement at the point (x, y) induced by unit concentrate force acting at the coordinate origin can be derived as,

$$G(x, y) = \frac{1}{2sE^*} \left[H_0 \left(\frac{\sqrt{x^2 + y^2}}{s} \right) - Y_0 \left(\frac{\sqrt{x^2 + y^2}}{s} \right) \right] \quad (2)$$

where $s = 2\tau_0/E^*$ is the characteristic length, $E^* = E/(1-\nu^2)$ is the reduced modulus, H_n and Y_n are the Struve function and the Bessel function of second kind of order n , respectively.

The FFT of this fundamental solution, $\hat{G}(\omega)$, is plotted in **Figure 2**, where $\omega = \sqrt{\omega_x^2 + \omega_y^2}$ is the magnitude of wave vector (ω_x is the angular frequency in the x -direction and ω_y is the

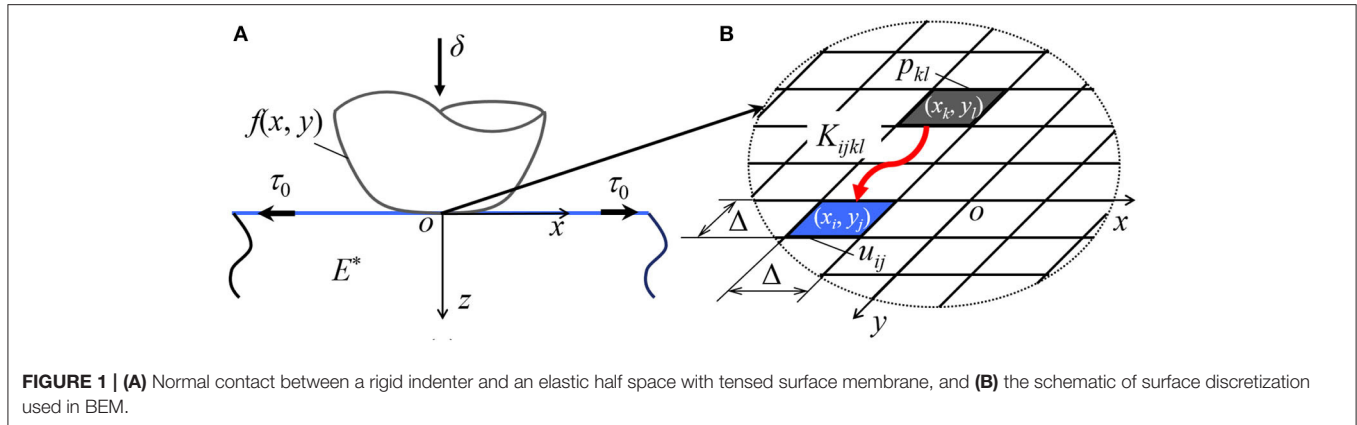


FIGURE 1 | (A) Normal contact between a rigid indenter and an elastic half space with tensed surface membrane, and **(B)** the schematic of surface discretization used in BEM.

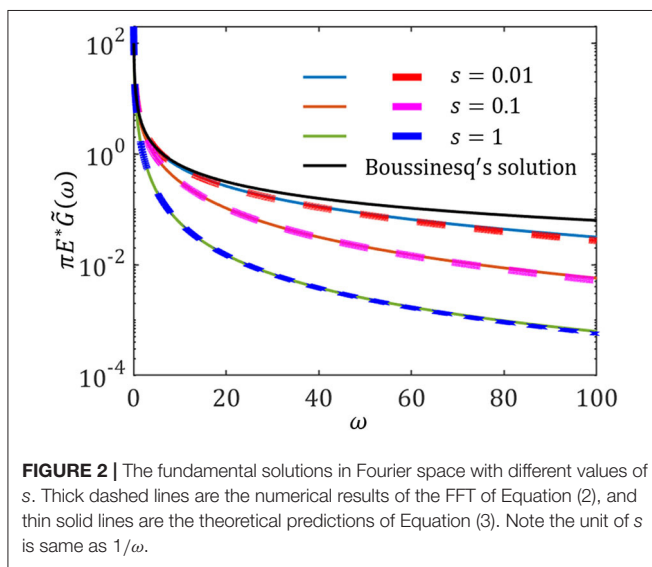


FIGURE 2 | The fundamental solutions in Fourier space with different values of s . Thick dashed lines are the numerical results of the FFT of Equation (2), and thin solid lines are the theoretical predictions of Equation (3). Note the unit of s is same as $1/\omega$.

angular frequency in the y -direction). It is found that the results are consistent with the theoretical prediction (Li and Popov, 2020),

$$\tilde{G}(\omega) = 2\pi / [\pi E^* (\omega + s\omega^2)] \quad (3)$$

which reduces to the Boussinesq's solution in Fourier space when the characteristic length s decreases to zero, $\tilde{G}(\omega) = 2 / (E^* \omega)$.

For the indentation of a rigid body with known profile $f(x, y)$ into a depth δ , see **Figure 1A**, the vertical displacement on the surface $u(x, y)$ has to fulfill the following condition:

$$\begin{cases} u(x, y) = \delta + f(x, y), & (x, y) \in \Omega \\ u(x, y) > \delta + f(x, y), & (x, y) \notin \Omega \end{cases} \quad (4)$$

where Ω is the domain of contact.

Assume that the positive contact pressure between the substrate and the indenter distributes as $p(x, y)$. Then, according

to the superposition principle, the vertical displacement generated by $p(x, y)$ can be expressed by

$$u(x, y) = \iint p(x', y') G(x - x', y - y') dx' dy' \quad (5)$$

which is exactly a two dimensional continuous convolution of fundamental solution and contact pressure distribution. As a result, substitution of Equation (5) into the displacement boundary condition Equation (4) leads to

$$\begin{aligned} \iint p(x', y') G(x - x', y - y') dx' dy' &= \delta \\ + f(x, y), & (x, y) \in \Omega \end{aligned} \quad (6)$$

Note that the inequality characterizing the non-overlapping condition outside the contact region is always accompanied with Equation (6) no matter which expression form is utilized. The resultant load P acting on the rigid indenter equals to the summation of the contact pressure over entire contact area. For a given indentation depth, the corresponding contact area and contact pressure are to be determined.

Next, this problem will be solved in the light of FFT based boundary element method. In contrast to finite element method, only part of substrate surface needs to be discretized. As shown in **Figure 1B**, squared elements with side length Δ are used to mesh the surface. Assuming that the pressure within each single element distributes uniformly, Equation (6) can be transformed into linear algebraic equations,

$$\sum_k \sum_l K_{ijkl} p_{kl} = \delta + f(x_i, y_j), \quad (x_i, y_j) \in \Omega \quad (7)$$

where

$$K_{ijkl} = \int_{x_k - \Delta/2}^{x_k + \Delta/2} \int_{y_l - \Delta/2}^{y_l + \Delta/2} G(x_i - x', y_j - y') dx' dy' \quad (8)$$

Note that the computational domain that is discretized into surface elements should include all potential contact

region. If the computational domain contains $N \times N$ elements, the influence coefficients K_{ijkl} have N^4 values. Using direct multiplication method has the complexity of $O(N^4)$. Instead, we prefer to apply the FFT technique to evaluate the vertical displacement for a given pressure, i.e., the left side of Equation (7). The complexity is then reduced to $O(N^2 \log N)$. It is worth mentioning that the multilevel multi-integration method suggested by Brandt and Lubrecht (1990) allows for the same computation acceleration.

The basic idea of the FFT based speed-up method is to interpret the left side of Equation (7) as a two-dimensional discrete convolution and applying the circular convolution theorem (Liu et al., 2000; Pohrt and Li, 2014). In this work, the influence coefficient matrix $(K_{kl})_{N \times N}$ is first prepared in real space by performing numerical integration of Equation (8) with $x_i = y_j = 0$. It should be pointed out that appropriate zero padding and wrap-around order operations to the original influence coefficient matrix $(K_{kl})_{N \times N}$ and pressure matrix $(p_{kl})_{N \times N}$ are necessary for the proper application of circular convolution theorem, which can avoid the periodic error (Liu et al., 2000). Next, FFT is implemented on the two matrices. The results can be easily obtained through element-wise multiplication of the influence coefficient matrix and pressure matrix in Fourier space. Then, performing an inverse FFT (IFFT) yields the vertical displacement caused by the pressure distribution $(p_{kl})_{N \times N}$ in real space. As a result, the problem becomes,

$$\text{IFFT} [\text{FFT}(\mathbf{K}) \circ \text{FFT}(\mathbf{p})] = \mathbf{h}, \text{ inside } \Omega \quad (9)$$

where \circ represents the element-wise product, and \mathbf{h} is given by the indenter shape and the indentation depth, $h_{ij} = \delta + f(x_i, y_j)$.

With the FFT based acceleration technique, this inverse problem can be solved using an iteration scheme based on conjugate gradient method (Liu et al., 2000; Polonsky and Keer, 2000; Pohrt and Li, 2014). The final obtained contact pressure $(p_{kl})_{N \times N}$ should be positive at all contact elements where the induced vertical displacement matches with the given rigid profile, and equal to zero in the non-contact region where the non-overlapping condition is always satisfied. Finally, the overall contact responses under a specified indentation loading δ are ready to be evaluated. The resultant force P applied on the indenter is obtained by summing up the forces at all discrete contacting elements. The real contact area A is computed by multiplying the number of contacting elements with the area of a single element Δ^2 . It should be pointed out that the contact area obtained by this direct summation method can be only accurate when the contacting surface is discretized by sufficiently refined grids. Area correction to continuum limit as done by Prodanov et al. (2014) is necessary. Inevitably, the BEM calculation with finer surface discretization would require more time cost. In addition, it is worth mentioning that Yastrebov et al. (2017) presented a correction route to compute accurate contact area with relatively coarse discretization for such kind of numerical method.

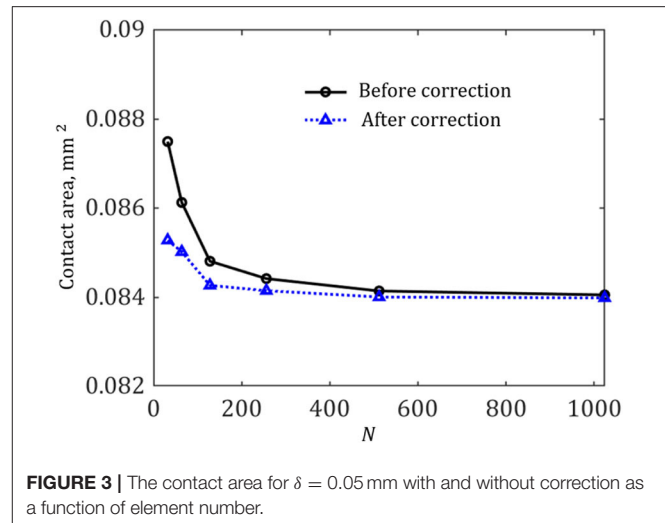


FIGURE 3 | The contact area for $\delta = 0.05$ mm with and without correction as a function of element number.

CALCULATION EXAMPLES

Contact With an Axisymmetric Parabolic Indenter

To validate the capability of the FFT-based BEM described in section Numerical method, we first consider a simple case that an axisymmetric parabolic indenter is pressed into a half space by a displacement δ . Numerical simulations are carried out for a soft material ($E^* = 10$ kPa), of which the surface is pre-tensed by a constant tension $\tau_0 = 1$ N/m (with negligible bending rigidity and in-plane stiffness). A square surface domain of edge length $2a_H$, where a_H is the contact radius predicted by Hertz theory, is meshed by $N \times N$ elements. The profile of the rigid indenter of curvature radius $R = 1$ mm is given by

$$f(x, y) = -\frac{1}{2R} (x^2 + y^2) \quad (10)$$

At first, the area in contact under different discretization level is examined to check the meshing dependent error in this BEM. In **Figure 3**, the contact area under a normal displacement $\delta = 0.05$ mm is plotted as a function of surface element number. It can be found that the area converges to a stable limit (i.e., the continuum limit) as the element number N increases. In this case, the maximum error is about 4% for the coarsest grid $N = 32$. The error gets sufficiently small ($< 0.5\%$) when N is larger than 256. Moreover, by using the correction method proposed by Yastrebov et al. (2017), it is found that the grid can be coarser to achieve same accuracy as that of finer meshing without correction. For convenience, we use the meshing grid of $N = 256$ to simulate the contacts of single asperity, which is already enough to get accurate results even without correction.

The influence of membrane/surface tension can be reflected by a dimensionless ratio of the characteristic length s (defined by $2\tau_0/E^*$) to the contact radius a . **Figure 4** displays the distributions of contact pressure for the cases of different ratio s/a . The variations of indentation load and indentation depth with respect to the ratio s/a are plotted in **Figures 5A,B**,

respectively. It is seen that our numerical results obtained from BEM calculations coincide well with existing semi-analytical solutions (Kim and Gouldstone, 2008; Long et al., 2017). The size dependent contact behavior is reproduced. Apparently, when the contact size a is comparable or smaller than the characteristic length, the contact response will essentially deviate from the classical Hertz contact theory. Compared with the prediction of Hertz theory, a higher indentation load or a larger indentation depth is needed to achieve a specific contact area because of the presence of membrane/surface tension.

Contact With an Elliptical Indenter

Let us further consider a smooth indenter with quadric surface, of which the principle curvatures (R_1, R_2) are unequal. By adjusting the x -axis and y -axis along the axes of principle curvature, the indenter shape can be expressed by

$$f(x, y) = -\left(\frac{x^2}{2R_1} + \frac{y^2}{2R_2}\right) \quad (11)$$

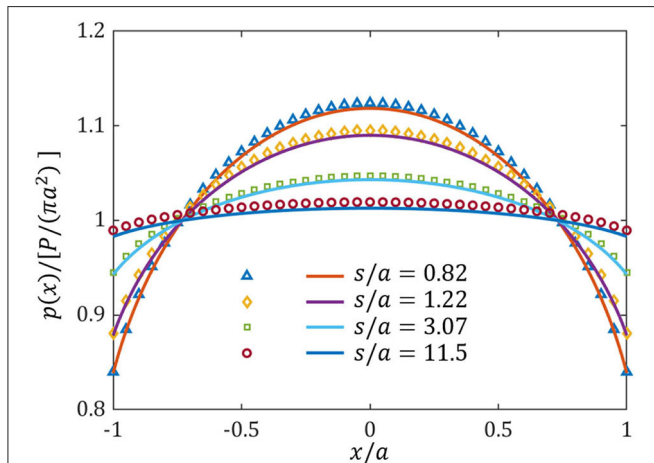


FIGURE 4 | Contact pressure distribution along x -axis direction. Solid lines are the semi-analytical solution from Kim and Gouldstone (2008), and scatter symbols are the numerical results from BEM.

For a given displacement $\delta = 0.05$ mm exerted on the rigid indenter of $R_1 = 1$ mm and $R_2 = 2$ mm, the distributions of normal pressure on the half space with different membrane/surface tension are shown in **Figure 6**. Based on the pressure distribution, the contact region can be determined, which is identified as an ellipse with semi-minor axis a_1 and semi-major axis a_2 . With the membrane/surface tension increasing, the elliptical contact area shrinks, and the distribution of contact pressure tends to be more uniform. According to the classical Hertz theory, the ratio of a_1/a_2 depends only on the ratio of principle curvatures of the indenter R_1/R_2 . However, it is noticed that the value of a_1/a_2 no longer keeps constant for a given indenter, but declines as the membrane/surface tension increases. In other words, the contact ellipse will be somewhat slender if the surface membrane is equi-biaxially tensed. This phenomenon can be explained by the size-dependent behavior caused by membrane/surface tension. Because the length along the minor axis of contact ellipse is smaller than that along major axis, the size effect on the reduction of contact dimension nearby minor axis would be relatively more remarkable.

Contact With a Rough Surface

With the validated BEM, we are also able to calculate the response of rough surface contact in the presence of membrane/surface tension. As illustrated in **Figure 7**, we consider the normal contact between a square rigid random rough surface and a soft elastic half space with tensed surface membrane. Assume that the rough surface is self-affine fractal and has the following power spectral density (PSD),

$$C_{2D}(q) = \begin{cases} C_0 q^{-2(H+1)}, & q_0 \leq q \leq q_s \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where C_0 is a constant, H is the Hurst exponent, q is the wave vector, $q_0 = 2\pi/L$ and q_s are the roll-off and cutoff wavenumbers, respectively.

Based on this power spectrum, a discrete rough surface can be generated by using the inverse Fast Fourier Transform (Pohrt and Popov, 2012). Then, the surface height data $f(x_i, y_j)$ over a uniform $N \times N$ grid are superimposed onto the bottom of a

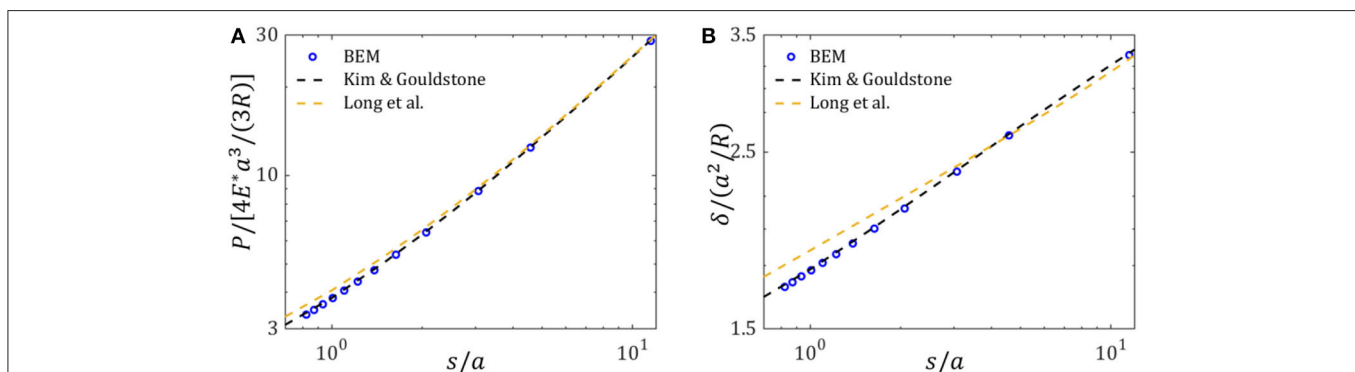


FIGURE 5 | Dependences of contact responses (normalized by the prediction of Hertz theory) on the ratio s/a : (A) indentation load; (B) indentation depth.

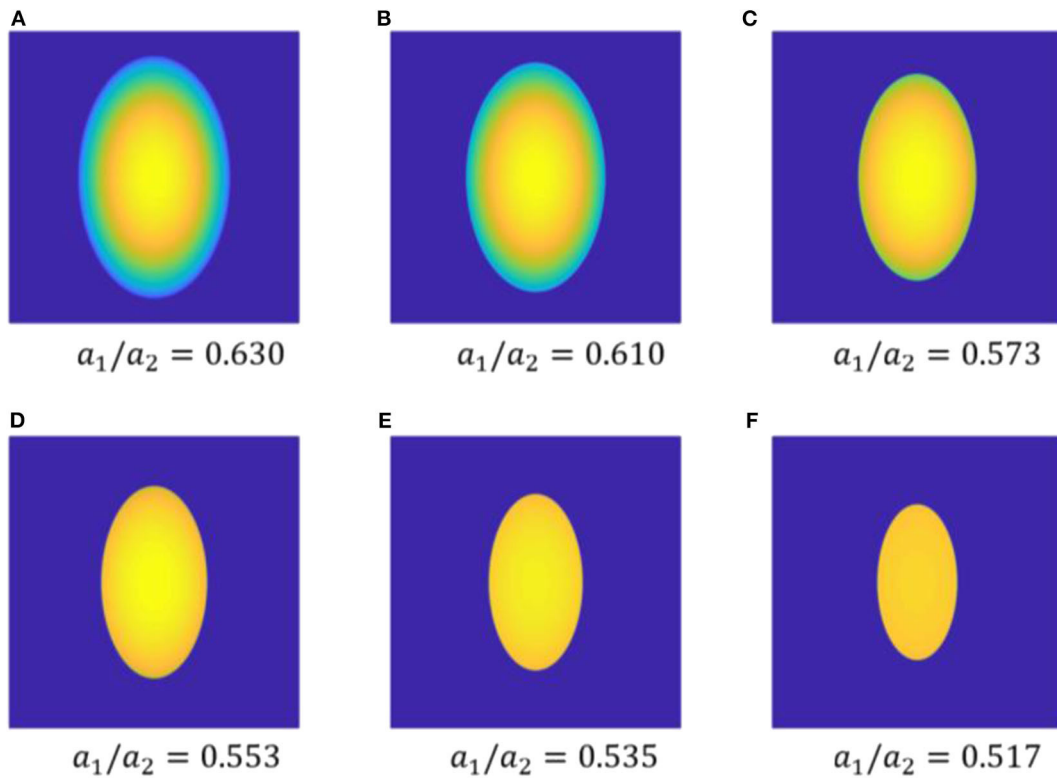


FIGURE 6 | Distributions of normal pressure of the contact between an elliptical indenter and a soft substrate with modulus $E^* = 10$ kPa and different membrane/surface tension: **(A)** $\tau_0 = 0$ N/m, **(B)** $\tau_0 = 0.1$ N/m, **(C)** $\tau_0 = 0.5$ N/m, **(D)** $\tau_0 = 1$ N/m, **(E)** $\tau_0 = 2$ N/m and **(F)** $\tau_0 = 5$ N/m.

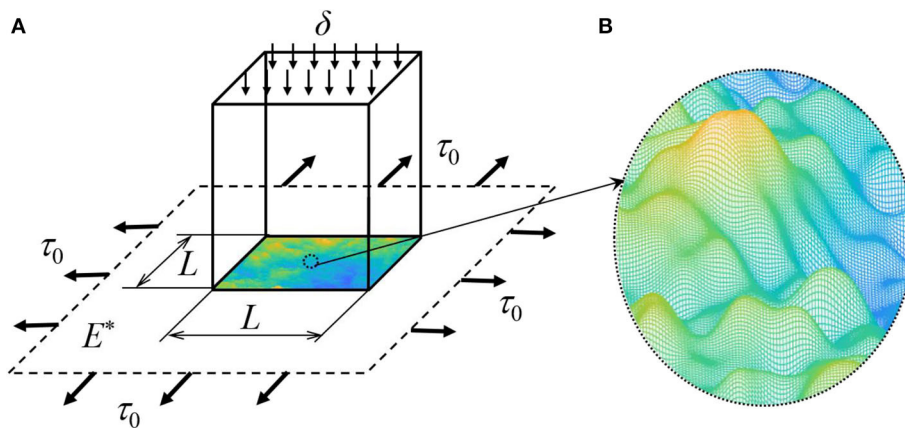
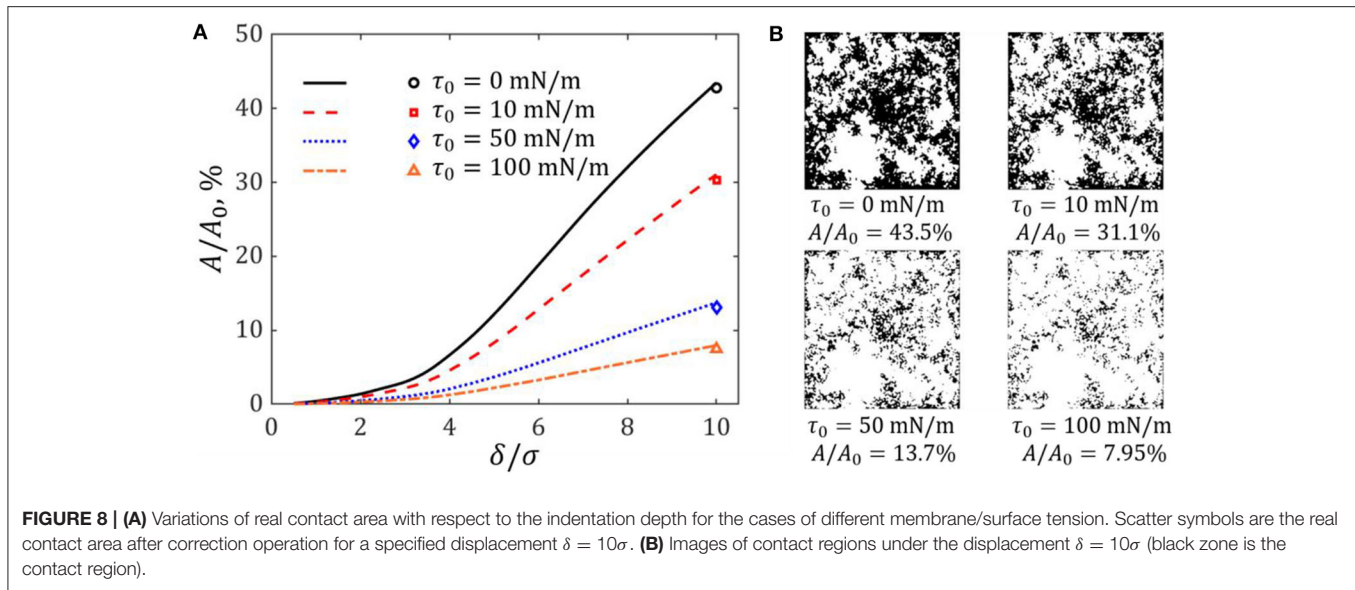


FIGURE 7 | (A) Indentation on the elastic half space by a square punch with fractal rough surface, and **(B)** the detail of generated rough surface.

square flat-ended indenter with length L . Note that the highest point of rough surface is initially put on the surface of half space. Accordingly, a square surface region on the half space $A_0 = L \times L$ that is corresponding to the projected area of indenter is meshed by $N \times N$ surface elements. Partial contact happens when the indenter moves downward by a displacement δ .

In general, the elastic contact response of rough surface should be dependent on the material property E^* and the

surface topography parameters including the root mean square (RMS) roughness σ , Hurst exponent H , the system size $L = 2\pi/q_0$, and the cut-off wavenumbers q_s . In this case with membrane/surface tension τ_0 , the contact response ϕ for a given indentation depth δ should be a function of these variables, $\phi = \phi(\tau_0, E^*, \sigma, H, L, q_s, \delta)$. Dimensional analysis reveals that the effects of membrane/surface tension can be reflected by a dimensionless parameter, $(\tau_0/E^*)/(\sigma^2/L)$. In this work, we would

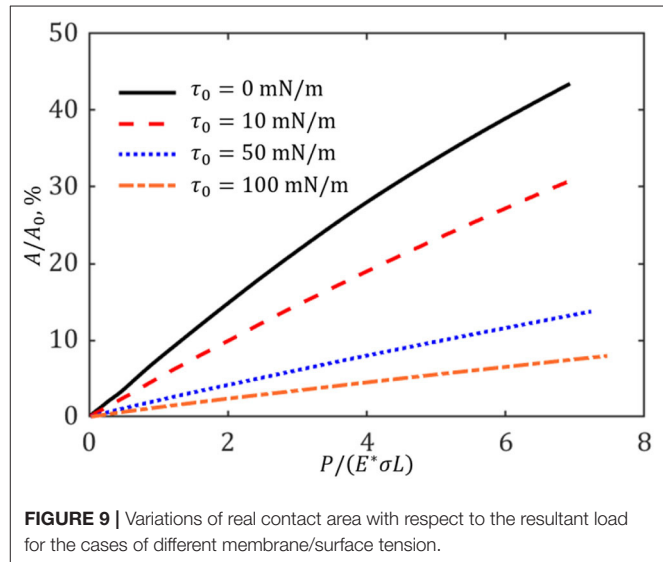


lay aside the role of surface topography property, which is controlled by the dimensionless parameters including σ/L , H , and σq_s , and just display the effects of membrane/surface tension.

An artificial rough surface ($N = 1,024$, $L = 1$ mm) with RMS roughness $\sigma = 0.01$ mm, $H = 0.7$, $q_0 = 2\pi/L$ and $q_s = 2\pi/(16L/N)$ is considered. Calculation examples are carried out for this rough surface in contact with a soft substrate with reduced modulus $E^* = 10$ kPa and different membrane tensions $\tau_0 = 0, 10, 50, 100$ mN/m. The real contact area A/A_0 and indentation load $P/(E^*\sigma L)$ are obtained for different indentation depth δ/σ .

Figure 8A shows the dependence of A/A_0 on δ/σ for the cases of different membrane/surface tension. For a given indenter displacement, the contact area significantly declines with the increasing of applied membrane/surface tension. For instance, when the indenter displacement $\delta = 10\sigma$, the fraction area of contact drops from 43.5% for a substrate with un-tensed surface to 7.95% for that having membrane/surface tension of 100 mN/m. The configurations of contact regions for $\delta = 10\sigma$ are exhibited in **Figure 8B**. Note that the real contact area computed by direct summation of contacting elements has been compared with that evaluated by the correction technique of Yastrebov et al. (2017). As shown in **Figure 8A**, the difference is small. Therefore, the contact area calculation for this level of discretization should be proper and accurate. In addition, the variation of A/A_0 with respect to $P/(E^*\sigma L)$ is displayed in **Figure 9**. The slope of the curve decreases as the value of τ_0 increases, which means that to generate a specified contact area needs a higher external indentation load for the case of larger membrane/surface tension. As a result, for a given substrate with determined modulus indented by a specific rough surface, the mean contact stress defined by the ratio of indentation load to the real contact area, P/A , can be increased by applying a certain membrane/surface tension on the substrate surface.

For this problem of rough surface contact, it should be pointed out that the analytical relations between the contact



response and the properties of substrate material and indenter rough surface have not been generalized in this paper. More numerical calculations for different rough surface and different substrate with different membrane/surface tension are required in order to achieve this point, which will be conducted in the future.

CONCLUSION

In summary, the FFT-based boundary element method is extended for normal contact of soft material, of which the surface is tensed by equi-biaxial tension. Three type of indenters compressing on the elastic half space with constant membrane/surface tension are considered. The calculation results show excellent agreement with related available solutions.

Typical size-dependent contact behavior is demonstrated. In particular, due to the presence of membrane/surface tension, the eccentricity of contact ellipse will be slightly increased for the contact with a rigid elliptical indenter. For a given substrate indented by a specific rigid indenter, the mean contact pressure defined by the ratio of indentation load to the real contact area can be increased by increasing membrane/surface tension on the substrate surface. This numerical method would be helpful for the deformation analysis of some soft systems with membrane/surface tension.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

REFERENCES

- Argatov, I., and Mishuris, G. (2016). An asymptotic model for a thin bonded elastic layer coated with an elastic membrane. *Appl. Math. Modell.* 40, 2541–2548. doi: 10.1016/j.apm.2015.09.109
- Argatov, I. I., and Sabina, F. J. (2012). Spherical indentation of a transversely isotropic elastic half-space reinforced with a thin layer. *Int. J. Eng. Sci.* 50, 132–143. doi: 10.1016/j.ijengsci.2011.08.009
- Brandt, A., and Lubrecht, A. A. (1990). Multilevel matrix multiplication and fast solution of integral equations. *J. Comput. Phys.* 90:348–70. doi: 10.1016/0021-9991(90)90171-V
- Bugnicourt, R., Sainsot, P., Dureisseix, D., Gauthier, C., and Lubrecht, A. A. (2018). FFT-based methods for solving a rough adhesive contact: description and convergence study. *Tribol. Lett.* 66:29. doi: 10.1007/s11249-017-0980-z
- Campaña, C., and Müser, M. H. (2006). Practical Green's function approach to the simulation of elastic semi-infinite solids. *Phys. Rev. B.* 74:075420. doi: 10.1103/PhysRevB.74.075420
- Gurtin, M. E., and Murdoch, A. I. (1975). A continuum theory of elastic material surfaces. *Arch. Rat. Mech. Anal.* 57, 291–323. doi: 10.1007/BF00261375
- Hajji, M. A. (1978). Indentation of a membrane on an elastic half space. *ASME J. Appl. Mech.* 45, 320–324. doi: 10.1115/1.3424295
- He, L. H., and Lim, C. W. (2006). Surface green function for a soft elastic half-space: influence of surface stress. *Int. J. Solids Struct.* 43, 132–143. doi: 10.1016/j.ijsolstr.2005.04.026
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. R. Soc. London A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Kim, J. H., and Gouldstone, A. (2008). Spherical indentation of a membrane on an elastic half-space. *J. Mater. Res.* 23, 2212–2220. doi: 10.1557/JMR.2008.0278
- Li, M., Zhang, H. X., Zhao, Z. L., and Feng, X. Q. (2020). Surface effects on cylindrical indentation of a soft layer on a rigid substrate. *Acta Mech. Sinica.* 36, 422–429. doi: 10.1007/s10409-020-00941-8
- Li, Q., and Popov, V. L. (2020). Non-adhesive contacts with different surface tension inside and outside the contact area. *Front. Mech. Eng.* doi: 10.3389/fmech.2020.00063
- Liu, S., Wang, Q., and Liu, G. (2000). A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. *Wear.* 243, 101–111. doi: 10.1016/S0043-1648(00)00427-0
- Long, J. M., Ding, Y., Yuan, W. K., Chen, W., and Wang, G. F. (2017). General relations of indentations on solids with surface tension. *ASME J. Appl. Mech.* 84:051007. doi: 10.1115/1.4036214
- Nogi, T., and Kato, T. (1997). Influence of a hard surface layer on the limit of elastic contact - Part I: Analysis using a real surface model. *J. Tribol.* 119, 493–500. doi: 10.1115/1.2833525
- Pohrt, R., and Li, Q. (2014). Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17, 334–340. doi: 10.1134/S1029959914040109

AUTHOR CONTRIBUTIONS

WY and GW conceived the study. WY carried out numerical simulations. All authors drafted and reviewed the manuscript.

FUNDING

This financial support from the National Natural Science Foundation of China (Grant No.11525209) was acknowledged. WY also thanks the support from the China Scholarship Council (CSC).

ACKNOWLEDGMENTS

The authors are grateful to Dr. Q. Li and Prof. V. L. Popov for helpful discussions on the BEM and providing results from a non-published manuscript.

- Pohrt, R., and Popov, V. L. (2012). Normal contact stiffness of elastic solids with fractal rough surfaces. *Phys. Rev. Lett.* 108:104301. doi: 10.1103/PhysRevLett.108.104301
- Pohrt, R., and Popov, V. L. (2015). Adhesive contact simulation of elastic solids using local mesh-dependent detachment criterion in boundary element method. *Facta Universitatis.* 13, 3–10.
- Polonsky, I. A., and Keer, L. M. (2000). Fast methods for solving rough contact problems: a comparative study. *ASME. J. Tribol.* 122, 36–41. doi: 10.1115/1.555326
- Prodanov, N., Dapp, W. B., and Müser, M. H. (2014). On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections, and reference data. *Tribol. Lett.* 53, 433–448. doi: 10.1007/s11249-013-0282-z
- Rey, V., Anciaux, G., and Molinari, J. F. (2017). Normal adhesive contact on rough surfaces: efficient algorithm for FFT-based BEM resolution. *Comput. Mech.* 60, 69–81. doi: 10.1007/s00466-017-1392-5
- Wang, G. F., and Feng, X. Q. (2007). Effects of surface stresses on contact problems at nanoscale. *J. Appl. Phys.* 101:013510. doi: 10.1063/1.2405127
- Yastrebov, V. A., Anciaux, G., and Molinari, J. F. (2017). On the accurate computation of the true contact-area in mechanical contact of random rough surfaces. *Tribol. Int.* 114, 161–171. doi: 10.1016/j.triboint.2017.04.023
- Zamir, E. A., and Taber, L. A. (2004). On the effects of residual stress in microindentation tests of soft tissue structures. *ASME J. Biomech. Eng.* 126, 276–283. doi: 10.1115/1.1695573
- Zhang, C. Y., and Zhang, Y. W. (2009). Extracting elastic properties and prestress of a cell using atomic force microscopy. *J. Mater. Res.* 24, 1167–1171. doi: 10.1557/jmr.2009.0121
- Zhang, M. G., Cao, Y. P., Li, G. Y., and Feng, X. Q. (2014). Spherical indentation method for determining the constitutive parameters of hyperelastic soft materials. *Biomech. Model. Mechanobiol.* 13, 1–11. doi: 10.1007/s10237-013-0481-4
- Zhou, S., and Gao, X. L. (2013). Solutions of half-space and half-plane contact problems based on surface elasticity. *Z. Angew. Math. Phys.* 64, 145–166. doi: 10.1007/s00033-012-0205-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Yuan and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multi-Scale Surface Roughness Optimization Through Genetic Algorithms

Paolo Cinat, Giorgio Gnecco* and Marco Paggi

IMT School for Advanced Studies, Lucca, Italy

OPEN ACCESS

Edited by:

Martin H. Müser,
Saarland University, Germany

Reviewed by:

Robert Jackson,
Auburn University, United States
Yoshitaka Nakanishi,
Kumamoto University, Japan
Giuseppe Carbone,
Politecnico di Bari, Italy

*Correspondence:

Giorgio Gnecco
giorgio.gnecco@imtlucca.it

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 23 December 2019

Accepted: 21 April 2020

Published: 15 May 2020

Citation:

Cinat P, Gnecco G and Paggi M
(2020) Multi-Scale Surface Roughness
Optimization Through Genetic
Algorithms. *Front. Mech. Eng.* 6:29.
doi: 10.3389/fmech.2020.00029

Artificial intelligence is changing perspectives of industries about manufacturing of components, introducing emerging techniques such as additive manufacturing technologies. These techniques can be exploited to manufacture not only precision mechanical components, but also interfaces. In this context, we investigate the use of artificial intelligence and in particular genetic algorithms to identify optimal multi-scale roughness features to design prototype surfaces achieving a target contact mechanics response. Exploiting an analogy with biology, the features of roughness at a given length scale are described through model profiles named chromosomes. In the present work, the mathematical description of chromosomes is firstly provided, then three genetic algorithms are proposed to superimpose and combine them in order to identify optimal roughness features. The three methods are compared, discussing the topological and spectral features of roughness obtained in each case.

Keywords: surface roughness, multivariate Weierstrass-Mandelbrot function, contact mechanics, optimization, genetic algorithms

1. INTRODUCTION

It is well-established in the literature that surface topology and texture are important for enhancing the tribological behavior of contacts. Therefore, optimization of topological features is considered as a research topic of paramount interest for industrial applications. For instance, optimization of the shape and position of periodic grooved textures has been investigated in Buscaglia et al. (2007) in relation to lubrication. Even more ambitious are the recent attempts to tailor roughness of engineering surfaces by controlling end-milling operations, as discussed in Zhang et al. (2007), or by cutting techniques as in Nouari et al. (2018). Artificial intelligence based on genetic algorithms (Zain et al., 2008) and artificial neural networks (Moghri et al., 2014) have been recently exploited to control milling operations and surface roughness manufacturing by material removal. As another strategy, additive manufacturing (Brettel et al., 2014) is opening new perspectives to produce surfaces with specified roughness (see e.g., Farina et al., 2016; Ko et al., 2019; Wüst et al., 2020).

In addition to tribology, the role of an interface between material constituents/phases is becoming progressively dominant over the one of bulk properties, thanks to the increasing trend in the miniaturization of components, and to the significant progress in the design of mechanical systems and materials, starting from the sub-micrometer scale. The interface has very different properties than the bulk, and is important in a variety of mechanical, physical, and chemical processes. Interfaces can have significant effects also on the properties of composite materials (Moroni et al., 2013; Guo et al., 2014, 2017; Kwon et al., 2019). Consequently, industrial applications often require the design of surface textures able to achieve given target responses and/or to enable

rapid manufacturing and morphology changes for in-line control of mechanical components, as in Bora et al. (2005) for Micro-Electro-Mechanical Systems (MEMS). Nature is also offering interesting perspectives for the identification of optimal surface topologies for selected applications, see e.g., the successful attempts to create artificial super-adhesives like Gecko's pads (Sherge and Gorb, 2001; Gao et al., 2005; Nosonovsky and Bushan, 2008).

In the present study, a theoretical multi-scale description of roughness is considered in relation to the Multivariate Weierstrass-Mandelbrot (MWM) function as a model for a fractal rough profile (Cinat et al., 2019). The aim of the research is to establish a mathematical and a computational framework for the robust identification of the profile model parameters that lead to identified topologies whose contact response is matching any requested real contact area-load and contact conductance-load curves that can be specified by the user. Although the authors are aware that not all the natural and manufactured surfaces do obey to the fractal scaling, as pointed out by Carr and Benzer (1991), Wen and Sinding-Larsen (1997), Borri and Paggi (2015), Xiaohan et al. (2017), and Gujrati et al. (2018), the choice here is motivated by the fact that the MWM function has been widely studied in relation to contact mechanics, and, for instance, we expect to find that optimization in the low frequency range is the major concern for controlling the thermal/electric contact conductance, while fine scale features are governing the real area of contact (see e.g., Ciavarella et al., 2000, 2004; Paggi and Barber, 2011). However, it has to be highlighted that, although the MWM model is herein adopted for validating identification predictions in relation to benchmark results established in the literature, the overall proposed computational procedure can be applied also to other profile or surface models as well, without specific restrictions.

Motivated by an analogy with biology, the expression *surface roughness genome* is herein used to denote the ensemble of parameters, the *genes*, associated with the set of elementary waves that describes the main features of roughness over multiple length scales. Although such a terminology imported from biology is not usually adopted in the contact mechanics community, it is introduced here for its common use in the research area of genetic algorithms. Within this framework, we optimize the parameters above to identify a suitable genome that produces a rough profile having a frictionless elastic normal contact mechanics response close to a requested target one. To achieve this goal, we propose various genetic algorithms, combining mechanical considerations and suitable optimization tools. The contact problem is solved in terms of the boundary element method (BEM) (Vakis et al., 2018; Paggi and Hills, 2020), based on the formulation validated in Bemporad and Paggi (2015). See also Bemporad and Paggi (2015) for a wide comparison of other BEM techniques that could also be effectively applied to solve the same contact problem.

Regarding the genetic algorithm research contribution, differently from the previous publication by Cinat et al. (2019) where a single algorithm was proposed to achieve that task, here we propose an extensive numerical comparison

of the performance of various algorithms with clear distinct features.

This article is structured as follows. In section 2, we describe the normal contact problem, the surface roughness genome, the features of a single length scale of roughness, and the roughness reconstruction over multiple length scales. In section 3, we propose three genetic algorithms, whose aim is to generate prototype profiles able to achieve given target contact mechanics responses. In section 4, the effectiveness of these algorithms is tested and discussed through a representative example, based on an artificial genome database. Finally, section 5 concludes the paper and discusses possible future work.

2. NORMAL CONTACT PROBLEM AND ROUGHNESS DESCRIPTION

According to Johnson (2003), the non-conforming contact problem between two rough surfaces is proved to be equivalent—under the assumption of linear elasticity—to the contact problem between a rigid rough surface and an elastic half-plane, provided with suitable effective elastic parameters. The reader is referred to Barber (2003) for details and for a proof of this equivalence. In this work, normal contact is controlled by imposing an approaching far-field closing displacement Δ , which corresponds to a rigid-body motion of the half-plane. Its value is computed from the tallest summit of the rough surface. Elastic interactions among asperities deform the half-plane. During the deformation process, the tallest summit of the profile remains in contact with the half-plane, while other asperities may loose such contact. Their deformation produces a normal contact traction distribution in the contact area A . Then, the average total contact pressure p is computed as the ratio between the sum of all the contact forces acting in this area, and the nominal contact area. Finally, the contact stiffness K is computed as the derivative of the contact pressure p with respect to the imposed displacement Δ . The reader is referred to Bemporad and Paggi (2015) for more details about the precise problem formulation and for a comparison of several techniques that can be exploited to make the unilateral contact constraints satisfied.

In the article, a surface roughness description is provided, based on several parameters. For each choice of such parameters, the contact problem above is solved by an application of the boundary element method (BEM), as in Bemporad and Paggi (2015).

2.1. Description of Multi-Scale Surface Roughness

In the following, one-dimensional rough profiles are considered. Likewise in Cinat et al. (2019), their height $Z(x)$, expressed as a function of the position x , is described in terms of the real part of the Multivariate Weierstrass-Mandelbrot function (MWM), which was proposed in Ausloos and Berman (1985) to describe stochastic processes having a larger number of features than in the original 2D formulation proposed earlier in Mandelbrot (1977). In the work, a rough profile $Z(x)$ is described parametrically as

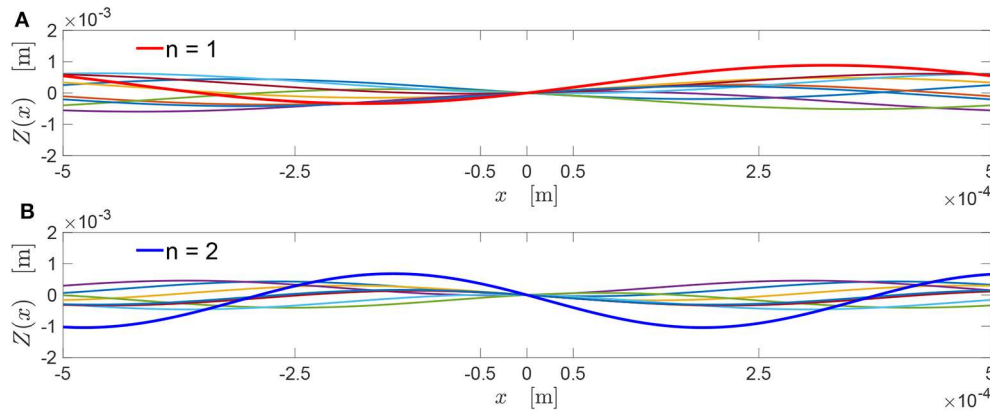


FIGURE 1 | (A) The chromosome $C_1(x)$ is visualized by the red line, and corresponds to the sum of the other colored co-sinusoids with $n = 1$. **(B)** The chromosome $C_2(x)$ corresponds to the set of co-sinusoids provided by the choice $n = 2$.

$$Z(x) = A \sqrt{\frac{\log_{10}(\gamma)}{\mathcal{M}}} \left(\frac{2\pi}{\lambda} \right)^{-\mathcal{H}} \sum_{m=1}^{\mathcal{M}} \sum_{n=-\infty}^{+\infty} \gamma^{-(n-1)\mathcal{H}} \left[\cos(\phi_{m,n}) - \cos\left(\frac{2\pi}{\lambda} \gamma^{n-1} x + \phi_{m,n} \right) \right]. \quad (1)$$

Each co-sinusoid in Equation (1) is defined by a unique combination of the parameters \mathcal{H} , \mathcal{A} , λ , $\gamma > 0$, and $\phi_{m,n} \in [0, 2\pi)$, whereas \mathcal{M} is a positive integer, which represents the number of ridges. In this framework, in analogy with biology, these parameters are also called *genes*. Thus, the *surface roughness genome* is the overall ensemble of genes providing the realization of a surface over multiple length scales.

2.2. Description of a Generic Length Scale of Roughness

The multi-scale description of profiles is governed by the genes \mathcal{H} , λ , γ . The gene γ rules the distance of consecutive wavelengths in the frequency domain, and \mathcal{H} their scaling in amplitude. The gene λ indicates a reference wavelength. The particular combination of genes associated with a fixed n identifies a rough profile $C_n(x)$, which is named *chromosome*, and is associated to the features of roughness at the fixed reference length-scale $\lambda_n = \lambda \gamma^{1-n}$.

In biology, a chromosome is a structure composed by some genes identifying specific features of the genome (King et al., 2006). Here, it identifies the features of roughness at the wavelength λ_n , as follows:

$$C_n(x) = A \sqrt{\frac{\log_{10}(\gamma)}{\mathcal{M}}} \left(\frac{2\pi}{\lambda} \right)^{-\mathcal{H}} \sum_{m=1}^{\mathcal{M}} \gamma^{-(n-1)\mathcal{H}} \left[\cos(\phi_{m,n}) - \cos\left(\frac{2\pi}{\lambda} \gamma^{n-1} x + \phi_{m,n} \right) \right]. \quad (2)$$

A profile $Z(x)$ is realized in an observation length L in $N = \frac{L}{\delta} + 1$ nodes, where δ is the distance between two consecutive nodes, i.e., the resolution, and L is a multiple of δ . Hence, the infinite

series in Equation (1) is replaced by a finite series, according to the observation scale L and the resolution δ chosen. The profile $Z(x)$ is realized from its longest to its shortest observed wavelength by summing a finite number n_c of chromosomes:

$$Z(x) = \sum_{n=n_s}^{n_f} C_n(x), \quad (3)$$

where n_s (starting index) and n_f (final index) refer to the chromosomes contributing with the longest and shortest wavelengths to the realization of the rough profile. It holds $n_c = n_f - n_s + 1$.

The realization of roughness over multiple length scales of observation is obtained by selecting different chromosomes, i.e., by varying n between n_s and n_f . The value n_s refers to the longest wavelength considered, and n_f refers to the shortest wavelength observed. In this way, the chromosomes contributing to a realization at a given length scale depend on n_s and n_f . Without loss of generality, the value $n_s = 1$ is assigned to the chromosome with longest wavelength referring to the coarsest realization of a surface. In more details, denoting by $\lfloor \cdot \rfloor$ the largest integer smaller than or equal to its argument, one has

$$\begin{cases} n_s = \lfloor \log_\gamma \left(\frac{\lambda}{L} \right) \rfloor + 1 \geq 1, \\ n_f = \lfloor \log_\gamma \left(\frac{\lambda}{\delta} \right) \rfloor + 1, \\ n_c = n_f - n_s + 1 = \lfloor \log_\gamma \left(\frac{\lambda}{\delta} \right) \rfloor. \end{cases} \quad (4)$$

To visualize the concept of chromosome, a simple example is presented in **Figure 1**. Here, two chromosomes are shown for two different values of n ($n = 1, 2$). In both cases, the parent $\mathcal{M} = 8$ co-sinusoids are shown in **Figures 1A,B** through the colored lines.

The chromosome $C_1(x)$ is assumed to have a wavelength equal to the sample wavelength, i.e., $\lambda_1 = L = 100 \mu\text{m}$. This chromosome, depicted by a thick red line in **Figure 1A**, is obtained by summing all the $\mathcal{M} = 8$ co-sinusoids with the associated value $n = 1$.

The chromosome $C_2(x)$, depicted by blue line in **Figure 1B**, is obtained as the sum of all the $\mathcal{M} = 8$ co-sinusoids with $n = 2$. Then, since it is imposed $\gamma = 1.5$, the chromosome $C_2(x)$ has $\lambda_2 = \frac{\lambda_1}{1.5} = 66.6 \mu\text{m}$.

Each of the two chromosomes in **Figure 1** maintains a co-sinusoidal shape with the same wavelength λ_n as the associated parent co-sinusoids, and with a phase angle θ . Its height field in Equation (2) can be also written as

$$C_n(x) = \mathcal{G}_n \left\{ g_{n,1} \left[1 - \cos \left(\frac{2\pi}{\lambda} \gamma^{n-1} x \right) \right] + g_{n,2} \sin \left(\frac{2\pi}{\lambda} \gamma^{n-1} x \right) \right\}. \quad (5)$$

In Equation (5), the amplitude parameter $\mathcal{G}_n = \mathcal{A} \sqrt{\frac{\log_{10}(\gamma)}{\mathcal{M}}} \left(\frac{2\pi}{\lambda} \right)^{-\mathcal{H}} \gamma^{-(n-1)\mathcal{H}}$ has been introduced, together with the two constants $g_{n,1}$ and $g_{n,2}$, which depend on the phases $\phi_{m,n}$ of the ridges with first index m :

$$g_{n,1} = \sum_{m=1}^{\mathcal{M}} \cos(\phi_{m,n}), \quad g_{n,2} = \sum_{m=1}^{\mathcal{M}} \sin(\phi_{m,n}). \quad (6)$$

A feasible mathematical model of a chromosome might be one in a form similar to that of the MWM function in Equation (1) with $\mathcal{M} = 1$ and a single value of n , obtained by introducing the parameters \mathcal{K}_n , $\theta_{n,1}$, and $\theta_{n,2}$:

$$C_n(x) = \mathcal{K}_n \left[\cos(\theta_{n,1}) - \cos \left(\frac{2\pi}{\lambda} \gamma^{n-1} x + \theta_{n,2} \right) \right]. \quad (7)$$

A particular case of Equation (7) is when the angles $\theta_{n,1}$ and $\theta_{n,2}$ coincide, say, with the same θ_n . In such a case, a chromosome is described exactly according to the MWM profile in Equation (1) with $\mathcal{M} = 1$ and a single value of n . The expressions of \mathcal{K}_n , θ_n in Equation (7) are obtained equating Equations (5) and (7), and are given by

$$\begin{cases} \mathcal{K}_n = \mathcal{G}_n \sqrt{\mathcal{M} + 2g_{n,3}}, \\ \theta_n = \arccos \left(\frac{g_{n,1}}{\sqrt{\mathcal{M} + 2g_{n,3}}} \right), \end{cases} \quad (8)$$

where the parameter $g_{n,3}$ is defined as follows:

$$g_{n,3} = \sum_{i=1}^{\mathcal{M}-1} \sum_{j=i+1}^{\mathcal{M}} \cos(\phi_{j,n} - \phi_{i,n}). \quad (9)$$

For its validity, Equation (8) requires the argument of $\arccos \left(\frac{g_{n,1}}{\sqrt{\mathcal{M} + 2g_{n,3}}} \right)$ to be between -1 and 1 , i.e.,

$$|g_{n,1}| \leq \sqrt{\mathcal{M} + 2g_{n,3}}. \quad (10)$$

In the above, the additional condition $\mathcal{M} + 2g_{n,3} \geq 0$ has not been imposed explicitly, since it always holds. Indeed, one can easily check that $\mathcal{M} + 2g_{n,3}$ is the square of the Euclidean norm of the

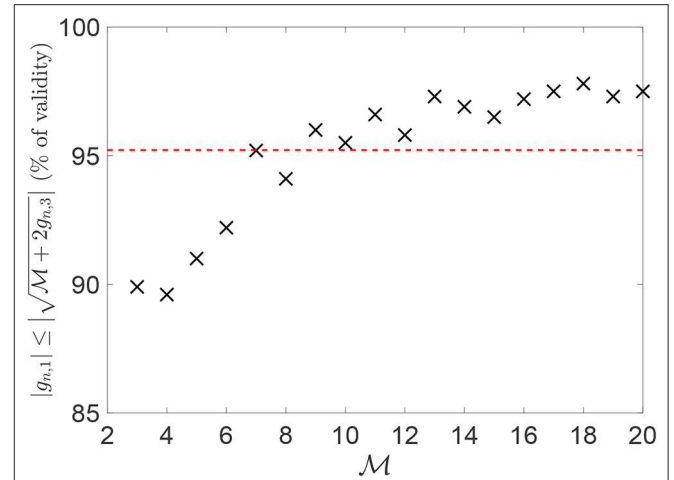


FIGURE 2 | Percentage of validity of Equation (10) over 1,000 random choices for the phases (uniformly and independently sampled between 0 and 2π), for each value of \mathcal{M} . The red line depicts the average percentage of validity.

vector with components $\left(\sum_{i=1}^{\mathcal{M}} \cos(\phi_{i,n}), \sum_{i=1}^{\mathcal{M}} \sin(\phi_{i,n}) \right)$. Of course, this norm is always larger than or equal to 0.

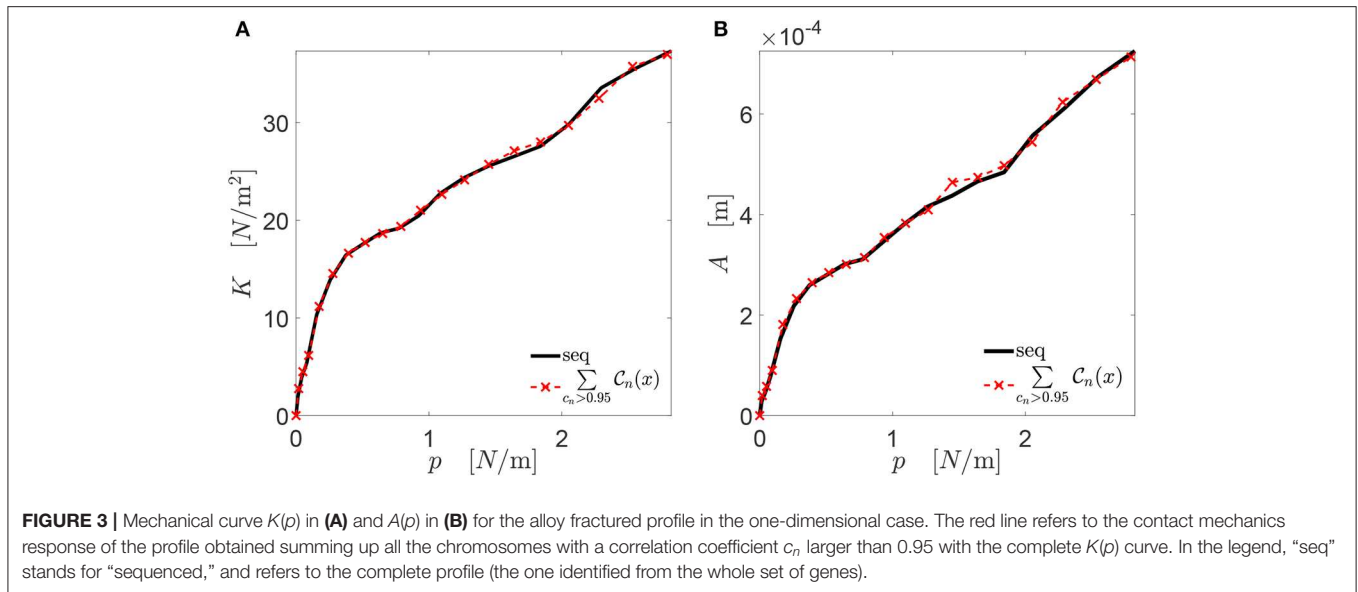
A numerical simulation has been conducted to verify the validity of condition (10). This has been evaluated for several values of \mathcal{M} , considering each time 1,000 random choices (uniformly and independently sampled between 0 and 2π) for the phases. **Figure 2** reports, for each such value of \mathcal{M} , the percentage of cases for which Equation (10) is satisfied. Such percentage ranges between 88.5 and 98%, with an average value of $\sim 95\%$.

2.3. Roughness Reconstruction Over Multiple Length Scales

The definition of chromosome allows the reconstruction of profiles $Z(x)$ over multiple approximate realizations of the same surface. This can be done by limiting the summation in Equation (3) to a subset of chromosomes $C_n(x)$.

For both the stiffness-load curve $K(p)$ and the area-load curve $A(p)$, only one subset of the chromosomes can be considered to represent the main features of the contact mechanics response. This is shown in **Figure 3** for a representative example coming from a fractured alloy interface (black line), considering both the $K(p)$ and $A(p)$ evolutions. The red dashed line is the contact mechanics response of the profile obtained by summing up chromosomes that have a $K(p)$ evolution with a correlation coefficient larger than 0.95 with the complete $K(p)$ one. Both the $K(p)$ and $A(p)$ evolutions of this profile overlap significantly with the ones corresponding to the complete profile, which is obtained when all the chromosomes are taken into account in the summation. This means that the interaction of chromosomes composing only one part of the complete genome is sufficient to approximate the stiffness-load curve $K(p)$ well, as previously deduced in Paggi and Barber (2011).

The selection of chromosomes outlined above is detailed in Algorithm 1. From the genome of a surface, one obtains the $K(p)$ evolution of the profile described by Equation (1) (Step



1). The mechanical curves $K_n(p)$ of all the n_c chromosomes composing the considered rough profile are iteratively extracted using the BEM algorithm for the frictionless elastic normal contact problem, which is solved via the Non-Negative Least Squares method proposed in Bemporad and Paggi (2015) (Steps 2–7). The correlation coefficients c_n between the extracted curves and the one $K(p)$ of the complete rough profile are calculated (see Step 6). Only chromosomes with a correlation coefficient higher than 0.95 are then retained in the final set U_c of chromosomes (see Step 8).

3. GENETIC ALGORITHMS TO IDENTIFY OPTIMAL PROFILES TO MATCH TARGET CONTACT MECHANICS RESPONSES

The goal of this section is to illustrate methods aimed to design a prototype profile able to achieve a target contact mechanics response $y_t(\xi)$ of a rough profile, where t stands for “target”. This contact mechanics response depends on the specific needs of the frictionless elastic normal contact problem and is discretized using n_t points, where n_t is the number of far-field displacements Δ imposed to the frictionless elastic normal contact problem (see Johnson, 2003; Bemporad and Paggi, 2015). This target evolution $y_t(\xi)$ can be, e.g., either the stiffness-load curve $K(p)$ or the contact area-displacement curve $A(\Delta)$. The unknown rough profile $Z_t(x)$ to be identified is discretized using $N = 512$ nodes for a length L . Then, the values of n_s and n_f are defined according to Equation (4). Here, without loss of generality and for simplicity, it is imposed $n_s = 1$ for each realization. The frictionless elastic normal contact problem is solved in $n_t = 20$ equi-spaced rigid body displacements from the topmost summit of the considered profile to its deepest valley.

In the following, the variable ξ is taken as the contact pressure p . Moreover, the gene \mathcal{A}_i of a generic genome (indexed by i in

Algorithm 1 | Chromosome selection

Input: genome of a profile, n_s , n_f , number n_t of far-field displacements to be imposed in BEM

Output: U_c : set of chromosomes determining the $K(p)$ evolution of the profile

- 1: $y_r \leftarrow$ BEM results ($K(p)$) for the reference profile given by the genome, n_s , and n_f
- 2: **for all** $n = n_s : n_f$ **do**
- 3: **for all** $j = 1 : n_t$ **do**
- 4: $y_n^j \leftarrow$ BEM results ($K_n^j(p)$)
- 5: **end for**
- 6: $c_n \leftarrow$ corr. coeff. (y_r, y_n)
- 7: **end for**
- 8: $U_c \leftarrow C_n(x)$ with $c_n > 0.95$

the database) is rescaled to match a given pressure requirement, $p \leq p_{\max}^t$, as follows:

$$\mathcal{A}_i \leftarrow \frac{p_{\max}^t}{p_{\max}^i} \mathcal{A}_i. \quad (11)$$

The maximum pressure p_{\max}^i is computed imposing a far-field displacement equal to the original profile amplitude. In such a way, the new profile shows a maximum pressure p_{\max}^t if a far-field displacement equal to the peak-valley amplitude is considered to solve the frictionless elastic normal contact problem.

Three different methods to design the prototype profile are proposed. All these methods start from a known database of genomes. In the first case, a genome is selected that has the closest contact mechanics response to y_t among all the genomes in the database. Its genes are then optimized to increase the similarity with the target contact mechanics response. In the second case,

two genomes are selected that match the target response in different imposed ranges Δp . Then, these genomes are combined using an optimized *cross-over* mechanism. The third and last case is similar to the second one but, instead of the complete genomes, two sets of chromosomes are mixed. These two sets are contained in two genomes that match the target response in different imposed ranges Δp .

3.1. Simple Optimization of Genes

In this section, the Simple Optimization of Genes (SOG) genetic algorithm is presented. Genes to be optimized belong to a genome producing a rough profile with a contact mechanics response very similar to the target one. The following similarity score

$$s_i = s(y_t, y_i) = 1 - \left\| \frac{y_t(\xi) - y_i(\xi)}{y_t(\xi)} \right\|_{\infty} \quad (12)$$

is defined to quantify how much the target contact mechanics response, $y_t(\xi)$, and the one $y_i(\xi)$ associated with the parent i -th rough profile, are similar. Here, $\|\cdot\|_{\infty}$ denotes the l_{∞} -norm, computed on the n_t points used to discretize the contact mechanics response. Due to Equation (12), the i -th curve coincides with the target one when $s_i = 1$ holds. Otherwise, it is close to the target one when $s_i \simeq 1$ holds.

The numerical steps performed to select the best profile from the set of N_D genomes in the given database are described in Algorithm 2. The similarity score in Equation (12) is computed for all the N_D rough profiles available in the database, rescaled according to Equation (11) (Step 2). The contact mechanics response $y_i(\xi)$ is computed via BEM, considering n_t different far-field displacements, from 0 to the new profile amplitude (Steps 3–5). Finally, $s_i = s(y_t, y_i)$ in Equation (12) is computed (Step 6).

Algorithm 2 | Similarity score extraction from a database of genomes

Input: target contact mechanics response y_t , database of N_D genomes (genes, p_{max}^i, n_t)

Output: $s_i(y_t, y_i)$ with $i \in \{1, \dots, N_D\}$

```

1: for all  $i = 1 : N_D$  do
2:    $\mathcal{A}_i \leftarrow$  Eq. (11)
3:   for all  $j = 1 : n_t$  do
4:      $y_i^j \leftarrow$  BEM results at  $\Delta_i^j$ 
5:   end for
6:    $s_i \leftarrow$  Eq. (12)
7: end for
```

Genes belonging to the three genomes with the three largest associated values of $s(y_t, y_i)$ are now optimized using the Globally Convergent Method of Moving Asymptotes (GCMMA) (Svanberg, 2002). This is an iterative optimization algorithm, which is often used in optimal design for mechanical problems (see Bacigalupo et al., 2016, 2017 for some of its recent applications to band gap optimization). The GCMMA algorithm replaces a nonlinearly constrained optimization problem by a sequence of approximating nonlinearly constrained optimization

subproblems, which are simpler to solve. In the work, the objective function has been chosen to be the square of the similarity score, i.e., $s^2(y_t, y_i)$, in order to increase its smoothness. The GCMMA iterative solution is obtained after a number n_{it} of steps, starting from an initial choice for the vector of optimization variables. In the case of the SOG, this initial choice is provided by Algorithm 2, which provides an initial value of the similarity score close to 1. In such a way, there is no risk to get a negative similarity score by locally maximizing its square.

To save CPU time, the number of variables in the optimization problem is limited as follows. The genes \mathcal{H} , λ , and γ determine the frequency spectrum, i.e., the interaction among different length scales. Therefore, they are not varied during the optimization, but fixed to their original values. The gene \mathcal{A} is varied at each optimization step to satisfy the pressure requirement, according to Equation (11). So, only the phases $\phi_{m,n}$ are considered as optimization variables. In the optimization, their values are constrained in the range between $\pm 10\%$ of their initial values, to preserve the main features of the original chromosomes. Moreover, only the phase genes $\phi_{m,n}$ of chromosomes that determine the main features of the $K(p)$ response are considered. Such genes are selected according to Algorithm 1.

Finally, the SOG algorithm is summarized in Algorithm 3. Starting with a profile scouting from the available database (Step 1), the three profiles whose contact mechanics responses are most similar to the target y_t are chosen (Step 2). The genes of each such genome are then optimized using the GCMMA algorithm (Steps 3–6), limiting the optimization variables only to the chromosomes determining the main features of the target y_t , as determined by Algorithm 2. The resulting optimized genomes are denoted by \hat{U}_1^i . Finally, among such genomes, the new genome is chosen as the one with the best (square of the) similarity score with respect to the target response (Step 7). In this last step, $\text{argmax}(f_i)_{i=1, \dots, n_1}$ denotes the index i associated with the largest f_i .

Algorithm 3 | Simple Optimization of Genes (SOG)

Input: target contact mechanics response y_t , genome database

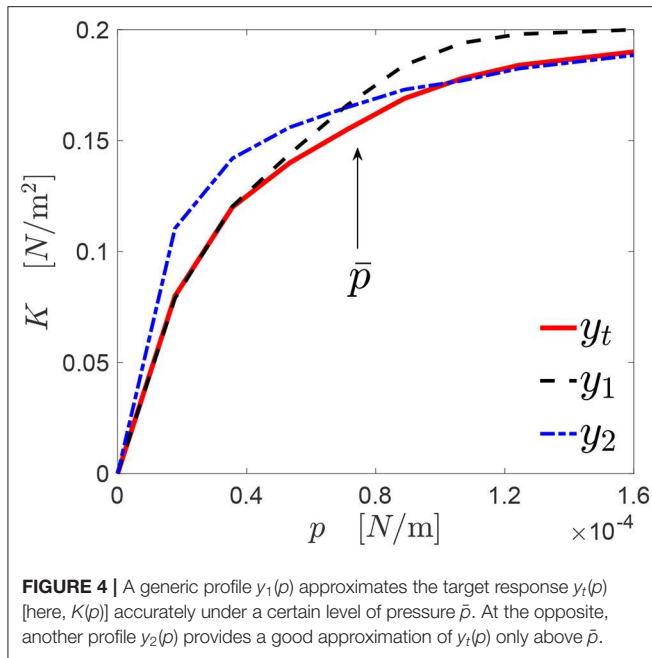
Output: new genome U_{SOG} with contact mechanics response close to y_t

```

1:  $s$  from Alg. 2
2:  $U_1 \leftarrow$  the three genomes with the three largest similarity scores (Eq. 12)
3: for all  $i = 1 : n_1$  do ( $n_1 = \text{card}(U_1)$ )
4:    $U_c^i \leftarrow$  Alg. 1 applied to  $U_1^i$ 
5:    $f_i \leftarrow s(y_t, y_i)$ ,  $\hat{U}_1^i$ , both from GCMMA( $U_c^i$ )
6: end for
7:  $U_{SOG} \leftarrow \hat{U}_1^{\text{argmax}(f_i)_{i=1, \dots, n_1}}$ 
```

3.2. Genome Cross-Over

In this section, the Genome Cross-Over (GCO) genetic algorithm is presented. In this method, different pairs of genomes are mixed, to obtain a new genome matching the target response $y_t(\xi)$. The best pair of genomes is chosen in relation to



their similarity scores in two specific ranges of the target response $y_t(\xi)$.

This concept is explained through **Figure 4**, where a target response $y_t(p)$ is shown by the red line. For the sake of explanation, two curves $y_1(p)$ (dashed black line) and $y_2(p)$ (dashed dot blue line) have been manually generated, which have a similarity score with respect to $y_t(p)$ equal to $s_1 \simeq 0.89$ and $s_2 \simeq 0.88$, respectively.

However, the curves $y_1(p)$ and $y_2(p)$ describe quite accurately the curve $y_t(p)$ in different ranges of pressures, if a suitable threshold pressure \bar{p} is defined. The value of \bar{p} can be chosen arbitrarily, or might be imposed by the particular problem formulation. In the specific case shown in the figure, in the interval $[0, \bar{p}]$, the curve y_1 represents with good accuracy y_t [$s_1(y_t, y_1) \simeq 0.97$]. The same happens in the interval $[\bar{p}, p_{\max}^t]$ for the curve y_2 [$s_2(y_t, y_1) \simeq 0.99$].

Then, it is reasonable to expect that a new profile obtained by combining the two genomes associated with y_1 and y_2 , respectively, should provide a contact mechanics response closer to y_t over the whole range of pressures. However, mixing these two genomes may also lead to a very different roughness organization. For this reason, the GCO iterative scheme checks if the new genome, obtained by mixing the genome of y_1 and the one of y_2 , is able to represent accurately y_t over the whole range of pressures, before using the GCMMA algorithm to increase the value of the similarity score.

The GCO structure is presented in Algorithm 4. According to the value of \bar{p} , two different sets U_1 and U_2 are identified from the database of genomes (Steps 1–2). The first set U_1 contains the genomes with a value of the similarity score (Equation 12) larger than 0.95 in the interval $[0, \bar{p}]$ (Step 3). The second set U_2 contains the genomes with a value of the similarity score

(Equation 12) larger than 0.95 in the interval $[\bar{p}, p_{\max}^t]$ (Step 4). All possible combinations of genomes from the two sets above are now considered, defining the set U_3 (Step 5–11). A new genome corresponds to each of these combinations, whose gene \mathcal{A} is rescaled according to Equation (11) (Step 8). Then, the value of the similarity score (Equation 12) is computed with respect to the target response (Step 9). The three new genomes showing the three largest values of the similarity score are used as inputs to the GCMMA algorithm, defining the set U_4 (Step 12). Also in this case, the number of genes to be optimized is limited, considering only the phases $\phi_{m,n}$ associated with the chromosomes that determine the main features of the $K(p)$ evolution of the parent curve (see Algorithm 1). The new genome U_{GCO} with the maximum value of the similarity is finally identified (Step 17).

3.3. Chromosomes Cross-Over

In this section, the Chromosomes Cross-Over (CCO) genetic algorithm is presented. We recall that, as observed in section 2, the main features of the contact mechanics response of a rough profile are determined by specific chromosomes of the genomes. This subdivision allows the introduction of a chromosomes selection step to reduce the number of variables to be optimized using the GCMMA algorithm, as presented before for the SOG (section 3.1) and the GCO (section 3.2).

In the CCO method, only chromosomes determining the main features of the contact mechanics responses of two different

Algorithm 4 | Genome Cross-Over (GCO)

Input: target contact mechanics response y_t , genome database, threshold pressure \bar{p}

Output: new genome U_{GCO} with contact mechanics response close to y_t

- 1: $s^{(1)}$ from Alg. 2, with similarity score computed in the interval $[0, \bar{p}]$
- 2: $s^{(2)}$ from Alg. 2, with similarity score computed in the interval $[\bar{p}, p_{\max}^t]$
- 3: $U_1 \leftarrow$ genomes with $s_i^{(1)} > 0.95$
- 4: $U_2 \leftarrow$ genomes with $s_i^{(2)} > 0.95$
- 5: **for all** $i_1 = 1 : n_1$ **do** ($n_1 = \text{card}(U_1)$)
- 6: **for all** $i_2 = 1 : n_2$ **do** ($n_2 = \text{card}(U_2)$)
- 7: $U_3^{(i_1, i_2)} \leftarrow U_1^{i_1} + U_2^{i_2}$
- 8: $U_3^{(i_1, i_2)}$ rescaled according to Eq. (11)
- 9: $s(i_1, i_2) \leftarrow s(y_t, y_{(i_1, i_2)})$ from Eq. (12) applied to $U_3^{(i_1, i_2)}$
- 10: **end for**
- 11: **end for**
- 12: $U_4 \leftarrow$ the three genomes in U_3 with the three largest similarity scores (obtained from s)
- 13: **for all** $i = 1 : n_4$ **do** ($n_4 = \text{card}(U_4)$)
- 14: $U_c^i \leftarrow$ Alg. 1 applied to U_4^i
- 15: $f_i \leftarrow s(y_t, y_i)$, \hat{U}_4^i , both from GCMMA(U_c^i)
- 16: **end for**
- 17: $U_{GCO} \leftarrow \hat{U}_4^{\arg\max(f_i)_{i=1, \dots, n_4}}$

genomes are mixed to match the target response y_t . These two sets of chromosomes come from genomes that have the largest values of the similarity score in specific ranges of the target response y_t , as done for the GCO method (see section 3.2). Then, in this case the GCMMA is applied to the complete set of chromosomes composing this new genomes, since they all determine its mechanical evolution.

The CCO iterative scheme is summarized in Algorithm 5. According to the value of the threshold pressure \bar{p} , two different sets U_1 and U_2 of reduced genomes are identified, starting from the given database (Steps 1–2). The first set U_1 is obtained as follows. First, one generates a subset of genomes that show a value of the similarity score (Equation 12) larger than 0.95 in the interval $[0, \bar{p}]$ (Step 3). Then, these genomes are reduced, limiting to those chromosomes that affect the $K(p)$ evolution significantly. Such chromosomes are selected according to Algorithm 1. The resulting reduced genomes form the set U_1 . The second set U_2 is obtained in a similar way as U_1 , but computing the similarity score in the interval $[\bar{p}, p_{\max}^t]$ (Step 4). All possible combinations of these reduced genomes from the two sets above are now considered, defining the set U_3 (Steps 5–11). A new genome corresponds to each of these combinations. Its amplitude gene \mathcal{A} is rescaled according to Equation (11), to match the pressure requirement (Step 8). Then, the value of the similarity score (Equation 12) is computed with respect to the target response (Step 9). The three new genomes showing the three largest values of the similarity score are used as inputs to the GCMMA algorithm, defining the set U_4 (Step 12). Only in the case of the CCO algorithm, the GCMMA algorithm is applied to all the genes of this new genome, as the size of the optimization problem has been already reduced in Steps 3–4. The new genome U_{CCO} with the maximum obtained value of the similarity score is finally identified (Step 16).

4. OPTIMAL GENOME TO MATCH A SPECIFIED CONTACT MECHANICS RESPONSE

In this section, the genetic algorithms described in section 3 are compared in their application to a representative example. Before doing that, the characteristics of the numerical genome database considered in the paper are reported. Then, the set-up of the numerical experiment proposed is introduced. Finally, the related results are discussed, focusing the attention on the characterization of the so-obtained new genomes.

4.1. Database of Genomes

A database of genomes is needed to apply all the genetic algorithms proposed in the paper. In the following, a small database is generated numerically, to show a representative example of this approach. To generate the database, the amplitude parameter is fixed to $\mathcal{A} = 1$, and the main wavelength is put equal to $\lambda = 849.42 \mu\text{m}$.

The number of ridges is set to $\mathcal{M} = 1$, to save computational time. This particular case corresponds to chromosomes described according to Equation (7) with $\theta_{n,1} = \theta_{n,2} = \theta_n$. As discussed

Algorithm 5 | Chromosomes Cross-Over (CCO)

Input: target contact mechanics response y_t , genome database, threshold pressure \bar{p}

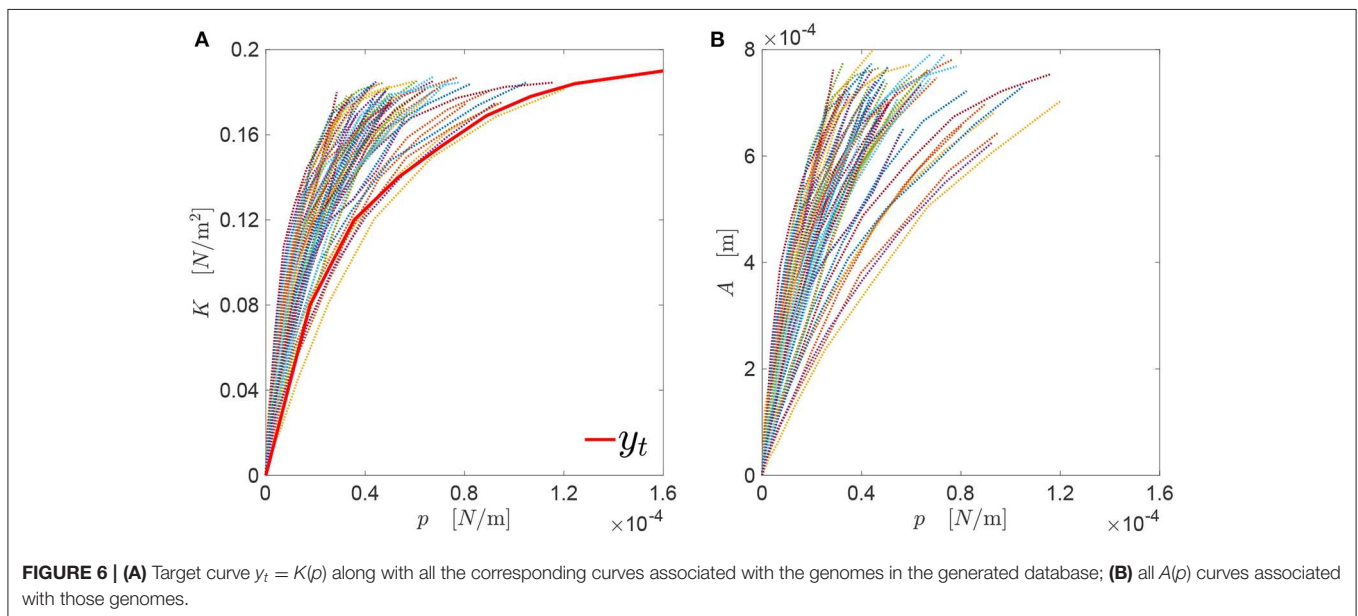
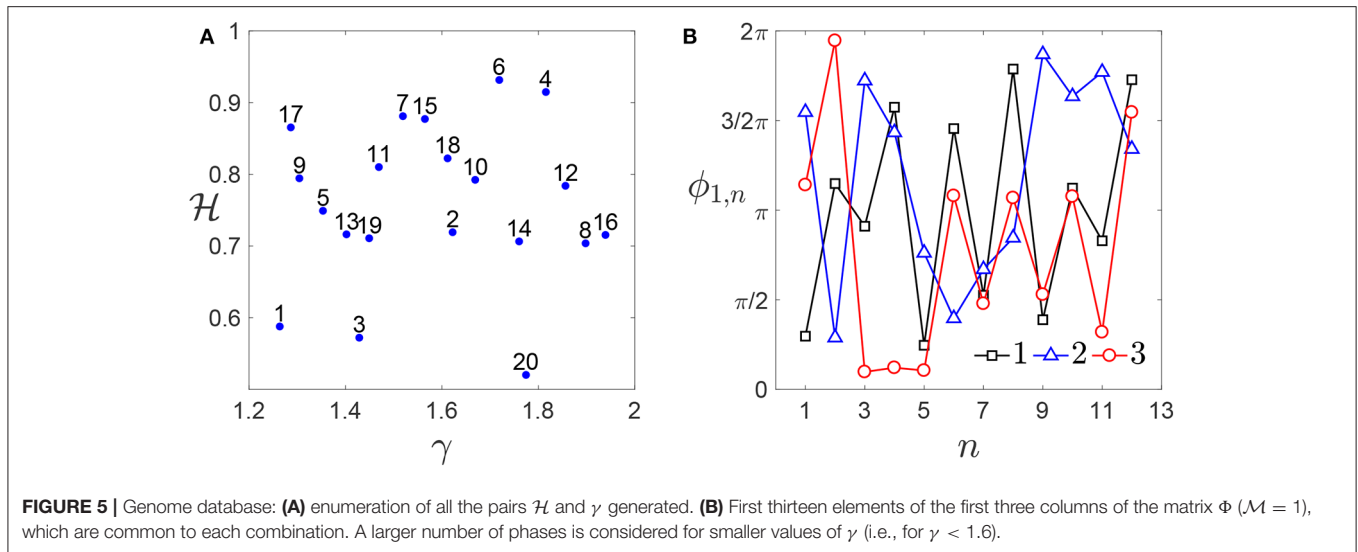
Output: new genome U_{CCO} with contact mechanics response close to y_t

- 1: $s^{(1)}$ from Alg. 2, with similarity score computed in the interval $[0, \bar{p}]$
- 2: $s^{(2)}$ from Alg. 2, with similarity score computed in the interval $[\bar{p}, p_{\max}^t]$
- 3: $U_1 \leftarrow C_n(x)$ from Alg. 1, for those genomes with $s_i^{(1)} > 0.95$
- 4: $U_2 \leftarrow C_n(x)$ from Alg. 1, for those genomes with $s_i^{(2)} > 0.95$
- 5: **for all** $i_1 = 1 : n_1$ **do** ($n_1 = \text{card}(U_1)$)
- 6: **for all** $i_2 = 1 : n_2$ **do** ($n_2 = \text{card}(U_2)$)
- 7: $U_3^{(i_1, i_2)} \leftarrow U_1^{i_1} + U_2^{i_2}$
- 8: $U_3^{(i_1, i_2)}$ rescaled according to Eq. (11)
- 9: $s(i_1, i_2) \leftarrow s(y_t, y_{(i_1, i_2)})$ from Eq. (12) applied to $U_3^{(i_1, i_2)}$
- 10: **end for**
- 11: **end for**
- 12: $U_4 \leftarrow$ the three genomes in U_3 with the three largest similarity scores (obtained from s)
- 13: **for all** $i = 1 : n_4$ **do** ($n_4 = \text{card}(U_4)$)
- 14: $f_i \leftarrow s(y_t, y_i)$, \hat{U}_4^i , both from GCMMA(U_4^i)
- 15: **end for**
- 16: $U_{CCO} \leftarrow \hat{U}_4^{\text{argmax}(f_i)_{i=1, \dots, n_4}}$

in section 2, this can be considered as representative of several situations for a rough profile. Each profile $Z_i(x)$ is then discretized using $N = 512$ nodes. The elastic modulus is imposed to be equal to $E = 1 \text{ MPa}$. Furthermore, $n_h = 20$ different pairs of values for \mathcal{H} and γ are considered. They are generated according to a Sobol' sequence (Niederreiter, 1992) (see Figure 5A), with \mathcal{H} ranging between 0.5 and 1.5, and γ between 1.4 and 2. The value of γ is chosen in such a way to keep small the number of frequencies composing the profile spectrum (see Equation 4), thus limiting the computational costs. Also the phase matrix Φ is generated according to a Sobol' sequence. This matrix is made of $n_\phi = 5$ columns, corresponding to 5 choices for the set of phases. The number of rows is equal to the maximum possible number of phases for these choices of the parameters. According to Equation (4), such number is obtained in correspondence of the smallest value of γ . Then, $N_D = n_h n_\phi$ rough profiles $Z_i(x)$ are generated, combining each pair (\mathcal{H}, γ) to each column of Φ , considering only the number of values needed for the phases.

4.2. Set-Up of the Numerical Experiment

In Figure 6, the contact mechanics response $y_t = K(p)$ is depicted through the black line. All the values are considered independent from E , as this is equal to unity. Also the $K(p)$ responses of all the genomes in the generated database are visualized in Figure 6. The target contact mechanics response y_t has been generated manually, and has a trend similar to the ones belonging to the database. The maximum pressure



reachable is imposed to be $p_{\max}^t = 1.6 \times 10^{-4}$ N/m. This value is larger than the ones of all the genomes in the database considering, for each profile, a far-field displacement equal to the profile amplitude.

For each algorithm, the stage just preceding the application of the GCMMA algorithm is referred to as “GCMMA (0).” For the SOG, it corresponds to the selection of the three best values obtained from the scouting of the database (Step 2 in Algorithm 3). For both the GCO and the CCO algorithms, it corresponds to the computation of the similarity scores of the new genomes determined after the cross-over (Step 12 in both Algorithms 4 and 5). For the SOG, the GCMMA algorithm is applied with $n_{it} = 5$, while for the GCO and CCO, only with $n_{it} = 3$. The output of this optimization step is denoted

by “GCMMA (1).” These different choices for n_{it} are made in order to have comparable computational times of about 1 minute. In such a way, it is possible to observe variations in the results in a sufficiently small time, enabling in-line control as a possible successive step. Moreover, in the numerical experiment, for the SOG and GCO algorithms, the GCMMA is also applied in a second optimization step, whose output is denoted as “GCMMA (2),” using $n_{it} = 2$ iterations. In this case, the genes to be optimized are the ones referred to the chromosomes excluded in the selection step (see Algorithm 1). For the CCO, this second step cannot be applied since these chromosomes are excluded in the first part of the algorithm (Steps 1–4 in Algorithm 5). However, to make the notation uniform for the three algorithms, a fictitious “GCMMA (2)”

output is introduced for the CCO, by duplicating the results of “GCMMA (1).”

4.3. Numerical Results

The values of the similarity score s given by Equation (12) for the genomes obtained by combination and optimization using the SOG, GCO, and CCO algorithms are reported in **Figure 7**. The threshold pressure used for both the GCO and CCO is set equal to $\bar{p} = 0.08$ [1/mm]. For each algorithm, solutions at different stages are reported in the figure. The first ones correspond to the three best solutions obtained in the first stage of each algorithm, before the application of the GCMMA (such solutions are represented in the figure by the triangle, circle, and cross, in increasing order). The figure shows that all the algorithms are quite efficient in matching the target contact mechanics response, achieving large values for the similarity score. The application of the GCMMA is beneficial for all the algorithms. However, the

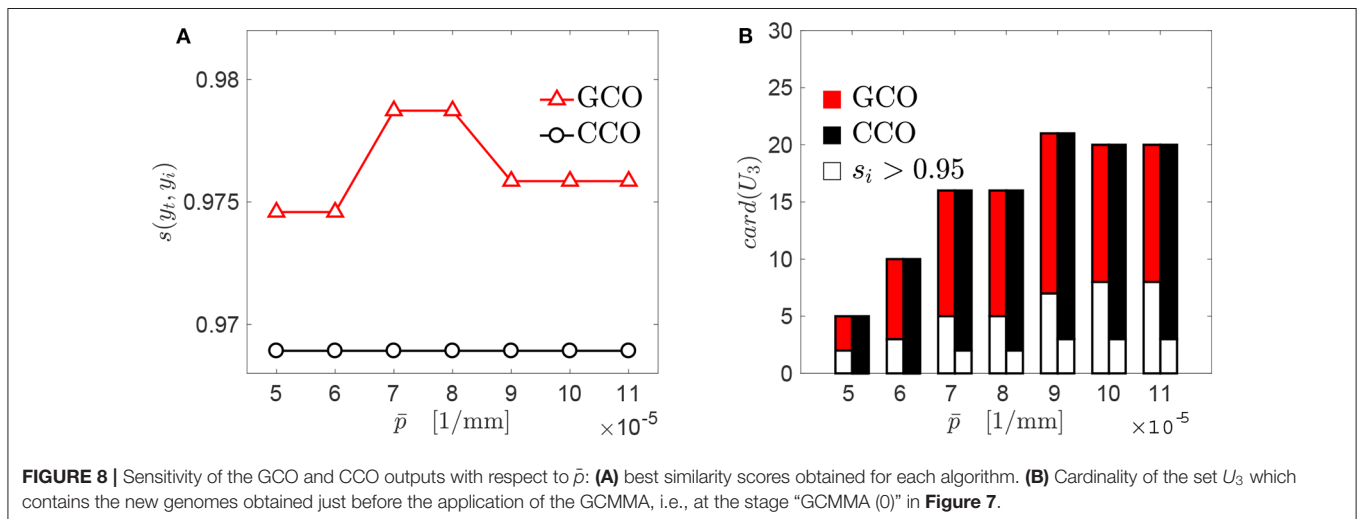
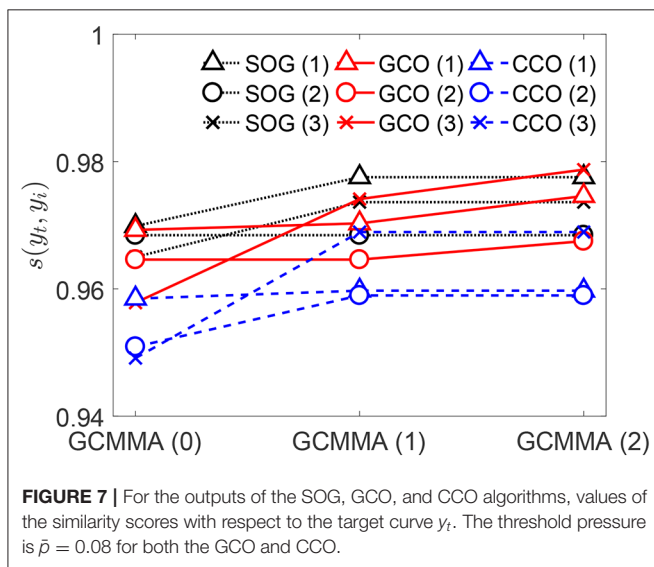
GCO and CCO algorithms might provide even better solutions by varying the threshold pressure \bar{p} .

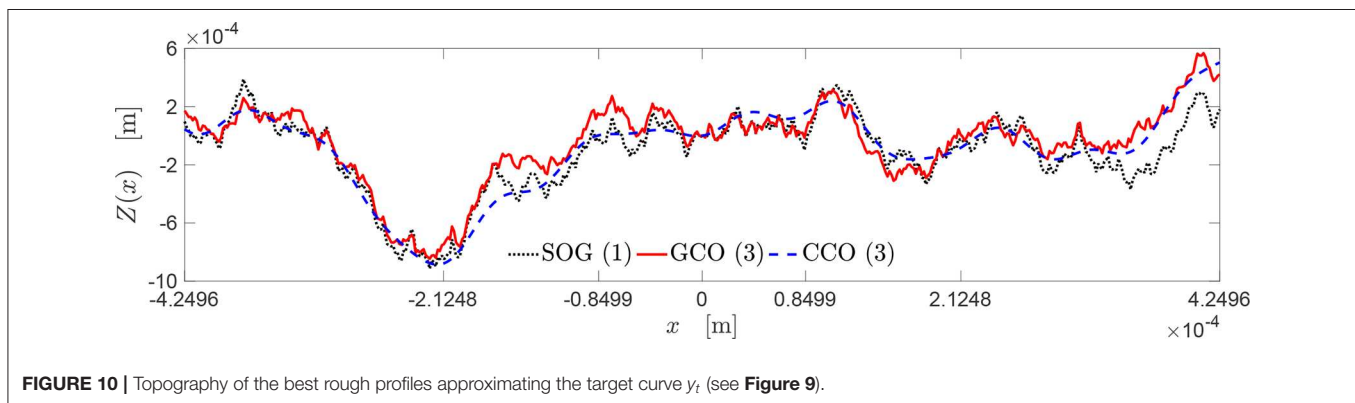
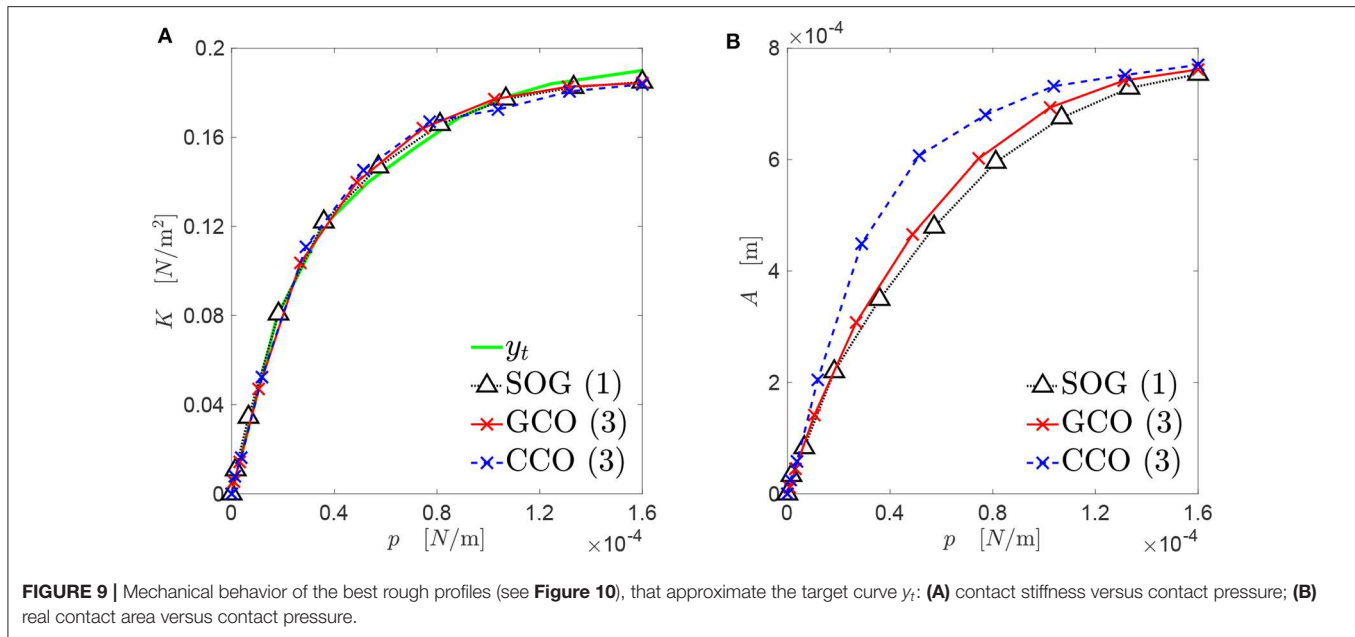
4.4. Effect of the Threshold Pressure \bar{p}

A threshold pressure \bar{p} has been introduced for the GCO and the CCO, to select individual genomes able to approximate locally the target curve with a good accuracy. Additional simulations have been made, to assess the sensitivity of the results with respect to such a parameter. Both the GCO and CCO algorithms have been applied with different values of \bar{p} . Such values have been chosen between 0.5×10^{-4} N/m and 1.1×10^{-4} N/m. A significant variation on the maximum value of the similarity score is observed for the GCO algorithm (see **Figure 8A**). On the contrary, the CCO algorithm seems to be not affected by the value of \bar{p} , and the new genomes obtained by that algorithm are composed by the same starting genomes, independently of the threshold pressure \bar{p} . This may be due to the fact that our investigation has been conducted starting from a small database of genomes. In **Figure 8B**, the cardinality of the set U_3 is presented, for both the GCO and CCO algorithms. This set contains the new genomes obtained after the cross-over of genomes/chromosomes. The white part indicates, for both algorithms, the number of genomes whose value of the similarity score with respect to the target curve y_t is larger than 0.95. The cardinality of the set U_3 varies significantly with \bar{p} , and is the same for both algorithms. Nevertheless, the GCO algorithm is more efficient in terms of the obtained similarity with the target curve.

4.5. Description of the Optimized Genomes Representing y_t

The best three new genomes obtained at the end of the three algorithms presented in this paper have contact mechanics responses overlapping significantly with the target curve (see **Figure 9A**). However, the three genomes present different evolutions of the bearing area, as depicted in **Figure 9B**. Only the $A(p)$ curves obtained by the SOG and GCO are similar. This may be due to the fact that, in the case of the CCO, high-frequency





features of the rough profile might have been neglected. This is clear from **Figures 10, 11**.

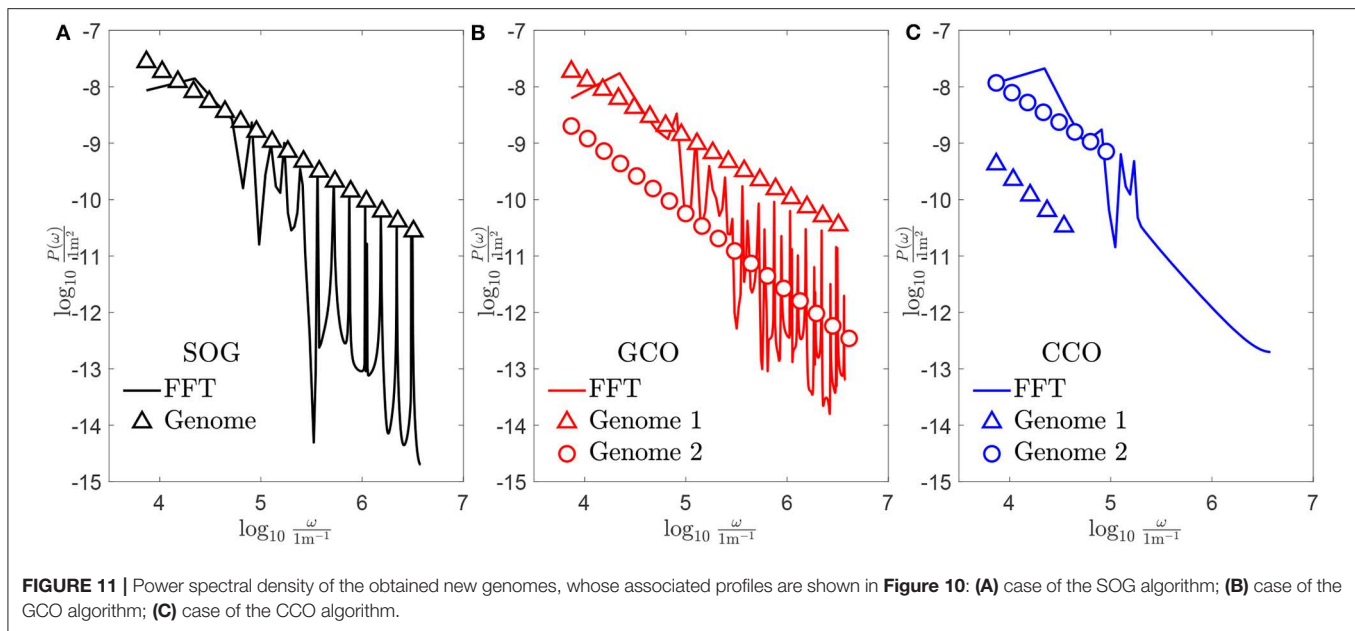
The topography of these rough profiles is presented in **Figure 10**. The profiles obtained by the SOG, GCO, and CCO algorithms are depicted, respectively, through a black dash-dot line, a red dashed line, and a blue continuous line. All these profiles have very similar geometrical features, regarding the locations of peaks and valleys. Moreover, it is interesting to notice that the profile provided by the CCO algorithm is a good approximation of the profile given by the GCO algorithm, which presents more high-frequency features.

To conclude, the discrete Power Spectral Density (PSD) $P(\omega)$ of the new obtained genomes is shown in **Figure 11**, and is represented by markers in both sub-figures. For each case, the continuous PSD function, which is obtained through the Fast Fourier Transform (FFT) filtering (Berry and Lewis, 1980), is shown by a continuous line. For the SOG (see **Figure 11A**), the peaks of the continuous PSD function match the discrete one accurately for high frequencies. No good matching is found for low frequencies. The same remark holds for the case of the GCO

(see **Figure 11B**). Here, the spectrum is more dense and, only for high frequencies, peaks of the continuous PSD function are located in the same positions of the genome. For both the SOG and the GCO, a consistent difference in the amplitude of $P(\omega)$ is found. Finally, the spectrum of the profile obtained by the CCO is composed by a small set of frequencies. The high-frequency part of the PSD is flat. In this case, the discrete power spectral density is nearly proportional to the one of the SOG in the low-frequency range. Since the GCO approximation of the target curve is obtained using a smaller set of chromosomes, it can be easily controlled for the case of in-line control of the rough profile morphology. However, in this case, the PSD is almost flat, and some peaks are found in frequencies where there are no genome components.

5. CONCLUSION

The main goal of this work was to provide a mathematical and computational methodology to identify new surface topologies



with given target contact mechanics responses, based on a multi-scale characterization of roughness. This was described by a superposition of a finite set of successive length scales of roughness, where each scale is associated with a particular form of the MWM function, named chromosome. Based on this representation, three genetic algorithms were proposed to combine profiles and identify the best one which meets the given target contact mechanics response.

The first genetic algorithm (Simple Optimization of Genes, SOG) optimizes the genes of a known genome. Embedded in the SOG, an iterative optimization algorithm (the GCMMA) was used to increase the similarity with the target contact mechanics response. To save computational time, optimization was performed only with respect to the dominant chromosomes of the genome, i.e., the ones for which the correlation coefficient between the associated contact mechanics response and the target one is above a given threshold.

The second genetic algorithm (Genome Cross-Over, GCO) crosses-over two different genomes that show a high similarity with respect to the target response in two given intervals determined by a threshold pressure \bar{p} . Also in this case, only dominant chromosomes were optimized to save computation time.

The third genetic algorithm (Chromosomes Cross-Over, CCO) consists in a cross-over of chromosomes that provide the main features of the contact mechanics responses of two genomes that well approximate the target curve in two reference intervals determined by the threshold pressure \bar{p} . In this case, GCMMA was applied to the complete new genome, since the number of optimization variables is smaller than for the other two approaches.

The three genetic algorithms proposed in this work generated profiles that almost fully reproduce the given $K(p)$ curve, when this is chosen as the target contact mechanics response. Also, they

have in common similar features regarding their topography, such as the locations of peaks and valleys. This characterization might be caused by the fact that a limited number of genomes was present in the database used for the numerical experiment reported in the paper. A much wider set of profile topologies is expected to be obtained by exploiting a significantly larger database, facilitating the discovery of optimal patterns/textures, based on the specific needs. Also, the investigation can be extended to all the genomes that provide a good approximation of the target curve, in order to identify some geometrical features that drive the elastic response of a rough profile. Once identified, these features can be taken into account in the optimization by adding specific chromosomes, to be controlled in the case of in-line profile morphing.

Finally, the authors would like to remark that the proposed approaches could be extended to the two-dimensional case, although this is expected to require a larger computational effort. Moreover, the use of micromechanical contact theories such as those compared in Zavarise et al. (2007) and Paggi and Ciavarella (2010) could be another interesting research direction. For instance, one could assume not to have the surface height field at all, but only a set of relevant statistical parameter inputs of micromechanical contact theories. In such a context, the search for the optimal topology to match a prescribed contact response would reduce to the identification of the statistical parameters based on the combination of the initial dataset and of the proposed genetic algorithms.

DATA AVAILABILITY STATEMENT

The dataset generated for this study is available on request to paolo.cinat@alumni.imtlucca.it.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Ausloos, M., and Berman, D. (1985). A multivariate Weierstrass-Mandelbrot function. *Proc. R. Soc. Lond. A* 400, 331–350. doi: 10.1098/rspa.1985.0083
- Bacigalupo, A., Gnecco, G., Lepidi, M., and Giambarotta, L. (2017). Optimal design of low-frequency band gaps in anti-tetrachiral lattice meta-materials. *Compos. Part B Eng.* 12, 341–359. doi: 10.1016/j.compositesb.2016.09.062
- Bacigalupo, A., Lepidi, M., Gnecco, G., and Giambarotta, L. (2016). Optimal design of auxetic hexachiral metamaterials with local resonators. *Smart Mater. Struct.* 25:054009. doi: 10.1088/0964-1726/25/5/054009
- Barber, J. (2003). Bounds on the electrical resistance between contacting elastic rough bodies. *Proc. R. Soc. Lond. A* 459, 53–66. doi: 10.1098/rspa.2002.1038
- Bemporad, A., and Paggi, M. (2015). Optimization algorithms for the solution of the frictionless normal contact between rough surfaces. *Int. J. Solids Struct.* 69–70, 94–105. doi: 10.1016/j.ijsolstr.2015.06.005
- Berry, M., and Lewis, Z. (1980). On the Weierstrass-Mandelbrot fractal function. *Proc. R. Soc. Lond.* 370, 459–484. doi: 10.1098/rspa.1980.0044
- Bora, C., Flater, E., Street, M., Redmond, J., Starr, M., Carpick, R., et al. (2005). Multiscale roughness and modeling of MEMS interfaces. *Tribol. Lett.* 19, 37–48. doi: 10.1007/s11249-005-4263-8
- Borri, C., and Paggi, M. (2015). Topological characterization of anti-reflective and hydrophobic rough surfaces: are random process theory and fractal modeling applicable? *J. Phys. D Appl. Phys.* 48:045301. doi: 10.1088/0022-3727/48/4/045301
- Brettel, M., Friederichsen, F., Keller, M., and Rosenberg, M. (2014). How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 perspective. *Int. J. Mech. Aerospace Indus. Mechatron. Eng.* 8, 37–44. Available online at: <https://publications.waset.org/9997144/pdf>
- Buscaglia, G., Ciuperca, I., and Jai, M. (2007). On the optimization of surface textures for lubricated contacts. *J. Math. Anal. Appl.* 335, 1309–1327. doi: 10.1016/j.jmaa.2007.02.051
- Carr, J., and Benzer, W. (1991). On the practice of estimating fractal dimension. *Math. Geol.* 23, 945–958. doi: 10.1007/BF02066734
- Ciavarella, M., Demelio, G., Barber, J., and Jang, Y. (2000). Linear elastic contact of the Weierstrass profile. *Proc. R. Soc. Lond. A* 456, 387–405. doi: 10.1098/rspa.2000.0522
- Ciavarella, M., Murolo, G., Demelio, G., and Barber, J. (2004). Elastic contact stiffness and contact resistance for the Weierstrass profile. *J. Mech. Phys. Solids* 52, 1247–1265. doi: 10.1016/j.jmps.2003.12.002
- Cinat, P., Paggi, M., and Gnecco, G. (2019). Identification of roughness with optimal contact response with respect to real contact area and normal stiffness. *Math. Probl. Eng.* 2019:7051512. doi: 10.1155/2019/7051512
- Farina, I., Fabbrocino, F., Colangelo, F., Feo, L., and Fraternali, F. (2016). Surface roughness effects on the reinforcement of cement mortars through 3D printed metallic fibers. *Compos. Part B Eng.* 99, 305–311. doi: 10.1016/j.compositesb.2016.05.055
- Gao, H., Wang, X., Yao, H., Gorb, S., and Arzt, E. (2005). Mechanics of hierarchical adhesion structures of geckos. *Mech. Mater.* 37, 275–285. doi: 10.1016/j.mechmat.2004.03.008
- Gujrati, A., Khanal, S. R., Pastewka, L., and Jacobs, T. D. B. (2018). Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales. *ACS Appl. Mater. Interfaces* 10, 29169–29178. doi: 10.1021/acsami.8b09899
- Guo, Q., Zhu, P., Li, G., Wen, J., Wang, T., Lu, D. D., et al. (2017). Study on the effects of interfacial interaction on the rheological and thermal performance of silica nanoparticles reinforced epoxy nanocomposites. *Compos. Part B Eng.* 116, 388–397. doi: 10.1016/j.compositesb.2016.10.081
- Guo, S.-J., Yang, Q.-S., He, X. Q., and Liew, K. M. (2014). Modeling of interface cracking in copper-graphite composites by MD and CFE method. *Compos. Part B Eng.* 58, 586–592. doi: 10.1016/j.compositesb.2013.10.042
- Johnson, K. (2003). *Contact Mechanics*. Cambridge, UK: Cambridge University Press.
- King, R., Stansfield, W., and Mulligan, P. (2006). *A Dictionary of Genetics, 7th Edn.* Oxford: Oxford University Press. doi: 10.1093/acref/9780195307610.001.0001
- Ko, K., Jin, S., Lee, S. E., Lee, I., and Hong, J.-W. (2019). Bio-inspired bimaterial composites patterned using three-dimensional printing. *Compos. Part B Eng.* 165, 594–603. doi: 10.1016/j.compositesb.2019.02.008
- Kwon, D.-J., Kim, J.-H., Kim, Y.-J., Kim, J.-J., Park, S.-M., Kwon, I.-J., et al. (2019). Comparison of interfacial adhesion of hybrid materials of aluminum/carbon fiber reinforced epoxy composites with different surface roughness. *Compos. Part B Eng.* 170, 11–18. doi: 10.1016/j.compositesb.2019.04.022
- Mandelbrot, B. (1977). *Fractals: Form, Chance, and Dimension*. San Francisco, CA: W. H. Freeman.
- Moghri, M., Madic, M., Omid, M., and Farahnakian, M. (2014). Surface roughness optimization of polyamide-6/nanoclay nanocomposites using artificial neural network: genetic algorithm approach. *Sci. World J.* 2014:485205. doi: 10.1155/2014/485205
- Moroni, F., Palazzetti, R., Zucchelli, A., and Pirondi, A. (2013). A numerical investigation on the interlaminar strength of nanomodified composite interfaces. *Compos. Part B Eng.* 55, 635–641. doi: 10.1016/j.compositesb.2013.07.004
- Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: SIAM. doi: 10.1137/1.9781611970081
- Nosonovsky, M., and Bushan, B. (2008). *Multiscale Dissipative Mechanisms and Hierarchical Surfaces: Friction, Superhydrophobicity, and Biomimetics*. Hoboken, NJ: Springer. doi: 10.1007/978-3-540-78425-8
- Nouari, M., Xiao, M., Shen, X., Ma, Y., Yang, F., Gao, N., et al. (2018). Prediction of surface roughness and optimization of cutting parameters of stainless steel turning based on RSM. *Math. Probl. Eng.* 2018:9051084. doi: 10.1155/2018/9051084
- Paggi, M., and Barber, J. (2011). Contact conductance of rough surfaces composed of modified RMD patches. *Int. J. Heat Mass Transf.* 54, 4664–4672. doi: 10.1016/j.ijheatmasstransfer.2011.06.011
- Paggi, M., and Ciavarella, M. (2010). The coefficient of proportionality k between real contact area and load, with new asperity models. *Int. J. Solids Struct.* 56, 1020–1029. doi: 10.1016/j.wear.2009.12.038
- Paggi, M., and Hills, D. A. (2020). *Modeling and Simulation of Tribological Problems in Technology, Vol. 593 of CISM Courses and Lectures*. Cham: Springer. doi: 10.1007/978-3-030-20377-1
- Sherge, M., and Gorb, S. (2001). *Biological Micro- and Nano- Tribology & Nature's Solutions*. Berlin: Springer.
- Svanberg, K. (2002). A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM J. Optim.* 12, 555–573. doi: 10.1137/S1052623499362822
- Vakis, A., Yastrebov, V., Scheibert, J., Nicola, L., Dini, D., Minfray, C., et al. (2018). Modeling and simulation in tribology across scales: an overview. *Tribol. Int.* 125, 169–199. doi: 10.1016/j.triboint.2018.02.005
- Wen, R., and Sinding-Larsen, R. (1997). Uncertainty in fractal dimension estimated from power spectra and variograms. *Math. Geol.* 29, 727–753. doi: 10.1007/BF02768900
- Wüst, P., Edelmann, A., and Hellmann, R. (2020). Areal surface roughness optimization of maraging steel parts produced by hybrid additive manufacturing. *Materials* 13, 1–17. doi: 10.3390/ma13020418
- Xiaohan, Z., Xu, Y., and Jackson, R. (2017). An analysis of generated fractal and measured rough surfaces in regards to their multi-scale structure and fractal dimension. *Tribol. Int.* 105, 94–101. doi: 10.1016/j.triboint.2016.09.036

- Zain, A. M., Haron, H., and Sharif, S. (2008). "An overview of GA technique for surface roughness optimization in milling process," in *2008 International Symposium on Information Technology*, Vol. 4, 1–6. doi: 10.1109/ITSIM.2008.4631925
- Zavarise, G., Borri-Brunetto, M., and Paggi, M. (2007). On the resolution dependence of micromechanical contact models. *Wear* 262, 42–54. doi: 10.1016/j.wear.2006.03.044
- Zhang, J., Chen, J., and Kirby, E. (2007). Surface roughness optimization in an end-milling operation using the Taguchi design method. *J. Mater. Process. Technol.* 184, 233–239. doi: 10.1016/j.jmatprotec.2006.11.029

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cinat, Gnecco and Paggi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Artificial Neural Network Architecture for Prediction of Contact Mechanical Response

Kalle Kalliorinne^{1*}, Roland Larsson¹, Francesc Pérez-Ràfols¹, Marcus Liwicki² and Andreas Almqvist¹

¹Division of Machine Elements, Luleå University of Technology, Luleå, Sweden, ²Embedded Intelligent Systems, Luleå University of Technology, Luleå, Sweden

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Li Chang,
The University of Sydney, Australia
Valentin L. Popov,
Technical University of Berlin,
Germany

*Correspondence:

Kalle Kalliorinne
kalle.kalliorinne@ltu.se

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 03 July 2020

Accepted: 30 November 2020

Published: 18 May 2021

Citation:

Kalliorinne K, Larsson R,
Pérez-Ràfols F, Liwicki M and
Almqvist A (2021) Artificial Neural
Network Architecture for Prediction of
Contact Mechanical Response.
Front. Mech. Eng 6:579825.
doi: 10.3389/fmech.2020.579825

Predicting the contact mechanical response for various types of surfaces is and has long been a subject, where many researchers have made valuable contributions. This is because the surface topography has a tremendous impact on the tribological performance of many applications. The contact mechanics problem can be solved in many ways, with less accurate but fast asperity-based models on one end to highly accurate but not as fast rigorous numerical methods on the other. A mathematical model as fast as an asperity-based, yet as accurate as a rigorous numerical method is, of course, preferred. Artificial neural network (ANN)-based models are fast and can be trained to interpret how in- and output of processes are correlated. Herein, 1,536 surface topographies are generated with different properties, corresponding to three height probability density and two power spectrum functions, for which, the areal roughness parameters are calculated. A numerical contact mechanics approach was employed to obtain the response for each of the 1,536 surface topographies, and this was done using four different values of the hardness per surface and for a range of loads. From the results, 14 *in situ* areal roughness parameters and six contact mechanical parameters were calculated. The load, the hardness, and the areal roughness parameters for the original surfaces were assembled as input to a training set, and the *in situ* areal roughness parameters and the contact mechanical parameters were used as output. A suitable architecture for the ANN was developed and the training set was used to optimize its parameters. The prediction accuracy of the ANN was validated on a test set containing specimens not seen during training. The result is a quickly executing ANN, that given a surface topography represented by areal roughness parameters, can predict the contact mechanical response with reasonable accuracy. The most important contact mechanical parameters, that is, the real area of contact, the average interfacial separation, and the contact stiffness can in fact be predicted with high accuracy. As the model is only trained on six different combinations of height probability density and power spectrum functions, one can say that an output should only be trusted if the input surface can be represented with one of these.

Keywords: artificial neural networks, contact mechanics, surface roughness, average interfacial separation, real area of contact

1 INTRODUCTION

Surface topography plays an extremely important role in processes such as wear, friction, lubrication, sealing, contact resistance, and heat conduction. This is due to that the roughness causes local contacts between the surfaces, as in mixed lubrication, thus governing how the surface deforms and behaves in contact and defining the boundary friction and the real area of contact. The surface topography may be characterized by a number of areal roughness parameters defined in ISO 25178, see ISO Central Secretary (2012). These parameters have, however, limited correlation to the real area of contact, as well as to friction and wear processes, especially if only a few out of the complete set of field parameters are considered.

By using computational contact mechanics, we can estimate the real area of contact for a surface with given topography and also show how the areal roughness parameters change inside the contact. Notice that, an accurate and reliable result requires a highly resolved surface topography measurement. Thereby, the mesh considered in the numerical solution procedure will have to be of equal resolution, which, in turn, increases the computational time significantly. From an engineering point of view, a non-iterative model that swiftly yields relatively exact predictions of contact mechanics parameters, such as the real area of contact, would therefore be highly desired.

The Greenwood and Williamson (GW) theory (Greenwood and Williamson, 1966; Greenwood and Tripp, 1970) has been and still is very frequently used. Note that it was Archard (1957) who laid the foundation for most of the (multi)asperity-based type of models known of today (Nayak, 1971; Onions and Archard, 1973; Bush et al., 1975; Bush et al., 1979; Carbone, 2009; Greenwood et al., 2011). The Persson contact mechanics theory (Persson, 2006; Yang and Persson, 2008) is also a frequently used tool. Although being highly useful models that provide insight and yield rapid predictions, they are based on assumptions, making them not always very accurate. The GW theory assumes that the asperities at the surfaces exhibit Gaussian probability distributions. The asperities are also assumed to deform independently of each other which leads to that GW theory is applicable only when the contact area is small (compared with the nominal contact area). Persson's theory assumes that the surfaces exhibit Gaussian height probability distributions and it considers interasperity coupling. Although Persson's theory might not be very accurate for small real area of contact, it applies well to study the complete contact (see e.g. (Almqvist et al., 2011)). The study by Müser et al. (2017) summarizes findings obtained with various kinds of models, including asperity-based ones and Persson's theory. Moreover, results from numerical brute force methods, all-atoms-based models, and experiments were presented as well. It was concluded that 1) rigorous numerical brute force approaches yield almost identical results on all properties, 2) Persson's theory, all-atom simulations, and experiments could be used to identify the correct trends, and almost exact numbers for some properties, and 3) asperity models predicted the real area of contact rather well and provided alternative interpretations for other properties. It would be very useful if it was possible to obtain a mathematical model for fast calculation, which is as accurate as

the rigorous models are, when predicting contact mechanics parameters such as real area of contact.

The ideal situation would be to describe surface topography by its height probability distribution and its power spectrum, which constitute the complete description. However, this complicates the analysis, and if a subset of the areal roughness parameters ISO Central Secretary (2012) would be sufficient, it would facilitate the analysis tremendously. In this study, we will present an artificial neural network (ANN)-based model. This model acts as a transfer function, taking areal roughness parameters as input and predicts the real area of contact and other contact mechanics parameters. A similar ANN-based approach has been used in contact mechanics before (see (Rapetto et al., 2009)). Other examples where ANN-based approaches have been used in tribology are Nasir et al. (2010), Nirmal (2010), Čirović et al. (2012), and Moder et al. (2018). If an ANN, which executes much faster than a computational contact mechanics approach, is well designed, trained, and tested, it can thus provide reasonably accurate predictions of tribological performance parameters very rapidly.

The idea with the present work is to generate thousands of surfaces by means of the method developed by Pérez-Ràfols and Almqvist (2019), and to compute parameters, such as the real area of contact and areal roughness parameters when these surfaces are pressed into contact with a flat rigid counter surface. To this end, we will use the computational contact mechanics approach presented by Almqvist et al. (2007), which was further developed by Sahlin et al. (2010).

The ANN is trained to find the relationship between the surfaces' original, the in-contact, that is, *in situ* areal roughness parameters and the contact mechanics parameters, for a range of loads, spanning from no load at all to a load that causes nearly as much as 50% real area of contact.

2 METHODS

This section presents, in a workflow order, the implementation of the ANN. It starts with describing surface topography generation, followed by preprocessing and a brief description of the contact mechanics approach adopted, and it ends with a presentation of the architecture of artificial neural network that was developed herein.

2.1 Surface Topography Generation

Training neural networks requires large data sets. Therefore, it is necessary to generate a wide range of different surface topographies. The surface randomization algorithm developed by Pérez-Ràfols and Almqvist (2019) was employed to randomly generate 2,022 surfaces topographies with given height probability distribution (HPD) and a power spectrum (PS). The HPD and PS can be mathematically modeled by classical distribution and spectrum functions, but they may also be obtained (and adapted) from measured surface topographies. In this work, mathematical models for Gaussian, bi-Gaussian and Weibull functions, and self-affine and exponential PS functions were used. The reader is referred to Pérez-Ràfols

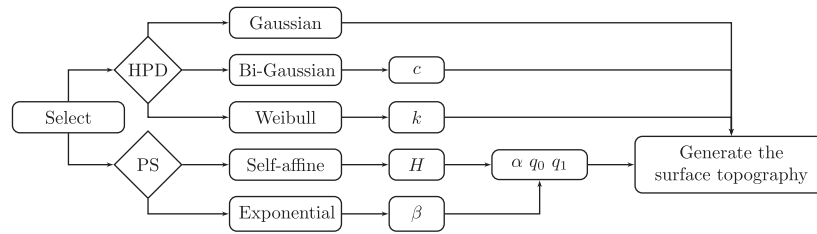


FIGURE 1 | Surface topography generation scheme.

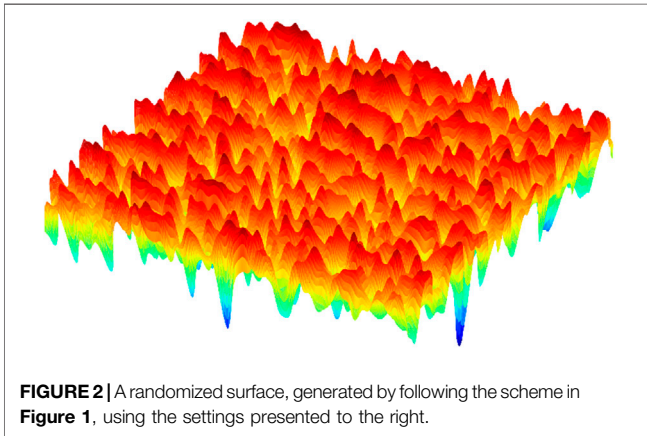


FIGURE 2 | A randomized surface, generated by following the scheme in **Figure 1**, using the settings presented to the right.

and Almqvist (2019) for a precise description of these HPD and PS. A surface topography generation selection scheme is depicted in **Figure 1**, where one type of HPD and PS is selected together with the corresponding shape-defining parameters, that is, c , k , H , β , α , q_0 , and q_1 . Remark that the HPDs are defined with zero mean value and unit standard deviation. With these constraints, the Gaussian HPD requires no input, while the bi-Gaussian and the Weibull distributions may be defined using one parameter, that is, c and k , respectively. Specifying the PS requires four parameters, that is, the Hurst exponent H for the self-affine and the parameter β , which defines the autocorrelation length $1/\beta$, for the exponential function, plus α for the anisotropy and the wave numbers (q_0, q_1) that specify the frequency bandwidth. A 256×256 mesh was considered affordable for the grand total of 7,602 elastoplastic contact mechanics simulations performed. The mesh limits the choice of the high frequency cutoff, and in order to resolve the shortest wavelength with at least eight nodes, it was chosen as $q_1 = 32$ (in terms of its wave number). This parameter was also kept constant when generating the sets of surfaces for training, testing, and validation. Thus, each surface is generated based on a pair of HPD and PS functions and five corresponding numerical parameters, except for the Gaussian which needs four.

Figure 2 shows an example of a generated surface, using the bi-Gaussian HPD model and the self-affine PS model. The corresponding parameter settings are displayed to the right.

The parameter space for the surface dataset used for training was defined by four equidistantly spaced values for each of the seven parameters (q_1 was kept constant). In this way, a wide and

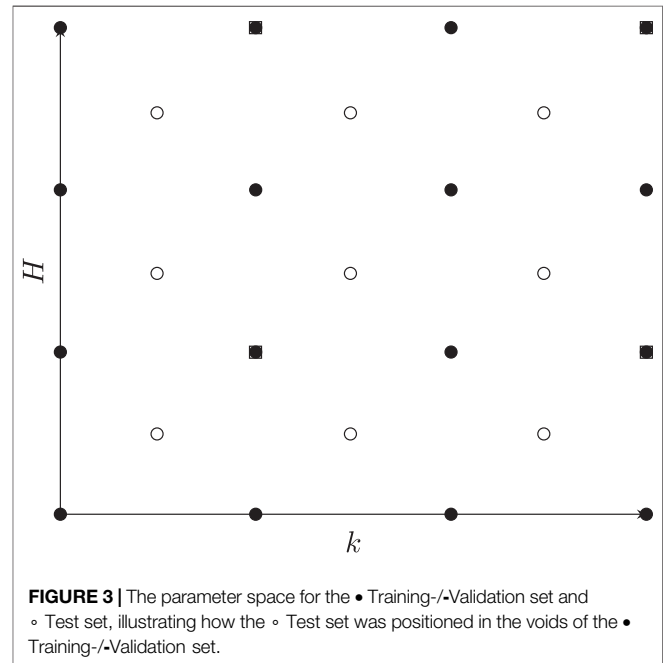


FIGURE 3 | The parameter space for the • Training-/Validation set and ◦ Test set, illustrating how the ◦ Test set was positioned in the voids of the • Training-/Validation set.

dense dataset range was obtained. A surface dataset for testing is also needed, and it is important that it is different from, but still within, the same parameter space as the training set. Notice that the validation set is a subset of the training set. The training and test sets, for a pair of parameters (k and H), are schematically illustrated in **Figure 3**. As the figure shows, the parameters in the test set are shifted a half step to be placed in the void of the training set. This ensures that the test set is located at the maximum Euclidean distance to the training set. The parameter space for the training range is specified in the table shown to the right in the figure. The training set contains 1,536 unique surfaces and the test set contains 486 unique surfaces. From the training set, 20% of the surfaces are transferred to a validation set, which is used to detect overfitting during training.

2.2 Preprocessing and Contact Mechanics

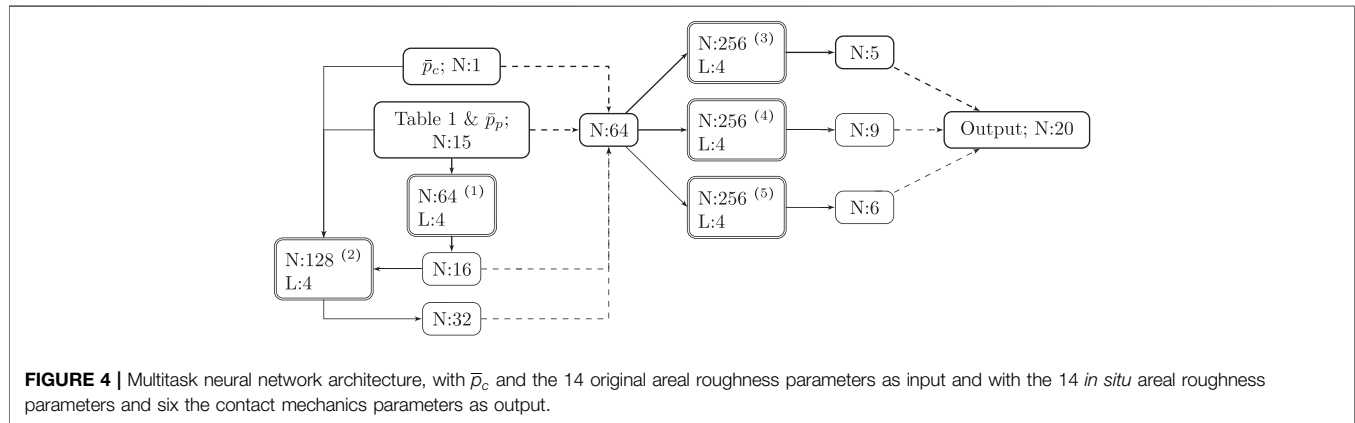
The areal roughness parameters in **Table 1** are calculated for all of the 2,022 surfaces, which are made dimensionless by scaling to exhibit unit rms roughness, that is, $S_q = 1$. This is done, in connection with the non-dimensionalization of the contact

TABLE 1 | Areal roughness parameters calculated according to ISO Central Secretary (2012).

Field type	\bar{S}_a	\bar{S}_q	\bar{S}_{sk}	\bar{S}_{ku}	\bar{S}_{dq}				
BAC type	\bar{S}_k	\bar{S}_{pk}	\bar{S}_{vk}	\bar{S}_{mr1}	\bar{S}_{mr2}	\bar{V}_{mp}	\bar{V}_{mc}	\bar{V}_{vw}	\bar{V}_{vc}

TABLE 2 | Contact mechanics parameters.

A_r	A_e	A_p	\bar{p}_{\max}	\bar{u}	\bar{K}
-------	-------	-------	------------------	-----------	-----------



mechanics model Almqvist and Pérez-Ràfols (2019), to increase the applicability of the ANN. The areal roughness parameters are, in fact, also calculated for the scaled surfaces *in situ*, that is, as the surfaces are loaded against a flat counter surface for a range of loads, spanning from no load at all to a load that causes up to 50% real area of contact. The loading sequence was defined in terms of the dimensionless nominal contact pressure \bar{p}_c , and the contact mechanics simulations were performed using the method presented by Almqvist et al. (2007); Sahlin et al. (2010) and the then utilized by Almqvist et al. (2011), Spencer et al. (2011), Spencer et al. (2013), Pérez-Ràfols et al. (2016), and Pérez-Ràfols et al. (2018).

The dimensionless areal roughness parameters (for the scaled surfaces) in **Table 1** are calculated according to ISO Central Secretary (2012). They are grouped as parameters of field type and of Bearing Area Curve (BAC) type. These parameters are the input for the ANN described in **Section 2.3**, with architecture illustrated in **Figure 4**. For more details on how to calculate the areal roughness parameters according to the standard, see Blateyron (2013). As a result of the contact mechanics simulations, the surfaces may be plastically deformed, and it is the hardness p_p that limits the maximum pressure the surface can exhibit before it yields plastically. Therefore, four equidistantly spaced values for \bar{p}_p in the range [20, 100] were used for the training set and three different values were chosen for the test set, in the same way as described for H and k in **Figure 3**. Thereby, there are in total 7,602 contact mechanics calculations performed. Out of these, 6,144 were used to train the ANN and 1,458 were used for testing.

The output from the contact mechanics calculations are the *in situ*, areal roughness parameters, and the six contact mechanics parameters in **Table 2**. These are the real area of contact to nominal contact area ratio $A_r = A_e + A_p$, the elastic- A_e and plastic contributions A_p to it, the dimensionless maximum

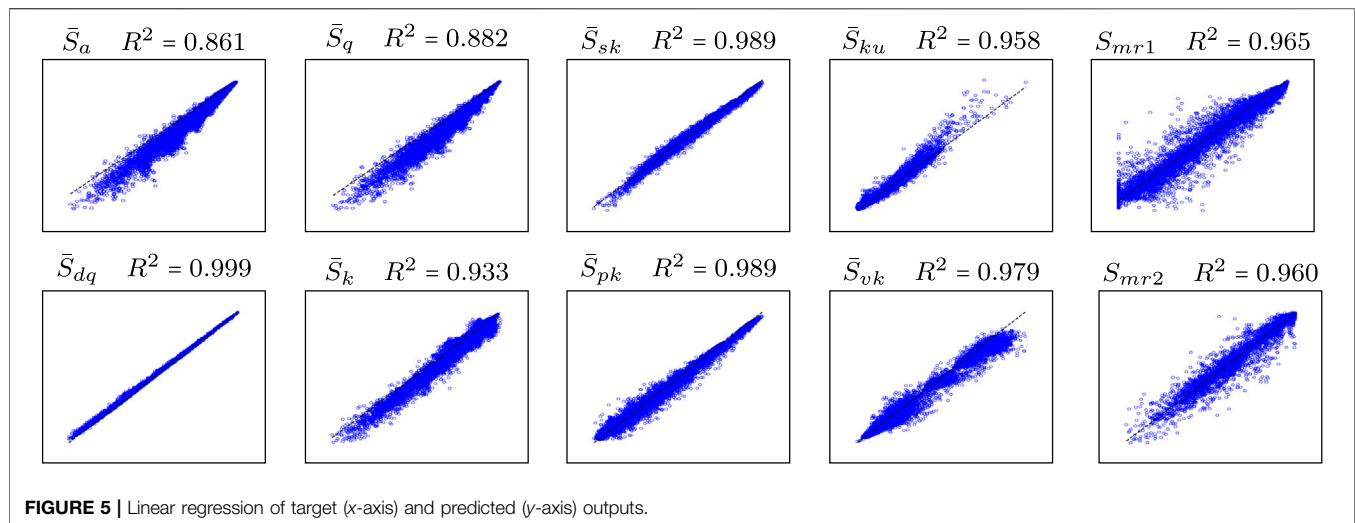
contact pressure $\bar{p}_{\max} = p_{\max}/E$, the dimensionless average interfacial separation $\bar{u} = u/h_r$, and the dimensionless contact stiffness $\bar{K} = Kh_r/E$.

2.3 The Artificial Neural Network

Here, the ANN architecture depicted in **Figure 4**, which is engineered to predict the contact mechanical response of surfaces represented by the areal roughness parameters (given in **Table 1**), will be described. The areal roughness parameters in **Table 1** and the dimensionless hardness \bar{p}_p are used as input for the ANN and it outputs the corresponding, *in situ*, areal roughness parameters and the six contact mechanics parameters in **Table 2**. As emphasized with double borders in **Figure 4**, the ANN consists of five subnetworks. These all have four fully connected layers, but a different amount of neurons per layer. The arrows with continuous lines indicate connections that are fully connected with weights, whereas arrows with dashed lines indicate just passing the data from one part of the network to another. A regular MSE loss function was adopted for the training procedure.

The first subnetwork⁽¹⁾, with 64 neurons, has rectifier (ReLU) activation functions ($f(x) = \max(0, x)$), and it takes the 14 areal roughness parameters in **Table 1** and the value of the hardness as input. The purpose of this network is to process the areal roughness parameters, without the influence of the contact pressure, to extract suitable input for the second subnetwork⁽²⁾ with 128 neurons. This subnetwork has sigmoid activation functions ($f(x) = 1/(1 + e^{-x})$), and it is fully connected to the 14 areal roughness parameters, the hardness, the dimensionless contact pressure, and the output of the first subnetwork.

The input to the first and second subnetworks and their output are assembled into one vector with 64 values. This vector is the input to each of the three parallel subnetworks⁽³⁾⁻⁽⁵⁾, which have softplus (smooth rectifying)



activation functions ($f(x) = \ln(1 + e^x)$) and 256 nodes each. As mentioned previously, the purpose of the ANN is to predict the 14 *in situ* areal roughness parameters as well as the six contact mechanics parameters in **Table 2**. To this end, each of the three parallel subnetworks output parameters grouped by its origin, that is, the five *in situ* areal roughness parameters of field type, the nine *in situ* areal roughness parameters related to the bearing area (or Abbott-Firestone) curve, and the six contact mechanics parameters.

3 RESULTS AND DISCUSSION

First, in **Section 3.1**, the performance of the ANN model will be evaluated with linear regression between predicted values and the output in the test dataset. Then, in **Section 3.2**, examples of how the predictions changes with the load will be presented and compared to the correct values.

3.1 Predicting Contact Mechanical Response

Herein, the test set, which contains 1,458 specimens that it has never seen before, is used to evaluate the ANN's predictive performance on surfaces for a whole range of loads. Depicted in **Figures 5, 6** are linear regression of all the predicted parameter values and the R^2 -value, that is,

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

where y is the target output, \hat{y} is the predicted output, and \bar{y} is the mean target output, is used as a measure of the accuracy. Overall, one can see that some parameters are predicted with extraordinary high accuracy, whereas a few are predicted with less precision. **Figure 5** reveals that there is a systematic error for the predictions of \bar{S}_a and \bar{S}_q , which both have relatively low R^2 -values. The reason for the low R^2 -values is because the absolute majority of predictions (for both \bar{S}_a and \bar{S}_q) are

underestimated. Among the output shown in **Figure 5**, the one with highest accuracy is the mean quadratic slope parameter \bar{S}_{dq} . Visually, the bearing area curve parameter S_{mr1} shows a quite large spread, while the R^2 -value is rather high. This is caused by a relatively small percentage predictions with large errors.

There is much that can be said about the results presented in **Figure 6**. One thing, which is nearly impossible not to notice, is the regression for the dimensionless maximum pressure \bar{p}_{max} , with wide-spread and a low R^2 -value. The reason for this is that the \bar{p}_{max} is a local event, while areal roughness parameters are averaged in some sense. In other words, prediction of a local quantity based on average terms is a complicated task, and the low accuracy is, therefore, to be expected. The more important outputs A_r , A_{re} , A_{rp} , \bar{u} , and \bar{K} are, fortunately, predicted with higher accuracy. For more details on the ANN's predictability, the reader is referred to next section.

3.2 Application

In this section, the accuracy of the predictions of the ANN for three different test specimens, taken from the test set that the network never has seen before, will be investigated over the whole range of loads considered when the test set was generated with the contact mechanics simulations. The test specimens are listed in **Tables 3, 4**, and they consist of the areal roughness parameters, corresponding to the surfaces topographies presented in **Figure 7**, combined with a value of the dimensionless hardness.

The predictions of the *in situ* dimensionless mean square height \bar{S}_q and skewness \bar{S}_{sk} are depicted in **Figure 8**, which also shows the correct values. This figure and **Figures 9–12** share the same legend in which the lines (continuous blue, dashed red, and dotted turquoise) are for the predictions, and the correct values are represented with the markers (round blue, round red, and cross turquoise). The ANN predicts both the \bar{S}_q and \bar{S}_{sk} output parameters for Specimen 1 (bi-Gaussian and self-affine) with best accuracy. The lowest accuracy was observed when predicting \bar{S}_q for Specimen 2 (Gaussian and self-affine), and the lowest accuracy was

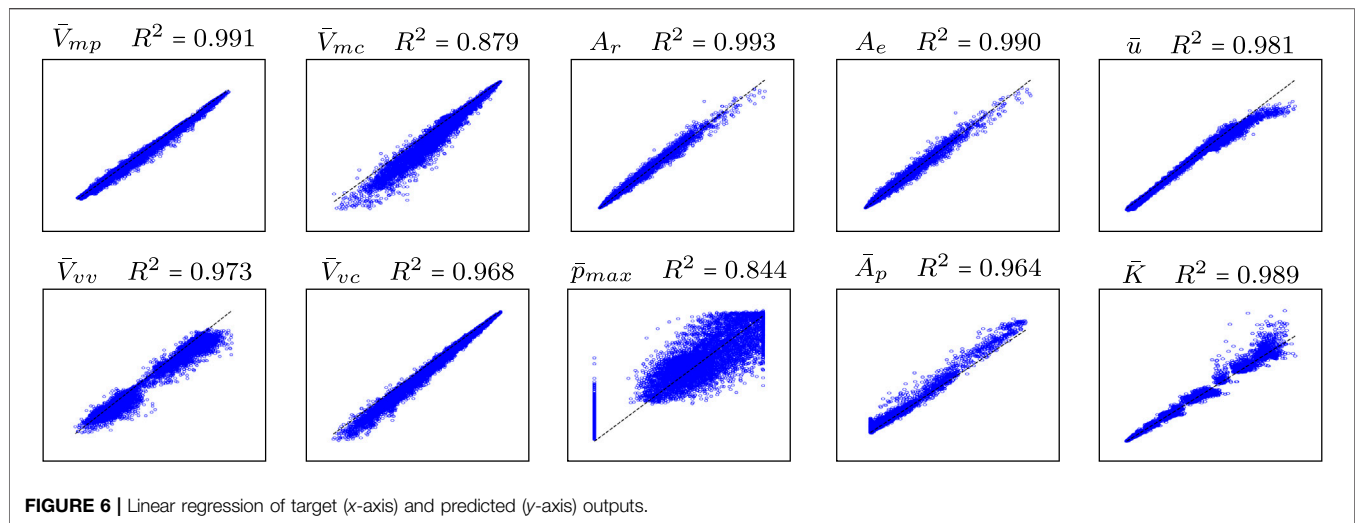
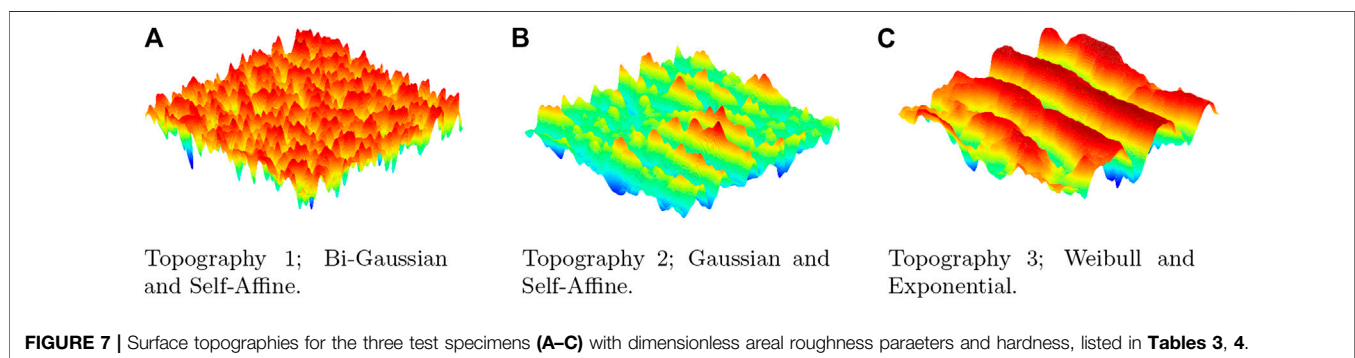


TABLE 3 | Test specimens containing the dimensionless areal roughness parameters, corresponding to the topographies depicted in **Figure 7** and a value of the dimensionless hardness: part 1—field type parameters.

	\bar{S}_a	\bar{S}_q	\bar{S}_{sk}	\bar{S}_{ku}	\bar{S}_{dq}
Specimen 1	0.792	1.000	0.832	3.863	82.254
Specimen 2	0.798	1.000	-0.013	2.967	51.310
Specimen 3	0.829	1.000	0.561	2.701	31.001

TABLE 4 | Test specimens containing the dimensionless areal roughness parameters, corresponding to the topographies depicted in **Figure 7** and a value of the dimensionless hardness: part 2—BAC type parameters and dimensionless hardness.

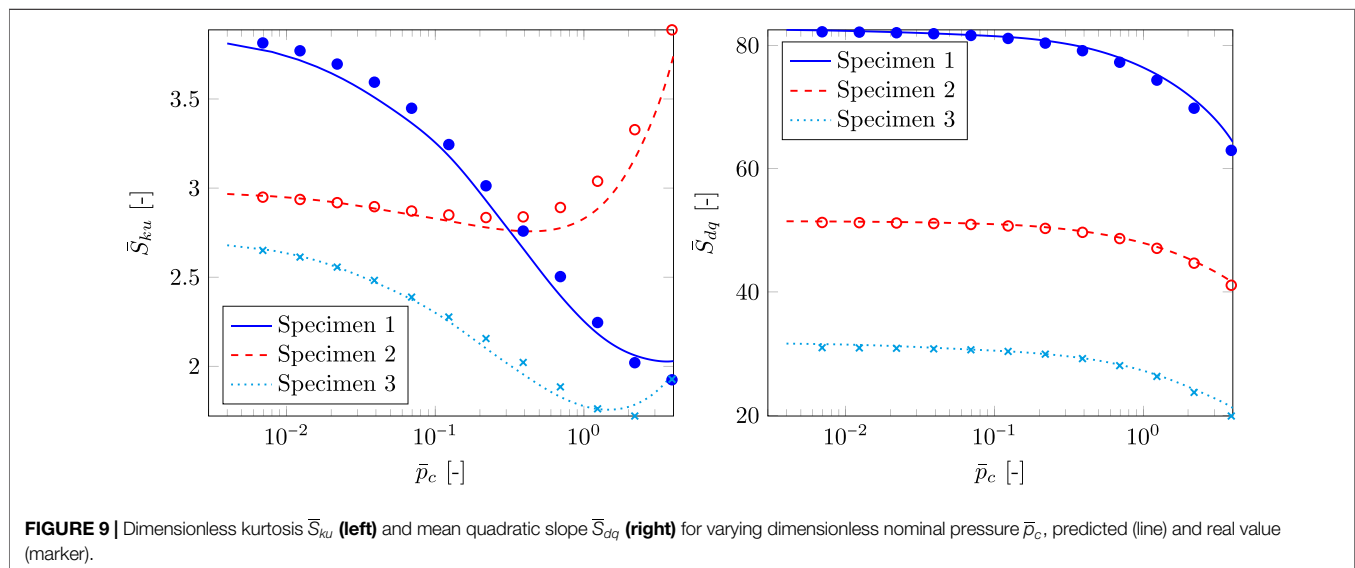
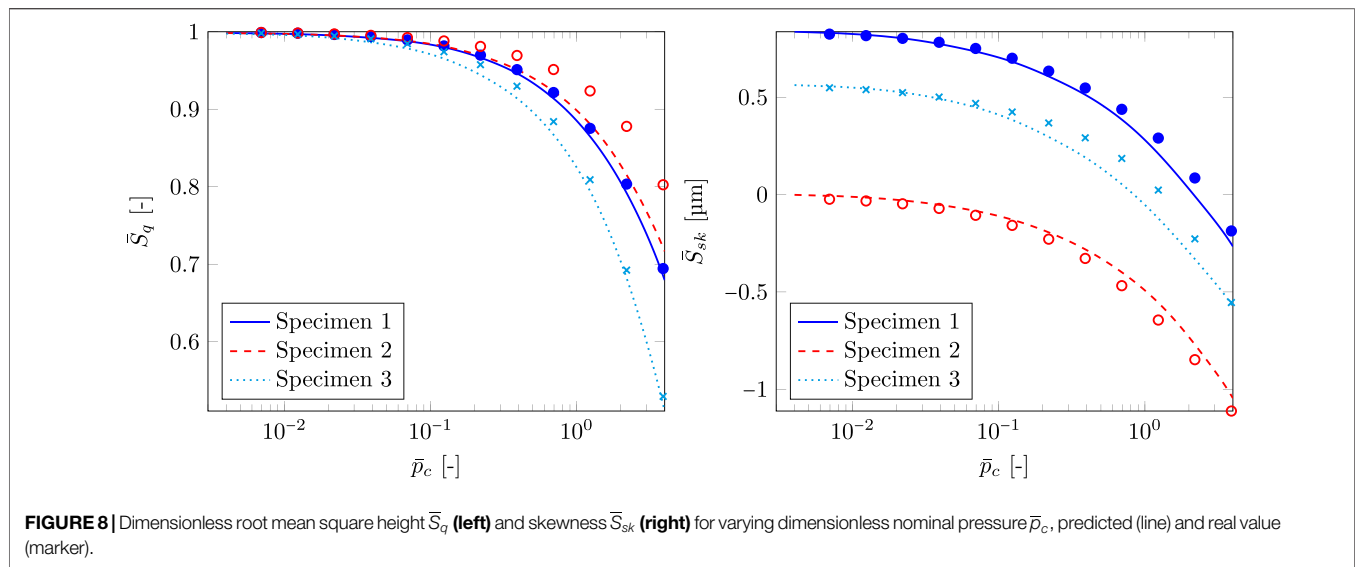
	S_{mr1}	S_{mr2}	\bar{S}_k	\bar{S}_{pk}	\bar{S}_{vk}	\bar{V}_{mp}	\bar{V}_{mc}	\bar{V}_{vv}	\bar{V}_{vc}	\bar{p}_p
Specimen 1	13.260	95.596	2.499	1.443	0.299	0.069	0.880	0.064	1.355	33.333
Specimen 2	9.877	89.899	2.567	0.937	0.953	0.047	0.911	0.112	1.215	60.000
Specimen 3	15.712	99.890	2.568	1.160	0.010	0.054	1.022	0.039	1.393	86.667



observed when predicting \bar{S}_{sk} for Specimen 3 (Weibull and exponential).

The predictions of the *in situ* dimensionless kurtosis \bar{S}_{ku} and mean quadratic slope \bar{S}_{dq} are depicted in **Figure 9**. For \bar{S}_{ku} , to the left in **Figure 9**, the accuracy is rather high and approximately the

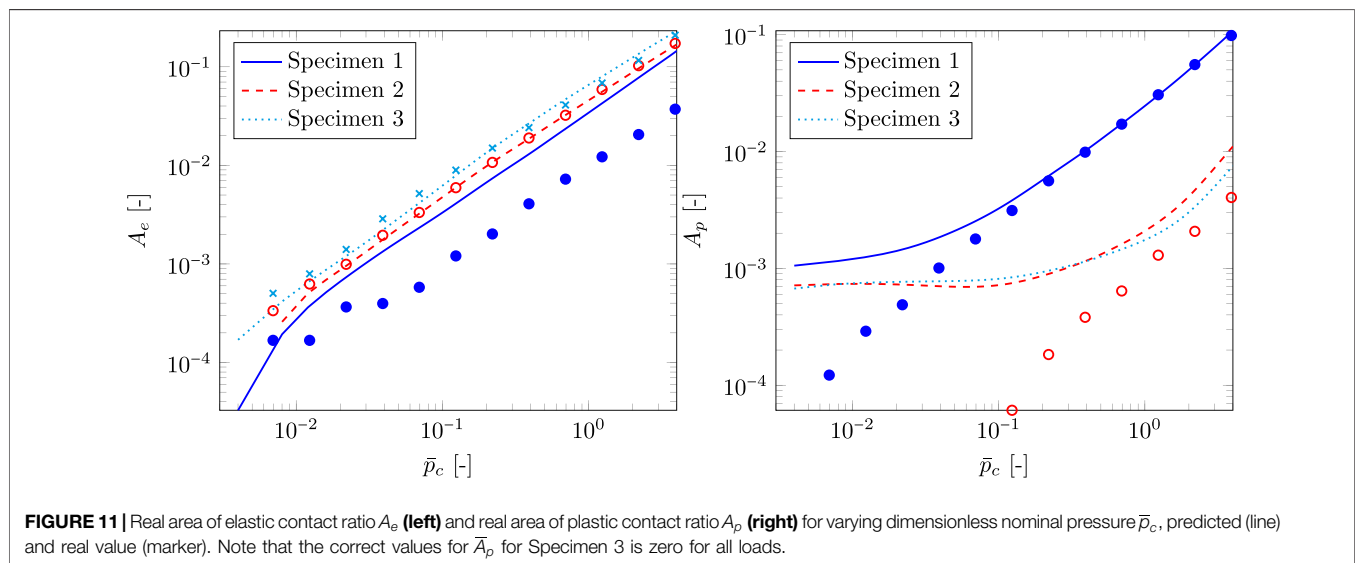
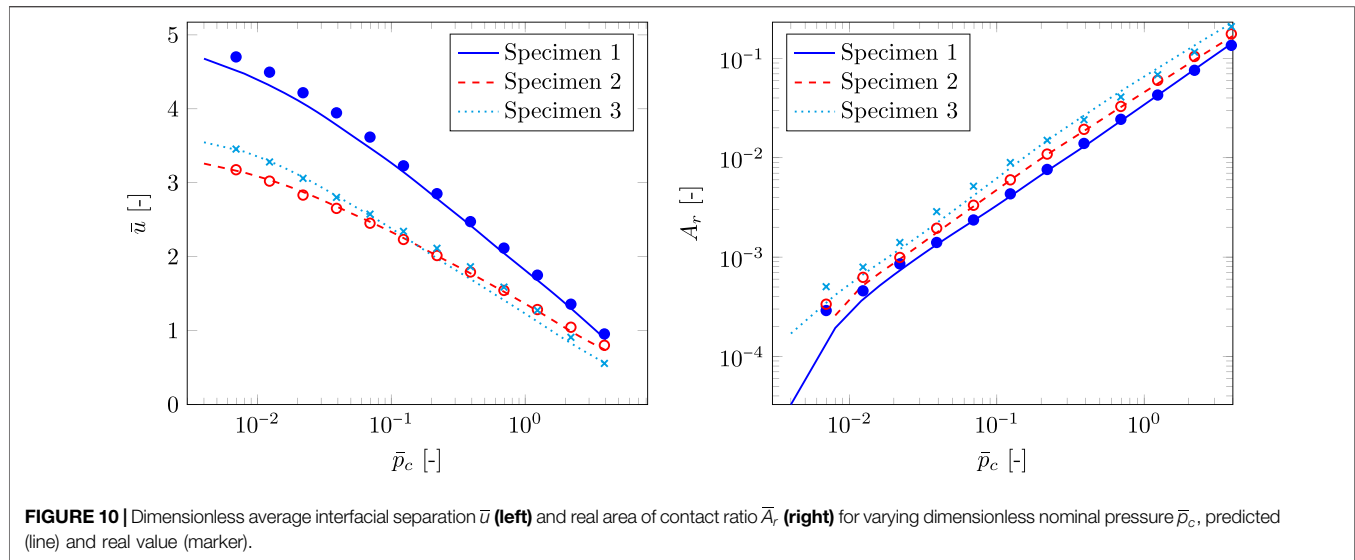
same for all three specimens. The variation of the *in situ* kurtosis is very complex, but still accurately captured by the ANN. To the right in **Figure 9**, it can be seen that the predictions of the *in situ* \bar{S}_{dq} are of very high accuracy, and this can be understood from the linear regression analysis presented in **Figure 5**.



The predictions of the dimensionless average interfacial separation \bar{u} and real area of contact ratio A_r are depicted in **Figure 10**. It is observed that the ANN very accurately predicts \bar{u} for Specimen 2 over the whole range of loads tested. The accuracy for Specimen 1 is not so high at low loads but really good for moderate and high loads, and it is vice versa for Specimen 3. Overall, the ANN's accuracy in predicting \bar{u} is good, which is required if the ANN would be employed in a mixed lubrication model like the one in Sahlin et al. (2010). As displayed in the right of **Figure 10**, the real area of contact ratio can be predicted with satisfactory accuracy for all but the lowest load, where it ideally should extrapolate $A_r \rightarrow 0$ as $\bar{p}_c \rightarrow 0$. Better performance could (most likely) have been obtained, by extending the training set to include more results for lower loads. Note that this would also require a higher mesh density than the 256×256 used presently.

The ANN is trained such that the areal roughness parameters for $\bar{p}_c = 0$ remains unchanged. The ANN is also trained such that the contact mechanics parameters are zero for $\bar{p}_c = 0$, except for the average interfacial separation, which is specified as the surface's maximum peak height.

The predictions of the elastic part of the real area of contact ratio \bar{A}_e (left) and plastic part of the real area of contact ratio A_p (right) are depicted in **Figure 11**. When looking at the predictions for \bar{A}_e , it is noticeable that there is a large error for Specimen 1; however, the other specimens are predicted with acceptable accuracy. From the results for \bar{A}_p presented to the right in **Figure 11**, it seems as if the ANN has qualitatively learned what the variation of \bar{A}_p would be. More precisely, that it is constant for low loads but that it starts to increase at some point. The reason for that it does not quantitatively capture the variation



correctly has probably to do with that a relative error for a large \bar{A}_p is much more significant than it is for a small \bar{A}_p , during the training procedure. Notice that the ANN predicts that Specimen 3 exhibits plastic deformation, but that the correct result is that the deformations are purely elastic for all loads considered ($\bar{A}_p = 0$ is not displayed on the log-scaled axis).

The predictions of the dimensionless maximum pressure \bar{p}_{\max} (left) and contact stiffness \bar{K} (right) are depicted in Figure 12. When looking at the predictions for \bar{p}_{\max} , one can see that there is a quite large error. This was also brought up in connection to the presentation of Figure 6. Most surfaces will be plastically deformed, and it seems as it would be fairly easy for the ANN to learn that the \bar{p}_{\max} will saturate at \bar{p}_p . Indeed, by looking at the predictions for Specimen 1 and 2, it is also clear that it has learned this. Specimen 1 that has the lowest \bar{p}_p is already plastically deformed at the smallest load in the range and Specimen 3 with

the highest \bar{p}_p is not plastically deformed at all. Specimen 2 does, however, exhibit \bar{p}_{\max} for an intermediate load in the range, and it can be observed that the ANN is able to predict that it will and that it is capable of capturing the position where it occurs. From Figure 12, it can also be observed that the contact stiffness, for all three specimens, can be predicted with quite high accuracy for moderate and high loads but that the accuracy decreases for lower loads.

4 CONCLUDING REMARKS

Two datasets containing a total of 2,022 different surface topographies were generated using the algorithm developed by Pérez-Ràfols and Almqvist (2019). Three different HPD functions and two different PS functions were obtained from

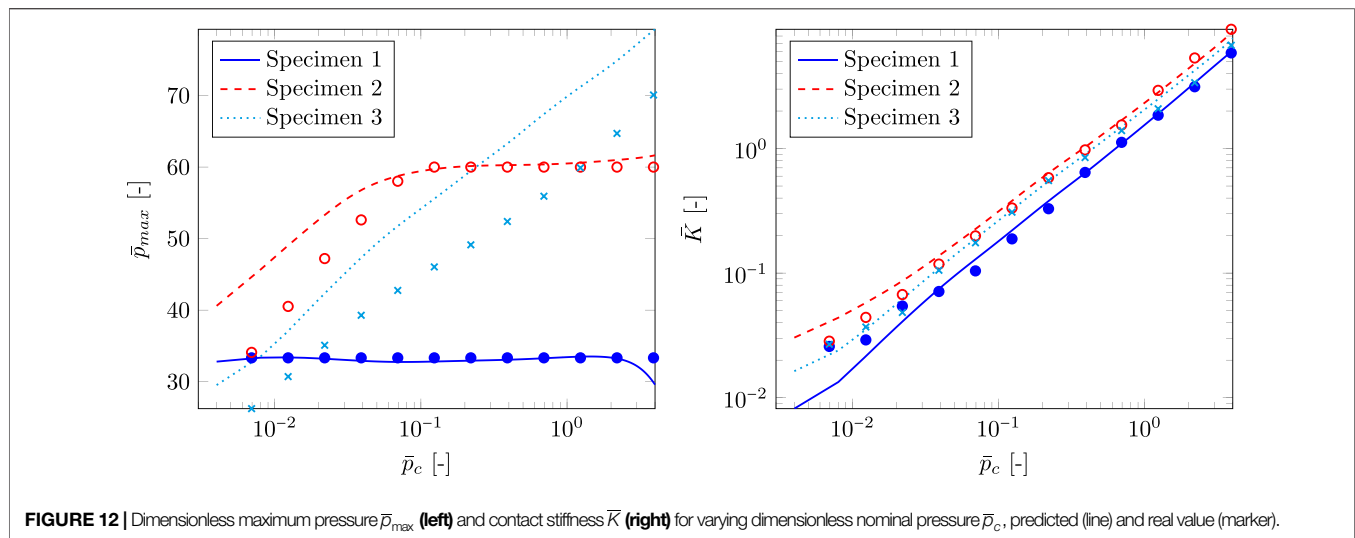


FIGURE 12 | Dimensionless maximum pressure \bar{p}_{max} (left) and contact stiffness \bar{K} (right) for varying dimensionless nominal pressure \bar{p}_c , predicted (line) and real value (marker).

Gaussian, bi-Gaussian, and Weibull HPD functions and self-affine and exponential PS functions, described with as few shape-defining parameters as possible. Fourteen areal roughness parameters were calculated for all surfaces in the dataset. Together with the surface indentation hardness and a given applied load, these 14 areal roughness parameters were used as input for the ANN.

A numerical elastoplastic contact mechanics approach, in which the hardness limits the maximum pressure the surface can exhibit before it yields plastically, was then employed to perform simulations of pressing each of the generated surfaces against a flat rigid counter surface for a sequence of loads. Since four values for the hardness were considered for the training set with 1,536 different surface topographies and three were considered for the test set with 486 topographies, a grand total of $4 \times 1536 + 3 \times 486 = 7602$ realizations were conducted. Out of these, 6,144 specimens were used for training and 1,458 were left for testing and validation. For each of these specimens, 14 *in situ* areal roughness parameters and six contact mechanics parameters were calculated for the sequence of loads that was also used as input for the ANN.

An architecture for an artificial neural network (ANN), which consisted of five different subnetworks, was designed and trained on the dataset. Linear regression was applied, and the R^2 -value was used to appreciate the correlation between the network prediction and the correct data. A few parameters were almost perfectly predicted, whereas other were predicted with large errors. According to the R^2 -values, the most important parameters, that is, the real area of contact ratio A_r , the dimensionless average interfacial separation \bar{u} , and contact stiffness \bar{K} were all predicted accurately by the ANN.

REFERENCES

- Almqvist, A., and Pérez-Ráfols, F. (2019). *Scientific Computing with Applications in Tribology: A Course Compendium*. Luleå, Sweden: Luleå University of Technology
- Almqvist, A., Campaña, C., Prodanov, N., and Persson, B. N. J. (2011). Interfacial Separation between Elastic Solids with Randomly Rough Surfaces: Comparison

Summing up, the ANN can be used to roughly appreciate the *in situ* behavior of various kinds of surface topographies, if the areal roughness parameters, the indentation hardness, and the nominal contact pressure are known. Some parameters, that is, the real area of contact ratio, the dimensionless average interfacial separation, and contact stiffness can actually be predicted with high accuracy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

KK has performed the main part of the development of the ANN-based model and has been coordinating the writing. RL contributed with expertise, to the discussions and was engaged in the writing. FR has been engaged in surface generation and contact mechanics and contributed to the writing. ML contributed with expertise, particularly regarding AI and machine learning, contributed to discussions and to the writing. AA has initiated the work and has been involved in all parts of it.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from VR (The Swedish Research Council): DNR 2019-04293.

between Theory and Numerical Techniques. *J. Mech. Phys. Sol.* 59 (11), 2355–2369. doi:10.1016/j.jmps.2011.08.004

Almqvist, A., Sahlin, F., Larsson, R., and Glavatskih, S. (2007). On the Dry Elastoplastic Contact of Nominally Flat Surfaces. *Tribology Int.* 40 (4), 574–579. doi:10.1016/j.triboint.2005.11.008

Archard, J. F. (1957). Elastic Deformation and the Laws of Friction. *Proc. R. Soc. Lond. A* 243 (1233):190–205. doi:10.1098/rspa.1957.0214

- Blateyron, F. (2013). *The Areal Field Parameters*. Berlin, Heidelberg: Springer Berlin Heidelberg, 15–43. doi:10.1007/978-3-642-36458-7_2
- Bush, A. W., Gibson, R. D., and Keogh, G. P. (1979). Strongly Anisotropic Rough Surfaces. *Trans. ASME. J. Lubrication Tech.* 101 (1), 15–20. doi:10.1115/1.3453271
- Bush, A. W., Gibson, R. D., and Thomas, T. R. (1975). The Elastic Contact of a Rough Surface. *Wear* 35 (1), 87–111. doi:10.1016/0043-1648(75)90145-3
- Ćirović, V., Aleksendrić, D., and Mladenović, D. (2012). Braking Torque Control Using Recurrent Neural Networks. *Proc. Inst. Mech. Eng. D: J. Automobile Eng.* 226 (6), 754–766. doi:10.1177/0954407011428720
- Carbone, G. (2009). A Slightly Corrected Greenwood and Williamson Model Predicts Asymptotic Linearity between Contact Area and Load. *J. Mech. Phys. Sol.*, 57(7):1093–1102. doi:10.1016/j.jmps.2009.03.004
- ISO Central Secretary (2012). *Geometrical Product Specifications (Gps)—Surface Texture: Areal—Part 2: Terms, Definitions and Surface Texture Parameters*. Geneva, CH: International Organization for Standardization. Available at: <https://www.iso.org/standard/42785.html>
- Greenwood, J. A., Putignano, C., and Ciavarella, M. (2011). A Greenwood & Williamson Theory for Line Contact. *Wear* 270, 332–334. doi:10.1016/j.wear.2010.11.002
- Greenwood, J., and Tripp, J. (1970). The Contact of Two Nominally Flat Rough Surfaces. *Proc. Inst. Mech. Eng.* 185, 625–634. doi:10.1243/pime_proc_1970_185_069_02
- Greenwood, J., and Williamson, J. (1966). Contact of Nominally Flat Surfaces. *Proc. R. Soc. Lond. A* 295, 300–319. doi:10.1098/rspa.1966.0242
- Müser, M. H., Dapp, W. B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, A. A., et al. (2017). Meeting the Contact-Mechanics Challenge. *Tribology Lett.* 65 (4), 118. doi:10.1007/s11249-017-0900-2
- Moder, J., Bergmann, P., and Grün, F. (2018). Lubrication Regime Classification of Hydrodynamic Journal Bearings by Machine Learning Using Torque Data. *Lubricants* 6 (4), 108. doi:10.3390/lubricants6040108
- Nasir, T., Yousif, B. F., McWilliam, S., Salih, N. D., and Hui, L. T. (2010). An Artificial Neural Network for Prediction of the Friction Coefficient of Multi-Layer Polymeric Composites in Three Different Orientations. *Proc. Inst. Mech. Eng. C: J. Mech. Eng. Sci.* 224 (2), 419–429. doi:10.1243/09544062jmes1677
- Nayak, P. R. (1971). Random Process Model of Rough Surfaces. *J. Lubrication Tech.* 93 (3), 398–407. doi:10.1115/1.3451608
- Nirmal, U. (2010). Prediction of Friction Coefficient of Treated Betelnut Fibre Reinforced Polyester (T-bfrp) Composite Using Artificial Neural Networks. *Tribology Int.* 43 (8), 1417–1429. doi:10.1016/j.triboint.2010.01.013
- Onions, R. A., and Archard, J. F. (1973). The Contact of Surfaces Having a Random Structure. *J. Phys. D: Appl. Phys.* 6 (3), 289–304. doi:10.1088/0022-3727/6/3/302
- Persson, B. N. J. (2006). Contact Mechanics for Randomly Rough Surfaces. *Surf. Sci. Rep.* 61, 201–227. doi:10.1016/j.surfrep.2006.04.001
- Pérez-Ràfols, F., and Almqvist, A. (2019). Generating Randomly Rough Surfaces with Given Height Probability Distribution and Power Spectrum. *Tribology Int.* 131, 591–604. doi:10.1016/j.triboint.2018.11.020
- Pérez-Ràfols, F., Larsson, R., and Almqvist, A. (2016). Modelling of Leakage on Metal-To-Metal Seals. *Tribology Int.* 94, 421–427. doi:10.1016/j.triboint.2015.10.003
- Pérez-Ràfols, F., Larsson, R., van Riet, E. J., and Almqvist, A. (2018). On the Loading and Unloading of Metal-To-Metal Seals: A Two-Scale Stochastic Approach. *Proc. Inst. Mech. Eng. Part J: J. Eng. Tribology* 232 (12), 1525–1537. doi:10.1177/1350650118755620
- Rapetto, M. P., Almqvist, A., Larsson, R., and Lugt, P. M. (2009). On the Influence of Surface Roughness on Real Area of Contact in Normal, Dry, Friction Free, Rough Contact by Using a Neural Network. *Wear* 266 (5), 592–595. doi:10.1016/j.wear.2008.04.059
- Sahlin, F., Larsson, R., Marklund, P., Lugt, P. M., and Almqvist, A. (2010). A Mixed Lubrication Model Incorporating Measured Surface Topography. Part 1: Theory of Flow Factors. *Proc. Inst. Mech. Eng. Part J: J. Eng. Tribology* 224 (4), 335–351. doi:10.1243/13506501jet658
- Spencer, A., Almqvist, A., and Larsson, R. (2011). A Semi-deterministic Texture-Roughness Model of the Piston Ring-Cylinder Liner Contact. *Proc. Inst. Mech. Eng. Part J: J. Eng. Tribology* 225 (6), 325–333. doi:10.1177/1350650110396279
- Spencer, A., Avan, E. Y., Almqvist, A., Dwyer-Joyce, R. S., and Larsson, R. (2013). An Experimental and Numerical Investigation of Frictional Losses and Film Thickness for Four Cylinder Liner Variants for a Heavy Duty Diesel Engine. *Proc. Inst. Mech. Eng. Part J: J. Eng. Tribology* 227 (12), 1319–1333. doi:10.1177/1350650113491244
- Yang, C., and Persson, B. N. J. (2008). Contact Mechanics: Contact Area and Interfacial Separation from Small Contact to Full Contact. *J. Phys. Condensed Matter* 20, 1–13. doi:10.1088/0953-8984/20/21/215214

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Kalliorinne, Larsson, Pérez-Ràfols, Liwicki and Almqvist. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

NOMENCLATURE

α anisotropy coefficient

S_a arithmetic mean height (μm)

$1/\beta$ auto-correlation length

u average interfacial separation (μm)

c bimodal shape parameter

K contact stiffness (N/m)

V_{mc} core material volume ($\mu\text{m}^3/\mu\text{m}^2$)

S_k core roughness depth (μm)

V_{vc} core void volume ($\mu\text{m}^3/\mu\text{m}^2$)

\bar{S}_a dimensionless arithmetic mean height S_a/h_r

\bar{u} dimensionless average interfacial separation u/h_r

\bar{K} dimensionless contact stiffness Kx_r/E

\bar{V}_{mc} dimensionless core material volume V_{mc}/h_r

\bar{S}_k dimensionless core roughness depth S_k/h_r

\bar{V}_{vc} dimensionless core void volume V_{vc}/h_r

\bar{p}_p dimensionless hardness $p_p/(Eh_r/x_r)$

\bar{S}_{ku} dimensionless kurtosis S_{ku}/h_r

\bar{p}_{\max} dimensionless maximum pressure $p_{\max}/(Eh_r/x_r)$

\bar{S}_{dq} dimensionless mean quadratic slope $S_{dq}x_r/h_r$

\bar{p}_c dimensionless nominal pressure $p_c/(Eh_r/x_r)$

\bar{V}_{mp} dimensionless peak material volume V_{mp}/h_r

\bar{S}_{pk} dimensionless reduced peak height S_{pk}/h_r

\bar{S}_{vk} dimensionless reduced valley depth S_{vk}/h_r

\bar{S}_q dimensionless root mean square height S_q/h_r

\bar{S}_{sk} dimensionless skewness S_{sk}/h_r

\bar{V}_{vv} dimensionless valley void volume V_{vv}/h_r

p_p hardness MPa

H Hurst exponent

S_{ku} kurtosis (μm)

q_0 long wavelength cutoff

S_{mr1} material ratio 1 –

S_{mr2} material ratio 2

p_{\max} maximum pressure MPa

S_{dq} mean quadratic slope ($\mu\text{m}/\text{mm}$)

p_c nominal pressure MPa

V_{mp} peak material volume ($\mu\text{m}^3/\mu\text{m}^2$)

A_r real area of contact ratio

A_e real area of elastic contact ratio

A_p real area of plastic contact ratio

S_{pk} reduced peak height (μm)

S_{vk} reduced valley depth (μm)

h_r reference height S_q (μm)

x_r reference length (mm)

S_q root mean square height (μm)

q_1 short wavelength cutoff

S_{sk} skewness (μm)

V_{vv} valley void volume ($\mu\text{m}^3/\mu\text{m}^2$)

k Weibull shape parameter

t worn shape parameter



Probabilistic, Fractal, and Related Techniques for Analysis of Engineering Surfaces

Feodor M. Borodich^{1,2}, Xiaoqing Jin¹ and Andrey Pepelyshev^{3*}

¹ College of Aerospace Engineering, Chongqing University, Chongqing, China, ² School of Engineering, Cardiff University, Cardiff, United Kingdom, ³ School of Mathematics, Cardiff University, Cardiff, United Kingdom

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of Berlin,
Germany

Reviewed by:

Julius Kaplunov,
Keele University, United Kingdom
Alexander Galybin,
Institute of Physics of the Earth (RAS),
Russia

*Correspondence:

Andrey Pepelyshev
pepelyshevan@cardiff.ac.uk

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 29 May 2020

Accepted: 02 July 2020

Published: 28 August 2020

Citation:

Borodich FM, Jin X and Pepelyshev A
(2020) Probabilistic, Fractal, and
Related Techniques for Analysis of
Engineering Surfaces.
Front. Mech. Eng. 6:64.
doi: 10.3389/fmech.2020.00064

In many engineering fields surface topography is of crucial importance solving problems of friction and other problems of tribology. A review of mathematical approaches for description of topography of engineering surfaces is presented. Firstly, we give a brief introduction to some of statistical parameters used for description of surface roughness. It is argued that although some of these parameters may be quite useful for specific engineering problems, a set of finite numbers of parameters cannot describe contact properties of rough surfaces. Then we discuss various models of surface roughness based on Gaussian models of the asperity heights. The results of application of various modern tests of normality for checking whether the distribution of the asperity heights is Gaussian, are presented. Further fractal models of roughness are discussed. Using fractal parametric-homogeneous (PH) surfaces, it is demonstrated that tribological properties of a rough surface cannot be characterized just by the fractal dimension of the surface. It is also shown that models based solely on the power-spectral density function (PSDF) are quite similar to fractal models and these models do not reflect tribological properties of surfaces. In particular, it is demonstrated that different profiles may have the same PSDF.

Keywords: roughness, power-spectral density, fractal, statistics, parametric homogeneity, contact problems

1. INTRODUCTION

The paper deals mainly with surfaces used in engineering practice that will be referred to as engineering surfaces. It is known that all engineering surfaces are rough (see e.g., Archard et al., 1975; Whitehouse, 2011) and therefore, contact between engineering surfaces is realized by a number of contact spots (see e.g., Zhuravlev, 1940, 2007; Holm, 1941; Goryacheva, 1998; Borodich, 2007). If the surface profile $z(x)$ is studied using Fourier decomposition, and the term 'roughness' is attributed to the short wavelength shapes, while the long wavelength shapes are referred to as "waviness" of the surface (see e.g., Morales-Espejel et al., 2000; Borodich and Bianchi, 2013). If the waviness is extracted from the surface profile then the rough surface may be considered as nominally flat (see e.g., Greenwood and Williamson, 1966). Roughness of engineering surfaces is a crucial factor for performance of tribological components. The energy dissipation during sliding of dry engineering surfaces and correspondingly, the friction are enormously influenced by the surface profile (see e.g., Borodich and Savencu, 2017). Here we present a critical review of some popular statistical, fractal and related techniques for modeling and analysis of the surface roughness.

One of the first attempts to employ statistical methods for description of surface roughness was presented by Abbott and Firestone (1933) who calculated the cumulative distribution function of

the surface heights. In tribology this parameter is called the Abbott-Firestone curve or the bearing area curve. Independently, Zhuravlev (1940) employed this parameter in his statistical model of contact between rough surfaces that were represented as collection of spherical protuberances having identical radii. He explained that the number of contacting spheres a specific height increases as the surfaces approach each other. In the English language literature this model is usually attributed to Greenwood and Williamson (1966).

After the bearing area curve parameter was introduced, there was a period that can be referred to as “the parameter rash” because a huge number of statistical parameters of roughness were introduced (Whitehouse, 1982). These characteristics were related to both the vertical distribution of heights and the horizontal distribution of the rough profiles (Nowicki, 1985).

The next step in surface roughness characterization was the idea that it may be modeled using theory of random processes. In 1953 Linnik and Khusu presented a seminar talk where they suggested to use graphs of a stationary Gaussian random process in order to describe surface roughness, see Linnik and Khusu (1954a) for detail, as well as Linnik and Khusu (1954b) and Khusu et al. (1975). Linnik and Khusu (1954a) suggested to study the following correlation function for the Gaussian random process

$$K(x) = K(0) \cdot e^{-\alpha|x|}, \quad (1)$$

where $K(0)$ and α are some parameters of the roughness. Independently, the same idea was introduced later by Whitehouse and Archard (1970). They presented an absolutely correct statement that if a profile $z(x)$ of a random rough surface is Gaussian then it can be fully described by a distribution of asperity heights and the correlation (auto-correlation) function of the process R . The auto-correlation is defined as

$$R(\delta) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T [z(x+\delta) - \bar{z}][z(x) - \bar{z}] dx \\ = \langle [z(x+\delta) - \bar{z}][z(x) - \bar{z}] \rangle \quad (2)$$

If one takes the Fourier transform of $R(\delta)$ then the power spectrum $G(\omega)$ or the power-spectral density function (PSDF) is obtained. If the signal frequency is denoted as ω then the PSDF is defined as

$$G(\omega) = \frac{2}{\pi} \int_0^\infty R(\delta) \cos \omega \delta d\delta \quad \text{and} \\ \bar{z} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T z(x) dx.$$

Developing the random signal approach, Sayles and Thomas (1978) presented experimental relations between wavelength and the scaled power spectral density for many different surfaces. They argued that the scaled spectral density functions of many surface profiles can be approximately presented as $G(\omega) = 2\pi \Lambda / \omega^2$. Sayles and Thomas (1978) referred to Λ as the surface topothesy. As Dr. Sayles said to one of the authors (FB), they never claimed that the real surfaces are fractal; in fact the fractal terminology to surface roughness description was triggered by

Berry and Hannay (1978) who presented a comment to Sayles and Thomas (1978) paper where they claimed that geometric properties of rough surfaces can be characterized by a new concept ‘fractal’ that was described in detail by Mandelbrot (1977).

Another important step in the promotion of the fractal approach to surface roughness description was the studies of the Weierstrass-Mandelbrot fractal function by Berry and Lewis (1980). Later the Weierstrass type functions were used by many researchers as a model of rough surfaces (see e.g., Roques-Carnes et al., 1988; Majumdar and Bhushan, 1990). For some period of time, the fractal models became very popular, there were even statements that “fractals are everywhere.” Fractal approach to surface topography were so popular that one could say that it became an “emperor” of many research areas. Speaking about fractal approaches in fracture mechanics Borodich (1999) argued that instead of careful presentations of the state-of-the-art, the papers dedicated to fractal analysis are often based on repetition of common myths about fractals (we call this as vulgar fractal approaches). Hence, in most of the papers dedicated to fractal approaches to fracture and surface topography, the state-of-the-street is ruling. Borodich (2002) listed examples introduced earlier by Borodich and Onishchenko (1993) and Borodich (1993) and reminded that “the fractal dimensions alone cannot characterize the features of contact.” The same situation is related to the papers dedicated to fractal approaches to surface topography, nevertheless there is still a stream of papers based on vulgar interpretation of the fractal approaches. One could mention here a statement by Mandelbrot (1998) that “fractals are not a panacea; they are not everywhere.” Using examples introduced by Borodich and his co-authors, we will show that quite often there is no meaning in fractal analysis of surface roughness.

Nowadays another tendency is quite popular, namely to describe rough surfaces using solely the PSDFs of surface topography. We argue that these papers are in essence an attempt to resurrect the fractal approach. Indeed, these papers contain usually a mixture of correct statements related to Gaussian processes and wrong statements based on attempts to extend the power-spectral analysis to non-Gaussian surfaces. In addition, these papers suffer often by employment of the ill-defined terms, such as the Hurst exponent, non-accurate statements about self-affinity of surface roughness and vulgar interpretation of fractal models. We will show that the power-spectral analysis applied to non-Gaussian surfaces is a kind of reformulation of the vulgar fractal approach. Using an analogy to Andersen’s tale about new clothes produced by cunning weavers, we can say that attempts to model surfaces solely by the use of the PSDF of its roughness are “the emperor’s new clothes.”

The paper is organized as follows. In section 2 probabilistic characteristics of rough surfaces are discussed. In section 3 we consider some approaches to modeling of surface roughness using graphs of random processes that in turn, assume that the asperity heights are normally distributed or they employ similar assumptions that involve normal distributions. We give a brief description of statistical methods employed for checking normality of distributions and some results of application of

these tests to roughness of engineering surfaces. In section 4 we discuss briefly the fractal approaches to surface roughness, in particular using a kind of parametric-homogeneous functions having fractal graphs, e.g., the Weierstrass-Mandelbrot functions, we clarify some common misinterpretations of statements by Berry and Lewis (1980). Finally in section 5 we discuss the models of surface roughness based solely on properties of the autocorrelation function or its Fourier transform (the PSDF). We argue that these models do not reflect tribological properties of surfaces. In particular, it is demonstrated that rather different profiles may have the same inclination of the PSDF in logarithmic coordinates, or they may have even the same PSDF. Thus, there is very small scientific value (if any) of a number of papers that model surface roughness only by its PSDF.

2. PROBABILISTIC CHARACTERISTICS OF SURFACE ROUGHNESS

If one considers a nominally flat surface and a plane perpendicular to the surface, then the surface profile is the cross-section between the plane and the rough surface. The rough profile may be presented as graph of a function $z(x)$. Let \bar{z} denote the mean profile line. If the origin level of the height measurements is taken at \bar{z} then

$$\frac{1}{2L} \int_{-L}^L [z(x) - \bar{z}] dx = 0.$$

Let us mention here several popular height parameters: R_{max} is the maximum height of the profile $z(x)$ defined on an interval $[-L, L]$, R_a is the arithmetical mean deviation of the surface, and the root mean square (rms) height R_q or σ^2 that is the square root of the mean square deviation with respect to the mean profile line $\bar{z} = 0$. The mathematical expressions for these parameters are given by

$$\begin{aligned} R_{max} &= \max_{x \in [-L, L]} z(x), \\ R_a &= \frac{1}{2L} \int_{-L}^L |z(x)| dx \approx \frac{\sum_{i=1}^n |z(x_i)|}{n}, \\ R_q &= \sigma = \left[\frac{1}{2L} \int_{-L}^L [z(x)]^2 dx \right]^{1/2}, \end{aligned} \quad (3)$$

where n is the number of points of measurements on the interval and $z(x_i)$ is the measured height at the interval point x_i . In addition, the arithmetic mean height R_z is often used for practical applications. This parameter may be calculated as the average distance between the five highest picks and the five lowest points of the profile, i.e.,

$$R_z = \frac{1}{5} \left[\sum_{i=1}^5 (z_i)_{max} - \sum_{i=1}^5 (z_i)_{min} \right]. \quad (4)$$

One can introduce the density probability function $\phi(z)$. This function shows the probability that the height $z(x)$ at a surface

point x is between z and $z + dz$. Then, the expressions for R_a and σ^2 in (3) is written as

$$R_a = \int_{-\infty}^{\infty} |z| \phi(z) dz, \quad \sigma^2 = \int_{-\infty}^{\infty} z^2 \phi(z) dz. \quad (5)$$

It is natural that the roughness parameters depend on the scale of considerations. For example, if $z_i = z(x_i)$ are measured with a stylus steps h then one can calculate the curvature of the profile peaks $\kappa(x_i)$

$$\kappa(x_i) = -(z_{i-1} - 2z_i + z_{i+1})/h^2.$$

However, it was found that the mean curvature varies depending on the sampling intervals (Greenwood, 1992). There was a hope that the fractal dimension could provide a scale independent parameter of surface roughness. It will be discussed later that actually this assumption was not justified.

Kragelsky (1948) published one of the first papers where he claimed that the roughness heights distribution is Gaussian (normal). Although Zhuravlev (1940) presented the general expressions of his model for arbitrary bearing area curve, his example that employed linear dependence as an approximation of the bearing area curve was criticized by Kragelsky (1948) who wrote that the normal distribution of heights should be used.

We will not list here a number of parameters used in literature on tribology to describe the surface roughness. Various attempts to describe surface topography using several statistical parameters of surface roughness are described in detail in many books and papers, see e.g., Khusu et al. (1975), Nowicki (1985), and Whitehouse (2011). Some of these parameters are useful but most are not (Whitehouse, 1982).

An example of a very useful parameter is the Abbott-Firestone curve of the surface heights (bearing area curve) or the cumulative distribution function $\Phi(z)$

$$\Phi(z) = \int_z^{\infty} \phi(t) dt. \quad (6)$$

For example, the $\Phi(z)$ was used by Zhuravlev (1940).

Let us demonstrate that for rough surfaces their contact properties are correlated with $\Phi(z)$ that is equal to the length of a horizontal slice of the surface profile at the level h . Sometimes this curve can be used to estimate the force acting of a rough solid penetrating into an elastic foundation. Indeed, if the characteristic size of contact region between a blunt punch and a thin elastic layer is larger than the layer thickness then the leading term of the asymptotic solution may be modeled as the Fuss-Winkler foundation (see e.g., Aleksandrov, 1963; Borodich et al., 2019b; Erbaş et al., 2019; and references therein). The Fuss foundation can be represented by a collection of independent springs attached to a rigid flat or as an elastic “mattress” (Winkler, 1867; Johnson, 1985) or as a punch acting on a liquid layer (Fus, 1801). Of course, there are restrictions on the use of the Fuss-Winkler foundation (Johnson, 1985; Popov, 2010). As Fus (more often his surname is written as Fuss) noted himself: “when a crumbly surface has a rigid substrate and does not have such depth as it should be for penetration of wheels into

it by laws of hydrostatic, therefore the angle GOR" (the angle between the wheel and the surface flat) "should not be determined just by actual penetration depth but rather by the depth of the rigid support" (Fus, 1801). The Fuss-Winkler foundation was employed to model contact between an elastic foundation and (i) a nominally flat C_B profile (Borodich and Mosolov, 1991, 1992); (ii) the rough hierarchical multi-level profile (Borodich and Onishchenko, 1993, 1999), and (iii) both nominally concave and convex fractal parametric-homogeneous punches (Borodich, 1998b).

The roughness parameters used may be formally divided in the following groups (Nowicki, 1985): (i) various parameters related to the shape of asperities, e.g., the rms curvature of the asperities; (ii) various parameters related to the asperity heights; (iii) horizontal parameters; (iv) parameters connected with amplitude of the asperities and their spatial extend, like the high spot count.

3. GAUSSIAN RANDOM PROCESSES AS MODELS OF ROUGH SURFACES

As it has been mentioned above, the surface roughness may be considered as graphs of random processes. The question if the process is Gaussian (normal) or not is crucial for estimation of validity of the models. Indeed, the overwhelming majority of papers on surface topography use statistical models of surface roughness based on explicit or implicit assumption of normality of roughness heights, see e.g., Fuller and Tabor (1975), Galanov (2011), and a discussion by Borodich et al. (2016).

3.1. Tests for Normality of Surface Roughness

There are many tests of normality (Thode, 2002). Each of these tests provides a quantitative estimation of proximity between the theoretical Gaussian distribution and an observed sample of measurements by producing the so-called p -value. The estimations are based on a particular test statistic. Borodich et al. (2016, 2019a) and Pepelyshev et al. (2018) used the most popular tests of normality to check normality of roughness of various surfaces. The following tests were employed: Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Cramer-von Mises (CVM), Shapiro-Wilk (SW), Shapiro-Francia (SF), Lilliefors (LF), and the Pearson. The p -value is a number that characterizes for the observed measurements, the significance at the scale $[0, 1]$ that the normality hypothesis is true. It is possible to nominate the acceptable significance level. In our tests it was 5%. If the p -value is more than this level then the hypothesis should be accepted. This subject is a current challenging task in contact mechanics description and comprehension of nano-related phenomena as highlighted by Carpick (2018).

Let us mention here a model of dry friction developed by Borodich and Savencu (2017). In this model molecular and chemical interactions are mainly connected to the nano-scale asperities, while mechanical interlocking between surfaces are connected to the micro-scale asperities. According to this model, for proper modeling of friction, one needs to get data about the surface roughness at both atomic/nano and at micro-scales.

Using the above mentioned tests of normality some data obtained for grinding surfaces was tested and results of the normality tests were negative at both nano and microscales (Borodich et al., 2016).

Then the test of normality were applied to surfaces of the epoxy resin replicas of polishing papers of various nominal asperity sizes. A white light interferometer (Zygo NewView 6000; Zygo Corporation, Middlefield, CT, USA) at a magnification of 50 was used for characterization of the surface roughness. The normality tests showed that the height distribution of the surfaces of nominal 0.3 and 1 μm are Gaussian (Pepelyshev et al., 2018). Finally, normality of roughness of carbon-based coatings deposited by direct current pulsed magnetron sputtering at two different substrate bias voltages was checked. The same as in the case of grinding surfaces, the roughness was measured at atomic/nano scales by AFM (Atomic Force Microscopy), while a profilometer was used for measuring of micro-asperity heights (Borodich et al., 2019b). It was found that surfaces at micrometer scale are normal. It is interesting to note that the AFM measurements with the 117 nm steps showed that the roughness of surfaces was Gaussian, while the AFM measurements with the 10 nm steps showed that the assumption of normality of roughness is not satisfied. This means that the use of the above mentioned statistical models of contact between nominally flat surfaces are justified only up to nanoscale, while their applicability at atomic and few nanometers scales may be questionable.

3.2. Stochastic Models of Surface Roughness

The normality of roughness is related to vertical distribution of profile heights. However, it is also very important to consider the horizontal distribution of the asperities. As it has been noted by Maugis (2000), two profiles may have the same peak height (local extrema) distributions, but the roughness may be rather different in the horizontal extension. If process is Gaussian then the horizontal properties may be fully described by the auto-correlation (correlation) function R of the process (2). Instead of R , one can use the PSDF (3), its Fourier transform $G(\omega)$.

Khusu et al. (1975) presented results of very detailed studies of Gaussian processes in applications to surface roughness. In addition to (1), they considered three other correlation functions. An extended list of various correlation functions was presented by Dette et al. (2015). However, we would like to emphasize again that the above results are effective if the roughness is Gaussian. If the roughness is not normal then the properties of the sample paths are not fully determined by the mean and covariance functions (see e.g., Ghosal and Van der Vaart, 2017).

It is known (see e.g., Aldous, 1989 and Bibby et al., 2005) that a non-Gaussian processes can be generated by the mean-reversing stochastic differential equation

$$dX(x) = -\theta(X(x) - \mu)dx + \sigma(X(x))dW_B(x), \quad (7)$$

where $x \geq 0$ and $W_B(x)$ is the standard Brownian motion (Wiener process). Choosing the appropriate value of the parameter μ and the function $\sigma(\cdot)$, we obtain a certain height distribution of the process $X(x)$, while the auto-correlation

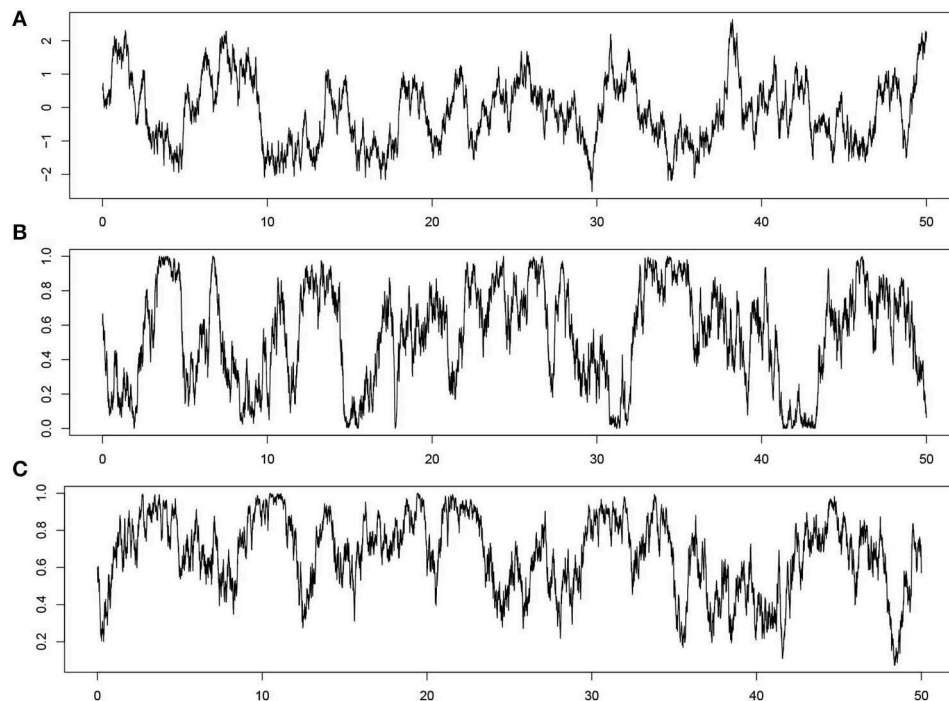


FIGURE 1 | The simulated profiles $X(x)$: **(A)** $\theta = 1$, $\mu = 0$, $\sigma(X) = 2$, and the height distribution of $X(x)$ is the Gaussian distribution; **(B)** $\theta = 1$, $\mu = 0.5$, $\sigma(X) = 2X(1 - X)$, the height distribution of $X(x)$ is the uniform distribution; **(C)** $\theta = 1$, $a = 2.2$, $b = 1.2$, $\mu = a/(a + b)$, $\sigma(X) = 2X(1 - X)/(a + b)$, the height distribution of $X(x)$ is the beta distribution with parameters a and b .

function is $\rho(x) = e^{-\theta|x|}$ and the power spectra is $G(\omega) = \frac{2}{\pi} \theta / (\theta^2 + \omega^2)$ for any choice of μ and $\sigma(\cdot)$. For example, the height distribution of $X(x)$ has the gamma density with parameters λ and β if $\mu = \beta/\lambda$ and $\sigma(X) = \sqrt{2\theta X/\lambda}$ and the beta density with parameters α and β if $\mu = \alpha/(\alpha + \beta)$ and $\sigma(X) = \sqrt{2\theta X(1 - X)/(\alpha + \beta)}$. **Figure 1** shows simulated profiles with various height distributions and the same auto-correlation function $R(x) = e^{-\theta|x|}$.

In general, using (7), one can construct non-Gaussian processes with flexible auto-correlation function as the sum

$$h(x) = \sum_{i=1}^m X_i(x),$$

where $X_1(x), \dots, X_m(x)$ are independent and each $X_i(x)$ is the solution of the stochastic differential equation

$$dX_i(x) = -\theta_i(X_i(x) - \phi_i\mu_i)dx + \sigma_i(X_i(x))dW_i(x),$$

$\phi_i > 0$ and $\sum_{i=1}^m \phi_i = 1$, see Bibby et al. (2005). Such the process $h(x)$ has the auto-correlation function and the power spectra

$$R(x) = \sum_{i=1}^m \phi_i e^{-\theta_i|x|}, \quad G(\omega) = \frac{2}{\pi} \sum_{i=1}^m \frac{\phi_i \theta_i}{\theta_i^2 + \omega^2}.$$

The above examples support the statement that the mean and covariance functions of a random non-Gaussian process do not determine the finite-dimensional distributions of a random process (Gusak et al., 2012).

4. FRACTALS APPROACHES TO SURFACE TOPOGRAPHY

Concepts of fractal, fractal geometry and fractal dimension (FD) were introduced by Mandelbrot (1975). However, he did not give any definition of fractals. Later he defined fractal sets in a metric space saying “a fractal will be defined as a set for which the Hausdorff-Besicovitch dimension strictly exceeds the topological dimension” (Mandelbrot, 1977). Finally he withdrew the definition and suggested to use the term “without a pedantic definition” (Mandelbrot, 1983). This was the reason for a comment by Greenwood (1992): “Mandelbrot is somewhat reluctant to define ‘fractals’ or ‘fractal dimension’ preferring to offer examples.” A popular description of many self-similar sets studied by fractal geometry, such as the Cantor staircase, von Koch snowflake, Sierpinski carpet, Menger sponge, along with Peano curve and other sets may be found in the book for school children published by Vilenkin (1968). Definitely Mandelbrot (1975) wrote his book under influence of Vilenkin’s book. Indeed, at least 20% of Figures presented by Mandelbrot (1975) had analogs in the book by Vilenkin (1968). The above mentioned sets are out the scope of mathematical programmes for schools and for non-mathematical specializations of universities because the sets are more irregular than sets considered in common courses on Euclidean geometry. To encourage children for studies of set theory, Vilenkin used many funny stories and terms, e.g., he used the term the Devil staircase to describe the Cantor

staircase. The Devil staircase term was later used and popularized by Mandelbrot (1983).

Because there is no generally accepted definition, we will define fractals as objects with a non-integer FD. Evidently, the term FD must be defined separately. One of the authors (FB) and his co-authors have already published several reviews related to fractals, in particular a review related to the use of fractal concepts in fracture mechanics (Borodich, 1999). There are reviews of application of fractal ideas in contact problems (Borodich and Onishchenko, 1999; Borodich, 2013c) and several articles about the use of fractal concepts in tribology (Borodich, 2013a,b; Borodich and Evans, 2013). Borodich introduced the C_B profile that is also known as the Cantor set model (Warren and Krajcinovic, 1996), and as the Cantor-Borodich profile, structure or fractal, see e.g., Abuzeid and Eberhard (2007), Soldatenkov (2015), and Thielen et al. (2016). At least two kind of contact problems can be solved for a punch described by the C_B profile (Borodich and Mosolov, 1991, 1992). The model was developed further by Borodich and Onishchenko (1993), Warren and Krajcinovic (1996), Plesha and Ni (2001), and others. Although Archard (1957) introduced the idea of hierarchical structure of roughness, there were no papers developing the idea of “bump on bump” structure of roughness until Borodich and Onishchenko (1993, 1999) introduced the multilevel hierarchical model of roughness. Now this idea of hierarchy of surface structures is quite popular due to discoveries of Gorb and his co-workers along with other experts in biological objects. In particular, the hierarchical structures were discovered in adhesive setae of geckos (Gao et al., 2005) and spiders (Schaber et al., 2019); water-repellent coatings of whip-spiders (Wolff et al., 2016); attachment discs of spiders (Wolff et al., 2015), and super-black snake skin (Spinner et al., 2013). Applications of fractal concepts to surface roughness were also discussed by Borodich and Galanov (2002) and by Borodich et al. (2016). Thus, here we will just remind some basic features of the fractal approach to surface roughness.

4.1. Mathematical and Physical Fractals

Unfortunately, many papers dedicated to applications of fractals do not provide definitions of used terminology. Often such papers are just a mess of vague discussions and non-justified statements. To clarify the subject, we split fractals into mathematical and physical (natural) fractals (see e.g., Borodich, 1997). Both mathematical and physical fractals use the concepts of a cover. This means that the object (set) is covered by cubes of size at most or equal to δ . Fractal geometry deals with mathematical fractals. The mathematical methods of fractal geometry are described in many books and papers (see e.g., Falconer, 1990; Tricot, 1995) where various FD are studied in application to mathematical objects. Various FDs are used in the studies, mainly the Hausdorff dimension and box-counting dimensions usually attributed to Minkowski, Bouligand, Pontrjagin, Schnirelman, and Kolmogorov. These FDs can be calculated by taking the limits when $\delta \rightarrow 0$. In addition, fractal geometry term is also quite often applied loosely to a collection of semi-empirical or empirical methods for estimations of the FDs of objects.

If real world or numerically simulated objects demonstrate the power-law of the number-radius relation then these objects are physical fractals. The power-law of the number-radius relation is

$$N(\delta) \sim \delta^{-D}, \quad \delta_* \leq \delta \leq \Delta_*, \quad N(R) \sim (R/\delta)^D, \quad r_* \leq R \leq R_* \quad (8)$$

where $N(\delta)$ is the number of cubes of size δ used to cover the object, D is a FD of the object, Δ_* and δ_* are the upper and lower cut-offs of the physical fractal law, respectively. The former relation of (8) is used when the cover size δ is varied and the object size R is fixed, while in the latter relation is used when the cover size δ is fixed and the object size R is varied. In this latter case R_* and r_* are the upper and lower cut-offs. The slope of linear approximation of $\ln(N(\delta_*))$ against $\ln(R)$ is used to estimate the value of D . The main distinction between these kinds of fractals is the following: the physical fractal behavior (8) is observed on a bounded region of scales only, while to study mathematical fractals one has to scale of consideration to zero limit.

4.1.1. Self-Similar Sets

Introduction to fractals starts often by presenting self-similar sets mentioned above (see examples given by Vilenkin (1968) and Mandelbrot, 1975). Due to such examples, there is a myth that all fractals are self-similar. However, self-similar sets are a very specific kind fractals. In general, self-similarity is not related to mathematical fractals. Their scaling properties are based on scaling of fractal measure or quasi-measure (see a discussion by Borodich and Feng, 2010, and Borodich, 2019), while for physical fractals their scaling properties are reflected by the relation (8).

4.1.2. Some Specific Features of Mathematical Fractals

Speaking about mathematical fractals, one has to specify the FD used. It is wrong to say that all FDs are equal to each other. For example, the Hausdorff dimension $\dim_H S$ of a set S may be not equal to its box-counting dimension $\dim_B S$. However it is known that $\dim_H S \leq \dim_B S$.

The statement that all nowhere differentiable curves are fractals is wrong. For example, Borodich (1998a) introduced the Moscow University function U_M that is nowhere differentiable, while it is a non-fractal curve and each finite subinterval of $[1, 2]$ contains an infinite number of reduced copies of the whole function, i.e., it is a self-similar curve.

A mathematical fractal curve has an infinite length. Borodich (1997) formulated the following paradox in application to fractal fracture: if the crack surface is modeled as a mathematical fractal and the concept of the Griffith surface energy is used, then the propagation of the fractal crack is impossible.

If a surface is smooth then it is not a mathematical fractal (see e.g., Falconer, 1990). Even if a mathematical fractal curve is continuous everywhere, it is nowherehas differentiable. Therefore, it is often very difficult to formulate a boundary-value problem for solids having fractal boundary (see a discussion by Borodich and Volovikov, 2000). Indeed, such nowhere differentiable profile do not have even normal vector.

4.1.3. Some Specific Features of Physical Fractals

If the FD is specified then it is convenient to use the fractional part of the FD D^* . Then the FDs of fractal profiles and surfaces are $1 + D^*$ and $2 + D^*$, respectively.

Sayles and Thomas (1978) presented the experimental relations between normalized PSDF and wavelength in logarithmic coordinates as a very impressive united line for many different surfaces. One could think that the relation spans from micrometers to many meters. However, Berry and Hannay (1978) noted that for each individual type of topography the span was over much smaller ranges. They argued that these experimental results could be represented as a united line $\lg(G) \sim \lg(\omega)$ due to a specific normalization of the results. Further, researchers from Jerusalem presented results of an analysis of data published by Physical reviews journals and showed that in average the physical fractals span about 1.5 orders of magnitude (Avnir et al., 1998). Mandelbrot (1998) argued that the published limited-range examples of power law correlations may be explained as *unfortunate side effects of enthusiasm, imperfectly controlled by refereeing*. However, the Jerusalem group disagreed with his arguments and believed that the limited-range relations for natural fractal are the dominant fractals observed in Nature (Biham et al., 1998). However, if the value of the FD is stable for less than two or three decades then fractal concept is not useful, as it was stated by Whitehouse (2001).

In general, the FD values obtained by various practical methods are not reliable (see a discussion by Borodich and Evans, 2013). For example, the power spectrum method is often employed for estimation of FD values of self-similar or self-affine fractals, in particular rough profiles (see e.g., Dubuc et al., 1989; Schmittbuhl et al., 1995). The method is based on a not very accurate statement *when the power spectrum obeys a power law $G(\omega) \sim c/\omega^\alpha$... it is reasonable to expect a signal with a $\omega^{-\alpha}$ power spectrum to have a graph of dimension $(5 - \alpha)/2$* (see Falconer, 1990). Normally the exponent α is in the range $1 < \alpha \leq 3$.

As Whitehouse (2001) noted there is a very small spread of the FD values obtained for surfaces produced by different manufacturing processes. In addition, there is no well-established algorithm for estimations of the fractal law cut-offs. Thus, the physical significance of the fractal approach is very limited. We can add that if one can attribute the fractal scaling for a small range that spans for just 1.5 or 2 decades then fractals do not provide a scale-independent description of surface roughness.

4.2. Parametric-Homogeneity

One of the authors (FB) and his co-authors introduced different types of fractal profiles that allowed to handle rough fractal surfaces rigorously. There were introduced : (i) the C_B profile (Borodich and Mosolov, 1991, 1992), (ii) multilevel Hierarchical profile (Borodich and Onishchenko, 1993, 1999), and (iii) graphs of parametric-homogeneous functions (Borodich, 1993, 1998a,b). The first two profiles were based on an iterative procedure and they has been already mentioned above, while PH-functions will be described below. Note that the general statements obtained by Borodich and his co-authors about the contact problems for all these three types of surfaces do not

depend on the statements about their FDs, i.e., they are valid for both fractal and non-fractal cases.

The same author (FB) introduced also the concept of parametric-homogeneity. The concept includes parametric quasi-homogeneous (PQH) transformations (in particular parametric-homogeneous (PH) transformations), corresponding functions, and PH- and PQH-sets. The PH and PQH-functions can be considered as a natural generalization of concepts of homogeneous and quasi-homogeneous functions. If the latter kind of functions is based on the use of the classical scaling with arbitrary positive scaling factor λ , PH-functions and PQH-functions are based on the use of discrete self-similarity with a fixed rescaling parameter p . Let \mathbb{Z} be the set of integer numbers. Then one can employ the discrete group of coordinate dilations $(\Gamma_{p^{ak}})$ apply the following PH-transformation

$$\Gamma_{p^{ak}} \mathbf{x} = (p^{k\alpha_1} x_1, \dots, p^{k\alpha_n} x_n), \quad p > 0, \quad k \in \mathbb{Z}.$$

4.2.1. Parametric Quasi-Homogeneous Functions

The parametric-quasi-homogeneous function of degree d and parameter p with weights $\alpha = (\alpha_1, \dots, \alpha_l)$ is denoted by $B_d: \mathbb{R}^l \rightarrow \mathbb{R}$ and it obeys the following scaling law for a positive parameter p , $p \neq 1$ and any $k \in \mathbb{Z}$

$$B_d(\Gamma_{p^{ak}} \mathbf{x}; p) = B_d(p^{k\alpha_1} x_1, \dots, p^{k\alpha_l} x_l; p) = p^{kd} B_d(\mathbf{x}; p),$$

The parameter should be unique in some domain (Borodich, 1998a,b). If $\alpha_1 = \dots = \alpha_l$ then PQH-functions is called a PH-functions. It is clear that if p is a parameter of the PH-transformation then p^k can be also used as a parameter. Hence, as the parameter, the least $p: p > 1$ is taken. The graphs of these functions can have various properties, in particular they can be smooth or fractal. The following fractal Weierstrass type functions and smooth sinusoidal log-periodic functions are examples of PH-functions

$$b_\beta(x, p) = \sum_{n=-\infty}^{\infty} p^{-\beta n} f(p^n x) \\ \text{or } b_0(x; p) = A \cos(2\pi \ln x / \ln p + \Phi),$$

where A and Φ are arbitrary constants and f is an arbitrary function. For any point x_0 , the PH-functions are repeated $b_d(p^k x_0; p) = p^{kd} b_d(x_0; p)$ in scaling form near all points $p^k x_0$, $k \in \mathbb{Z}$.

4.2.2. Weierstrass-Mandelbrot Functions

Mandelbrot (1977) generalized the Weierstrass function, whose graph is continuous everywhere and differentiable nowhere, and introduced the complex value Weierstrass-Mandelbrot (W-M) function $W(x)$ and its particular real case $C(x; p)$

$$W(x; p) = \sum_{n=-\infty}^{\infty} p^{-\beta n} (1 - e^{ip^n x}) e^{i\phi_n}, \\ C(x; p) = \sum_{n=-\infty}^{\infty} p^{-\beta n} (1 - \cos(p^n x)), \quad p > 1, 0 < \beta < 1 \quad (9)$$

where ϕ_n are arbitrary phases. Properties of these functions were studied in detail by Berry and Lewis (1980). The box dimension of $C(x; p)$ graphs is $D = 2 - \beta$. There is no rigorous mathematical proof that its Hausdorff dimension is the same.

It is evident that $C(x; p)$ is a particular PH-function (9) of degree β . As it has been mentioned above, the graph of $C(x; p)$ function was used often for modeling rough profiles.

In addition, Berry and Lewis (1980) were first who studied the discrete power spectrum $G(\omega)$ of the W graphs (9) and approximated it by a continuous spectrum $G(\omega) \sim \omega^{-(5-2D)}$. This is in agreement with the above mentioned statement by Falconer (1990).

4.2.3. Self-Affine Scaling and the Hurst Exponent Terms

Many papers dedicated to studies of surface roughness mention the concept of self-affine fractals without any definition of the concept. However, one should be very careful using such terms as self-affine scaling. Actually, it is known from mathematical courses that mapping on a plane is self-affine if it is a one-to-one mapping whose images of any three collinear points in turn belong to a line. Under this mapping the angle between two straight lines is normally not preserved. A particular case of such mapping is the quasi-homogeneous (QH) or anisotropic coordinate dilation

$$x' = e^{\alpha_1} x, \quad y' = e^{\alpha_2} y \quad (10)$$

where α_1 and α_2 are weights, $\lambda_x = e^{\alpha_1}$ and $\lambda_y = e^{\alpha_2}$ are scaling factors. This means that if a coordinate system x, y was Cartesian one then the x', y' system is also rectangular. However, papers on fractals reduce usually the term self-affine mapping to the above particular QH case.

In 1980 the concept of fractal was very novel. Perhaps for popularization of the concept, they used some explanations and statements that were not mathematically rigorous. For example, studying W-M functions, they wrote *The fractal nature of W implies that repeated magnification of its graph reveals ever-finer levels of detail.... The levels of detail are self-similar under an affine scaling in which the x axis is stretched by a factor p and W axis by p^{2-D}* [note we have changed γ and t used by Berry and Lewis (1980) to p and x , respectively]. Unfortunately, these statements of Berry and Lewis (1980) were misinterpreted and used in many papers as generally accepted and mathematically rigorous statements applicable to all so-called self-affine fractals.

There was an attempt by Mandelbrot (1986) to develop further the concept of self-affinity. He formulated several statements about self-affine fractals that were not supported by any strict definition of these fractals. In essence, his statements are: there are both local and global scaling for self-affine fractals and each version of FD for self-affine fractals has a local and a global value, separated by a crossover. However, these statements have no mathematical meaning.

Schmittbuhl et al. (1995) wrote that the scaling (10) means that the heights h are homogeneous functions of the scaling factors (putting $\lambda_x = \lambda_y$)

$$h(\lambda_x x, \lambda_x y) = \lambda_x^H h(x, y)$$

where the homogeneity exponent H is the self-affine exponent, or the Hurst exponent. Unfortunately, the statement by Schmittbuhl et al. (1995) about the homogeneity of the fractal graphs is not very accurate, while the meaning of the Hurst exponent is ill-defined. It would be better if they would say that the fractal roughness to obey a discrete dilation homogeneity, i.e., it obeys the PH-law. Indeed, the scaling properties of $C(x; p)$ function are connected with the discrete group of coordinate dilation

$$C(p^k x; p) = p^{(2-D)k} C(x; p). \quad (11)$$

Although this scaling is quite often attributed to FD of the $C(x; p)$ graph, saying that $H = 1 - D^*$, these statements are wrong because the scaling properties are governed by the degrees d of PH-functions. (Borodich, 1998a,b) showed that he can construct PH-functions with prescribed global trend $h \sim x^d$ and given value of FD, while the Hausdorff and box-counting FDs of the Weierstrass type functions preserve their values under action of quasi-homogeneous transformation. For example, taking constants A and $0 < \epsilon < 1$ one can construct PH-functions $b_\beta(x, p) = C(x; p)$, $b_0(x; p)$, $b_1(x; p)$ and $b_2(x; p)$

$$\begin{aligned} b_0(x; p) &= x^{-\beta} C(x; p), & b_1(x; p) &= Ax [1 + \epsilon b_0(x; p)], \\ b_2(x; p) &= Ax^2 [1 + \epsilon b_0(x; p)] \end{aligned} \quad (12)$$

have the same values of the Hausdorff and box dimensions and absolutely different trends. On the other hand, if one uses the above self-affine terminology then $b_\beta(x, p)$ and $b_2(x; p)$ are self-affine fractals, while $b_1(x; p)$ is self-similar fractal because

$$b_1(p^k x; p) = p^k b_1(x; p).$$

It follows from the above discussion that Hurst exponent is an ill-defined term and it is not connected to FDs of the PH-graphs.

A statement by Mandelbrot (1983) is often cited. Mandelbrot (1983) wrote that coastlines are not circles, clouds are not spheres, and mountains are not cones, assuming that they have to be modeled as fractals. Borodich and Onishchenko (1999) extended Mandelbrot's statement saying that roughness of real bodies is not a mathematical fractal and all these geometrical objects: spheres, cones, circles as well as fractals are only mathematical idealizations of complex shapes of natural objects. Mathematical and physical fractals should not be confused. Certainly one can employ mathematical fractal as a possible model that reflect the power-law number-radius relation or PSDF of a natural object within a bounded interval of scales. However, the obtained problem may be a very complicated.

5. POWER SPECTRAL DENSITY FUNCTION APPROACHES TO ROUGH SURFACES

The last 20 years or so the PSDF approach became very popular in tribology community (see e.g., Persson, 2006). In fact, the authors of the papers reduce the properties of rough surfaces just to their PSDF $G(\omega)$. This approach was criticized by Borodich (2002). He wrote: *two punches having the same fractal surface but*

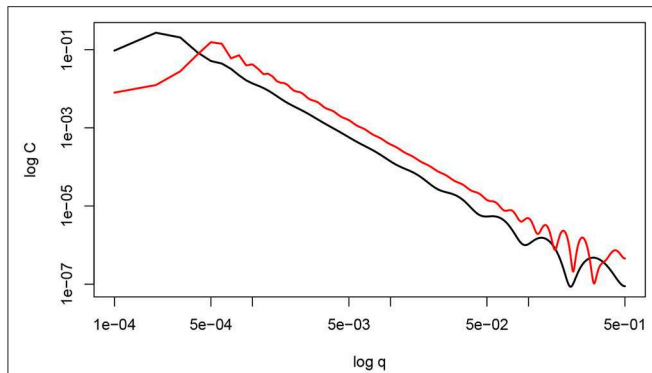


FIGURE 2 | The power spectra of the profile $h(x) = \sin(2\pi \ln(x)/\ln(p))$ for $x \in [0.0001, 1]$ with $p = 2$ (black) and $p = 1.3$ (red).

situated either above or below the surface show usually different asymptotics in both load-displacement and load-area relations. As an example he considered C_B profile and solutions presented by Borodich and Onishchenko (1993). Then Borodich et al. (2016) showed that any surface and its replica have the same power spectral density, therefore one cannot characterize contact properties of rough surfaces just by surface “spectrum.” Indeed, one can consider an absolutely flat smooth surface having many sharp dents and its complementary replica (an inverted replica): a surface having many sharp asperities of the surface roughness. Evidently, they gave absolutely different contact properties (see e.g., examples by Borodich and Onishchenko, 1993).

Moreover quite often the PSDF approach is reduced to the so-called self-affine surfaces by employment of the Hurst exponent term. As we have discussed above the term is ill-defined and such papers just create “new clothes” for the vulgar fractal approach.

Using examples, we will show below that the many common statements about rough non-Gaussian surfaces are wrong or just meaningless.

5.1. Smooth Functions May Have a Power-Law PSDF

Let us show that connections of FDs and the slopes of PSDFs in logarithmic coordinates are meaningless. As an example, consider a very simple and smooth PH-function, namely a sin log-periodic function. **Figure 2** shows the power spectra for $h(x) = \sin(2\pi \ln(x)/\ln(p))$ with $x \in [0.0001, 1]$.

It is clear that the function is non-fractal, nevertheless its power spectra in logarithmic coordinates located along a straight line within a bounded range of scales.

5.2. PSDFs Are Almost the Same for Different Truncated Weierstrass-Mandelbrot Functions

Let us consider two truncated Weierstrass-Mandelbrot functions

$$C_t(x; p) = \sum_{n=-N}^N p^{-(2-D)n} (1 - \cos(p^n x)),$$

$$A(x; p) = \sum_{n=-N}^N (-1)^n p^{-(2-D)n} \sin(p^n x),$$

where N is large natural number, $p > 1$ and $D \in (1, 2)$. In addition, let us consider a function $x^{D-2} C_t(x; p)$. If $N \rightarrow \infty$ then $C_t(x; p) \rightarrow C(x; p)$ and D is their box-dimension, however they have different trends.

Let us fix the values $p = 1.5$ and $D = 1.5$ and vary N . Let us calculate numerically the PSDFs for obtained graphs. The results are presented in logarithmic coordinates in **Figure 3**.

One can see that the average PSDFs are approximately the same while the functions are rather different.

Now let us fix the values $p = 1.5$, and $N = 100$ and vary D and calculate numerically the PSDFs for obtained functions. The results are presented in logarithmic coordinates in **Figure 4**.

Note that for $N = 100$, the graphs of $C_t(x; p)$ and $x^{D-2} C_t(x; p)$ are very close to graphs of PH-functions of degrees $d = 2 - D$ and $d = 0$, respectively. We know that these functions have the same FD D and different trends (see a discussion above), while their PSDFs are very similar to each other.

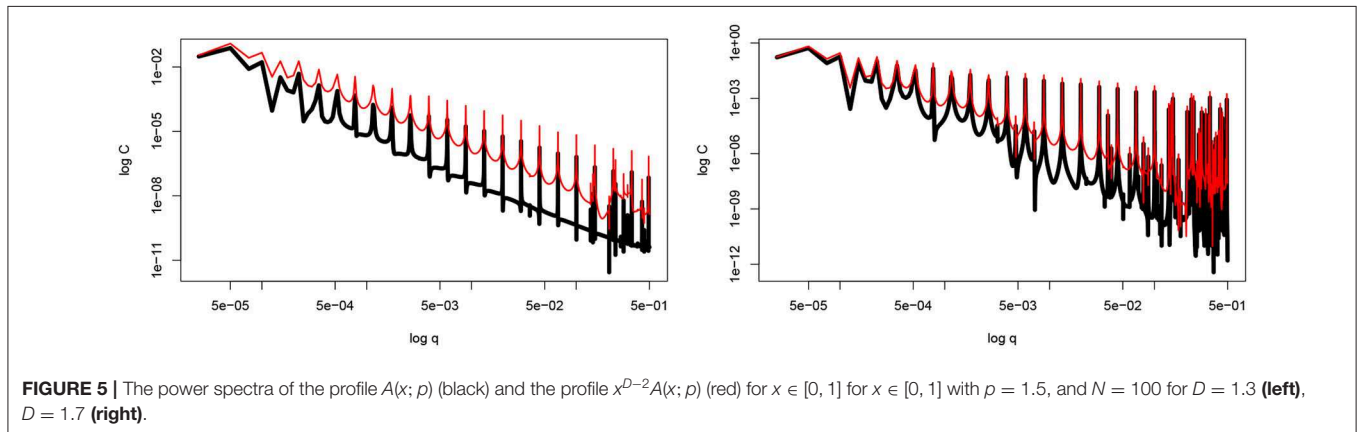
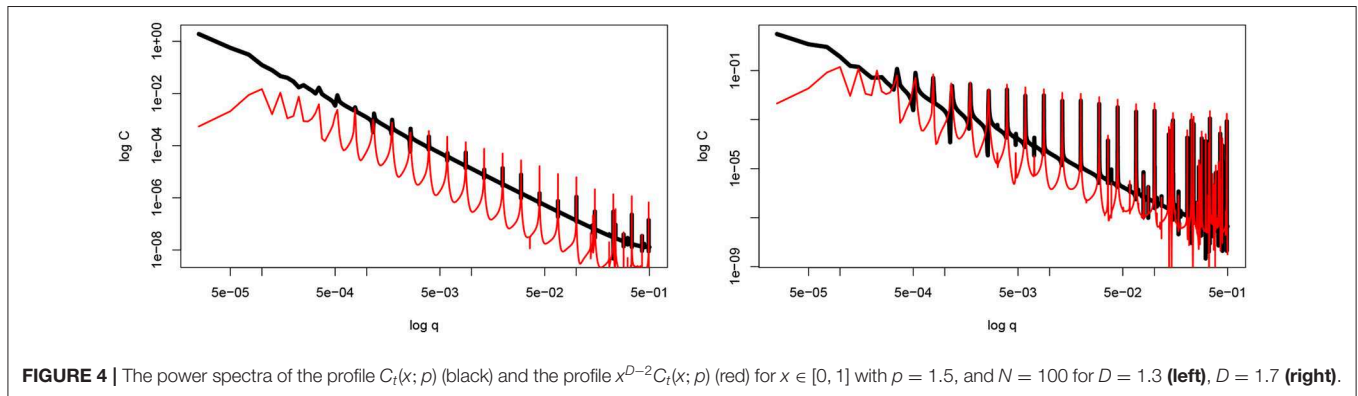
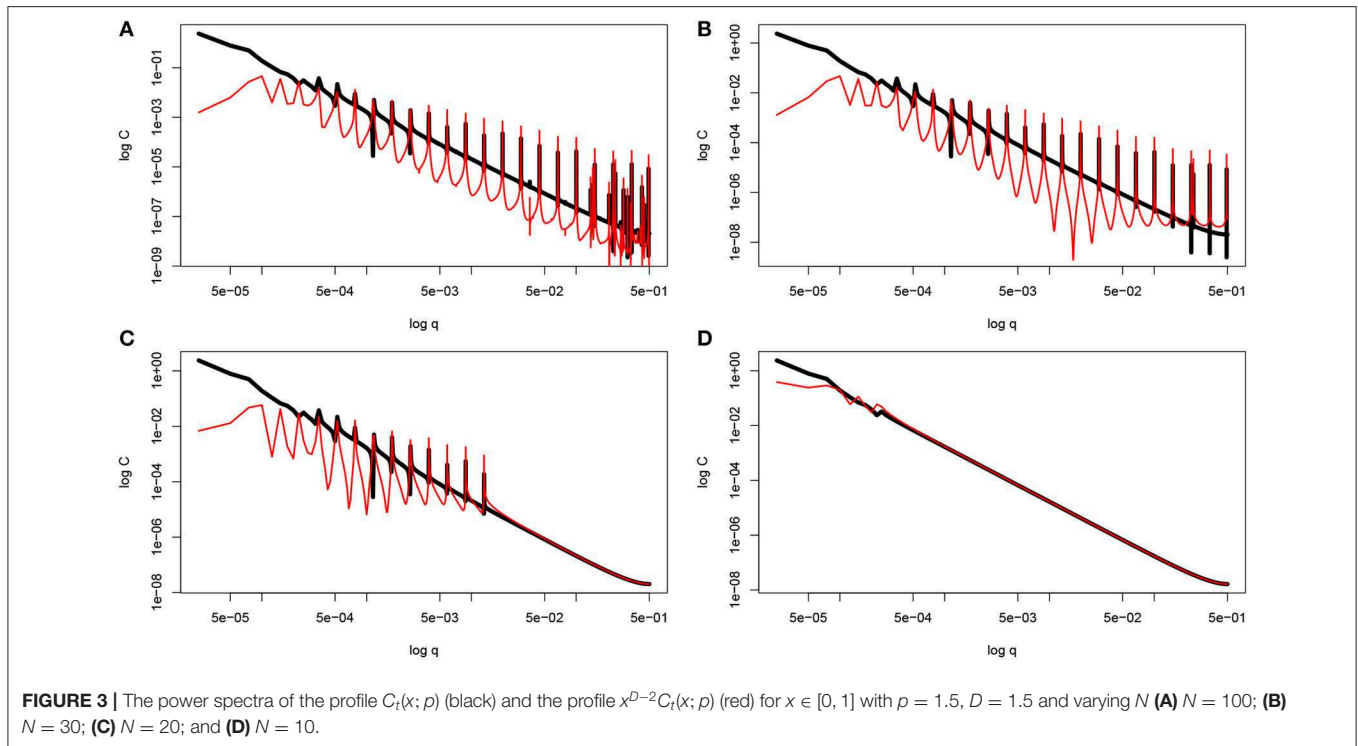
The same procedure can be applied to the profiles $A(x; p)$ and $x^{D-2} A(x; p)$. The results are shown in **Figure 5**. The results lead to the same conclusion as above.

5.3. PSDF Does Not Characterize Tribological Properties of Polished Surfaces

It is known that the tribological properties of gears can be greatly extended by improved surface finish (Krantz et al., 2001). This fact was employed by Harris et al. in a series of papers dedicated to the use of modern hard carbon-based coatings. Indeed, these hard carbon-based films can remove protuberances on the steel counterparts, reducing the values of high intensity stresses and the total number the stress concentration points. On the other hand, due to polishing of the coating asperities, the rate at which the coatings abrade steel declines as a power-law of the cycle numbers, i.e., very rapidly. The coated surface become super smooth even after just 500 cycles (see SEM images by Borodich et al., 2003). Let us consider the following thought experiment. An intact surface is modeled by a Gaussian function $f_1(x)$ shown in **Figure 1A** (this assumption is in accordance with roughness measurements of intact surfaces by Pepelyshev et al., 2018 and Borodich et al., 2019a). Then the asperities of the surface have been polished away and the roughness of the polished parts are described by $0.1f_2(x)$ where the profile $f_2(x)$ is shown in **Figure 1B**. Hence, the full polished surface is described by $f_3(x) = \min[f_1(x), 0.1f_2(x)]$ shown in **Figure 6**.

Figure 7 shows the power spectra for the profile $f_1(x)$ (black), the profile $f_2(x)$ (red), and the “polished” profile $f_3(x) = \min[f_1(x), 0.1f_2(x)]$ (blue).

These three profiles have the different outlook and tribological properties, however have the same exponent 1.9 in the power spectra, i.e., $G(\omega) \sim 1/\omega^{1.9}$.



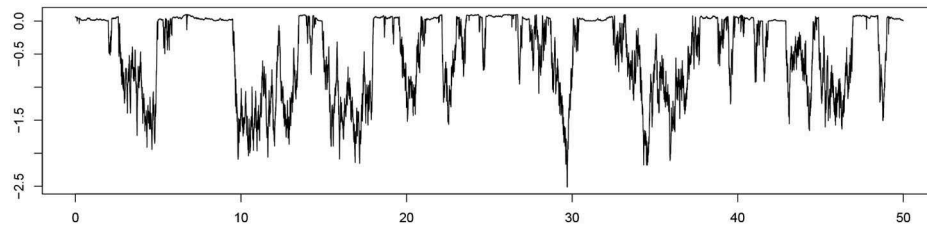


FIGURE 6 | The “polished” profile $f_3(x) = \min[f_1(x), 0.1f_2(x)]$, where the profile $f_1(x)$ is shown in **Figure 1A** and the profile $f_2(x)$ is shown in **Figure 1B**.

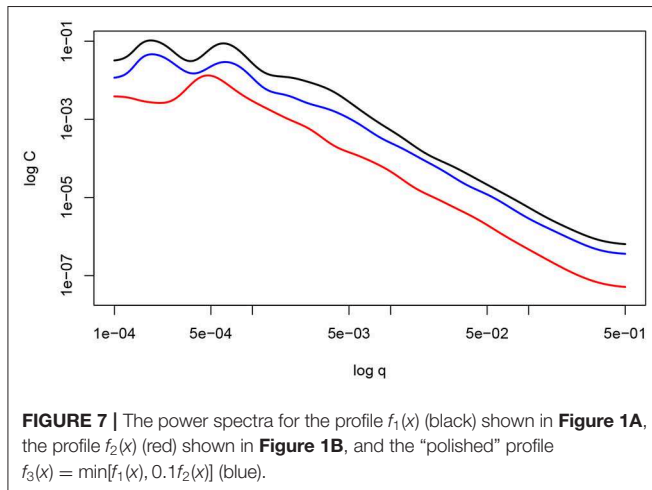


FIGURE 7 | The power spectra for the profile $f_1(x)$ (black) shown in **Figure 1A**, the profile $f_2(x)$ (red) shown in **Figure 1B**, and the “polished” profile $f_3(x) = \min[f_1(x), 0.1f_2(x)]$ (blue).

6. CONCLUSIONS

A review of mathematical approaches for description of topography of engineering surfaces has been presented. It has been shown that in spite of huge amount of parameters used to characterize the surface topography, only some of the parameters are quite useful. However, their use is rather bounded, e.g., these parameters may be helpful at meso or even microscales, however they could be useless at the nanoscale.

There are many models of random processes, however just the case of Gaussian processes is well-elaborated. Some statistical methods employed for checking normality of distributions of asperity heights are reviewed. The preliminary results showed that the intact surfaces are quite often Gaussian at both micro and nanoscales, while the grinding surfaces are not normal.

Then the fractal approaches to surface roughness are discussed. Some common misinterpretations of statements by Berry and Lewis (1980) are also discussed. Some disadvantages of the fractal approaches and commonly reported wrong statements about fractals are listed. It is argued that the practical usefulness

of the fractal approaches is rather doubtful. One should not expect that the employment of a mathematical fractal model of a rough surface will provide considerable advantages. In fact, such models may cause many mathematical difficulties. Thus, a strict approach to fractal modeling may substitute a difficult problem to another more difficult than the original one. Further, dimensions of physical fractals cannot be used as scale independent parameters. One should provide proper explanations of the fractal concepts used, otherwise this could lead to misinterpretation of the results. Unfortunately, quite often the answer to the question by Jelinek et al. (1998): *Is there meaning in fractal analyses?*, is “No.”

Finally the models of surface roughness based solely on properties of the auto-correlation function or its Fourier transform (the PSDF) are discussed. It has been noted that the PSDF approach to non-Gaussian surfaces are in essence the new clothes for the vulgar fractal approaches. We argue that these PSDF models do not reflect tribological properties of surfaces. In particular, it is demonstrated that rough profiles may have the same slopes of the PSDF in logarithmic coordinates, while they have rather different tribological properties. Thus, there is a bounded scientific value (if any) of a number of papers based on the PSDF approach.

AUTHOR CONTRIBUTIONS

FB: substantial contributions to the conception of the work, analysis, and interpretation of published data for the work. XJ: analysis and interpretation of published data for the work. AP: statistical analysis of the results and numerical simulations for the work. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors were grateful to support of Chongqing University enabled one of the authors (FB) to visit the University in 2019. One of the authors (XJ) would like to acknowledge the support from the National Science Foundation of China (Grant Nos. 11932004 and 51875059).

REFERENCES

- Abbott, E. J., and Firestone, F. A. (1933). Specifying surface quality: a method based on accurate measurement and comparison. *Mech. Eng.* 55, 569–572.
- Abuzeid, O. M., and Eberhard, P. (2007). Linear viscoelastic creep model for the contact of nominal flat surfaces based on fractal geometry: standard linear solid (SLS) material. *J. Tribol.* 129, 461–466. doi: 10.1115/1.2736427
- Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. New York, NY: Springer-Verlag.
- Aleksandrov, V. M. (1963). On the solution of certain contact problems of the theory of elasticity. *PMM J. Appl. Math. Mech.* 27, 1490–1494.
- Archard, J. F. (1957). Elastic deformation and the laws of friction. *Proc. R. Soc. A* 243, 190–205.
- Archard, J. F., Hunt R. T., and Onions, R. A. (1975). “Stylus profilometry and the analysis of the contact of rough surfaces,” in *Proceedings of the IUTAM Symposium*, eds A. D. de Pater and J. J. Kalker (Delft: Delft University Press), 282–303.
- Avnir, D., Biham, O., Lidar, D., and Malcai, O. (1998). Is the geometry of nature fractal? *Science*. 279, 39–40.
- Berry, M. V., and Hannay, J. H. (1978). Topography of random surfaces. *Nature* 273:573.
- Berry, M. V., and Lewis, Z. V. (1980). On the Weierstrass-Mandelbrot fractal function. *Proc. R. Soc. A* 370, 459–484.
- Bibby, B. M., Skovgaard, I. M., Sorensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli* 11, 281–299. doi: 10.3150/bj/1116340291
- Biham, O., Malcai, O., Lidar, D. A., and Avnir, D. (1998). Fractality in nature: response. *Science*. 279, 785–786.
- Borodich, F. M. (1993). Similarity properties of discrete contact between a fractal punch and an elastic medium. *C. R. Ac. Sc. (Paris) Ser.* 2 316, 281–286.
- Borodich, F. M. (1997). Some fractal models of fracture. *J. Mech. Phys. Solids* 45, 239–259.
- Borodich, F. M. (1998a). Parametric homogeneity and non-classical self-similarity. I. Mathematical background. *Acta Mech.* 131, 27–45.
- Borodich, F. M. (1998b). Parametric homogeneity and non-classical self-similarity. II. Some applications. *Acta Mech.* 131, 47–67.
- Borodich, F. M. (1999). Fractals and fractal scaling in fracture mechanics. *Int. J. Fracture* 95, 239–259.
- Borodich, F. M. (2002). Comment on “Elastoplastic contact between randomly rough surfaces.” *Phys. Rev. Lett.* 88:069601. doi: 10.1103/PhysRevLett.88.069601
- Borodich, F. M. (2007). Translation of historical paper. Introduction to V. A. Zhuravlev’s historical paper: ‘on the question of theoretical justification of the Amontons-Coulomb law for friction of unlubricated surfaces’. *Proc. Inst. Mech. Eng. J. J. Eng. Tribol.* 221, 893–894.
- Borodich, F. M. (2013a). “Fractal geometry,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (New York, NY: Springer), 1258–1264.
- Borodich, F. M. (2013b). “Fractal nature of surfaces,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (New York, NY: Springer), 1264–1269.
- Borodich, F. M. (2013c). “Fractal contact mechanics,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (New York, NY: Springer), 1249–1258.
- Borodich, F. M. (2019). Development of Barenblatt’s scaling approaches in solid mechanics and nanomechanics. *Phys. Mesomech.* 22, 73–82. doi: 10.1134/S1029959919010119
- Borodich, F. M., and Bianchi, D. (2013). “Surface synthesis based on surface statistics,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (New York, NY: Springer), 3472–3478.
- Borodich, F. M., Brousseau, E., Clarke, A., Pepelyshev, A., and Sanchez-Lopez, J. C. (2019a). Roughness of deposited carbon-based coatings and its statistical characteristics at nano and microscales. *Front. Mech. Eng.* 5, 73–82. doi: 10.3389/fmech.2019.00024
- Borodich, F. M., and Evans, H. P. (2013). “Fractal characterization of surfaces,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (New York, NY: Springer), 1246–1249.
- Borodich, F. M., and Feng, Z. (2010). Scaling of mathematical fractals and box-counting quasi-measure. *ZAMP* 61, 21–31. doi: 10.1007/s00033-009-0010-6
- Borodich, F. M., and Galanov, B. A. (2002). Self-similar problems of elastic contact for non-convex punches. *JMPS* 50, 2441–2461. doi: 10.1016/S0022-5096(02)00031-5
- Borodich, F. M., Galanov, B. A., Perepelkin, N. V., and Prikazchikov, D. A. (2019b). Adhesive contact problems for a thin elastic layer: Asymptotic analysis and the JKR theory. *Math. Mech. Solids* 24, 1405–1424. doi: 10.1177/1081286518797378
- Borodich, F. M., Keer, L. M., and Harris, S. J. (2003). Self-similarity in abrasiveness of hard carbon-containing coatings. *J. Tribol. Trans. ASME* 125, 1–7. doi: 10.1115/1.1509773
- Borodich, F. M., and Mosolov, A. B. (1991). Fractal contact of solids. *Sov. Phys. Tech. Phys.* 36, 995–997.
- Borodich, F. M., and Mosolov, A. B. (1992). Fractal roughness in contact problems. *PMM J. Appl. Math. Mech.* 56, 681–690.
- Borodich, F. M., and Onishchenko, D. A. (1993). Fractal roughness for problem of contact and friction (the simplest models). *J. Friction Wear* 14, 452–459.
- Borodich, F. M., and Onishchenko, D. A. (1999). Similarity and fractality in the modelling of roughness by a multilevel profile with hierarchical structure. *Int. J. Solids Struct.* 36, 2585–2612.
- Borodich, F. M., Pepelyshev, A., and Savencu, O. (2016). Statistical approaches to description of rough engineering surfaces at nano and microscales. *Tribol. Int.* 103, 197–207. doi: 10.1016/j.triboint.2016.06.043
- Borodich, F. M., and Savencu, O. (2017). “Hierarchical models of engineering rough surfaces and bioinspired adhesives,” in *Bio-Inspired Structured Adhesives*, eds L. Heepe, S. Gorb, and L. Xue (Springer), 179–219. doi: 10.1007/978-3-319-59114-8_10. Available online at: <https://www.springer.com/gp/book/9783319591131>
- Borodich, F. M., and Volovikov, A. Y. (2000). Surface integrals for domains with fractal boundaries and some applications to elasticity. *Proc. R. Soc. Lond. A* 456, 1–23. doi: 10.1098/rspa.2000.0506
- Carpick, R. W. (2018). The contact sport of rough surfaces. *Science* 359:38. doi: 10.1126/science.aag1814
- Detle, H., Pepelyshev, A., and Zhigljavsky, A. (2015). “Design for linear regression models with correlated errors,” in *Handbook of Design and Analysis of Experiments*, eds A. Dean, M. Morris, J. Stufken, and D. Bingham (Boca Raton, FL: CRC Press; Taylor and Francis Group), 237–276.
- Dubuc, B., Quiniou, J. F., Roques-Carnes, C., Tricot, C., and Zucker S. W. (1989). Evaluating the fractal dimension of profiles. *Phys. Rev. A* 39, 1500–1512.
- Erbaş, B., Aydın, Y. E., and Borodich, F. M. (2019). Indentation of thin elastic films glued to rigid substrate: Asymptotic solutions and effects of adhesion. *Thin Solid Films* 683, 135–143. doi: 10.1016/j.tsf.2019.05.038
- Falconer, K. J. (1990). *Fractal Geometry: Mathematical Foundations and Applications*. Chichester: John Wiley.
- Fuller, K. N. G., and Tabor, D. (1975). The effect of surface roughness on the adhesion of elastic solids. *Proc. R. Soc. Lond. A* 345, 327–342.
- Fus, N. (1801). “An experience of theory on resistance by roads of various types to four-wheeled and two-wheeled carriages with definition of circumstances for which ones of these carriages are more useful than others,” in *Academic Papers Selected From the First Volume of “Activities of Academy of Sciences” under title “Nova Acta Academiae Scientiarum Imperialis Petropolitanae”, Part 1* (in Russian) (Saint Petersburg), 373–422.
- Galanov, B. A. (2011). Models of adhesive contact between rough elastic bodies. *Int. J. Mech. Sci.* 53, 968–977. doi: 10.1016/j.ijmecsci.2011.07.010
- Gao, H., Wang, X., Yao, H., Gorb, S. N., and Arzt, E. (2005). Mechanics of hierarchical adhesion structures of geckos. *Mech. Mater.* 37, 275–285. doi: 10.1016/j.mechmat.2004.03.008
- Ghosal, S., and Van der Vaart, A. (2017). *Fundamentals of Nonparametric Bayesian Inference*. Vol. 44. Cambridge: Cambridge University Press.
- Goryacheva, I. G. (1998). *Contact Mechanics in Tribology*. Dordrecht: Kluwer Academic Publishers.
- Greenwood, J. A. (1992). “Problems with surface roughness,” in *Fundamentals of Friction: Macroscopic and Microscopic Processes*, eds I. L. Singer and H. M. Pollock (Boston, MA: Kluwer), 57–73.
- Greenwood, J. A., and Williamson, J. B. P. (1966). Contact of nominally flat surfaces. *Proc. R. Soc. Lond. A* 370, 300–319.
- Gusak, D., Kukush, A., Kulik, A., Mishura, Y., and Pilipenko, A. (2012). *Theory of Stochastic Processes*. New York, NY: Springer-Verlag.

- Holm, R. (1941). *Die technisfische Physik der elektrischen Kontakte*. Berlin: Springer-Verlag.
- Jelinek, H., Jones, C. L., and Warfel, M. D. (1998). Is there meaning in fractal analyses? *Complex. Int.* 98, 144–149.
- Johnson, K.L. (1985). *Contact Mechanics*. Cambridge, UK: Cambridge University Press.
- Khusu, A. P., Vitenberg, Y. R., and Palmov, V. A. (1975). *Roughness of Surfaces: Theoretical Probabilistic Approach*. Moscow: Nauka
- Kragelsky, I. V. (1948). Static friction between two rough surfaces. *Bul. USSR Acad. Sci. Div. Tech. Sci.* 10, 1621–1625.
- Krantz, T. L., Alanou, M. P., Evans, H. P., and Snidle, R. W. (2001). Surface fatigue lives of case-carburized gears with an improved surface finish. *ASME J. Tribol.* 123, 709–716.
- Linnik, Yu. V., and Khusu, A. P. (1954a). Mathematical and statistical description of unevenness of surface profile at grinding. *Uspekhi Matematich. Nauk* 9:255.
- Linnik, Yu. V., and Khusu, A. P. (1954b). Mathematical and statistical description of unevenness of surface profile at grinding. *Bul. USSR Acad. Sci. Div. Techn. Sci.* 20, 154–159.
- Majumdar, A., and Bhushan, B. (1990). Role of fractal geometry in roughness characterization and contact mechanics of surfaces. *J. Tribol.* 112, 205–216.
- Mandelbrot, B. B. (1975). *Les Objects Fractals: Forme, Hasard et Dimension*. Paris: Flammarion.
- Mandelbrot, B. B. (1977). *Fractals: Form, Chance, and Dimension*. San Francisco, CA: W. H. Freeman.
- Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature*. New York, NY: Freeman.
- Mandelbrot, B. B. (1986). “Self-affine fractal sets, I: the basic fractal dimensions,” in *Fractals in Physics*, eds L. Pietronero and E. Tosatti (Amsterdam), 3–15. Available online at: <https://www.elsevier.com/books/fractals-in-physics/pietronero/978-0-444-86995-1>
- Mandelbrot, B. B. (1998). Is nature fractal? *Science* 229:783.
- Maugis, D. (2000). *Contact, Adhesion and Rupture of Elastic Solids*. Berlin: Springer-Verlag.
- Morales-Espejel, G. E., Venner, C. H., and Greenwood, J. A. (2000). Kinematics of transverse real roughness in elastohydrodynamically lubricated Oline contacts using Fourier analysis. *Proc. Inst. Mech. Eng. J. J. Eng. Trib.* 214, 523–534. doi: 10.1243/1350650001543395
- Nowicki, B. (1985). Multiparameter representation of surface roughness. *Wear* 102, 161–176.
- Pepelyshev, A., Borodich, F. M., Galanov, B. A., Gorb, E. V., Gorb, S. N. (2018). Adhesion of soft materials to rough surfaces: experimental studies, statistical analysis and modelling. *Coatings* 8:350. doi: 10.3390/coatings8100350
- Persson, B. N. L. (2006). Contact mechanics for randomly rough surfaces. *Surf. Sci. Rep.* 61, 201–227. doi: 10.1016/j.surfrep.2006.04.001
- Plesha, M. E., and Ni, D. (2001). Scaling of geological discontinuity normal load-deformation response using fractal geometry. *Int. J. Num. Anal. Meth. Geomech.* 25, 741–756. doi: 10.1002/nag.149
- Popov, V. L. (2010). *Contact Mechanics and Friction*. Berlin: Springer.
- Roques-Carnes, C., Wehbi, D., Quiniou, J. F., and Tricot, C. (1988). Modelling engineering surfaces and evaluating their non-integer dimension for application in material science. *Surf. Topogr.* 1, 435–443.
- Sayles, R. S., and Thomas, T. R. (1978). Surface topography as a nonstationary random process. *Nature* 271, 431–434.
- Schaber, C. F., Flenner, S., Glisovic, A., Krasnov, I., Rosenthal, M., Stieglitz, H., et al. (2019). Hierarchical architecture of spider attachment setae reconstructed from scanning nanofocus X-ray diffraction data. *J. R. Soc. Interface* 16:20180692. doi: 10.1098/rsif.2018.0692
- Schmittbuhl, J., Vilotte, J.-P., and Roux, S. (1995). Reliability of self-affine measurements. *Phys. Rev. E* 51, 131–147.
- Soldatenkov, I.A. (2015). Calculation of friction for indenter with fractal roughness that slides against a viscoelastic foundation. *J. Frict. Wear* 36, 193–196. doi: 10.3103/S1068366616060131
- Spinner, M., Kovalev, A., Gorb, S. N., and Westhoff, G. (2013). Snake velvet black: hierarchical micro- and nanostructure enhances dark colouration in *Bitis rhinoceros*. *Sci. Rep.* 3, 1–8. doi: 10.1038/srep01846
- Thielen, S., Magyar, B., and Piros, A. (2016). Reconstruction of three-dimensional turned shaft surfaces with fractal functions. *Tribol. Int.* 95, 349–357. doi: 10.1016/j.triboint.2015.11.028
- Thode, H. C. (2002). *Testing for Normality*. New York, NY: Marcel Dekker.
- Tricot, C. (1995). *Curves and Fractal Dimension*. Berlin: Springer-Verlag.
- Vilenkin, N. Ya. (1968). *Stories About Sets*. New York, NY; London: Academic Press.
- Warren, T. L., and Krajcinovic, D. (1996). Random Cantor set models for the elastic perfectly plastic contact of rough surfaces. *Wear* 196, 1–15.
- Whitehouse, D. J. (1982). The parameter rash—is there a cure? *Wear* 83, 75–78.
- Whitehouse, D. J. (2001). Fractal or fiction. *Wear* 249, 345–353. doi: 10.1016/S0043-1648(01)00535-X
- Whitehouse, D. J. (2011). *Handbook of Surface and Nanometrology*. Boca Raton, FL: CRC Press.
- Whitehouse, D. J., and Archard, J. F. (1970). The properties of random surfaces of significance in their contact. *Proc. R. Soc. Lond. A* 316, 97–121.
- Winkler, E. (1867). *Die Lehre von der Elastizität und Festigkeit, mit Besonderer Rücksicht auf ihre Anwendung in der Technik, für Polytechnische Schulen, Bauakademien, Ingenieure, Maschinenbauer, Architekten, etc.* Prague: Verlag H. Dominicus.
- Wolff, J. O., Grawe, I., Wirth, M., Karstedt, A., and Gorb, S. N. (2015). Spider’s super-glue: thread anchors are composite adhesives with synergistic hierarchical organization. *Soft Matter* 11, 2394–2403. doi: 10.1039/c4sm02130d
- Wolff, J. O., Schwaha, T., Seiter, M., and Gorb, S. N. (2016). Whip spiders (Amblypygi) become waterrepellent by a colloidal secretion that selfassembles into hierarchical microstructures. *Zool. Lett.* 2:23. doi: 10.1186/s40851-016-0059-y
- Zhuravlev, V. A. (1940). On question of theoretical justification of the Amontons-Coulomb law for friction of unlubricated surfaces. *Zh. Tekh. Fiz.* 10, 1447–1452.
- Zhuravlev, V. A. (2007). On question of theoretical justification of the Amontons-Coulomb law for friction of unlubricated surfaces. *Proc. Instn. Mech. Engrs. Part J: J. Eng. Trib.* 221, 894–898. doi: 10.1243/13506501JET176

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Borodich, Jin and Pepelyshev. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effect of Structural Parameters on the Relative Contact Area for Ideal, Anisotropic, and Correlated Random Roughness

Yunong Zhou and Martin H. Müser*

Lehrstuhl für Materialsimulation, Department of Materials Science and Engineering, Saarland University, Saarbrücken, Germany

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of Berlin,
Germany

Reviewed by:

Luciano Afferrante,
Politecnico di Bari, Italy
Lucia Nicola,
University of Padova, Italy

*Correspondence:

Martin H. Müser
martin.mueser@mx.uni-saarland.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 15 May 2020

Accepted: 24 June 2020

Published: 07 August 2020

Citation:

Zhou Y and Müser MH (2020) Effect of
Structural Parameters on the Relative
Contact Area for Ideal, Anisotropic,
and Correlated Random Roughness.
Front. Mech. Eng. 6:59.
doi: 10.3389/fmech.2020.00059

The relative contact area between an initially flat, adhesion- and frictionless, linearly elastic body and a variety of rough, rigid counterbodies is studied using Green's function molecular dynamics. The indenter's height profiles range from ideal random roughness through roughness with a moderate amount of correlation to periodically repeated, single-asperity indenters having perfect phase coherence. At small reduced pressures, $p^* \equiv p/(E^*\bar{g}) \ll 1$, sufficiently large systems are consistent with a linear $a_c = \kappa p^*$ relation. Here p is the pressure, E^* is the contact modulus, \bar{g} the root-mean-square height gradient, and κ a unitless proportionality coefficient. However, the parameter \bar{g} must be evaluated over the real contact area for the linear relation to hold if the random roughness is correlated or the interfacial dimension reduced. No single unitless structural parameter—including the Nayak parameter—correlates in a significant fashion with κ .

Keywords: contact mechanics, simulation, theory, random roughness surface, contact area

1. INTRODUCTION

The understanding of how the relative contact area a_r depends on pressure in nominally flat, linearly elastic contacts has made great progress in the past two decades (Persson, 2001; Müser et al., 2017). Many advanced simulation studies support the view that it increases linearly with pressure from non-zero but very small a_r up to $a_r \approx 0.1$, at least in the limiting and much investigated case of randomly rough, self-affine surfaces (Hyun et al., 2004; Campañá and Müser, 2007; Carbone and Bottiglione, 2008; Putignano et al., 2012; Prodanov et al., 2013). The height spectrum $C(q)$ of a self-affine surface obeys power law scaling with the magnitude of the wavevector q over several decades (Majumdar and Tien, 1990; Palasantzas, 1993; Persson, 2014; Jacobs et al., 2017), i.e., $C(q) \propto q^{-2(1+H)}$, where H is called the Hurst exponent. In the limit of ideal random roughness, the phases of Fourier transforms of the surface height are independent random numbers that are uniformly distributed on $(0, 2\pi)$. In this random-phase approximation, linearity between contact area and load can be rationalized with Persson's contact mechanics theory (Persson, 2001; Persson et al., 2004), which, in addition to predicting the area-load relation reasonably well, finds a highly accurate pressure-dependence of the interfacial stiffness along with accurate distribution functions of the interfacial separation (Lorenz and Persson, 2008; Almqvist et al., 2011; Campañá et al., 2011; Dapp et al., 2012; Prodanov et al., 2013; Afferrante et al., 2018) and correct spatial stress correlations (Campañá et al., 2008; Persson, 2008), none of which bearing-area models are able to achieve.

While it is sometimes argued that there must be a rigorous linear relation between contact area and load for random surfaces (in the thermodynamic limit), there are several indications for this linearity not to be strict. When the load is so small that contact is made only in a more-or-less connected patch near the highest point, relations for randomly rough surfaces may no longer hold (Pohrt et al., 2012). This, however, can be rationalized as a finite-size effect and be accounted for in Persson theory (Pastewka et al., 2013). However, even when several asperities were in contact, deviations from linearity had been noted by Yastrebov et al. (2015). Unfortunately, their calculations were conducted at an essentially fixed ratio of system size and short-wavelength cutoff (with varying roll-off wavelengths) so that small logarithmic corrections to linear relations between contact area and load may not be particularly telling. So far, the strongest support for deviations from linearity were reported in two carefully conducted simulation studies by Nicola and coworkers (van Dokkum et al., 2018; Salehani et al., 2020). They found quite remarkable logarithmic corrections to linear laws in large, one-dimensional adhesionless contacts (van Dokkum et al., 2018) as well as clearly sublinear scaling for two-dimensional, adhesive surfaces (Salehani et al., 2020). However, the latter may have been in a regime with non-negligible adhesive hysteresis, in which case non-linearity between a_r and p is unavoidable. Thus, even these two latter works do not explicitly challenge the view that a linear dependence of contact area on load is an excellent approximation for linearly elastic, sufficiently large, two-dimensional elastic bodies squeezed against a randomly rough, non-adhesive indenter.

An unsolved question is why different Hurst exponents lead to slightly different proportionality coefficients in the equation

$$a_r = \frac{\kappa p}{E^* \bar{g}}, \quad (1)$$

where p is the nominal contact pressure, E^* the contact modulus, and \bar{g} the root-mean-square height gradient. Yastrebov et al. (2015) found that the proportionality coefficient κ correlates with the Nayak parameter (Nayak, 1971), Φ_N at fixed p^* . As we will argue in this work, Φ_N can be seen as the lowest-order unitless, scalar parameter describing a surface topography other than \bar{g} , however, only if the random-phase approximation (rpa) is satisfied. The Nayak parameter is mentioned in many contact-mechanics studies. However, none of them that we are aware of and that claim the proportionality coefficient κ to depend on the Nayak parameter has the numerical sophistication of Yastrebov et al.'s works (Yastrebov et al., 2015, 2017). Nonetheless, they only addressed surfaces with ideal random roughness, i.e., surfaces obeying the rpa, and, as already mentioned above, the ratio of system size and short-wavelength cutoff was kept in a relatively narrow range. Therefore, the Nayak parameter may merely have strongly correlated with other, relevant (stochastic) quantifiers, but not have causally determined the $\kappa(H)$ dependence.

In fact when leaving the realm of ideally rough surfaces, it does not appear plausible that the Nayak parameter strongly affects κ . To see why consider the exact solution to a contact problem (no friction, no adhesion, small-slope approximation) between an

initially smooth, linearly elastic manifold squeezed down against a rough, rigid substrate fixed in space. In a thought experiment, modify the substrate's height profile in the non-contact zones such that all moved points remain below the elastic body. Its equilibrium displacement field and contact points at this and smaller pressures will be the same as for the original substrate, however, the Nayak parameter can have changed by orders of magnitude between the original and the modified indenter. In another thought experiment, consider an indenter with blunt peaks (facing the elastic body) and steep valleys. At a given load, it will have a relatively large contact area. If the indenter's height profile is flipped around, the Nayak parameter remains the same but now the contact area is strongly diminished as the elastic body is now in contact with sharp peaks. Thus, in one thought experiment, the Nayak parameter changed its value significantly, while the contact area remained unaffected. In the other one, the situation reversed: the contact area changed while the Nayak parameter didn't. For these reasons, there can be scarcely a broadly applicable correlation between contact area and Nayak parameter, at least once the random-phase approximation has been abandoned.

The two thought experiments reveal that relevant stochastic topographic quantifiers may only depend on functions that are defined in the true contact zones. As desired, such a quantifier would remain unaffected in the first thought experiment when a_r does not change, while both would adopt new values in the second thought experiment. Dimensional analysis of the specified contact problem reveals that the root-mean-square gradient plays a central role for the contact area, even in the absence of the random-phase approximation (Prodanov et al., 2013). In fact, one of us (Müser, 2017) demonstrated that Equation (1) also holds for indenters having a harmonic height profile (asperity height decreases as a powerlaw with distance from a symmetry axis) if \bar{g} is averaged over the true contact area. In this case, a_r is no longer strictly linear in p since \bar{g} is a function of load or pressure. Similarly, line contacts between two-dimensional solids were found to obey Equation (1) but only when \bar{g} was averaged over the true contact (van Dokkum et al., 2018). Thus, the effective κ is a function of the bluntness of the indenting tip, where bluntness would be defined by the functional form of the indenters rather than by numerical prefactors.

In this work, we investigate what structural parameters affect the proportionality coefficient κ , or more generally, the area-pressure relation. Toward this end, we study the contact mechanics of height profiles with anisotropy and with phase correlation in the Fourier transforms of the height profiles. This correlation is realized by warping rpa height profiles such that peaks are smoothed and valleys are roughened or vice versa, whereby the rpa is destroyed. Anisotropy will also be considered. We then explore to what extent the results on relative contact area correlate with unitless scalar parameters describing the topographic features of the surfaces. These parameters include the Nayak parameter but also a variety of other parameters, which take constant values for rpa surfaces but not for surfaces with correlated (random) roughness.

The remainder of this paper is organized as follows: Model and method are introduced in section 2. The theory is introduced

in section 3, which includes the systematic construction of topographic order parameters beyond the Nayak parameter. Results are presented in section 4, while conclusions are drawn in section 5.

2. MODEL AND METHOD

As mentioned in the introduction, we are concerned with contacts between an initially flat, linearly elastic body in frictionless contact with a rigid, rough counterface indenting from below. Interactions between the two bodies consist of a non-overlap constraint. This model has been used extensively since Hertz.

The contact problem is solved with Green's function molecular dynamics (GFMD), which is a frequently employed and described boundary-value method (Campaña and Müser, 2006; Pastewka et al., 2012; Kajita, 2016; van Dokkum and Nicola, 2019). Here we use it in the so-called FIRE-GFMD variant (Zhou et al., 2019).

Most surfaces simulated in this work are randomly rough. The default or rather starting surfaces have the isotropic, self-affine roughness described in the introduction. In addition, it is assumed that ideal self-affine scaling only exists for wave vectors $q_r \leq q \leq q_s$, where q_r is the roll-off wave vector and q_s is the short-wavelength cutoff. Thus, the used spectra are

$$C(q) = C(q_r) \times \begin{cases} f_{\text{roll}} & \text{if } q < q_r \\ (q/q_r)^{-2(1+H)} & \text{if } q_r < q < q_s \\ 0 & \text{else,} \end{cases} \quad (2)$$

where f_{roll} is a Boolean variable, which takes the value of one if true and of zero if false. f_{roll} is false by default and only set to true when mentioned explicitly. Thus, by default, $\lambda_r = 2\pi/q_r$ plays the role of a long wavelength cutoff rather than of a rolloff. In almost all cases, the linear dimension of the simulation cell \mathcal{L} is still chosen larger than λ_r in order to average implicitly over different random realizations, which can become relevant for large Hurst exponents at small relative contact areas.

The elastic body is discretized into elements having a linear dimension Δa , which is sufficiently small compared to π/q_s in order for the continuum limit to be reached. In fact, different discretizations are considered and results for all observables are extrapolated to the continuum limit ($\varepsilon_c \equiv \Delta a/\lambda_s \rightarrow 0$), using a Richardson extrapolation as described in Prodanov et al. (2013). For the contact area, we find a linear dependence in ε_c of the continuum corrections rather than the previously identified $\varepsilon_c^{2/3}$ dependence (Campaña and Müser, 2007; Prodanov et al., 2013). Unfortunately, we no longer possess the old configuration files so that analyzing the origin of the discrepancy cannot be clarified. While our GFMD code has been rewritten twice after the original single-authored code by Campana (version 2 by Müser, Dapp, and Prodanov while version 3 was by Müser with Zhou and Wang), we now find a linear $a(\varepsilon_c)$ scaling even when using version 2, which was used, for example, in the contact-mechanics challenge. Although, corrections turn out differently than before, the extrapolated values between old and new results matched whenever tested.

For the so-called fractal correction, i.e., the excess contact area arising due to the ratio $\varepsilon_f = \lambda_s/\lambda_r$ being finite, we also find different scaling than before (Prodanov et al., 2013), i.e., the exponent α_f of the correction $\varepsilon_f^{\alpha_f}$ is found to vary between zero (implying approximately logarithmic corrections) and unity. Precise values are difficult to determine due to stochastic scatter. One reason for why we no longer find a unique value for α_f is that we now consider contact areas in the continuum limit before exploring the limit $\varepsilon_f \rightarrow 0$. However, since we find α_f to be always equal or less to unity, extrapolation with $\alpha_f = 1$ from finite ε_f to zero should never result in an overcorrection.

Deviations from isotropy are realized by using an effective “Peklenik” wavenumber

$$q_P = \left(\frac{\gamma_P^2 q_x^2 + q_y^2 / \gamma_P^2}{\sqrt{\gamma_P^4 + 1/\gamma_P^4}} \right)^{1/2} \quad (3)$$

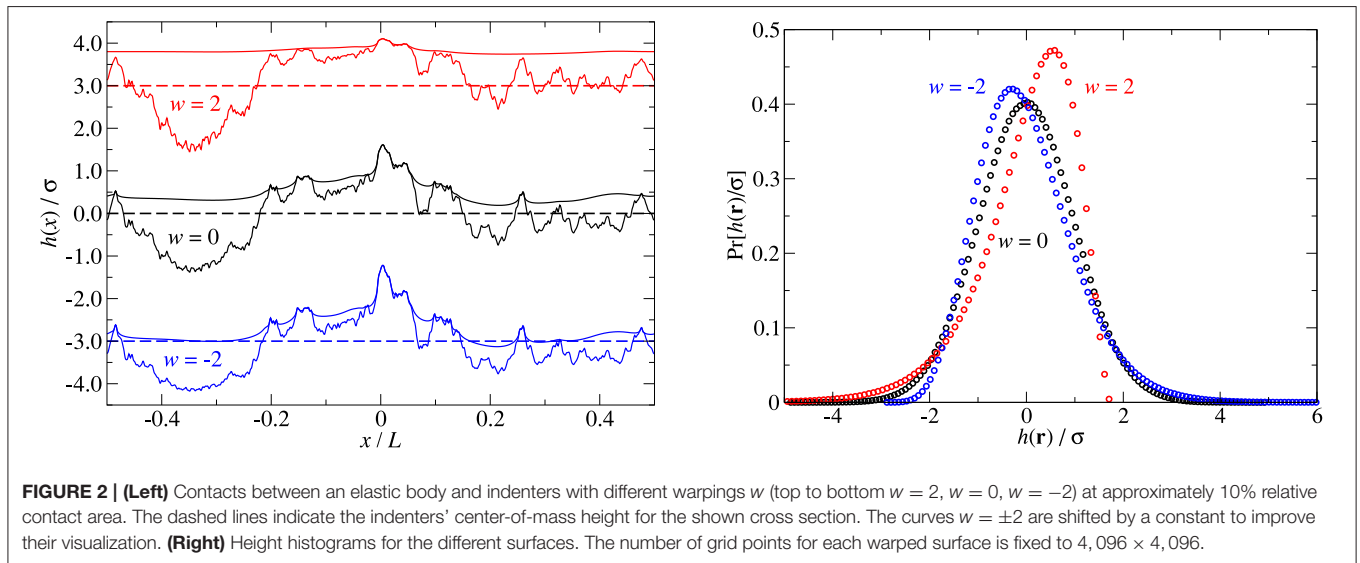
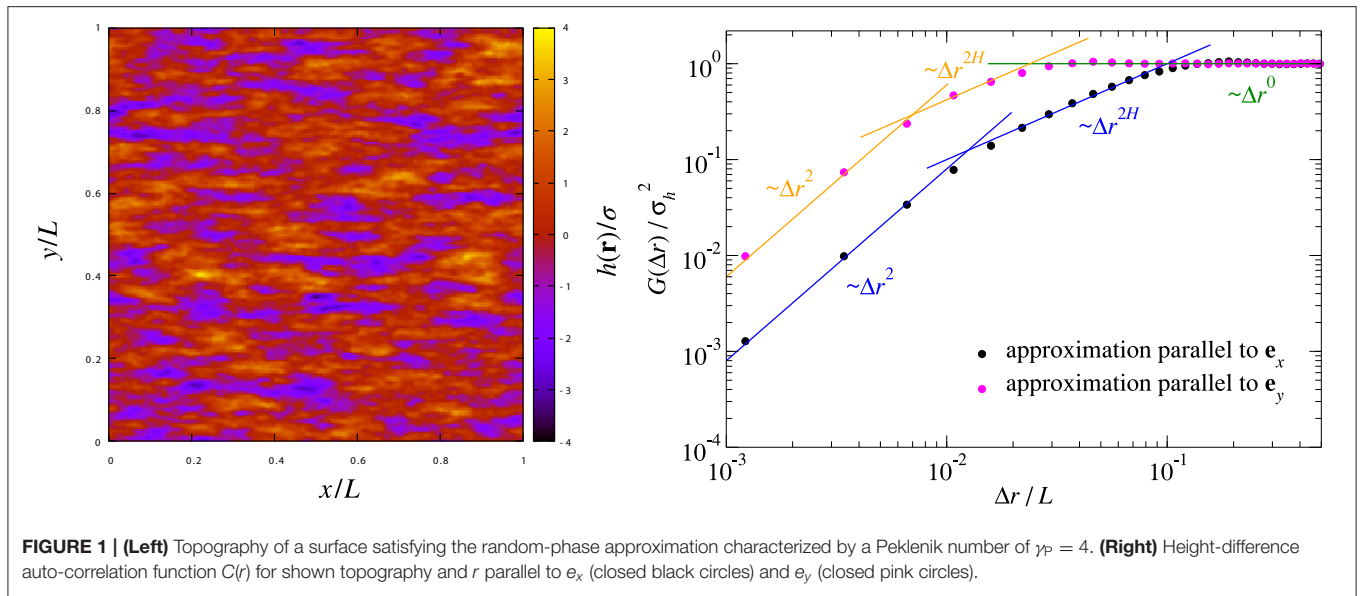
as the variable in the height spectrum rather than the true wave number q . The Peklenik parameter satisfies $0 < \gamma_P < \infty$. Surfaces are stochastically isotropic if $\gamma_P = 1$. Anisotropy increases with increasing $(\ln \gamma_P)^2$. The such produced height profiles reveal preferred directions or “grooves” in Figure 1, which makes them somewhat more similar to scratched surfaces than topographies with isotropic random roughness.

Phase coherence can be destroyed through a warping transformation. A warped height is defined by

$$h_w = \begin{cases} h_{\min} + (h - h_{\min}) \left(\frac{h - h_{\min}}{h_{\max} - h_{\min}} \right)^w & \text{if } w \geq 0 \\ h_{\max} - (h_{\max} - h) \left(\frac{h_{\max} - h}{h_{\max} - h_{\min}} \right)^{|w|} & \text{else,} \end{cases} \quad (4)$$

where $h_{\min} \leq \text{Min}\{h(\mathbf{r})\}$ and $h_{\max} \geq \text{Max}\{h(\mathbf{r})\}$. Here h indicates the height of the original (rpa) height profile at a given position \mathbf{r} and h_w is the height at a given location of the warped surface. At $w = 0$, height profiles remain unchanged. For $w > 0$, peaks are blunted and valleys sharpened, while the opposite is achieved with $w < 0$, as can be seen in the cross-section of the height-profile shown in Figure 2. In the current study, $h_{\min} = 2 \text{Min}\{h(\mathbf{r})\} - \text{Max}\{h(\mathbf{r})\}$ and $h_{\max} = 2 \text{Max}\{h(\mathbf{r})\} - \text{Min}\{h(\mathbf{r})\}$ were used.

Phase correlation through our warping procedure is reflected by the observation that the height histograms are systematically skewed for $w \neq 0$. In contrast, the height histograms of rpa surfaces are symmetric, except for finite-size effects, because each height profile has the same likelihood to occur as its mirror image in which valleys turn to peaks and vice versa. In addition, the expectation value of the root-mean-square gradient changes with height, i.e., it is much reduced in the blunted parts of the surface compared to the roughened parts. In contrast, for (sufficiently large) rpa surfaces, \bar{g} is independent of the height, except near the highest and lowest heights in a finite sample. In future topographic determinations of random surfaces with correlation, it would be very interesting to know to what extent the rms-height gradient changes with height. This would probably constitute an important quantity to be targeted



for computer-generated surfaces. To mimic scratched surfaces realistically, it will probably be required to assume a scale or q -dependent Peklenik parameter (Candela et al., 2011) in addition to warping and to consider additional refinements. However, our impression is that the $w = 2$ surface already contains some features of polished surfaces, i.e., relatively deep valleys and smoothed tops.

In principle, it would have also been possible to consider phase correlation by limiting the values of the phases of $\tilde{h}(\mathbf{q})$ to a range $-\varphi_m \leq \varphi \leq \varphi_m$ defined by a minimum/maximum allowed phase φ_m for the phases of all $\tilde{h}(\mathbf{q})$, while keeping the height spectra $C(q)$ unchanged. We found this method not to be helpful, because it lead to indenters with sharp thorns. In the limit of $\varphi_m \rightarrow 0$, essentially single-point indenters are obtained. Rather than investigating those, we will consider indenters with

harmonic height profiles satisfying

$$h(r) = \frac{R}{n} \left(\frac{r}{R} \right)^n, \quad (5)$$

where R is of unit length, while r is the in-plane distance of a point from the (closest) symmetry axis of an individual indenter, which we place into the center-of-mass of the simulation cell. The values of n used in this study were restricted to $n = 1.5$ (sharp indenter), $n = 2$ (Hertzian geometry), and $n = 4$ (blunt indenter). Analytical solution for stress and displacement fields (in the contact area and its vicinity) and thus for the dependence of contact area on load are known when the contact radii are negligibly small compared to the linear dimension of the simulation cell.

3. THEORY

3.1. Phenomenological Generalization of Persson Theory

Persson theory for linear elastic contacts between ideally randomly rough surfaces predicts a relative contact area of

$$a_r = \operatorname{erf} \left\{ \sqrt{2} p_0 / E^* \bar{g} \right\}, \quad (6)$$

where \bar{g} is the root-mean-square height gradient averaged over the entire surface, i.e.,

$$\bar{g}^2 = \int d^2 q C(q) q^2, \quad (7)$$

which is the expression in the derivation of Persson theory.

One of us noticed that Equation (6) also holds for smooth indenters with harmonic height profiles if \bar{g} is replaced with \bar{g}_c , where the latter is not determined over the entire surface but only over the true contact. Thus, the range of applicability of Equation (6) appears to extend when re-expressing it through

$$a_r = \operatorname{erf} \left(\sqrt{\pi} \kappa_c p_c^* / 2 \right), \quad (8)$$

where κ_c is the proportionality coefficient between relative contact area and the reduced pressure defined by $p_c^* = p_0 / (E^* \bar{g}_c)$. When omitting the index “c” in κ_c , we mean to refer to the proportionality coefficient when the reduced pressure is normalized to \bar{g} rather than to \bar{g}_c .

In fact, when taking Persson theory literally, it does ask the question how the rms-height changes in a given point of contact (described using a small resolution of the surface topography) when short-wavelength roughness is added to the description of the contact problem. Using the full spectrum to estimate this increase in rms-roughness can be seen as an approximation, which might be possible to correct in future work. Some possibilities are discussed in the conclusions.

The standard Persson theory predicts a value of $\kappa = \sqrt{8/\pi} \approx 1.596$. In the formulation of the theory, it appears to us as if no distinction is made—or must be—between κ and κ_c , at least as long as pressures lie in a range in which $a_r \propto p$, that is, if p is sufficiently large for contact to spread out over clearly disconnected contact patches but still small enough so that a_r is less than, say, 0.05. Careful simulations of sufficiently large systems find a relatively weak dependence of κ on the Hurst roughness exponent, e.g., $\kappa(H \approx 0.8) \lesssim 2$ and $\kappa(H \approx 0.3) \gtrsim 2$. (Hyun et al., 2004; Campañá and Müser, 2007; Hyun and Robbins, 2007; Putignano et al., 2012; Prodanov et al., 2013; Afferrante et al., 2018). Analytical results for the Hertz contact result in $\kappa_c \approx 1.666$, see also Equation (28) and the more detailed discussion of periodically repeated, smooth indenters in section 4.5.

3.2. Dimensionless Parameters: Nayak and Beyond

Since the relative contact area is a number, it can only be a function of other unitless parameters (p^* , Φ_1 , Φ_2 , ...) itself,

where p^* is a reduced pressure, e.g., $p/(E^* \bar{g})$. Moreover, each variable in the set $\{\Phi\}$ must obey the law of dimensional analysis. Specifically, they must remain invariant under the scaling transformation $\mathbf{r} \rightarrow s \cdot \mathbf{r}$ and $z \rightarrow t \cdot z$, since the relation $a_{\text{rel}} = a_{\text{rel}}(p/E^* \bar{g})$ already reflects the correct dependencies on s and t (Prodanov et al., 2013). Moreover, real contact is destroyed predominantly due to roughness at small wavelengths. Thus, the Φ_i should not depend on parameters that are defined exclusively by parameters from the height-distribution also known as Abbott Feierstone curve.

Before constructing the respective unitless parameters, it is in place to discuss the square-gradient term. When determining \bar{g} over a periodically repeated contact, $\langle (\nabla h)^2 \rangle$ is identical to $-\langle \delta h \Delta h \rangle$, as can be seen by integration in parts. However, this equality no longer holds for partial contact. Defining

$$(\bar{g}_c')^2 = -\langle \delta h \Delta h \rangle_c, \quad (9)$$

the dependence of a_{rel} on parameters depending on height profiles in the contact then becomes

$$a_{\text{rel}} = a_{\text{rel}}[p/(E^* \bar{g}_c), \bar{g}_c/\bar{g}_c', \{\Phi_c\}]. \quad (10)$$

Ultimately, a_{rel} is a functional of the height topography. As such, there should exist a dependence of a_{rel} that does not necessitate parameters averaged over the real contact. However, those might necessitate non-local terms, or, alternatively, high-order derivatives of $h(\mathbf{r})$. The latter may not be well defined when the surface profile or its (higher-order) derivatives are not well-defined, as is the case, for example, for a conical indenter. Thus, the following discussion assumes surface height profiles to be smooth and “well behaved.”

For the construction of relevant parameters, it is useful to keep in mind three symmetry relations. First, a mirror inversion ($\mathbf{r} \rightarrow -\mathbf{r}$) leaves the contact area unchanged. This is why each derivative with respect to a spatial coordinate must appear an even number of times in the construction of an invariant. Second, each measure should be rotationally invariant and reduce to a scalar. This is automatically achieved when representing derivatives with the Einstein summation convention, which requires every index (enumerating an in-plane coordinate) occurring twice in a product to be summed over. To use it effectively, we use it jointly with the index notation, in which case h_α^2 indicates the square-height gradient $(\nabla h) \cdot (\nabla h)$ and $h_{\alpha\alpha}$ the Laplacian Δh . However, \bar{g} will keep indicating the rms height gradient $\sqrt{\langle h_\alpha^2 \rangle}$. Third, the invariants may not change on a rigid, vertical translation of the surface $h(\mathbf{r}) \rightarrow h(\mathbf{r}) + h_0$. This is why only $\delta h = h - \langle h \rangle$ can appear in the invariants. The lowest-order invariant obeying these rules that we could identify are given by

$$\Phi_1 = \frac{\langle \delta h h_\alpha^2 \rangle}{\sqrt{\langle h_\alpha^2 \rangle} \bar{g}^2} \quad (11)$$

$$\Phi_2 = \frac{-\langle \delta h^2 h_{\alpha\alpha} \rangle}{2 \sqrt{\langle \delta h^2 \rangle} \bar{g}^2} \quad (12)$$

$$\Phi_3 = \frac{-\langle \delta h^2 \rangle \langle h_{\alpha\alpha} \rangle}{2 \sqrt{\langle \delta h^2 \rangle} \bar{g}^2}. \quad (13)$$

Before constructing the next parameters, the allowed values for the parameter Φ_1 to Φ_3 will be discussed. Φ_1 and Φ_2 are identical when averaged over a periodically repeated surfaces (as can be seen again by integration in parts) but not when they are determined over partial contact, in which case the index “c” would be added. For periodically repeated surfaces, the parameter Φ_3 is automatically identical to zero but generally different from zero when averaged over partial contact. This is because the mean curvature disappears for a periodically repeated surface, while the curvature must average to a negative number for partial contact (assuming the elastic body indents the profile from above).

The values of Φ_1 and Φ_2 averaged over a single rpa surface may be finite. However, averaging these means over various disorder realization will make them disappear, as any surface realization $h(\mathbf{r})$ has the same probability (density) to occur as $-h(\mathbf{r})$. Thus, Φ_1 and Φ_2 —as well as any other parameter, in which the symbol h appears an odd number of times as a factor—should be small when determined over a single rpa surface realization, in particular when the roll-off domain is sufficiently large.

When averaged over partial contact and/or over surfaces violating the rpa, Φ_1 and Φ_2 may and usually do take finite values. This is why we call them symmetry allowed in **Table 1**. For the remaining parameters, we will no longer state the rationale for why terms are symmetry allowed or forbidden, as all relevant arguments have been mentioned or can be easily derived using the relations for Gaussian random numbers summarized in section 3.3. **Table 1** contains our conclusions on each parameter constructed in this work.

Additional parameters in which numerator and denominator are second order in the derivative but higher order in h can be constructed. They will be considered up to the lowest order needed beyond the rms-height gradient, in which the parameters do not disappear in case of the random-phase approximation. This includes the parameters

$$\Phi_4 = \frac{\langle \delta h^2 h_{\alpha}^2 \rangle}{\langle \delta h^2 \rangle \bar{g}^2} \quad (14)$$

$$\Phi_5 = \frac{-\langle \delta h^3 h_{\alpha\alpha} \rangle}{3 \langle \delta h^2 \rangle \bar{g}^2} \quad (15)$$

For rpa-surfaces, Φ_4 is automatically equal to unity and for all periodically repeated surfaces, $\Phi_4 = \Phi_5$.

Finally, we consider parameters in which the order of the derivatives is increased from two to four while the order in the height is kept as small as possible. Three of quite a few resulting (irreducible) possibilities are

$$\Phi_6 = \frac{1}{\bar{g}^4} \langle \delta h^2 \rangle \langle h_{\alpha\alpha} h_{\beta\beta} \rangle \quad (16)$$

$$\Phi_7 = \frac{2}{3 \bar{g}^4} \langle \delta h^2 h_{\alpha\alpha} h_{\beta\beta} \rangle \quad (17)$$

$$\Phi_8 = \frac{1}{\bar{g}^4} \left\{ \langle h_{\alpha}^2 h_{\beta}^2 \rangle - 3 \langle h_{\alpha}^2 \rangle \langle h_{\beta}^2 \rangle + \langle h_{\alpha} h_{\beta} \rangle^2 \right\} \quad (18)$$

$$\Phi_9 = \frac{2}{\bar{g}^4} \left\langle h_{\alpha} h_{\beta} - h_{\gamma}^2 \delta_{\alpha\beta} / 2 \right\rangle^2, \quad (19)$$

where $\delta_{\alpha\beta}$ is the Kronecker-delta symbol.

TABLE 1 | Values of parameters averaged over a full, periodically repeated surface (prs) if the random-phase approximation (rpa) is valid and when it is not valid (n-rpa).

Full prs	rpa	n-rpa
Φ_1	ϵ	Allowed
Φ_2	Φ_1	Φ_1
Φ_3	0	0
Φ_4	$1 - \epsilon$	Allowed
Φ_5	Φ_4	Φ_4
Φ_6	allowed	Allowed
Φ_7	$\approx \Phi_6$	Allowed
Φ_8	ϵ	Allowed
Φ_9	ϵ	ϵ
	if isotropic	If isotropic

The word “allowed” indicates that a finite value is symmetry allowed. The number ϵ implies that the result averages to zero after an ensemble average over many surface realizations and that it should be small for an individual instantiation.

The parameter Φ_6 is nothing but the Nayak parameter Φ_N , up to a multiplicative constant of 2/3. It is frequently interpreted as a measure for the spectral width. We chose the prefactor such that Φ_6 and Φ_7 are equal to unity for single-wave-vector roughness. The parameter Φ_7 is a generalization of the Nayak parameter. For randomly rough, rpa surface, its expectation value is close to but less than Φ_6 . Thus, both parameters tend to infinity as the ratio $\epsilon_f = \lambda_s/\lambda_r$ becomes small, that is, with ϵ_f^{-2H} . However, for (strongly) correlated random roughness Φ_7 takes much greater values than Φ_6 , just as Φ_4 starts to substantially exceed unity, because the factorization of the various terms (see also section 3.3) no longer holds once the rpa is no longer obeyed.

The parameter Φ_8 plays the role of a generalized height gradient cumulant. It is constructed such that it takes the value of zero when the fourth-order cumulant of the surface slope s parallel to any in-plane unit vector \mathbf{n} takes the value of zero if it is distributed normally, i.e., when $c_{4,\mathbf{n}} = \langle s^4 \rangle - 3 \langle s^2 \rangle^2$ disappears for every \mathbf{n} . This parameter is implicitly symmetrized with respect to its mirror images in the xz and yz planes so that $\langle s \rangle = 0$ follows. Note that Φ_8 being small is a necessary but not a sufficient criterion for every $c_{4,\mathbf{n}}$ to disappear. It is only sufficient if the surfaces are stochastically isotropic.

Finally, Φ_9 is a measure for anisotropy. It takes the values of zero and one in the limits of ideal isotropic and ideal anisotropy, respectively, where, for the latter, surfaces are perfectly smooth along one spatial direction. Assuming the Peklenik number to be independent of the wavevector, Φ_9 can be easily shown to be identical to $(\gamma_p^2 - 1/\gamma_p^2)^2/2$. As is the case for some other parameters too, Φ_9 is not identical to zero for an individual surface realization, but only averages to zero after taking sufficiently many surface realizations.

We conclude this section by stating that the Nayak parameter is the lowest-order scalar structural parameter that the proportionality coefficient κ can depend on if the surface is isotropic and satisfies the random-phase approximation. All

other parameters of similar or smaller order in height are either identical to zero, or their expectation value is zero, or they strongly correlate with the Nayak parameter.

3.3. Evaluation of Fourth-Order Invariants

For (isotropic) randomly rough surfaces, invariants being fourth order in height and fourth order in derivatives are the leading-order, scalar structural parameters that can affect the proportionality coefficient κ . Of particular interest should be those that—unlike the Nayak parameter—cannot be reduced to products of invariants being second order in height. Yet, the evaluation of fourth-order expressions is commonly done using Wick's theorem (Wick, 1950), which, applied to the current problem, translates to

$$\begin{aligned} C_4(\mathbf{q}_1, \dots, \mathbf{q}_4) &\equiv \langle \tilde{h}(\mathbf{q}_1) \tilde{h}(\mathbf{q}_2) \tilde{h}(\mathbf{q}_3) \tilde{h}(\mathbf{q}_4) \rangle \\ &\approx \langle \tilde{h}(\mathbf{q}_1) \tilde{h}(\mathbf{q}_2) \rangle \langle \tilde{h}(\mathbf{q}_3) \tilde{h}(\mathbf{q}_4) \rangle \\ &\quad + \langle \tilde{h}(\mathbf{q}_1) \tilde{h}(\mathbf{q}_3) \rangle \langle \tilde{h}(\mathbf{q}_2) \tilde{h}(\mathbf{q}_4) \rangle \\ &\quad + \langle \tilde{h}(\mathbf{q}_1) \tilde{h}(\mathbf{q}_4) \rangle \langle \tilde{h}(\mathbf{q}_2) \tilde{h}(\mathbf{q}_3) \rangle, \end{aligned} \quad (20)$$

whereby expectation values of fourth-order expressions are reduced to products of height spectra, since

$$\langle \tilde{h}(\mathbf{q}_m) \tilde{h}(\mathbf{q}_n) \rangle \propto C(|\mathbf{q}_m|) \delta(\mathbf{q}_m + \mathbf{q}_n), \quad (21)$$

where $C(q)$ is the height spectrum. Equation (20) is exact for Gaussian random variables of mean zero.

4. RESULTS

4.1. On the Accurate Calculation of Contact Area and the Proportionality Coefficient κ

In Yastrebov et al. (2017), Yastrebov et al. claimed to ensure an unprecedented accuracy in computation of the true contact. If n_c is the number of contact points, n_{cl} the number of contact or rather contact-line points, which are in contact but have at least one non-contact point as nearest neighbor, then the relative contact area was estimated with

$$a_r = \frac{n_c - \alpha n_{cl}}{n_{tot}}, \quad (22)$$

where n_{tot} is the total number of points into which the surface is discretized and $\alpha = (\pi - 1 + \ln 2)/24$.

While we find the proposed correction to be quite useful in order to get an astoundingly accurate first estimate of the continuum correction for isotropic, rpa surfaces, we believe that it can at best be on par with a properly executed Richardson extrapolation even in that limiting case. The reason is that the numerical coefficient $\alpha \approx 0.11811$ can scarcely be universal even if the special form of writing it as $(\pi - 1 + \ln 2)/24$ might convey a profound mathematical reason for its specific value. To argue why the proposed method cannot surpass a Richardson extrapolation in the asymptotic limit, let us assume

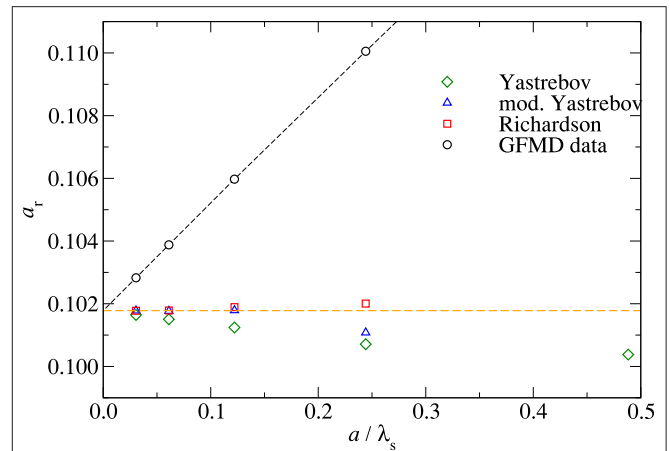


FIGURE 3 | Comparison of different extrapolation schemes for the relative contact area a_r to the continuum limit $\varepsilon_c \equiv a/\lambda_s \rightarrow 0$ for an rpa surface specified by $H = 0.8$, $p^* = 0.05$, $\lambda_r/\mathcal{L} = 0.5$, and $\lambda_s/\mathcal{L} = 0.008$. In the modified Yastrebov extrapolation, the prefactor α in Equation (22) was chosen such that the extrapolated contact area at a discretization $a\lambda_s$ and at $2a\lambda_s$ gave the same estimate for the contact area. Dashed lines are drawn to guide the eye.

that the leading order-correction were indeed proportional to the number of contact-line points within the contact. This number would ultimately scale with a/λ_s , because the fractal nature of the contact seizes to exist in this limit, so that the contact line acquires the (fractal) dimension of unity, in which case $n_{cl} \propto \lambda_s/a$. This linear scaling of the leading-order corrections to the contact area would be picked up by a Richardson extrapolation and the proportionality coefficient would automatically adjust to the exact value and not only to a value, which is very good but not exact.

The proposed correction scheme can only outperform a Richardson extrapolation if higher-order corrections happened to be incorporated into it. This, in fact, might be the case, as is revealed in Figure 3 by the accurate values from the Yastrebov extrapolation at large ε_c . However, since the exact value α must depend on the specifics of a simulation (artificial geometries can be easily constructed for which $\alpha \equiv 0$ at large loads), it can only be exact in isolated cases of measure zero so that it is generally asymptotically inferior to Richardson. As such, a claim of having provided data with unprecedented accuracy with this method is somewhat hazardous given that previous work used a Richardson extrapolation while employing ratios of $\varepsilon_c = a/\lambda_s$, $\varepsilon_f = \lambda_s/\lambda_r$, and $\varepsilon_t = \lambda_r/\mathcal{L}$, which were simultaneously all greater than the data having the purportedly unprecedented accuracy.

The ideas of Yastrebov's extrapolation can be modified by incorporating the spirit of a Richardson extrapolation into it: if the numbers of contact and contact-line points are known at two different values of a/λ_s , the parameter α could be adjusted such that the Yastrebov extrapolation gives the same estimate for both discretizations. Asymptotically, this modification leads to an improvement, as can be seen in Figure 3, however, it is usually a disimprovement for $a/\lambda_s \geq 1/4$. To terminate our discussion

on the Yastrebov extrapolation, we note that we found it to work equally well for warped surfaces as for rpa surfaces and that our final estimates for the relative contact area were always very close, say, within 0.1%, to using a rigorous extrapolation with the value that we obtained with a Yastrebov extrapolation using

$$\alpha = 0.11 \quad (23)$$

and a discretization of $a/\lambda_s \lesssim 1/4$. This value of α typically deteriorates the quality of the contact-area estimate for $a = \lambda_s/2$ but improves it overall otherwise.

Finally, given that the extrapolation used by Yastrebov and coworkers may lead to different errors at different pressures, the following question remains: What is the origin of their observed logarithmic correction of κ on pressure? Is it the lack of a rigorous extrapolation to the continuum ($\varepsilon_c = a/\lambda_s \rightarrow 0$), the fractal ($\varepsilon_f = \lambda_s/\lambda_r \rightarrow 0$), or the thermodynamic ($\varepsilon_t = \lambda_r/\mathcal{L} \rightarrow 0$) limit? Or is it actually true? After all and despite our criticism of their self appraisal, we see their data to be of similar quality as that of most other cutting-edge works, which had unanimously come to the conclusion of κ acquiring a constant value at small p , but which either simply took the values at the smallest value of ε_c without further extrapolation or that made Richardson extrapolations, which we can no longer reconstruct. Moreover, one-dimensional systems show robust logarithmic corrections in κ at small p (van Dokkum et al., 2018) even for very large sizes when contact is also made far away from the highest point. (In the preparation of this manuscript, we verified those claims on 1D systems for sizes exceeding 2^{20} grid points.)

The last limit that needs to be taken to zero is the pressure $p \rightarrow 0$, i.e., the proportionality coefficient κ should be computed with the smallest possible errors. To do so, we compute κ as a pressure-dependent function through the equation

$$a_r = \operatorname{erf}(\sqrt{\pi\kappa p^*}/2), \quad (24)$$

rather than through a_r/p^* or $\partial a_r/\partial p^*$, because Equation (24) accounts for some of the low-pressure non linearities, as can be appreciated in Figure 1 of Dapp et al. (2014). The zero-pressure limit can only be taken in a meaningful manner in conjunction with a proper size scaling as will be discussed next.

4.2. Isotropic rpa Surfaces

4.2.1. Does κ Have a Low-Pressure Limit?

κ cannot have a finite low-pressure limit, if it is defined or computed according to

$$\kappa = \left(\prod_x \lim_{\varepsilon_x \rightarrow 0} \right) \lim_{p \rightarrow 0} \kappa(p), \quad (25)$$

where the various limits (defined in the previous section with x taking the “values” c for continuum, f for fractal, and t for thermodynamic) are taken in arbitrary order. The reason is that for any finite system at exceedingly small pressures, only the highest (meso-scale) asperity is in contact, which (in the realm of continuum mechanics) will ultimately be Hertzian, in which case κ tends to infinity.

The interesting question is whether a_r can be proportional to p over an extended range in pressure with no or negligible logarithmic corrections to the proportionality coefficient κ if both $\varepsilon_{t,f}$ are sufficiently small while ε_c is properly taken to zero at each pressure. Thus, mathematically speaking, we are interested in a limit

$$\kappa = \lim_{p \rightarrow 0} \lim_{\varepsilon_{t,f} \rightarrow 0} \frac{a_r}{p^*}, \quad (26)$$

since experimental systems, to which simulations should compare, tend to have much larger values of $\varepsilon_{t,f}$ than can be realized in a computer simulation. This limit is certainly not approached if the product $\varepsilon_t \varepsilon_f$ is taken to be effectively constant while having varying discretization errors, as in Yastrebov et al. (2017). Keeping all but one $\varepsilon_{t,f,c}$ constant—as done by Prodanov et al. (2013)—is only an unsatisfactory improvement, because the discretization corrections probably decrease as ε_f decreases due to the increase of the characteristic patch size. The cardinal mistake made by Prodanov et al. (2013) was to assume that leading errors to κ are sums of terms $c_x \varepsilon_x^{\nu_x}$ with $x \in (c, t, f)$ and $0 < \nu_x$, while they could also be of a more general form and involve, e.g., products of powers of different ε_x .

To explore the question raised at the beginning of the previous paragraph in a somewhat more meaningful way than before, we ran simulations in which ε_f and/or ε_t were decreased simultaneously with decreasing pressure p according to

$$\varepsilon = \varepsilon_{\text{ref}} (p/p_{\text{ref}})^{\frac{1}{4}}. \quad (27)$$

Results are shown in **Figure 4**, for which we chose (arbitrarily) $\varepsilon_f = 1/32$ and $\varepsilon_t = 1/2$ at a reference pressure of $p^* = 0.2$. It reveals that κ increases quite noticeably with decreasing pressure for all three $H = 0.3$ systems, while it essentially plateaus for the two $H = 0.8$ systems in which ε_f is decreased as p decreases. While the $H = 0.3$ might also plateau at even smaller p^* , a qualitative difference between $H = 0.3$ and $H = 0.8$ would remain: The curves in which both ε_t and ε_f are decreased with decreasing p^* lead to small κ for $H = 0.8$ but to large κ for $H = 0.3$, which become even larger for fixed ε_f .

The reason for the different trends can be potentially linked to the distribution of contact patch areas and the characteristic contact patch size A_c , which we define to be the expected patch size that a randomly picked contact point belongs to. The three $H = 0.3$ cases and the $H = 0.8$ case with fixed ε_f , all belong to situations, in which the characteristic contact areas are rather small: A_c increases only logarithmically with ε_f for $H < 0.5$ (Müser and Wang, 2018) so that large patches are not possible, even when ε_f is small. Likewise, only small contact patches can arise for $H = 0.8$ at pressures well below the percolation threshold if ε_f is fixed to a constant value as large as $\varepsilon_f = 1/32$. In contrast, large contact patches can arise even at small pressures, if ε_f is small and $H > 0.5$. Large patches are needed for a linear $a_r(p)$ dependence at small pressures, as can be rationalized qualitatively from bearing-area models.

In order to explore the question further, if κ can have a well-defined limit when being deduced with the meaningful limit defined in Equation (26), we compare the $\kappa(p)$ relation with two

sets of $\varepsilon_{t,f}$: (i) a small system with $\varepsilon_t = 1$ and $\varepsilon_f = 1/32$ and (ii) a larger systems in which both ε_t and ε_f were decreased by a factor of four to $\varepsilon_t = 1/4$ and $\varepsilon_f = 1/128$. **Figure 5** reveals that the pressure sensitivity of κ reduces quite noticeably with increasing system size for $H = 0.8$ but not for $H = 0.3$, which, however, has a relatively minor logarithmic dependence of κ on p^* to begin with.

The trends revealed in **Figure 5** are consistent with those of **Figure 4**. Increasing system size, i.e., decreasing simultaneously ε_t and ε_f , reduces κ and its pressure sensitivity for $H = 0.8$ but not for $H = 0.3$. Our trends are also roughly consistent with the observations by Prodanov et al. (2013), who proposed the

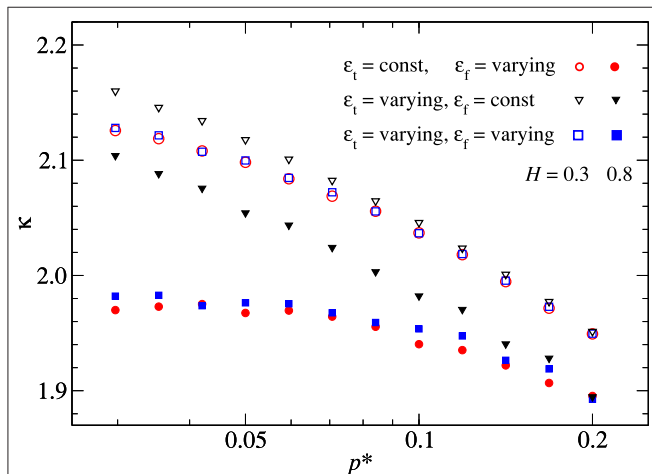


FIGURE 4 | Proportionality coefficient κ as defined in Equation (27) for two Hurst exponents and three different choices of how $\varepsilon_{t,f}$ change with pressure p . The term “const” relates to $\varepsilon_t = 1/2$ and $\varepsilon_f = 1/32$. These two values are the reference values at $p^* = 0.2$. The term “varying” indicates that the respective ε is scaled according to Equation (27). To give an example, the scaling makes a varying ε_t go down to $1/64$ and a varying ε_t to 1.4 at $p^* = 0.0125$. Results were averaged over up to 100 random realizations per data point.

existence of a well-defined value for κ for $H = 0.8$ but not for $H = 0.3$.

While we already speculatively linked the different trends for $H = 0.3$ and $H = 0.8$ to the way how a characteristic contact patch size changes with decreasing ε_f , it remains to be explained why A_c increases only logarithmically with ε_f for $H < 0.5$ but algebraically for $H > 0.5$. One possibility could be that most of the elastic energy (in full contact) is stored in long-wavelength modes for $H > 0.5$ but in short-wavelength modes for $H < 0.5$. If this argument were true, $H = 0.5$ could be the dividing line for the different types of behaviors. We therefore repeated the analysis shown in **Figure 5** for $H = 0.5$. Unfortunately, the results leave us somewhat uncertain. Significantly more work must probably be done to characterize the transition between the different scaling behaviors rigorously.

To conclude this section on the much investigated isotropic, rpa surfaces, we expect κ to have a well defined zero-pressure limit, in the sense of Equation (26), which does not depend upon how it is approached. For small Hurst exponents, i.e., for $H > 0.5$, it might not exist and/or it might depend on how $\varepsilon_f \rightarrow 0$ is approached, e.g., it could take different values when reaching it with constant $\varepsilon_f/\varepsilon_t$ than with constant $\varepsilon_f/\sqrt{\varepsilon_t}$. At the same time, we somewhat expect that the value of $\kappa \approx 2.5$ predicted by the more sophisticated bearing-area models, such as that by Bush et al. (1975), might provide a (potentially rigorous) upper bound for κ when the limit $\varepsilon_t \rightarrow 0$ is taken before $\varepsilon_f \rightarrow 0$. In this case, the surface roughness on scales exceeding the roll-off wavelength is white noise so that the underlying mean-field model of elasticity might not be detrimental.

4.2.2. Effect of Nayak and Related Parameters on κ

Yastrebov et al. (2017) proposed the proportionality coefficient κ to decrease logarithmically with the Nayak parameter, at least for self-affine, rpa surfaces. However, two aspects of this claim and their data strike us as troublesome. First, κ is implicitly predicted to become negative for very large Nayak parameters, which is physically meaningless. Second, their data points seem

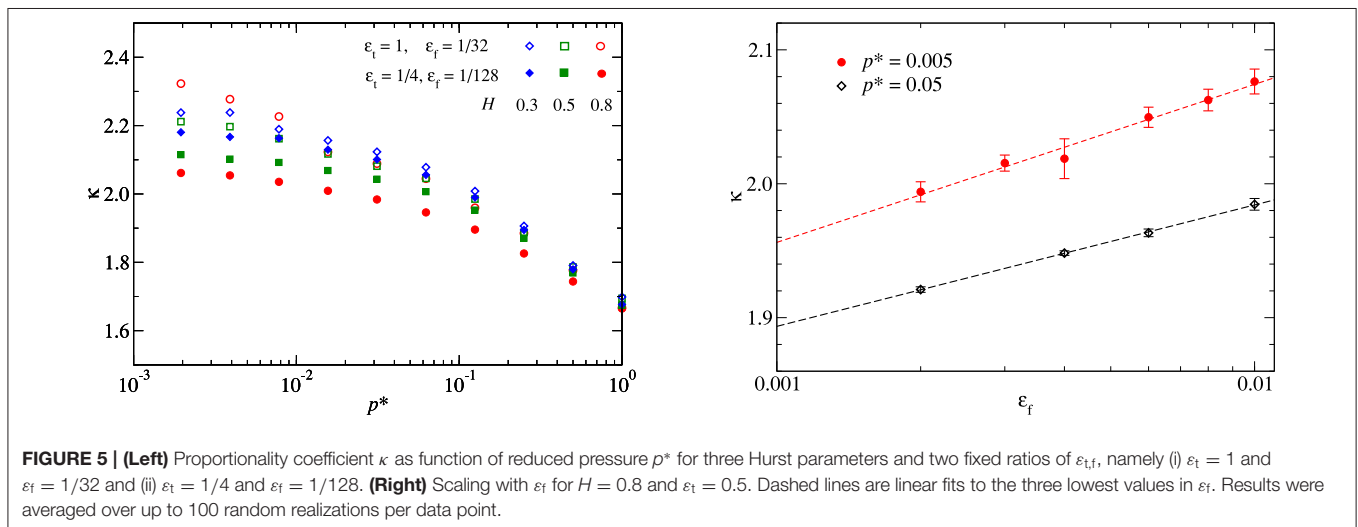
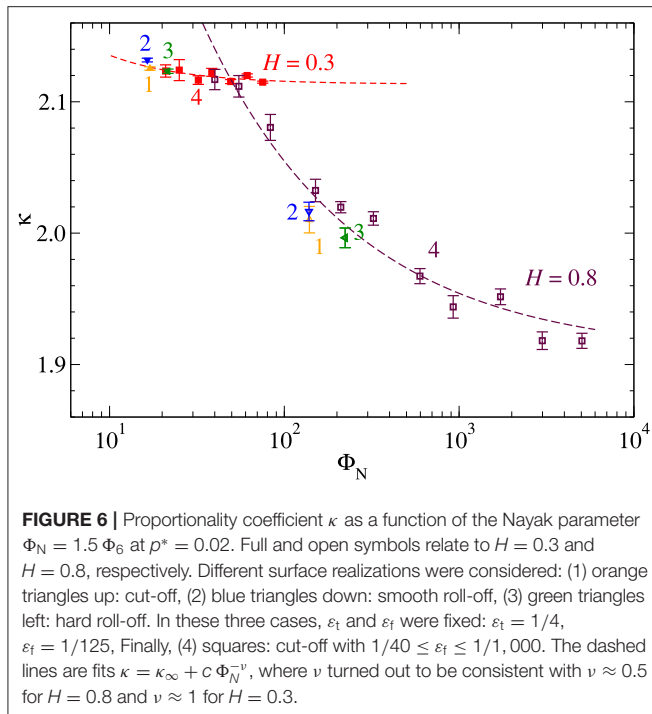


FIGURE 5 | (Left) Proportionality coefficient κ as function of reduced pressure p^* for three Hurst parameters and two fixed ratios of $\varepsilon_{t,f}$, namely (i) $\varepsilon_t = 1$ and $\varepsilon_f = 1/32$ and (ii) $\varepsilon_t = 1/4$ and $\varepsilon_f = 1/128$. (Right) Scaling with ε_f for $H = 0.8$ and $\varepsilon_t = 0.5$. Dashed lines are linear fits to the three lowest values in ε_f . Results were averaged over up to 100 random realizations per data point.



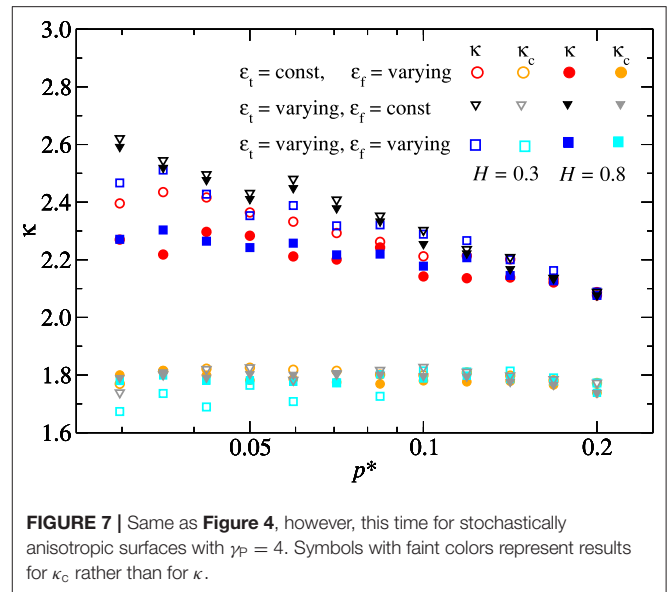
to be partially inconsistent, e.g., in their Figure 10D, the points $(\Phi_N, \kappa) = (700, 1.93)$ and $(70, 2.05)$ should be moved to $(70, 1.93)$ and $(700, 2.05)$, respectively. The Nayak parameter Φ_N would then be consistent not only with our own calculations but also with the values that Yastrebov et al. reported themselves in their Figure 1. Once these two data points are corrected, the logarithmic relation seems to be much less convincing than with improperly plotted data.

To provide an independent test of the extent with which the Nayak parameter affects the proportionality coefficient κ , we investigated a broader range of surfaces than before and contrast surfaces with cutoff to those with smooth and abrupt rollofs. **Figure 6** summarizes the results, which were averaged over up to 400 random realizations per data point.

Our data appears to be consistent with a $\kappa(\Phi_N) = \kappa(\infty) - c_N \Phi_N^{-\nu_N}$ relation rather than with a logarithmic $\kappa(\Phi_N)$ dependence. If the original data had been plotted without taking averages, the $H = 0.3$ might have scattered sufficiently much to make them appear as a continuation of the $H = 0.8$ data. Two markedly different dependencies of κ on Φ_N are obtained for $H = 0.3$ and $H = 0.8$. Thus, κ cannot be reduced to be a function of Φ_N (and p^*) alone.

4.3. Anisotropic rpa Surfaces

To address the question how anisotropy affects the relative contact area, we repeated the simulations presented in **Figure 4** with a Peklenik number of $\gamma_P = 4$. Results are shown in **Figure 7**. It can be seen that anisotropy enhances the pressure dependence of κ at fixed values of $\varepsilon_{t,f}$ compared to stochastically isotropic surfaces with $\gamma_P = 1$. This may not be particularly surprising in light of the observation that one-dimensional surfaces have



logarithmic corrections to the $\kappa(p)$ relation, even for $H > 0.5$. By using a Peklenik number clearly differing from unity, roughness dominates along one spatial dimension, whereby the spatial dimension can be argued to have been reduced to a certain extent.

Figure 7 also finds that $\kappa(H = 0.8)$ is not very pressure sensitive when ε_f is decreased with pressure so that for macroscopic systems, in which ε_f is two to three orders of magnitude smaller than in simulations, the pressure sensitivity is marginally small. However, due to anisotropy, κ is noticeably increased with respect to the isotropic case. When Peklenik numbers differ very much from unity, different laws may apply as the surface's dimensionality has effectively changed from two to one. In the limit of one-dimensional surfaces, the pressure sensitivity of κ at small p^* has been convincingly established not only for small H but also for $H = 0.8$ (van Dokkum et al., 2018).

As a final comment on the non-isotropic rpa surfaces, we note that κ_c , whose definition of reduced pressure uses the root-mean-square height gradient \bar{g} averaged over contact only, has a rather weak dependence on p^* . Values are generally close to 1.8. Interestingly, the ordering of the points are essentially in reverse order compared to the analysis in which \bar{g} was averaged over entire surfaces.

The smallest κ_c occurs for the smallest Hurst exponent, which can be rationalized as follows: For $H = 0.3$, roughness exists predominantly at short wavelengths and contact patches are rather small compared to $H = 0.8$. The coarse-grained, or rotationally averaged height profile of an individual meso-scale asperity is therefore blunter for $H = 0.8$ (where many valleys are sampled over) than for $H = 0.3$ (containing essentially only high peaks). Neglecting the fact that contact is more and more partial in such a mesoscale asperity with distance from the highest point, Equation (28) finds that blunter tips have larger κ_c . The extreme limit at small p^* should be a Hertzian asperity, for which $\kappa_c \approx 1.666$. The blunter the profile, the larger κ_c .

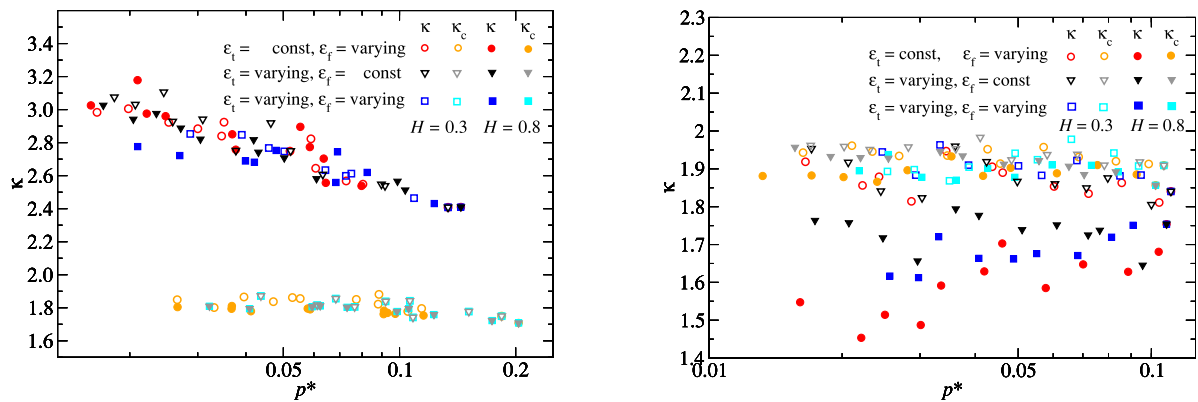


FIGURE 8 | Same as **Figure 4**, however, this time for a height warped surfaces: **(Left)** blunt indenting peaks and steep valleys ($w = 2$) **(Right)** sharp peaks and shallow valleys ($w = -2$).

4.4. Isotropic Height-Warped Surfaces

The assumption of the random-phase approximation is subject to the serious and legitimate criticism not only of Persson's contact mechanics theory but on a large body of simulation studies using ideal, rpa surfaces. Quite a few numerical studies use a Weierstrass profile, which has (perfect) phase correlation, while reproducing a height autocorrelation function (ACF) being similar to experimental ACFs. However, the Weierstrass profile lacks any visual similarity to experimental surfaces as demonstrated in **Figure 2** of Müser (2018). We therefore believe that the warping method proposed in section 2, while probably still being far from ideal, reproduces the stochastic properties of correlated surfaces in a significantly more realistic fashion than a Weierstrass-function based height profile.

Figure 8 reveals that the increase of κ with decreasing pressure is much accentuated for a positive warping exponent w , in which case the peaks of the indenting asperities are flattened compared to the valleys. This times $\kappa(H = 0.8)$ even increases with decreasing p^* when ε_f is reduced with decreasing pressure. However, for the opposite case of $w < 0$ leading to sharp asperities and shallow valleys, κ is found to decrease with pressure. This statement also holds in certain pressure ranges, when ε_f is not scaled according to Equation (27) but kept constant.

When defining the root-mean-square height gradient and thus the reduced pressure with respect to the true contact area, κ_c turns out yet again to be rather insensitive to the reduced pressure and to be around 1.8 for small p_c^* . Correlating the respective values of κ_c with the structural parameters, which are symmetry-allowed and finite, has remained unsuccessful so far.

4.5. Periodically Repeated Smooth Indenters

When indenters are periodically repeated, each indenter carries the same load. If the linear contact dimension is small compared to the period, that is, at small applied external pressures p_0 , a similar relation between contact area and load or pressure must be obtained as if the indenter were isolated. As already mentioned

in the introduction, it has been noticed recently (Müser, 2017) that the corresponding asymptotic low-pressure relation for periodically repeated indenters with harmonic height profiles can be cast in terms of Equation (8). The prefactor κ_c can be calculated analytically, specifically

$$\kappa_c(n) = \sqrt{\frac{\pi}{n}} \frac{\Gamma(\frac{n}{2} + \frac{3}{2})}{\Gamma(\frac{n}{2} + 1)}, \quad (28)$$

where $\Gamma(\bullet)$ represent the gamma function. The numerical data shown in **Figure 9** confirms the analytical results at low pressures. Errors in the relative contact become noticeable only once a_r exceeds 0.3, but they remain below 25%. The high-pressure asymptotics can also be described reasonably well with Equation (8), however, the value for κ_c needs to be decreased. Since the approach to full contact is the quite special case of a conical dimple for the periodically repeated indenters, it will not be investigated any further in this study.

5. DISCUSSIONS AND CONCLUSIONS

In this study, various structural parameters determining the relative contact area a_r in a contact between a rough surface and a linearly elastic counterbody were investigated. The focus was laid on the questions if a_r is linear in pressure p and inversely proportional to the root-mean square height gradient \bar{g} for small reduced pressures, defined as $p^* = p/(E^*\bar{g})$, and if yes, what structural parameters determine the proportionality coefficient κ .

One of the difficulties to determine whether or not a_r is linear in p^* at small p^* is that taking the limit $p^* \rightarrow 0$ properly is not a simple task, because the ratios of roll-off wavelength and system size, $\varepsilon_t = \lambda_r/\mathcal{L}$ (size scaling), as well as the ratio of short-wavelength cutoff and roll-off wavelength, $\varepsilon_f = \lambda_s/\lambda_r$ (fractal scaling), have to be made systematically small. While Prodanov et al. (2013) emphasized the necessity of such a finite-size and fractal scaling, they failed to take that limit properly themselves, as they kept the ratio of linear mesh size and short-wavelength cutoff, $\varepsilon_c = a/\lambda_s$, fixed. In this study,

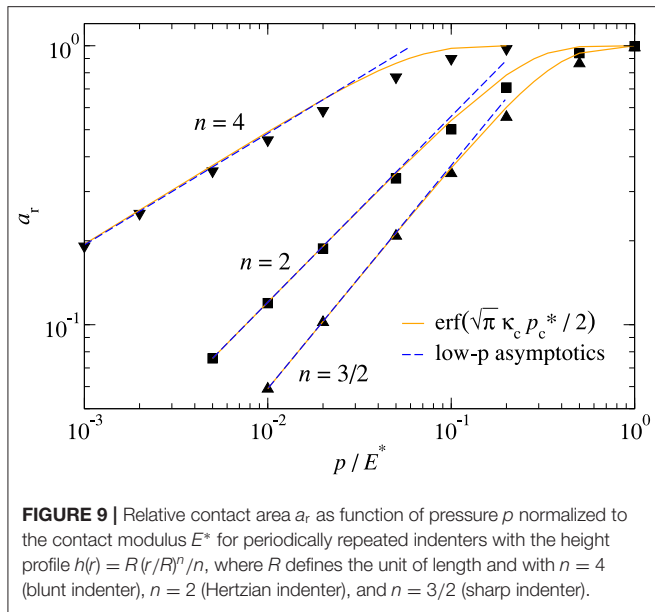


FIGURE 9 | Relative contact area a_r as function of pressure p normalized to the contact modulus E^* for periodically repeated indenters with the height profile $h(r) = R(r/R)^n/n$, where R defines the unit of length and with $n = 4$ (blunt indenter), $n = 2$ (Hertzian indenter), and $n = 3/2$ (sharp indenter).

we extrapolated results for a specific surface configuration to ε_c first and then computed contact area while taking ε_t and ε_f to zero simultaneously or by taking ε_f to zero while keeping ε_t constant.

The last type of analysis, in which effectively $\varepsilon_c = 0$, while ε_t is kept constant and p^* is set to a (very) small value, deserves particular attention. When doing so, only a single meso-scale asperity contact (SMAC) will ultimately remain in contact for very small ε_f and Hurst exponents $H > 0.5$. This is because typical contact patch sizes increase algebraically with decreasing ε_f for $H > 0.5$ (Müser and Wang, 2018). In an individual SMAC, which can be described as a single asperity with microscale roughness added to it, linearity between load and contact area is well satisfied (Pastewka and Robbins, 2016) and the (complementary) contact area in it can be well-determined using Persson theory (Müser, 2016). As such, linearity between load and contact area in a macroscopic system should arise automatically. However, it remains to be understood why one-dimensional (1D) interfaces behave differently from 2D interfaces, since 1D interfaces do not obey a linear $a_r(p^*)$ relationship at small pressures (van Dokkum et al., 2018), even for very large systems. Moreover, we are not yet certain about the small-pressure $a_r(p^*)$ relationship for $H < 0.5$, although our current analysis supports the finding (Yastrebov et al., 2015, 2017) that the relationship may have indeed non-negligible logarithmic corrections in p^* . They might be the consequence of the small, logarithmic growth of characteristic contact patch sizes with decreasing ε_f for $H < 0.5$. At the same time, we wonder if κ computed in the thermodynamic limit can systematically exceed predictions of the more advanced bearing-area models such as Bush, Gibson, and Thomas (BGT) (Bush et al., 1975). Thus, although we believe to have furthered the rigor with which κ is computed, we expect that the final answer to how κ has to be computed in the thermodynamic limit still needs to be found.

The discussion in the last paragraph as well as the analysis conducted in the results section of this work reveals already at the present level of rigor that the Nayak parameter Φ_N has no stringent direct correlation with the proportionality coefficient κ , which would allow the function $\kappa(p^*, H, \varepsilon_t, \varepsilon_f)$ to be reduced to a smaller number of variables, such as, $\kappa = \kappa(p^*, \Phi_N, \varepsilon_t)$. This in turn emphasizes once more that a rigorous understanding of contact mechanics necessitates a spectral analysis of the height profiles. Knowledge of the Nayak parameter and even more so of simple distribution functions of asperity heights and geometries is unavailing.

While Persson theory cannot (yet) be used either to explain why different Hurst exponents lead to different κ , it does allow deviations of $\kappa(p^*)$ from linearity to be rationalized for both finite systems and surfaces violating the random-phase approximation. The basic version of Persson theory assumes that the elastic body “feels” the full root-mean-square gradient (averaged over the entire surface) as soon as the elastic body hits the rough substrate. However, for any *finite* surface, a certain fraction must be in contact before the root-mean-square gradient and other stochastic parameters, such as the kurtosis, approach their “true” mean values. While this fraction decreases with system size, \bar{g} (typically) remains below its asymptotic value for finite rpa surfaces at small a_r so that (according to Persson theory and simulations presented in this work) a_r turns out larger than in the thermodynamic limit. In the case of correlated random roughness, the situation is more complex, since the rms-gradient, kurtosis, etc., can be functions of the height even when surfaces are in the thermodynamic limit. A possible correction of Persson theory for this case could be to identify the rms-gradient of the $a_r \times 100\%$ top- (or bottom) most heights and use this value to determine the reduced pressure p_c^* , which would then satisfy Equation (8) reasonably well. To some extent, this would constitute a somewhat dissatisfactory compromise between Persson theory and bearing-area models, since it is not the top- (or bottom) most, say, 20% of the peaks that are in contact at 20% relative contact area, as is implicitly assumed in bearing-area models. However, this is the simplest correction that comes to our mind at this point of time. It is certainly much less tedious to work out than the systematic corrections to Persson theory relying on a cumulant expansion of the short- but finite-range repulsive interactions between indenter and elastic body (Müser, 2008).

In full simulations, \bar{g} can be averaged over the true contact area and no compromise between bearing-area models and Persson theory needs to be made. In all investigated randomly-rough contacts, we find a close-to-linear relation between a_r and p_c^* , i.e., when averaging the rms height gradient only over the true contact even if the original $a_r(p)$ deviates clearly from linearity. In these simulations, we find κ_c to lie in the relatively narrow range satisfying $\kappa_c \approx 1.8 \pm 0.1$. This value for κ_c is only slightly larger than the value of 1.6 predicted by Persson theory but clearly below the value of 2.5 predicted by BGT (Bush et al., 1975). Thus, the range of validity of Persson theory could be substantially expanded if the approximation of using the full rms-height gradient were replaced with an accurate estimate of the mean rms-height gradient in the true contact.

It is certainly justified to consider many parts of this study to be more of a mathematical exercise rather than an attempt to model any specific real-world problem. And anyone wondering why there has to be yet another paper on κ can be assured to have our full sympathy. However, if poorly conducted studies of contact problems lead to the conclusion that a legitimate theory is brought into discredit and these studies receive many citations as evidence for the theory to be flawed, it should be in place to demonstrate that a carefully conducted analysis supports the theory.

REFERENCES

- Afferrante, L., Bottiglione, F., Putignano, C., Persson, B. N. J., and Carbone, G. (2018). Elastic contact mechanics of randomly rough surfaces: an assessment of advanced asperity models and Persson's theory. *Tribol. Lett.* 66:75. doi: 10.1007/s11249-018-1026-x
- Almqvist, A., Campaña, C., Prodanov, N., and Persson, B. (2011). Interfacial separation between elastic solids with randomly rough surfaces: comparison between theory and numerical techniques. *J. Mech. Phys. Solids* 59, 2355–2369. doi: 10.1016/j.jmps.2011.08.004
- Bush, A., Gibson, R., and Thomas, T. (1975). The elastic contact of a rough surface. *Wear* 35, 87–111. doi: 10.1016/0043-1648(75)90145-3
- Campaña, C., and Müser, M. H. (2006). Practical green's function approach to the simulation of elastic semi-infinite solids. *Phys. Rev. B* 74:075420. doi: 10.1103/PhysRevB.74.075420
- Campaña, C., and Müser, M. H. (2007). Contact mechanics of real vs. randomly rough surfaces: a Green's function molecular dynamics study. *Europhys. Lett.* 77:38005. doi: 10.1209/0295-5075/77/38005
- Campaña, C., Müser, M. H., and Robbins, M. O. (2008). Elastic contact between self-affine surfaces: comparison of numerical stress and contact correlation functions with analytic predictions. *J. Phys.* 20:354013. doi: 10.1088/0953-8984/20/35/354013
- Campaña, C., Persson, B. N. J., and Müser, M. H. (2011). Transverse and normal interfacial stiffness of solids with randomly rough surfaces. *J. Phys.* 23:085001. doi: 10.1088/0953-8984/23/8/085001
- Candela, T., Renard, F., Bouchon, M., Schmittbuhl, J., and Brodsky, E. E. (2011). Stress drop during earthquakes: effect of fault roughness scaling. *Bull. Seismol. Soc. Am.* 101, 2369–2387. doi: 10.1785/0120100298
- Carbone, G., and Bottiglione, F. (2008). Asperity contact theories: do they predict linearity between contact area and load? *J. Mech. Phys. Solids* 56, 2555–2572. doi: 10.1016/j.jmps.2008.03.011
- Dapp, W. B., Lücke, A., Persson, B. N. J., and Müser, M. H. (2012). Self-affine elastic contacts: percolation and leakage. *Phys. Rev. Lett.* 108:244301. doi: 10.1103/PhysRevLett.108.244301
- Dapp, W. B., Prodanov, N., and Müser, M. H. (2014). Systematic analysis of Persson's contact mechanics theory of randomly rough elastic surfaces. *J. Phys.* 26:355002. doi: 10.1088/0953-8984/26/35/355002
- Hyun, S., Pei, L., Molinari, J.-F., and Robbins, M. O. (2004). Finite-element analysis of contact between elastic self-affine surfaces. *Phys. Rev. E* 70:026117. doi: 10.1103/PhysRevE.70.026117
- Hyun, S., and Robbins, M. O. (2007). Elastic contact between rough surfaces: effect of roughness at large and small wavelengths. *Tribol. Int.* 40, 1413–1422. doi: 10.1016/j.triboint.2007.02.003
- Jacobs, T. D. B., Junge, T., and Pastewka, L. (2017). Quantitative characterization of surface topography using spectral analysis. *Surface Topogr.* 5:013001. doi: 10.1088/2051-672X/aa51f8
- Kajita, S. (2016). Green's function nonequilibrium molecular dynamics method for solid surfaces and interfaces. *Phys. Rev. E* 94:033301. doi: 10.1103/PhysRevE.94.033301
- Lorenz, B., and Persson, B. N. J. (2008). Interfacial separation between elastic solids with randomly rough surfaces: comparison of experiment with theory. *J. Phys.* 21:015003. doi: 10.1088/0953-8984/21/1/015003
- Majumdar, A., and Tien, C. L. (1990). Fractal characterization and simulation of rough surfaces. *Wear* 136, 313–327. doi: 10.1016/0043-1648(90)90154-3
- Müser, M., and Wang, A. (2018). Contact-patch-size distribution and limits of self-affinity in contacts between randomly rough surfaces. *Lubricants* 6:85. doi: 10.3390/lubricants6040085
- Müser, M. H. (2008). Rigorous field-theoretical approach to the contact mechanics of rough elastic solids. *Phys. Rev. Lett.* 100:055504. doi: 10.1103/PhysRevLett.100.055504
- Müser, M. H. (2016). On the contact area of nominally flat hertzian contacts. *Tribol. Lett.* 64:14. doi: 10.1007/s11249-016-0750-3
- Müser, M. H. (2017). On the linearity of contact area and reduced pressure. *Tribol. Lett.* 65:129. doi: 10.1007/s11249-017-0912-y
- Müser, M. H. (2018). Response to “comment on meeting the contact-(mechanics) challenge.” *Tribol. Lett.* 66:38. doi: 10.1007/s11249-018-0986-1
- Müser, M. H., Dapp, W. B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T. A., et al. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65:118. doi: 10.1007/s11249-017-0900-2
- Nayak, P. R. (1971). Random process model of rough surfaces. *J. Lubricat. Technol.* 93, 398–407. doi: 10.1115/1.3451608
- Palasantzas, G. (1993). Roughness spectrum and surface width of self-affine fractal surfaces via the k-correlation model. *Phys. Rev. B* 48, 14472–14478. doi: 10.1103/PhysRevB.48.14472
- Pastewka, L., Prodanov, N., Lorenz, B., Müser, M. H., Robbins, M. O., and Persson, B. N. J. (2013). Finite-size scaling in the interfacial stiffness of rough elastic contacts. *Phys. Rev. E* 87:062809. doi: 10.1103/PhysRevE.87.062809
- Pastewka, L., and Robbins, M. O. (2016). Contact area of rough spheres: large scale simulations and simple scaling laws. *Appl. Phys. Lett.* 108:221601. doi: 10.1063/1.4950802
- Pastewka, L., Sharp, T. A., and Robbins, M. O. (2012). Seamless elastic boundaries for atomistic calculations. *Phys. Rev. B* 86:075459. doi: 10.1103/PhysRevB.86.075459
- Persson, B. N. J. (2001). Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115:3840. doi: 10.1063/1.1388626
- Persson, B. N. J. (2008). On the elastic energy and stress correlation in the contact between elastic solids with randomly rough surfaces. *J. Phys.* 20:312001. doi: 10.1088/0953-8984/20/31/312001
- Persson, B. N. J. (2014). On the fractal dimension of rough surfaces. *Tribol. Lett.* 54, 99–106. doi: 10.1007/s11249-014-0313-4
- Persson, B. N. J., Albohr, O., Tartaglino, U., Volokitin, A. I., and Tosatti, E. (2004). On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *J. Phys.* 17, R1–R62. doi: 10.1088/0953-8984/17/1/R01
- Pohrt, R., Popov, V. L., and Filippov, A. E. (2012). Normal contact stiffness of elastic solids with fractal rough surfaces for one- and three-dimensional systems. *Phys. Rev. E* 86:026710. doi: 10.1103/PhysRevE.86.026710
- Prodanov, N., Dapp, W. B., and Müser, M. H. (2013). On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections, and reference data. *Tribol. Lett.* 53, 433–448. doi: 10.1007/s11249-013-0282-z
- Putignano, C., Afferrante, L., Carbone, G., and Demelio, G. (2012). The influence of the statistical properties of self-affine surfaces in elastic contacts: a numerical

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MM designed the study and wrote the manuscript. YZ run the simulations. MM and YZ jointly analyzed the data and jointly produced the figures. All authors contributed to the article and approved the submitted version.

- investigation. *J. Mech. Phys. Solids* 60, 973–982. doi: 10.1016/j.jmps.2012.01.006
- Salehani, M. K., van Dokkum, J., Irani, N., and Nicola, L. (2020). On the load-area relation in rough adhesive contacts. *Tribol. Int.* 144:106099. doi: 10.1016/j.triboint.2019.106099
- van Dokkum, J. S., and Nicola, L. (2019). Green's function molecular dynamics including viscoelasticity. *Modell. Simul. Mater. Sci. Eng.* 27:075006. doi: 10.1088/1361-651X/ab3031
- van Dokkum, J. S., Salehani, M. K., Irani, N., and Nicola, L. (2018). On the proportionality between area and load in line contacts. *Tribol. Lett.* 66:115. doi: 10.1007/s11249-018-1061-7
- Wick, G. C. (1950). The evaluation of the collision matrix. *Phys. Rev.* 80, 268–272. doi: 10.1103/PhysRev.80.268
- Yastrebov, V. A., Anciaux, G., and Molinari, J.-F. (2015). From infinitesimal to full contact between rough surfaces: evolution of the contact area. *Int. J. Solids Struct.* 52, 83–102. doi: 10.1016/j.ijsolstr.2014.09.019
- Yastrebov, V. A., Anciaux, G., and Molinari, J.-F. (2017). The role of the roughness spectral breadth in elastic contact of rough surfaces. *J. Mech. Phys. Solids* 107, 469–493. doi: 10.1016/j.jmps.2017.07.016
- Zhou, Y., Moseler, M., and Müser, M. H. (2019). Solution of boundary-element problems using the fast-inertial-relaxation-engine method. *Phys. Rev. B* 99:114103. doi: 10.1103/PhysRevB.99.144103

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Zhou and Müser. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Continuum Model Analysis of QCM Nanotribological Data to Obtain Friction Coefficients for 304SS Contacts Lubricated by Water and TiO₂ Nanoparticle Suspensions

Caitlin M. Seed, Biplav Acharya and Jacqueline Krim*

Department of Physics, North Carolina State University, Raleigh, NC, United States

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Guoxin Xie,
Tsinghua University, China
Qunyang Li,
Tsinghua University, China
Motohisa Hirano,
Hosei University, Japan

*Correspondence:

Jacqueline Krim
jkrim@ncsu.edu

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 01 June 2020

Accepted: 27 July 2020

Published: 08 September 2020

Citation:

Seed CM, Acharya B and Krim J
(2020) Continuum Model Analysis of
QCM Nanotribological Data to Obtain
Friction Coefficients for 304SS
Contacts Lubricated by Water and
TiO₂ Nanoparticle Suspensions.
Front. Mech. Eng. 6:72.
doi: 10.3389/fmech.2020.00072

We report a study of the response of a Quartz Crystal Microbalance (QCM) to rubbing contacts in air, water and aqueous suspensions of 40 nm TiO₂ nanoparticles. Measurements were performed with a contact comprised of 3 close-packed 304SS ball bearings situated symmetrically about the center of a 304SS QCM electrode with 2 nm rms roughness. Two continuum methods were employed to infer macroscale friction coefficients μ employing QCM nanotribological data recorded in the Cattaneo-Mindlin (CM) slip regime at vibrational amplitudes that varied between 1 and 17 nm. The “slope” Method 1 involved sweeps of the QCM amplitude of vibration as ball bearings were held in continuous contact with the oscillating electrode. The “contact” Method 2 obtained μ by analyzing the shifts in frequency and bandwidth that occur at a fixed u_0 to solve for μ . when ball bearings were brought in and out of contact with the QCM’s electrode. The results for dry and water lubricated contacts compared favorably with macroscale friction coefficients reported in the literature. The model failed to adequately describe contacts lubricated with the NP suspension, but its continuum nature did not appear to be the dominant factor underlying failure. The failure was more likely attributable to either a lack of a CM slip regime when NP were present at the interface and/or the fact that the amplitude of vibration was close in size to the individual NP contacting regions, in violation of a key underlying assumption of the model.

Keywords: friction, QCM, cattaneo-mindlin, nanotribology, contact mechanics

INTRODUCTION

Quartz Crystal Microbalance (QCM) studies of tribological contacts (Rodahl and Kasemo, 1996; Laschitsch and Johannsmann, 1999; Brizmer et al., 2007; Johannsmann, 2007; Dawson et al., 2009; Krim, 2012; M’bougui et al., 2014; Borovsky et al., 2017) at both the macro and nanoscale are of increasing importance in a wide range of nano-technological and energy-related applications (Braiman et al., 2003; Kim and Kim, 2009; Zhang and Li, 2010; Krim, 2012; Hsu et al., 2014). Literature reports in this area have, to date, focused primarily on either macroscale or nanoscale contacts, and a remaining challenge in the field of tribology is to establish linkage between studies performed at wide-ranging length and time scales. One approach to establishing this linkage is to perform studies that examine the limits of macroscopic continuum mechanics, by probing where

continuum methods break down as the atomic scale is approached. Nanoparticle (NP) additives are attractive candidates for such studies, as they significantly alter friction and wear at macroscopic length scales when present in a lubricated contact (Yu et al., 2016).

We report here a study of the response of QCMs with 304 Stainless Steel (304SS) sensing electrodes to rubbing contact with 304SS ball bearings, employing a continuum model to analyze the response in air, water and an aqueous TiO_2 suspensions. The QCM technique yields real-time nanotribological information at macroscale contact velocity amplitudes, typically 0.1–1 m/s. It is comprised of a thin piezoelectric quartz crystal driven at its resonance frequency by applying an AC voltage to thin metal film electrodes deposited onto each of its major faces. Contact forces and/or changes in the environment surrounding the QCM's surface electrode produce shifts in both its resonance frequency f , and quality factor Q , and the response of these quantities to probes, biological agents, and/or granular materials has been extensively studied (Borovsky et al., 2001, 2007; Krim, 2012; Dultsev and Nekrasov, 2018; Dirri et al., 2019; Han et al., 2019; Wang et al., 2019). Analysis approaches for inferring macroscopic friction coefficients from such measurements (Berg and Johannsmann, 2003; Berg et al., 2003; Leopoldes and Jia, 2010; Hanke et al., 2013; Vlachová et al., 2015; Borovsky et al., 2017, 2019) have been largely developed within the context of continuum contact models. The models must eventually break down as nanoscale length scales are approached, where systems become less uniform in nature and/or as contact regions become more poorly defined (**Figure 1**).

Vlachová et al. (2015) recently utilized a Cattaneo-Mindlin (CM) slip scenario to relate contact stiffness obtained from QCM measurements to macroscale friction coefficients, employing an analysis approach developed by Johannsmann and coworkers for QCM's loaded with spherical contacts (Berg and Johannsmann, 2003; Berg et al., 2003; Hanke et al., 2013). The CM scenario is an approximation that allows for a simplified analysis framework for two curved surfaces that are pressed with a normal load F_N into contact and then sheared tangentially (Etsion, 2010; Paggi et al., 2014; Ciavarella, 2015). Such bodies commonly remain stuck to each other in one part of the contact while slipping in other regions. The CM approximation treats the contact as rigid indenter pressed into a flat elastic surface, and assumes that sliding friction in the slip region is governed by a "Coulomb" friction law, $F_f = \mu_k F_N$.

Berg and Johannsmann (2003), Berg et al. (2003), Hanke et al. (2013), and Vlachová et al. (2015) analysis method is based on both the linear and non-linear contact mechanics regimes associated with partially slipping contacts undergoing small amplitude reciprocal motion. It assumes that the contact area diameter is both independent of the tangential loading force and much smaller than the loading sphere diameter. Vlachová et al. obtained realistic values for friction coefficients μ_k by analyzing QCM frequency and bandwidth shifts, Δf and $\Delta\Gamma = (f/2)\Delta(Q^{-1})$, as a function of the amplitude of

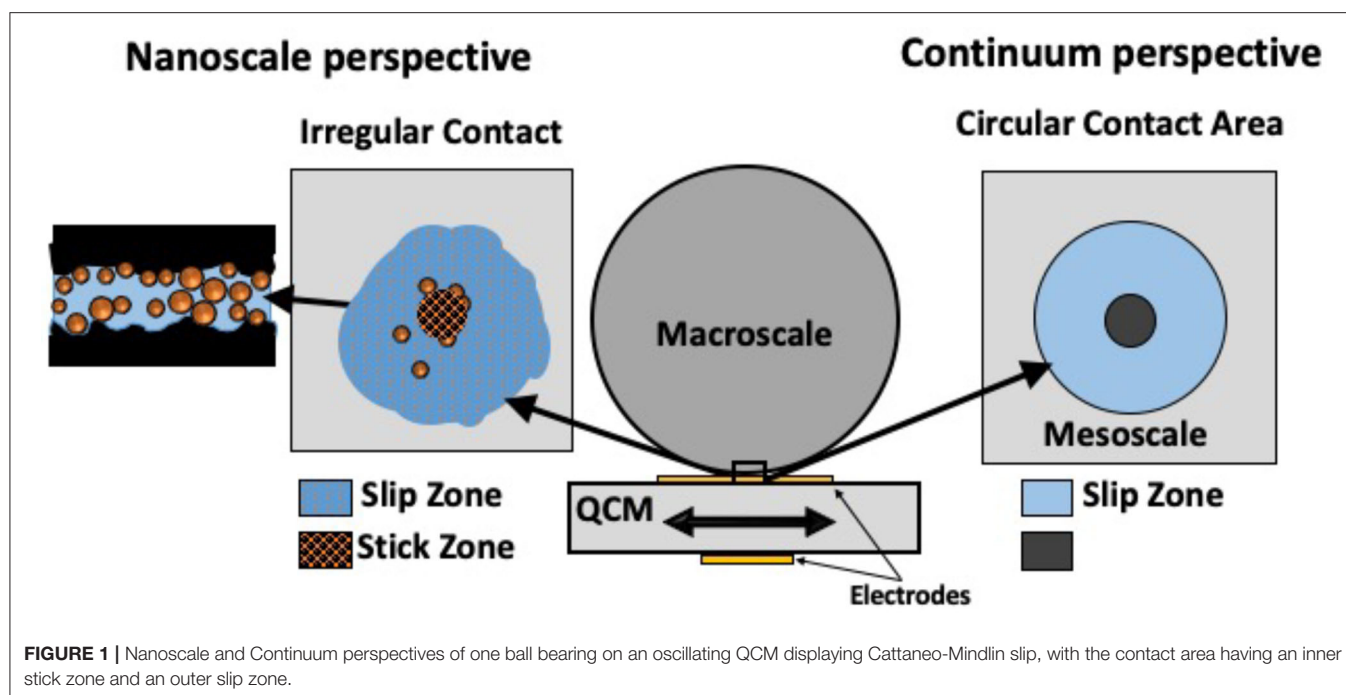
vibration for glass spheres in contact with silica or polymer-coated QCM electrodes. The values were not, however, compared to conventional macroscopic measurements of μ_k on the same materials combinations. Seed et al. (2020) very recently employed the Vlachová et al. (2015), Berg and Johannsmann (2003), Berg et al. (2003), and Hanke et al. (2013) analysis approach to infer macroscale friction coefficients μ for QCMs loaded with stainless steel spherical contacts, and also measured μ in conventional macroscale setups for the same materials combinations. They found that the values agreed well when the dependence of $\Delta\Gamma$ on vibrational amplitude was utilized. Overall, Seed et al. validated the combined assumptions of the continuum analysis methods, and demonstrated that the Vlachová et al. (2015), Berg and Johannsmann (2003), Berg et al. (2003), and Hanke et al. (2013) approach was viable for linkage macro and nanoscale tribological measurements (Seed et al., 2020).

The study reported by Seed et al. was performed in air on unlubricated stainless steel contacts, and the analysis was performed from a continuum perspective (Seed et al., 2020). We employ it here to examine both dry and lubricated contacts, with a particular focus on the degree to which such a continuum model can be applied to a complex liquid-nanoparticle-solid interface. TiO_2 suspensions were selected for study, as they have been reported on extensively in the tribological literature and are known to reduce friction in stainless steel contacts at both the nanoscale and macroscale (Glavatskih and Höglund, 2008; Ali et al., 2016; Acharya et al., 2019a; Wu et al., 2020). TiO_2 suspensions are also of interest in tribotronics ["active" or "smart" control of friction by combining mechanical elements with electronics (Glavatskih and Höglund, 2008; Zhang and Wang, 2016; Krim, 2019)], since they are readily repositioned by external fields for purposed of tuning of friction levels (Acharya et al., 2019b; Krim, 2019). In addition, the suspensions are of interest in energy storage applications utilizing flow batteries, where interfacial tribological properties are of importance (Wang et al., 2007; Sen et al., 2017).

MATERIALS AND METHODS

Measurements were performed in air, deionized (DI) water, and DI water with suspensions of 0.67 wt% TiO_2 nanoparticles in an apparatus depicted schematically in **Figure 2**. The apparatus was comprised of a Ball Bearing (BB) configuration with three 304SS bearings arranged in a close packed triangular array centered on the QCM upper (liquid facing) electrode. The BB's made contact with the upper electrode at a distance of 2.29 mm from the center, which fell within the perimeter of the lower electrode and therefore within an area of active oscillation. The system weighed $F_N = 0.282$ N, distributed across the three contacts so as to load 0.094 N onto each of them. This loading configuration has a Hertzian contact pressure of 245 MPa in air and a contact radius of 11.05 μm under the assumption of a circular contact (Popov, 2010). Measurements were performed in reciprocal motion with maximum sliding speed amplitudes ranging between 0.03 and 0.22 m/s (liquid) or 0.53 m/s (air) at the point of contact. Under these conditions, the system falls within

Abbreviations: QCN, Quartz Crystal Microbalance; CM, Cattaneo-Mindlin; SS, Stainless Steel; BB, Ball Bearing.



the boundary lubrication regime where direct contact between asperities supports the load (Lu et al., 2006; Yamaguchi and Hokkirigawa, 2016). The liquid nonetheless penetrates the gaps in contact associated with roughness, allowing introduction of NP. For the liquid measurements, 0.6 mL of either DI water or the TiO_2 suspension was added to evenly cover the QCM, resulting in the tripod BBs being slightly more than half submerged. The associated buoyant force of 0.00086 N was significantly smaller than the normal force, but was nonetheless included in the model calculations. The system components were rigidly attached to each other, and, in contrast to the system employed by Vlachová et al. were close packed. Differences in the results associated with the different experiment geometries are however expected to be minimal, as the tangential forces and local contact geometries were very similar.

Materials

Anatase TiO_2 (titania) nanoparticles with density, Zeta potential, and radius, respectively, $=4.23 \text{ g/cm}^3$, -32.7 mV and 20 nm were obtained from US Research Nanomaterials (stock number: US7071 Houston, TX 77084, USA) at concentration $\sim 20 \text{ wt\%}$ and were diluted to a concentration of 0.67 wt\% . The NP suspension, as prepared, had a pH of 7.7 and density of 1.00367 g/cm^3 . NP aggregation was not observed, but some sedimentation was observed for samples left unstirred for periods of 12–24 h. The sedimentation was readily reversible by stirring and sonication and the suspensions were stirred and sonicated for 10–15 min immediately before each experimental run, with measurements being completed within 60 min thereafter. All data were recorded within 48 h of initial preparation of the suspension.

Bulk 0.2 g 5/32" diameter Grade 100 (G100) 304SS bearings with maximum rms surface roughness $5.0 \mu\text{m}$ (127 nm) were attached to the underside of a slider, as depicted in **Figure 2**. They were glued onto the holder using Gorilla Super Glue (Gorilla Glue Company, Sharonville, OH, USA).

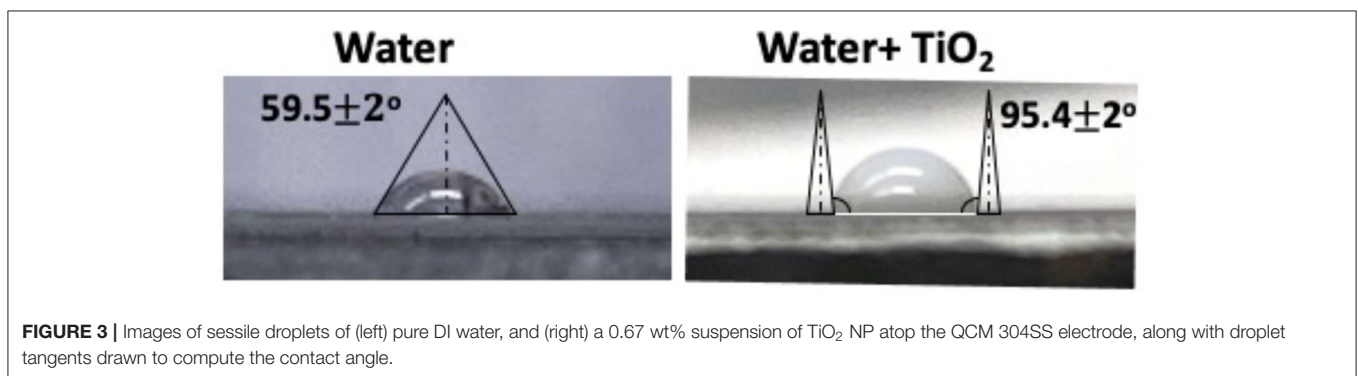
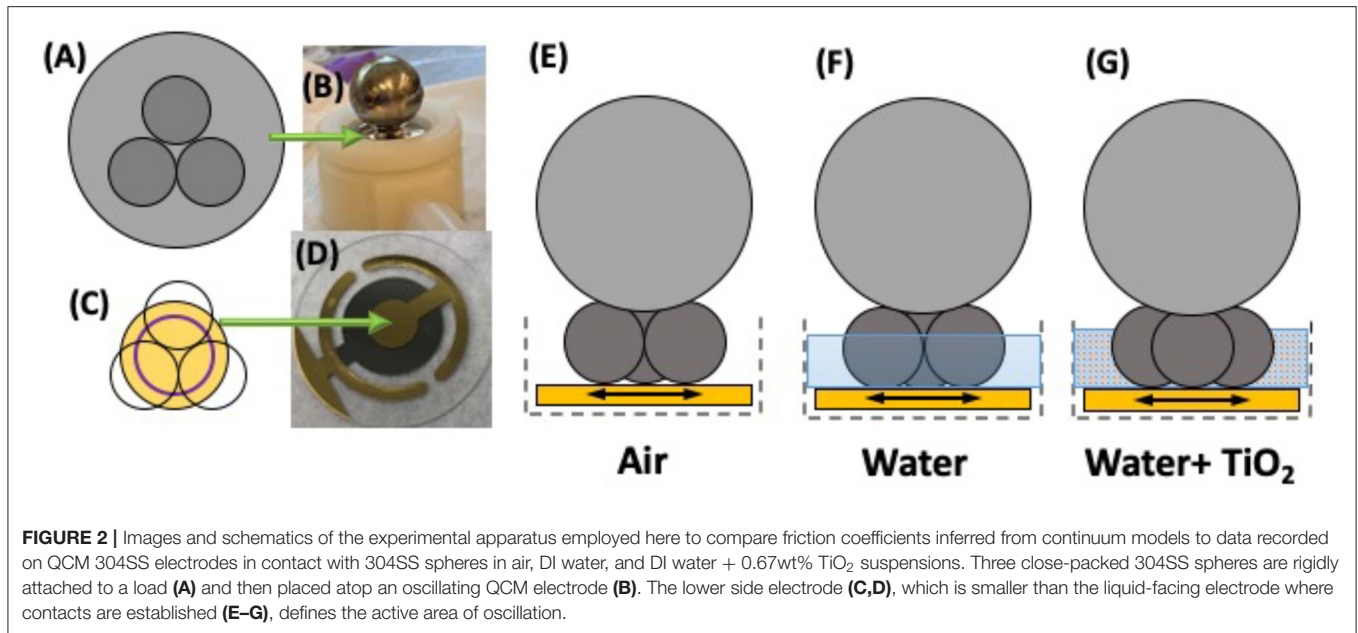
The QCM crystals employed for the studies were AT-cut, 5 MHz, one-inch diameter QCM crystals with 304SS electrodes (FILTECH, Inc. part no. QM1022, Boston, MA). The crystals had half-inch diameter surface electrodes made of 304SS and quarter inch backside electrodes made of Au. The 304SS electrodes consisted of 200 nm 304SS films deposited atop 100 nm thick Au films deposited on 50 nm thick Ni adhesion layers, with rms surface roughness $2 \pm 1 \text{ nm}$ (Acharya et al., 2018).

Methods

Contact Angle

Contact angle measurements were performed as an independent indication of NP suspension potential for lubricity. The measurements were performed on the actual samples used for the measurements, as contact angles, and also lubricating properties, are highly dependent on surface roughness, NP concentrations, NP size and interfacial chemical compositions (Wu et al., 2020). Contact angles were measured for both DI water and TiO_2 suspensions by using a pipette to place a $10 \mu\text{L}$ drop onto the 304SS QCM electrode. An image was used to determine the contact angle on both sides of the droplets by drawing a right triangle where the base was the surface from the edge of the droplet to the center, and a line was drawn tangent to the droplet edge, forming the hypotenuse. The value for contact angle displayed in **Figure 3** is the average of the right and left angles.

The TiO_2 suspension exhibited a higher contact angle than that for pure water, indicating that a significantly larger slip



length along with a lower friction coefficient is to be expected for contacts completely immersed in the liquid (Ellis et al., 2003; Acharya et al., 2019b). We note that this is distinct from the wetting and spreading of lubricant droplets, which also play major roles in the lubrication of dynamic mechanical systems (Wu et al., 2020).

QCM Continuum Model Analysis

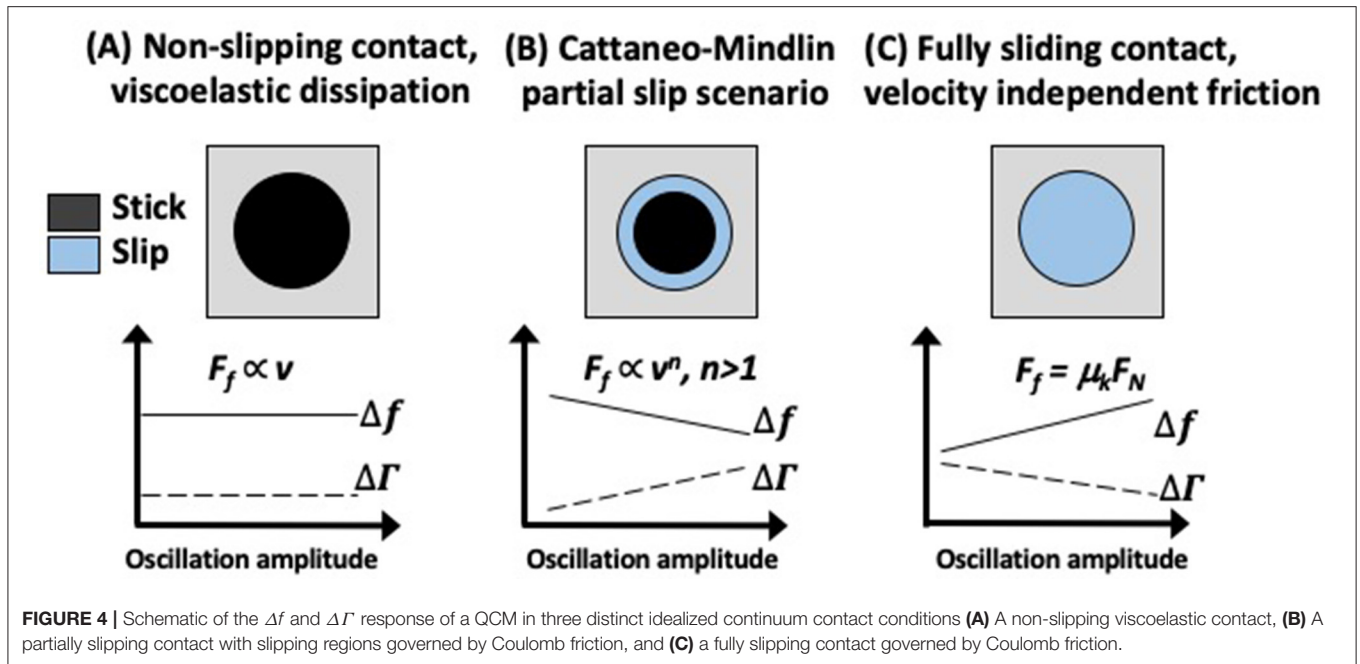
Data recorded here were analyzed according to the methods described in Berg and Johannsmann (2003), Berg et al. (2003), Hanke et al. (2013), Vlachová et al. (2015), and Seed et al. (2020), which are applicable to the CM partial slip approximation in a small amplitude reciprocal motion (Figure 1). For reciprocal motion, and as the amplitude of motion increases, the partial slip regime in this model commences at the contact's edge in the form of thin annulus. The annulus progressively grows and the system transitions to full gross slippage for a sufficiently large applied tangential force (Figure 4).

Values for μ in the slipping region of the contact are inferred from the shifts in resonance frequency, Δf , and half bandwidth at

half height (“bandwidth,” for short), $\Delta \Gamma$, of the QCM oscillating at MHz resonance frequency upon contact of a sphere with the surface (We note here that the quality factor Q , damping parameter D_f , and dissipation D , (all unitless) are all in common use to represent system dissipation. They are related to $\Delta \Gamma$, which has units of Hz, as The model developed by Johannsmann and coworkers treats the system as a composite mechanical resonator consisting of a QCM in contact with a fixed load, with contact comprised of a dashpot in parallel with a spring. On a basic level, Δf and $\Delta \Gamma$ are proportional to the in-phase and out-of-phase component of the area-averaged periodic tangential stress at the resonator surface (Berg et al., 2003), and can be inferred from expressions for the tangential force and contact area. The time dependent tangential force at the contact is written as Hanke et al. (2013):

$$F_x(t) = \kappa u(t) + \xi v(t), \quad (1)$$

where u is the tangential displacement, κ is the contact stiffness, v is the tangential sliding speed and ξ is a linear drag coefficient. For a viscoelastic contact small amplitude cyclic motion with



amplitude u_o and tangential velocity amplitude $v_o = i\omega u_o$, the contact area A_{eff} is constant, all forces depend linearly on displacement, and the complex frequency shift associated with the presence of the contact can be expressed as Laschitsch and Johannsmann (1999) and Hanke et al. (2013):

$$\Delta f + i\Delta\Gamma = \frac{n_p \kappa}{2n\pi^2 A_{eff} Z_q} (\kappa + i\omega\xi), \quad (2)$$

where A_{eff} is the effective area of the resonator, κ is the contact stiffness, n is the overtone order, F_N is the load, and $Z_q = 8.8 \times 10^6 \text{ kg}/(\text{m}^2\text{s})$ is the shear wave impedance. The quantities in Equation (2) are amplitude independent, and therefore both Δf and $\Delta\Gamma$ are constant in amplitude (**Figure 4A**). For the case of full slip governed by a velocity independent Coulomb friction law, Δf trending linearly upward and $\Delta\Gamma$ trends linearly downward as the amplitude increases (**Figure 4C**) (Vittorias et al., 2010).

For the case of a composite contact that is partially slipping, $F_x(t)$ is not time harmonic, and the analysis must encompass both linear and non-linear contact mechanics regimes with proper matching of boundary conditions. The analysis developed in Berg and Johannsmann (2003), Berg et al. (2003), Hanke et al. (2013), Vlachová et al. (2015) reveals that in this regime Δf trends downward with amplitude while $\Delta\Gamma$ trends upward (**Figure 4B**). To emphasize the linear trends, the expressions in Vlachová et al. (2015), Equation 11 can be rewritten as Seed et al. (2020):

$$\Delta f(u_o) = h - bu_o \quad (3)$$

$$\Delta\Gamma(u_o) = \frac{4}{3\pi} bu_o \quad (4)$$

$$h = \frac{n_p \kappa}{2n\pi^2 A_{eff} Z_q} \quad (5)$$

$$b = \frac{n_p \kappa^2}{6n\pi^2 A_{eff} Z_q F_N \mu}, \quad (6)$$

Where n_p is the number of BB contacts, μ is the coefficient of friction, and u_o is the amplitude of oscillation at the center of the QCM (Vlachová et al., 2015). The latter quantity is inferred from Hanke et al. (2013):

$$u_o = \frac{\alpha d_q}{2\omega e_{26} A_{eff}} i, \quad (7)$$

where the parameter α is a numerical factor set equal to 1 since the QCM is modeled as a piezoelectric parallel plate, d_q is the thickness of the QCM, $e_{26} = 9.65 \times 10^{-2} \text{ C/m}^2$ is the piezoelectric stress coefficient, i is the current, and ω is the angular resonant frequency of the QCM (Hanke et al., 2013). In addition to the above formulation, a heuristic term representing losses linear in stress $\Delta\Gamma_{off}$ can be added to the term in Equation (4), which offsets the curve by a constant (Vlachová et al., 2015).

The value for the friction coefficient μ is inferred via two distinct methods, which were referred to by Seed et al. as “Method 1” and “Method 2,” referred to herein as “slope Model 1” and “contact Model 2.” The slope Method 1 substitutes the slope of $\Delta\Gamma$ vs. u_o , as determined from an amplitude sweep measurement, into Equation (4) to solve for b . The y intercept of a linear fit to the Δf vs. u_o data meanwhile yields a value for h . The values for b and h are then substituted into Equations (5, 6) to obtain μ . The contact Method 2 utilizes the Δf and $\Delta\Gamma$ shifts that occur at a fixed u_o to solve for μ .

Both methods must be evaluated in the CM slip regime, which is a basic assumption of the model, and predicts a negative (positive) linear slope in frequency (bandwidth) with amplitude. Friction coefficients were therefore evaluated only in amplitude

regions exhibiting these linear trends, and those regions were denoted as CM slip regimes. The value of h was determined from the intercept of the Δf vs. u_o linear fit in the CM slip regime, while the value for b was obtained from the $\Delta \Gamma$ vs. u_o slope in the same regime. Since Equation 4 has a zero intercept, a requirement was in general enforced when finding the $\Delta \Gamma$ data fit, with the slope of this line used as the value of b . This condition was relaxed however for fits to the contacts lubricated by the NP suspension, as the nature of the contact geometry might be expected to change for low vibrational amplitudes.

The friction coefficient, which is a kinetic friction coefficient as determined by Seed et al. is obtained by solving Equations (5, 6) for μ in terms of h and b (Seed et al., 2020):

$$\mu = \frac{2\pi^2 n A_{eff} Z_q}{3 n_p F_N b} h^2 \quad (8)$$

For the present studies, a value of $A_{eff} = 3.165 \times 10^{-5} \text{ m}^2$ was employed for the effective area of oscillation, corresponding to the area of the smaller electrode, along with $f_0 = 5 \text{ MHz}$ for the fundamental ($n = 1$) resonant frequency, $n_p = 3$ for the number of BBs in contact with the QCM, and the acoustic impedance value $Z_q = 8.8 \times 10^6 \text{ kg/(m}^2\text{s)}$.

For the contact Method 2, the measured values of Δf and $\Delta \Gamma$ at a fixed oscillation amplitude in the CM Slip allow evaluation of μ by direct substitution of Equation (3) into Equation (4), followed by some rearrangement of terms:

$$\mu = \frac{4\beta\kappa^2}{9\pi F_N \Delta \Gamma} u_o \quad (9)$$

where β and κ are given by:

$$\beta = \frac{n_p}{2n\pi^2 A_{eff} Z_q}, \quad \kappa = \frac{\Delta f + \frac{3\pi}{4} \Delta \Gamma}{\beta} \quad (10)$$

It is noted here that the contact Method 2 is straightforward in that it requires only a single measurement of Δf and $\Delta \Gamma$ to obtain a value for μ . The analysis is only valid however if measurement is performed at an amplitude of vibration where CM slip is occurring. This can be confirmed by imaging the contact zone or alternatively performing a local amplitude sweep that confirms a negative (positive) and linear slope for Δf ($\Delta \Gamma$) vs. u_o in the region being studied.

Quartz Crystal Microbalance Apparatus and Data Recording

Measurements were performed with no contacting sphere by first installing the QCM into a Teflon holder and allowing the frequency to stabilize in air or, for the case of liquid measurements, in liquid. The tripod BB contacts were then loaded onto the QCM electrode while the system response was recorded, beginning with the minimum amplitude of vibration, and continuing on as the amplitude of vibration was increased. Variations in amplitude were achieved by introducing electrical resistance, ranging from 1,190 Ω to 0 Ω , between the QCM electrode and the oscillator circuit. This procedure was starting

with 1,190 Ω and subsequently reduced every 60 s until a value of 0 Ω was attained. A baseline resonance frequency was established by using the same process for changing the QCM amplitude on the same QCM without the BB load, and for immersed contacts in the case of the lubricated measurements. Plots of Δf and $\Delta \Gamma$ vs. u_o were then generated. The values from these curves at equal values for u_o were used to solve for μ via the contact Method 2. To take into account the fact that the amplitude of vibration at the contact points was lower in magnitude than at the center of the electrode, the amplitude at the contacts was modeled as a Gaussian decay, and the central maximum was calculated from Equation 11, according to Martin and Hager (1989), Josse and Lee (1998), and Lu et al. (2004):

$$u = u_o e^{-\frac{ar^2}{R_e^2}} \quad (11)$$

where $a = 2$ is an estimated constant, $r = 2.29 \text{ mm}$ is the distance from the center, and $R_e = 3.175 \text{ mm}$ is the radius of the active oscillation region of QCM electrode. Slopes of the plots were then utilized to evaluate μ .

A QCM100 (Stanford Research Systems, Sunnyvale, CA, USA) system with the included oscillator driving circuitry, controller, and Teflon sample holder was employed to record the QCM measurements. A frequency counter (HP 53181A, Keysight Technologies, Santa Rosa, CA) and a multimeter (Keithley 2000 Series, Tektronix, Inc., Beaverton, OR) were used to measure the QCM frequency and conductance voltage via a LabView (National Instruments, Austin, TX) data acquisition system. Conductance voltage V_c was converted to motional resistance R according to $R = 10^{(4-V_c/5)} - 75^1$, and shifts in R were converted to shifts in Γ according to Johannsmann et al. (2009):

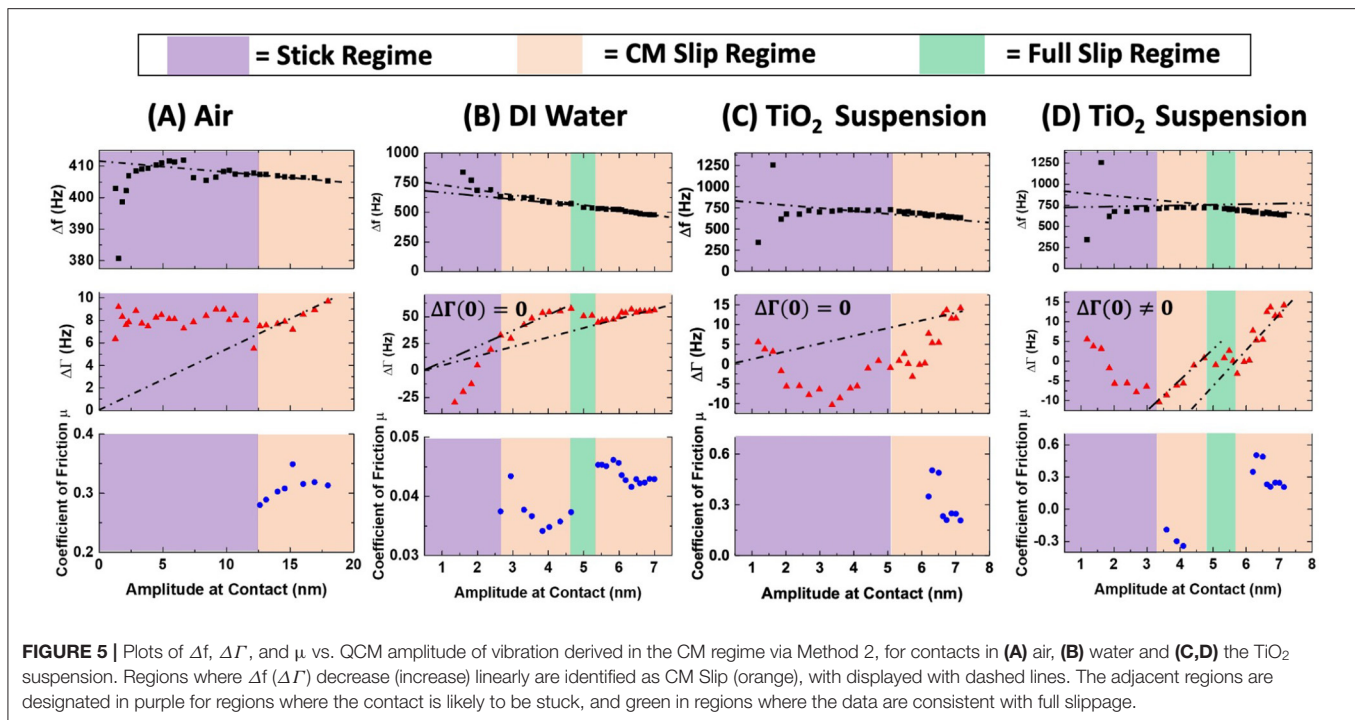
$$\Delta \Gamma = \frac{16 A_{eff} f_0^3 Z_q d_{26}^2}{\pi} \Delta R \quad (12)$$

where f_0 is the fundamental resonant frequency ($n=1$ for these studies), $Z_q = 8.8 \times 10^6 \text{ kg/(m}^2\text{s)}$ is the acoustic impedance, and $d_{26} = 3.1 \times 10^{-12} \text{ m/V}$ is the piezoelectric strain coefficient.

RESULTS

Figure 5 presents amplitude sweep data sets recorded for the response of the QCM in continuous contact with the tripod in the three environments studied. Values for μ displayed in lower panel of Figure 5 were evaluated in the respective CM regimes using the contact Method 2. All data are plotted vs. the amplitude at the actual contact (i.e., not the amplitude at the center of electrode) and range from 1 to 17 nm for air and 1 to 7 nm for the liquids, which, respectively, correspond to velocity amplitudes of 0.03–0.53 m/s and 0.03–0.22 m/s. Notably, the maximum amplitude in all cases was <40 nm diameter of the TiO₂ NP, >2 nm rms roughness of the 304SS electrodes but <127 nm maximum rms roughness of the 304SS BB's. Regions of the amplitude sweep that are consistent CM slip are shaded orange and designated

¹QCM200 Quartz Crystal Microbalance Digital Controller - QCM25 5 MHz Crystal Oscillator, Revision 2.



as “CM slip.” The amplitude region falling between zero and the first CM slip regime is shaded purple and arbitrarily labeled the “stick regime” since the contacts are likely to transition to being stuck at the lowest amplitudes. For the lubricated contacts, there are multiple regions that are consistent with CM slip (e.g., **Figure 4B**) that are separated by regions consistent with full slip (e.g., **Figure 4C**). For the water lubricated contacts the $\Delta\Gamma$ vs. amplitude data in the CM slip regimes trend toward zero at zero amplitude. It is difficult to identify a region of clear CM slip for the NP suspensions, as there is no extended region where the $\Delta\Gamma$ data exhibits a positive linear slope that extrapolates to zero.

Values for friction coefficients obtained inferred from the data displayed in **Figure 5** in the CM slip regimes are presented in **Table 1**, as well as the amplitude of vibration at which the transition from stick to a CM partial slip condition is designated. The slope Method 1 values employed the slope and intercept values from the fits displayed as black dotted lines to the $\Delta\Gamma$ data in **Figure 5**. The contact Method 2 value are inferred from the values of Δf and $\Delta\Gamma$ at fixed amplitudes, as denoted in the table. Since the NP suspension $\Delta\Gamma$ data do not trend toward zero at zero amplitude, the slope Method 1 was also applied by employing fits that were either required or not required to have a zero intercept. Relaxing the zero-intercept condition resulted in a significantly lower value for μ that was also more realistic, as discussed next.

DISCUSSION

Prior measurements employing the same experimental configuration have indicated the amplitudes inferred from

Equation 11 may be underestimated, causing the values for friction coefficients reported in **Table 1** to be slightly underestimated by $\sim 25\%$ (Seed et al., 2020). Inspection of the results in **Figure 5** and **Table 1** for contacts in air, nonetheless reveals values of $\mu \sim 0.3$, which are in close agreement in close agreement with conventional ball on disk macroscale measurements (Seed et al., 2020). The continuum model therefore successfully treats this system, irrespective of the fact that the amplitude of vibration is clearly nanoscale in extent.

For the case of contacts immersed in water, the Method 1 and 2 values of $\mu \sim 0.04$ in region 1 agree closely, and are close for the second region, at $\mu \sim 0.045$ & 0.064 . The values are also of the same order as conventional ball on disk macroscale measurements for water-immersed 304SS contacts (Curtis et al., 2017), where $\mu \sim 0.1$ but falling slightly lower. Overall, the continuum model can be considered to successfully treat this system, with the QCM data revealing both a lower friction coefficient and amplitude of transition from stick to partial slip that has been reduced to less than half of the value in air.

The analysis methods yield results that are far more problematic for contact immersed in the TiO_2 suspension. TiO_2 suspensions are known to lower the macroscopic friction coefficient for pure water by as much as 20–50% (Ali et al., 2016; Pardue et al., 2018; Acharya et al., 2019a; Wu et al., 2020), and this is not reflected in the **Table 1** values. The Method 1 and 2 values of μ are moreover in clear disagreement with each other, and clear regions of CM slip are difficult to define. Realistic values for the friction are obtained by relaxing the zero-intercept condition. Without relaxed conditions, the model clearly fails upon introduction of NPs. The zero-intercept

TABLE 1 | Friction coefficients inferred from the **Figure 5** data by two analysis methods, along with the transition amplitude of vibration from stick to partial slip.

	Air	DI Water	0.67 wt% TiO ₂ $\Delta\Gamma(u=0)=0$	0.67 wt% TiO ₂ $\Delta\Gamma(u=0)\neq 0$
Slope method 1 first region	0.29 ± 0.01	0.038 ± 0.001	0.47 ± 0.05	0.054 ± 0.004
Slope method 1 second region	–	0.064 ± 0.001	–	0.092 ± 0.01
Contact method 2 first region	0.31 @ $u = 14.6$ nm	0.037 @ $u = 2.6$ nm	0.21 @ $u = 2.7$ nm	–0.14 @ $u = 3.4$ nm
Contact method 2 second region		0.045 @ $u =$ 5.5 nm		0.21 @ $u = 6.7$ nm
Stick to first CM slip amplitude	12.5 nm	2.6 nm	6 nm	3.4 nm

condition ignores viscous dissipation while calculating μ , which by the fits shown in **Figure 5**, is an appropriate assumption to make for air and DI water. However, the fit is poor in the case of TiO₂, suggesting that this assumption may no longer be applicable. It is possible that the presence of NP lowers the dissipation at low oscillation amplitudes, replacing the contact with a more complex granular interface. Once this condition is relaxed, however, the values still do not show the expected reductions to friction, indicating that there are still difficulties with applying the model.

One possible explanation for the continued analysis failure is that the NPs possibly change the mechanics at the contacts, which may violate some of the underlying assumptions. One possibility of NPs affecting assumptions is that the CM slip regime as defined may no longer be present if rolling and/or slipping NP are present at the interface. Additionally, the amplitude of vibration can no longer be assumed to be much smaller than the size of the contact, since the contact may be comprised of 40 nm sized objects. Other model assumptions that could become problematic by the addition of NPs include the sphere being in contact with a ridged, flat surface, ignoring surface roughness, and assuming the contact area and slip annulus are uniform circles (Hanke et al., 2013). These assumptions could also be exacerbated by the potential formation of NP film formation at the contact points and/or the embedding of NPs into the surfaces. A visible residue was observed on the QCM after the TiO₂ trial which remained even after rinsing the QCM with acetone and DI water, but we were unable to confirm the presence of embedded NP in the film by imaging. Complications to the contact introduced by the NPs could have a notable effect on the model because $\Delta\Gamma$ is equal to the area under the force-displacements loop divided by u_0 , where the force displacement loop will have different shapes depending on what slip regime the contact is in Hanke et al. (2013).

The sensitivity of $\Delta\Gamma$ to the contact shape and conditions could also explain the behavior observed in $\Delta\Gamma$ in the stick regime. The small variations in $\Delta\Gamma$ at small amplitudes in air are likely due to small shifts and departures from the ideal model of perfect stick. This does not readily explain the large departures from expectation seen in the liquid trials at small amplitudes. The addition of liquids into the system resulted in low amplitude $\Delta\Gamma$ being largely negative with the shape either mostly concave down (DI) or concave up (TiO₂). Since

this same behavior was not seen in the air experiments, it is likely the behavior is caused by the liquids. It is possible that at low amplitude, the observed $\Delta\Gamma$ behavior may be dominated by any energy dissipation in the liquid or behavior such as capillary action which might change the liquid barrier between contacts.

From another point of view, one might argue that the continuum model is highly successful even for the case of the TiO₂ suspension. Given that it is only applicable in CM partial slip regimes, and the lack of a clear CM slip trend in the Δf and $\Delta\Gamma$ vs. u data may indicate that there simply is no CM slip regime present in this complex interface, the model still manages to produce a value for TiO₂ close to water. Since the CM slip regime will potentially result in the formation of wear particles (Fouvry et al., 1998), and these wear particles do not appear to have a negative impact on the model as seen in the air and water cases, then it is possible that the wear particles formed in the CM regime are less numerous than the NPs or that the wear particles were small enough to fall in the roughness gaps while the 40 nm NPs were not. Therefore, a follow-up study with varying sized NP would be of interest to determine the size of the particulate at which, if any, the model might become more applicable. In particular, if a collection of poorly coupled collection of sliding or rolling particles were pushed into the gaps, then the continuum behavior might be recovered. In contrast, if the smaller particles formed the contact itself, then the model assumption of an amplitude of vibration being far smaller than the NP contact size would be even less applicable.

SUMMARY AND CONCLUSIONS

Using a continuum analysis model, the response of a QCM to tripod 3[4SS/304SS rubbing contacts in air, water and a TiO₂ suspension has been employed to evaluate friction coefficients. The results for dry and water lubricated contacts compare favorably with friction coefficient measurements reported in the literature for the same materials' combinations. The model appears less applicable to the case of contacts lubricated with the NP suspension. The continuum nature of the model does not however appear to be the dominant factor underlying its

failure to apply to NP lubricated suspensions. The failure is likely attributed to a lack of a CM slip regime when NP are present at the interface, e.g., in the form of a poorly coupled collection of sliding or rolling particles, and/or the fact that the amplitude of vibration is no longer far smaller than the NP contact size regions.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

REFERENCES

- Acharya, B., Pardue, T. N., Avva, K. S., and Krim, J. (2018). In situ, real time studies of thermal reaction film formation temperatures for iron and 304ss surfaces immersed in 5% tricresyl phosphate in base oil. *Tribol. Int.* 126, 106–115. doi: 10.1016/j.triboint.2018.04.034
- Acharya, B., Pardue, T. N., Su, L., Smirnov, A. I., Brenner, D. W., and Krim, J. (2019a). Nanotribological performance factors for aqueous suspensions of oxide nanoparticles and their relation to macroscale lubricity. *Lubricants* 7:49. doi: 10.3390/lubricants7060049
- Acharya, B., Seed, C. M., Brenner, D. W., Smirnov, A. I., and Krim, J. (2019b). Tuning friction and slip at solid-nanoparticle suspension interfaces by electric fields. *Scient. Rep.* 9:18584. doi: 10.1038/s41598-019-54515-1
- Ali, M. K., Xianjun, H., Mai, L., Qingping, C., Turkson, R. F., and Bicheng, C. (2016). Improving the tribological characteristics of piston ring assembly in automotive engines using Al_2O_3 and TiO_2 nanomaterials as nano-lubricant additives. *Tribol. Int.* 103, 540–554. doi: 10.1016/j.triboint.2016.08.011
- Berg, S., and Johannsmann, D. (2003). High speed microtribology with quartz crystal resonators. *Phys. Rev. Lett.* 91:145505. doi: 10.1103/PhysRevLett.91.145505
- Berg, S., Prellberg, T., and Johannsmann, D. (2003). Nonlinear contact mechanics based on ring-down experiments with quartz crystal resonators. *Rev. Sci. Instrum.* 74, 118–126. doi: 10.1063/1.1523647
- Borovsky, B., Booth, A., and Manlove, E. (2007). Observation of microslip dynamics at high-speed microcontacts. *Appl. Phys. Lett.* 91:114101. doi: 10.1063/1.2784172
- Borovsky, B., Krim, J., Syed Asif, A., and Wahl, K. J. (2001). Measuring nanomechanical properties of a dynamic contact using an indenter probe and quartz crystal microbalance. *J. App. Phys.* 90:639. doi: 10.1063/1.1413493
- Borovsky, B. P., Bousse, C., O'Neill, C., and Sletten, L. R. (2017). An integrated force probe and quartz crystal microbalance for high-speed microtribology. *Tribol. Lett.* 65:148. doi: 10.1007/s11249-017-0933-6
- Borovsky, B. P., Garabedian, N. T., McAndrews, G. R., Wieser, R. J., and Burris, D. L. (2019). Integrates QCM-microtribometry: friction of single-crystal MoS_2 and gold from $\mu\text{m/s}$ to m/s . *ACS Appl. Mater. Interfaces*. 11, 40961–40969. doi: 10.1021/acsami.9b15764
- Braiman, Y., Barhen, J., and Protopopescu, V. (2003). Control of friction at the nanoscale. *Phys. Rev. Lett.* 90:094301. doi: 10.1103/PhysRevLett.90.094301
- Brizmer, V., Kligerman, Y., and Etsion, I. (2007). A model for junction growth of a spherical contact under full stick condition. *J. Tribol.* 129, 783–790. doi: 10.1115/1.2772322
- Ciavarella, M. (2015). Transition from stick to slip in Hertzian contact with “Griffith” friction: The Cattaneo-Mindlin problem revisited. *J. Mech. Phys. Solids*. 84, 313–324. doi: 10.1016/j.jmps.2015.08.002
- Curtis, C. K., Marek, A., Smirnov, A. I., and Krim, J. (2017). A comparative study of the nanoscale and macroscale tribological attributes of alumina and stainless steel surfaces immersed in aqueous suspensions of positively or negatively charged nanodiamonds. *Beil. J. Nanotechnol.* 8, 2045–2059. doi: 10.3762/bjnano.8.205
- Dawson, B. D., Lee, S. M., and Krim, J. (2009). Tribo-induced melting transition at a sliding asperity contact. *Phys. Rev. Lett.* 103:205502. doi: 10.1103/PhysRevLett.103.205502
- Dirri, F., Palomba, E., Longobardo, A., Zampetti, E., Saggin, B., and Scaccabarozzi, D. (2019). A review of quartz crystal microbalances for space applications. *Sens. Actu. A Phys.* 287, 48–75. doi: 10.1016/j.sna.2018.12.035
- Dultsev, F. N., and Nekrasov, D. V. (2018). Treatment of the resonance curve recorded during measurement of the signal of particle rupture from the QCM surface. *Sens. Actu. B Chem.* 15, 70–75. doi: 10.1016/j.snb.2018.04.029
- Ellis, J. S., McHale, G., Hayward, G. L., and Thompson, M. (2003). Contact angle-based predictive model for slip at the solid-liquid interface of a transverse-shear mode acoustic wave device. *J. App. Phys.* 94, 6201–6207. doi: 10.1063/1.1619195
- Etsion, I. (2010). Revisiting the cattaneo-mindlin concept of the interfacial slip in tangentially loaded compliant bodies. *J. Tribol.* 132:20801. doi: 10.1115/1.4001238
- Fouvry, S., Kapsa, P., and Vincent, L. (1998). Developments of fretting sliding criteria to quantify the local friction coefficient evolution under partial slip condition. *Tribol. Series*. 34, 161–172. doi: 10.1016/S0167-8922(98)80071-0
- Glavatskih, S., and Höglund, E. (2008). Tribotronics—towards active tribology. *Tribol. Int.* 41, 934–939. doi: 10.1016/j.triboint.2007.03.001
- Han, T., Nag, A., Mukhopadhyay, S. C., and Xu, Y. (2019). Carbon nanotubes and its gas-sensing applications: a review. *Sens. Actu. A Phys.* 291, 107–143. doi: 10.1016/j.sna.2019.03.053
- Hanke, S., Petri, J., and Johannsmann, D. (2013). Partial slip in mesoscale contacts: dependence on contact size. *Phys. Rev. E*. 88:032408. doi: 10.1103/PhysRevE.88.032408
- Hsu, S., Ying, C., and Zhao, F. (2014). The nature of friction: a critical assessment. *Friction* 2, 1–26. doi: 10.1007/s40544-013-0033-z
- Johannsmann, D. (2007). Studies of contact mechanics with the QCM. *Springer Ser Chem Sens Biosens*. 5, 151–170. doi: 10.1007/5346-026
- Johannsmann, D., Reviakine, I., and Richter, R. (2009). Dissipations in films of adsorbed nanospheres studied by quartz crystal microbalance (QCM). *Anal. Chem.* 81, 8167–8176. doi: 10.1021/ac901381z
- Josse, F., and Lee, Y. (1998). Analysis of the radial dependence of the mass sensitivity for modified-electrode quartz crystal resonators. *Anal. Chem.* 70, 237–247. doi: 10.1021/ac970603z
- Kim, H. J., and Kim, D. E. L. (2009). Nano-scale friction: a review. *Precis. Eng. Manuf.* 10, 141–151. doi: 10.1007/s12541-009-0039-7
- Krim, J. (2012). Friction and energy dissipation mechanisms in adsorbed molecules and molecularly thin films. *Adv. Phys.* 61:3, 155–323. doi: 10.1080/00018732.2012.706401
- Krim, J. (2019). Controlling friction with external electric or magnetic fields: 25 examples. *Front. Mech. Eng.* 5:22. doi: 10.3389/fmech.2019.00022
- Laschitsch, A., and Johannsmann, D. (1999). High frequency tribological investigations on quartz resonator surfaces. *J. Appl. Phys.* 85:3759. doi: 10.1063/1.369745
- Leopoldes, J., and Jia, X. (2010). Transverse shear oscillator investigation of boundary lubrication in weakly adhered films. *Phys. Rev. Lett.* 105:266101. doi: 10.1103/PhysRevLett.105.266101

AUTHOR CONTRIBUTIONS

CS, BA, and JK conceived the design of the experiment and constructed the apparatus. CS performed the experimental measurements and data analysis. CS and JK composed and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the US National Science Foundation, Materials Genome Initiative Project #DMR1535082.

- Lu, F., Lee, H. P., and Lim, S. P. (2004). Quartz crystal microbalance with rigid mass partially attached on electrode surfaces. *Sens. Actu. A Phys.* 112, 203–210. doi: 10.1016/j.sna.2004.01.018
- Lu, X., Khonsari, M. M., and Gelinck, E. R. (2006). The Stribeck curve: experimental results and theoretical prediction. *J. Tribol.* 128, 789–794. doi: 10.1115/1.2345406
- Martin, B. A., and Hager, H. E. (1989). Velocity profile on quartz crystals oscillating in liquids. *J. Appl. Phys.* 65:2630. doi: 10.1063/1.342772
- M'boungui, G., Semail, B., Giraud, F., and Jimoh, A. A. (2014). Development of a novel plane piezoelectric actuator using Hamilton's principle based model and Hertz contact theory. *Sens. Actu. A Phys.* 217, 116–123. doi: 10.1016/j.sna.2014.06.026
- Paggi, M., Pohrt, R., and Popov, V. L. (2014). Partial-slip frictional response of rough surfaces. *Sci Rep.* 4:5178. doi: 10.1038/srep05178
- Pardue, T. N., Acharya, B., Curtis, C. K., and Krim, J. (2018). A tribological study of Γ -Fe₂O₃ nanoparticles in aqueous suspension. *Tribol. Lett.* 66:130. doi: 10.1007/s11249-018-1083-1
- Popov, V. L. (2010). "Chapter 2: qualitative treatment of contact problems—normal contact without adhesion," in *Contact Mechanics and Friction: Physical Principles and Applications* (Berlin; Heidelberg: Springer Verlag), 9–25. doi: 10.1007/978-3-642-10803-7_3
- Rodahl, M., and Kasemo, B. (1996). On the measurement of thin liquid overlayers with the quartz-crystal microbalance. *Sens. Actu. A Phys.* 54, 448–456. doi: 10.1016/S0924-4247(97)80002-7
- Seed, C. M., Acharya, B., Andrus, R., and Krim, J. (2020). Correlation of high frequency QCM sphere-plate stiffness measurements with macroscopic frictional contacts in thin film and bulk stainless steel materials. *Sens. Actu. A Phys.* 36:111913. doi: 10.1016/j.sna.2020.111913
- Sen, S., Chow, C. M., Moazzen, E., Segre, C. U., and Timofeeva, E. V. (2017). Electroactive nanofluids with high solid loading and low viscosity for rechargeable redox flow batteries. *J. Appl. Electrochem.* 47, 593–605. doi: 10.1007/s10800-017-1063-4
- Vittorias, E., Kappl, M., Butt, H. J., and Johannsmann, D. (2010). Studying mechanical microcontacts of fine particles with the quartz crystal microbalance. *Powder Technol.* 203, 489–502. doi: 10.1016/j.powtec.2010.06.011
- Vlachová, J., König, R., and Johannsmann, D. (2015). Stiffness of sphere-plate contacts at MHz frequencies: dependence on normal load, oscillation amplitude, and ambient medium. *Beils. J. Nanotechnol.* 6, 845–856. doi: 10.3762/bjnano.6.87
- Wang, J., Polleux, J., Lim, J., and Dunn, B. (2007). Pseudocapacitive contributions to electrochemical energy storage in TiO₂ (Anatase) nanoparticles. *J. Phys. Chem. C* 111, 14925–14931. doi: 10.1021/jp074464w
- Wang, L., Junkuo, G., and Xu, J. (2019). QCM formaldehyde sensing materials: design and sensing mechanism. *Sens. Actu. B Chem.* 293, 71–82. doi: 10.1016/j.snb.2019.04.050
- Wu, H., Jia, F., Li, Z., Lin, F., Huo, M., Huang, S., et al. (2020). Novel water-based nanolubricant with superior tribological performance in hot steel rolling. *Int. J. Extr. Manuf.* 2:025002. doi: 10.1088/2631-7990/ab82fe
- Yamaguchi, T., and Hokkirigawa, K. (2016). Friction and wear properties of PEEK resin filled with RB ceramics particles under water lubricated condition. *Tribol. Online* 11, 653–660. doi: 10.2474/troll.11.653
- Yu, X., Zhou, J., and Jiang, Z. (2016). Developments and possibilities for nanoparticles in water-based lubrication during metal processing. *Rev. Nanosci. Nanotechnol.* 5, 136–163. doi: 10.1166/rnn.2016.1072
- Zhang, C., and Wang, Z. L. (2016). Tribotronics—A new field by coupling triboelectricity and semiconductor. *Nano Today* 11, 521–536. doi: 10.1016/j.nantod.2016.07.004
- Zhang, X., and Li, Q. (2010). Enhancement of friction between carbon nanotubes: an efficient strategy to strengthen fibers. *ACS Nano* 4, 292–316. doi: 10.1021/nn901515j

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Seed, Acharya and Krim. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Meso–Macro Coupled Analysis of Pressure-Dependent Friction of Rubber

Shingo Ozaki^{1*}, Keishi Mieda¹, Satoru Maegawa² and Ken Nakano³

¹ Faculty of Engineering, Yokohama National University, Yokohama, Japan, ² Department of Electrical and Mechanical Engineering, Nagoya Institute of Technology, Nagoya, Japan, ³ Faculty of Environment and Information Science, Yokohama National University, Yokohama, Japan

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Varvara Romanova,
Institute of Strength Physics and
Materials Science (ISPM SB
RAS), Russia
Francesco Massi,
Sapienza University of Rome, Italy

*Correspondence:

Shingo Ozaki
s-ozaki@ynu.ac.jp

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 16 March 2020

Accepted: 15 April 2020

Published: 06 May 2020

Citation:

Ozaki S, Mieda K, Maegawa S and
Nakano K (2020) Meso–Macro
Coupled Analysis of
Pressure-Dependent Friction of
Rubber. *Front. Mech. Eng.* 6:24.
doi: 10.3389/fmech.2020.00024

To deepen the understanding of the frictional sliding phenomena of rough surfaces, advances are required in numerical analysis methods at various spatial scales. In this study, to examine the microscopic behavior of a rough asperity contact corresponding to a bulk contact on the macroscopic scale, a loop-type meso–macro coupled analysis scheme is proposed. A mesoscale numerical model and a macroscopic friction model are required for the proposed multi-scale analysis. A friction model was adopted based on the multipoint contact model for the mesoscale model, and the pressure- and state-dependent elastoplastic analogy friction model was used for the macroscale model. In the proposed meso–macro coupled analysis, the parameter set for the elastoplastic analogy friction model was first identified via a numerical friction test using the mesoscale multipoint contact model assuming various conditions. Then, a macroscale finite element analysis incorporating the elastoplastic analogy friction model was performed for the macroscopic analysis of contact between a rough rubber hemisphere and a smooth plate. Here, the information from the mesoscale rough surfaces were reflected in the macroscale finite element analysis. Finally, a mesoscale localization analysis was performed in which the macroscopic histories of several typical locations were obtained by finite element analysis and used as boundary conditions for the mesoscale model. It is suggested that the microscopic sliding process of rough surfaces represented by the finite element analysis can be examined using the proposed method.

Keywords: multiscale analysis, real contact area, roughness, friction model, FEM

INTRODUCTION

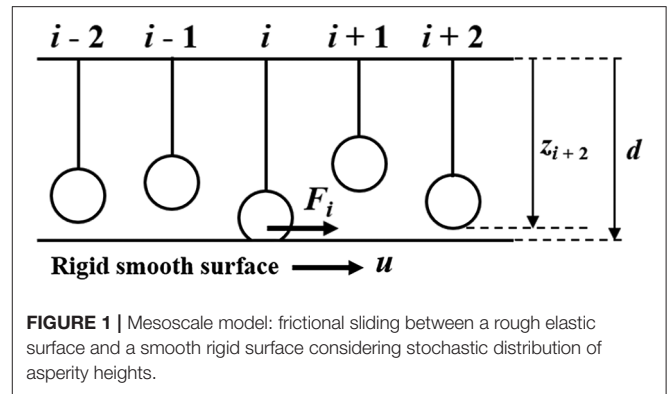
Friction is an important physical phenomenon in mechanical engineering. For example, friction between sliding parts of machines not only accounts for the majority of energy loss, but also causes failure. Meanwhile, rubber materials are often used for mechanical components where a frictional contact occurs. It is widely known that rubber friction can be categorized into two main components: adhesion friction and hysteresis friction (Tabor, 1960; Schallamach, 1971; Fuller and Tabor, 1975; Roberts, 1992; Persson, 2001). According to Roberts and Thomas (1975) and Persson and Volokitin (2006), adhesion friction is reported to be dominant when coarse rubber slides on a smooth hard surface. The focus of this study was therefore the adhesion friction of rubber.

For soft rubber, the distance between real contact points decreases as the contact load increases, and the mutual interference between the contact points saturates the real contact area (Johnson et al., 1985; Manners, 2000; Persson et al., 2002; Hyun et al., 2004; Yang and Persson, 2008; Maegawa et al., 2015). In this process, since the ratio of the real contact area A_r to the apparent contact area A_a increases, the relationship between the frictional force and the contact load is no longer proportional; thus, the friction coefficient exhibits a complex pressure dependency (Schallamach, 1971; Persson et al., 2002; Yang and Persson, 2008; Maegawa et al., 2015). Therefore, to deepen the understanding of the frictional sliding phenomena of rubber materials, advances in numerical analysis methods on various spatial scales are required in addition to experimental observations.

The invention of the atomic force microscope in the 1985 made it possible to directly measure frictional phenomena occurring at the atomic (nano) and molecular (micro) scales (Binnig et al., 1986), and several molecular dynamics methods were developed around the same time. Meanwhile, macroscale numerical analysis methods include the finite element and boundary element methods. At present, many commercial software packages are equipped with a standard analysis function for frictional contact problems (Kikuchi and Oden, 1988; Laursen, 2001; Wriggers, 2003). Furthermore, various phenomenological rate- and state-dependent friction models have been proposed for more advanced analyses (Dieterich, 1972, 1979; Ruina, 1983; Hashiguchi and Ozaki, 2008; Ozaki and Hashiguchi, 2010; Ozaki et al., 2012, 2013, 2020). In between the nano/microscale and macroscale, multipoint asperity level contact theories take the form of mesoscale models. Thus, studies focusing on each scale are being actively pursued, and a great deal of knowledge has already been accumulated as a result.

Meanwhile, multiscale analysis methods that connect analyses at each scale are also required. A multiscale analysis makes it possible to examine the elementary behavior of asperities on a microscopic scale corresponding to macroscopic frictional contact behavior. However, in practice, contact surfaces that need to be engineered often have uncertainties due to roughness; additionally, they have no periodicity or regularity. This makes it difficult to apply bi-directionally coupled (strong coupling) multiscale analysis methods, such as the homogenization method used in computational solid mechanics, to frictional contact problems. Therefore, it is necessary to use a multiscale analysis method in another framework that can overcome this issue.

In this study, a loop-type coupled analysis scheme is proposed that bridges the mesoscale and macroscale domains with reference to the multiscale uncoupled analysis used in studies of crystal structures that affect the strength of metals (Watanabe and Terada, 2010). Specifically, the “mesoscale multipoint contact model” and the “macroscale finite element analysis model” are linked via the “rate-, state-, and pressure-dependent friction model” proposed by the authors (Ozaki et al., 2020). Based on this scheme, the analysis results at each scale can be mutually expanded, and a multiscale understanding of the frictional sliding phenomena becomes possible. The procedure for implementing the proposed analysis method is summarized below. Note



that in this study, the rate dependency of rubber friction is not considered in order to investigate the pressure-dependent behavior and the effect of roughness upon it.

The proposed loop-type coupled analysis method procedure is as follows:

1. Perform a “numerical friction test” using a mesoscale frictional contact model (mesoscale analysis) at the scale of the representative contact area.
2. Carry out parameter fitting of the macroscale friction model by using the results obtained in Step 1.
3. Implement the macroscale friction model in a finite element model and perform a “macroscale analysis” of various frictional contact problems using the previously determined parameters.
4. Select arbitrary evaluation points (nodes or elements) on the contact surface of the macroscale finite element model, and perform a “localization analysis” at the mesoscale using the histories of contact stress and sliding velocity as boundary conditions. Then, study the elementary behavior on the mesoscale corresponding to the macroscale analysis results.

FRICITION MODELS

The mesoscale and macroscale friction models are briefly explained in this section.

Mesoscale Model

In this study, the statistical model of asperities in contact based on Greenwood and Williamson (1966) was adopted as the mesoscale analysis model. **Figure 1** shows a schematic of this analysis model. This model corresponds to multi-point contact between an elastic rough surface and a rigid smooth surface. As shown in the figure, there are N asperities on the upper surface, and the height of asperity i is z_i . Assuming that the distance between the rough surface and the smooth surface is d , the compression of asperity i in the normal direction, δ_i , is given as follows:

$$\delta_i = \begin{cases} z_i - d & \text{for } z_i > d \\ 0 & \text{for } z_i \leq d \end{cases} \quad (1)$$

Here, it is assumed that the deformation of the asperity follows Hertz's contact theory. Thus, the radius of contact a_i , the contact area A_i , and the normal load w_i supported by the asperity can be defined in the following manner:

$$a_i = \sqrt{\beta \delta_i} \quad (2)$$

$$A_i = \pi a_i^2 \quad (3)$$

$$w_i = \frac{4E}{3(1-\nu^2)} \sqrt{\beta \delta_i^3} \quad (4)$$

where E and ν are the elastic modulus and Poisson's ratio of the asperity, respectively, and β is the radius of curvature at the tip of the asperity. Further, the tangential contact stiffness k_i is given by the following equation using the shear modulus, G , and Poisson's ratio, ν :

$$k_i = \frac{8Ga_i}{2-\nu} \quad (5)$$

The friction force F_i is expressed by the product of the tangential contact stiffness k_i and the tangential displacement u when

sticking, and by the product of the shear strength τ and the contact area A_i when sliding. That is,

$$F_i = \begin{cases} k_i u & \text{for } k_i u < A_i \tau \\ A_i \tau & \text{for } k_i u \geq A_i \tau \end{cases} \quad (6)$$

Macroscale Model

In the proposed loop-type coupled multiscale analysis scheme, the capabilities of the macroscale friction model are important. In this study, the rate-, state-, and pressure-dependent elastoplastic analogy friction model previously proposed by the authors was used as the macroscale friction model (Ozaki et al., 2020). This model can rationally describe basic frictional sliding characteristics such as the smooth transition of static-kinetic frictions, time-dependent recovery of static friction, velocity-weakening of frictional resistance, and pressure dependency.

TABLE 1 | Conditions of the numerical friction test using the statistical model of asperities in contact.

Number of trials	10	Analysis range	1×1
[–]		[mm ²]	
Young's modulus of Rubber	1.5	Poisson's ratio of Rubber	0.49
E [MPa]		ν [–]	
Average asperity height	35	Sliding velocity	0.1
Z_{ave} [μm]		v [mm/s]	
Maximum height	20.0, 56.5	Shear strength	0.5
R_z [μm]		τ [MPa]	
Asperity radius	10	Interval of asperity	20
β [μm]		[μm]	
Normal stress	0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.40, 0.50		
f_n [MPa]			

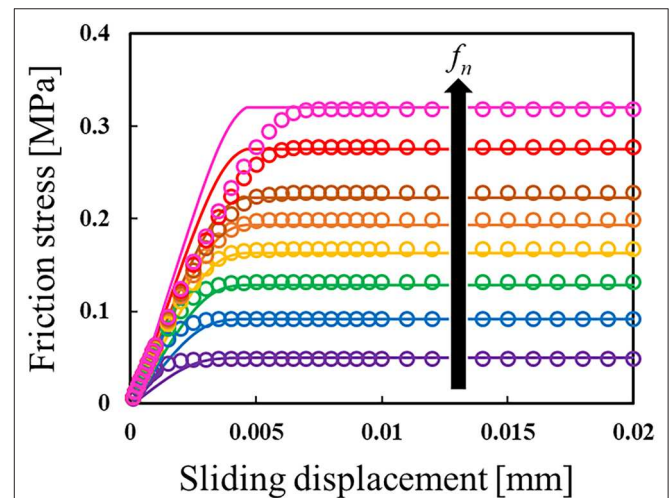


FIGURE 3 | Variation in friction stress with sliding displacement under eight levels of constant normal stress, where $R_z = 20.0 \mu\text{m}$. The solid lines show results obtained by a numerical friction test (NFT), while the open circles show the fitting results from the macroscale friction model.

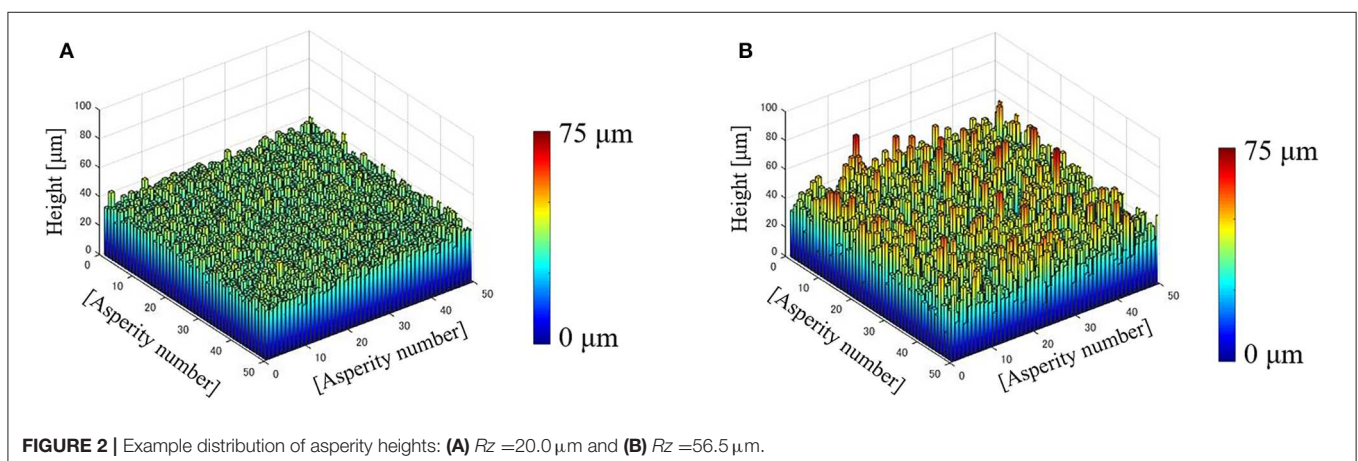


FIGURE 2 | Example distribution of asperity heights: (A) $R_z = 20.0 \mu\text{m}$ and (B) $R_z = 56.5 \mu\text{m}$.

As rate dependency was not considered in this study, the state and pressure dependencies are the focus of this section. The relationship between the contact stress rate $\dot{\mathbf{f}}$ and sliding velocity $\bar{\mathbf{v}}$ is defined in the following manner (Hashiguchi and Ozaki, 2008; Ozaki and Hashiguchi, 2010; Ozaki et al., 2012, 2013, 2020):

$$\dot{\mathbf{f}} = \mathbf{C}^{ep} \bar{\mathbf{v}} \quad (7)$$

where the second-order elastoplastic contact stiffness tensor is given as:

$$\mathbf{C}^{ep} \equiv \alpha_n \mathbf{n} \otimes \mathbf{n} + \alpha_t (\mathbf{I} - \mathbf{n} \otimes \mathbf{n}) - \mathbf{t} \otimes \frac{\alpha_t \alpha_n \mathbf{n} R \tau S'_r - \alpha_t \alpha_t \mathbf{t}}{\alpha_t + U \tau S_r} \quad (8)$$

Here, α_n and α_t are the elastic contact stiffness moduli in the normal and tangential directions, respectively; \mathbf{I} , \mathbf{n} , and \mathbf{t} are the unit tensor, unit normal vector, and unit tangential vector, respectively; and \mathbf{n} and \mathbf{t} are defined as follows:

$$\mathbf{n} \equiv \frac{\mathbf{f}_n}{\|\mathbf{f}_n\|}, \quad \mathbf{t} \equiv \frac{\mathbf{f}_t}{\|\mathbf{f}_t\|} \quad (9)$$

where \mathbf{f}_n and \mathbf{f}_t are the normal and tangential contact stress vectors, respectively, and hold the relationship of $\mathbf{f} = \mathbf{f}_n + \mathbf{f}_t$.

In this study, $R (0 \leq R \leq 1)$ is defined as the state variable to describe the microscopic sliding before gross sliding, and is called the normal-sliding ratio. The following function was adopted for

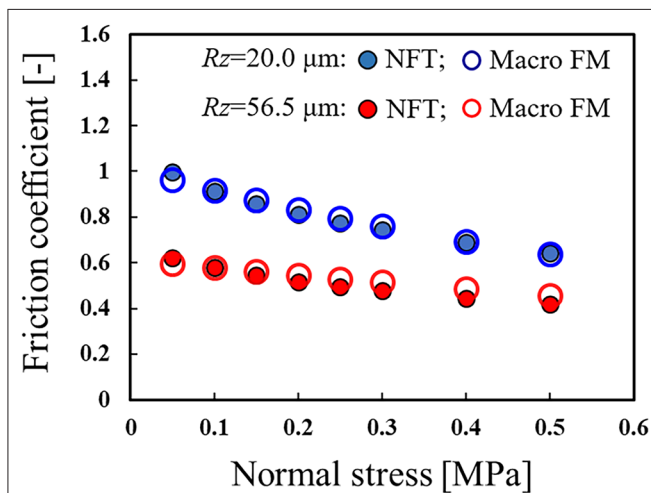


FIGURE 4 | Relationship between the steady state ratio of stress (friction coefficient) and normal stress. Here, the closed circles show the results of the numerical friction test (NFT), while the open circles show the results of the macroscale friction model (Macro FM).

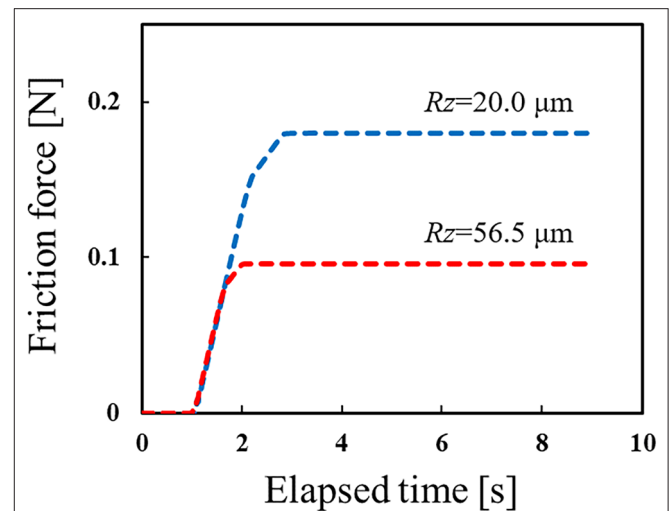


FIGURE 6 | Variation of friction force on rubber hemisphere over time obtained by finite element analysis under two roughness conditions. Here, the vertical pressing displacement of the rubber hemisphere is 0.25 mm.

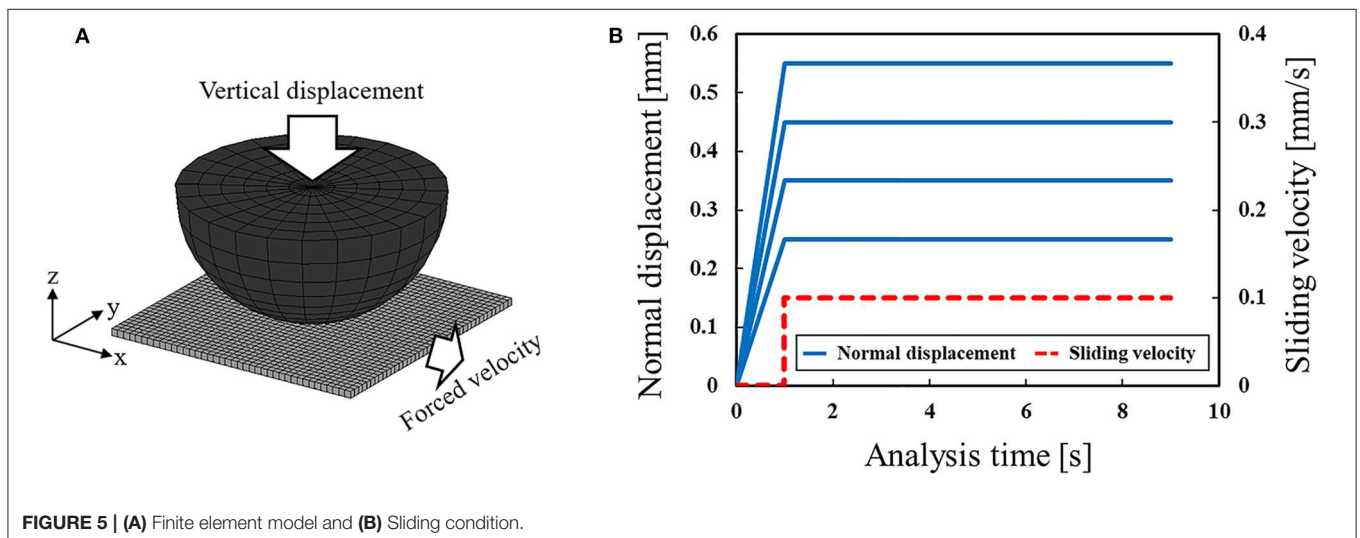


FIGURE 5 | (A) Finite element model and (B) Sliding condition.

the evolution law of the normal-sliding ratio:

$$\dot{R} = U(R) \|\bar{v}^p\| \quad (10)$$

where

$$U(R) = r \cot \left\{ \left(\frac{\pi}{2} \right) R \right\} \quad (11)$$

Here, r is the parameter for microscopic sliding and \bar{v}^p is the plastic (irreversible) sliding velocity.

Meanwhile, to describe the pressure dependency, the variable $S_r = A_r/A_a$ is introduced in the model. Here, A_r is the real contact area and A_a is the apparent contact area. The pressure

dependency of the real contact area is assumed as follows:

$$S_r = 1 - \exp(-b \|\mathbf{f}_n\|) \quad (12)$$

$$S'_r = \frac{\partial S_r}{\partial \|\mathbf{f}_n\|} \quad (13)$$

where b is the parameter for pressure dependency.

MULTISCALE ANALYSIS

In this study, the contact between a rough rubber hemisphere made of cross-linked polydimethylsiloxane (PDMS) and a smooth plate made of polymethylmethacrylate (PMMA) (Maegawa et al., 2015) was adopted as the analysis target. Although PDMS hemispheres with different surface roughness values were assumed in the analysis, a quantitative comparison is beyond the scope of this study as the purpose of this study was to propose a meso-macro coupled analysis method. In the following sections, the details of each step of the proposed analysis method are described. Note that it was assumed that the PMMA plate was a rigid body in the analysis because the Young's modulus of PMMA is sufficiently large compared to that of PDMS.

Numerical Friction Test

This section describes the numerical friction test using the statistical model of asperities in contact, as shown in Figure 1. The heights of asperity, z_i , were assumed to correspond to two levels of surface roughness. Here, the heights of asperity were set to follow a normal distribution.

Table 1 lists the conditions evaluated in the numerical friction tests, in which two levels of roughness were controlled by setting the maximum height R_z , where the values of R_z correspond to 6σ of each normal distribution. In addition, eight levels of normal stress (pressure) were evaluated to examine the pressure dependency. In the analysis, the target normal stress f_n was set,

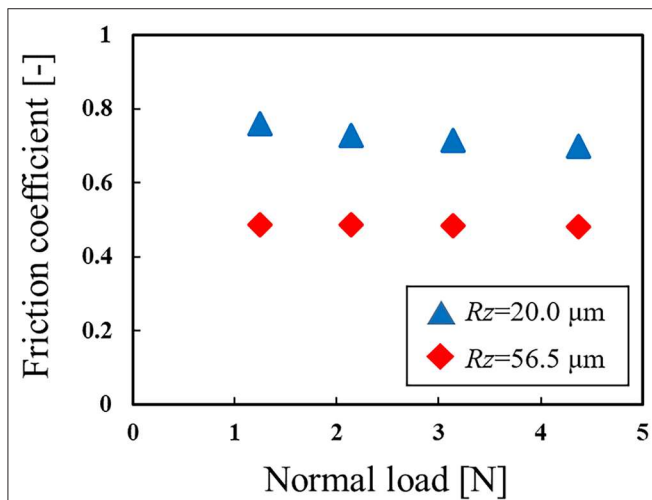


FIGURE 7 | Relationship between the friction coefficient and normal stress at the bottom of the rubber hemisphere, obtained by finite element analysis.

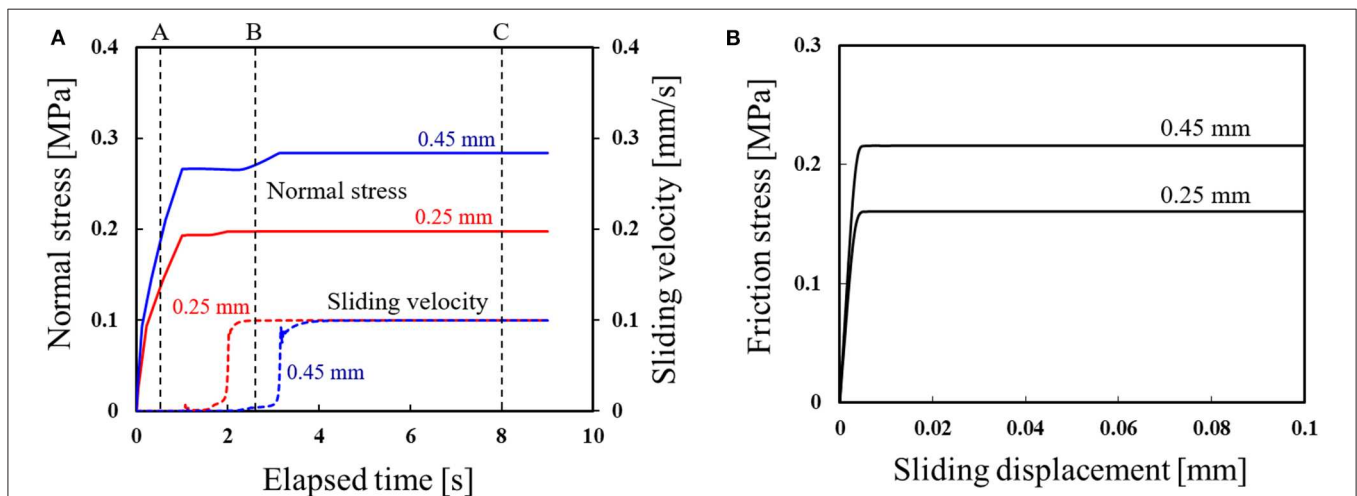
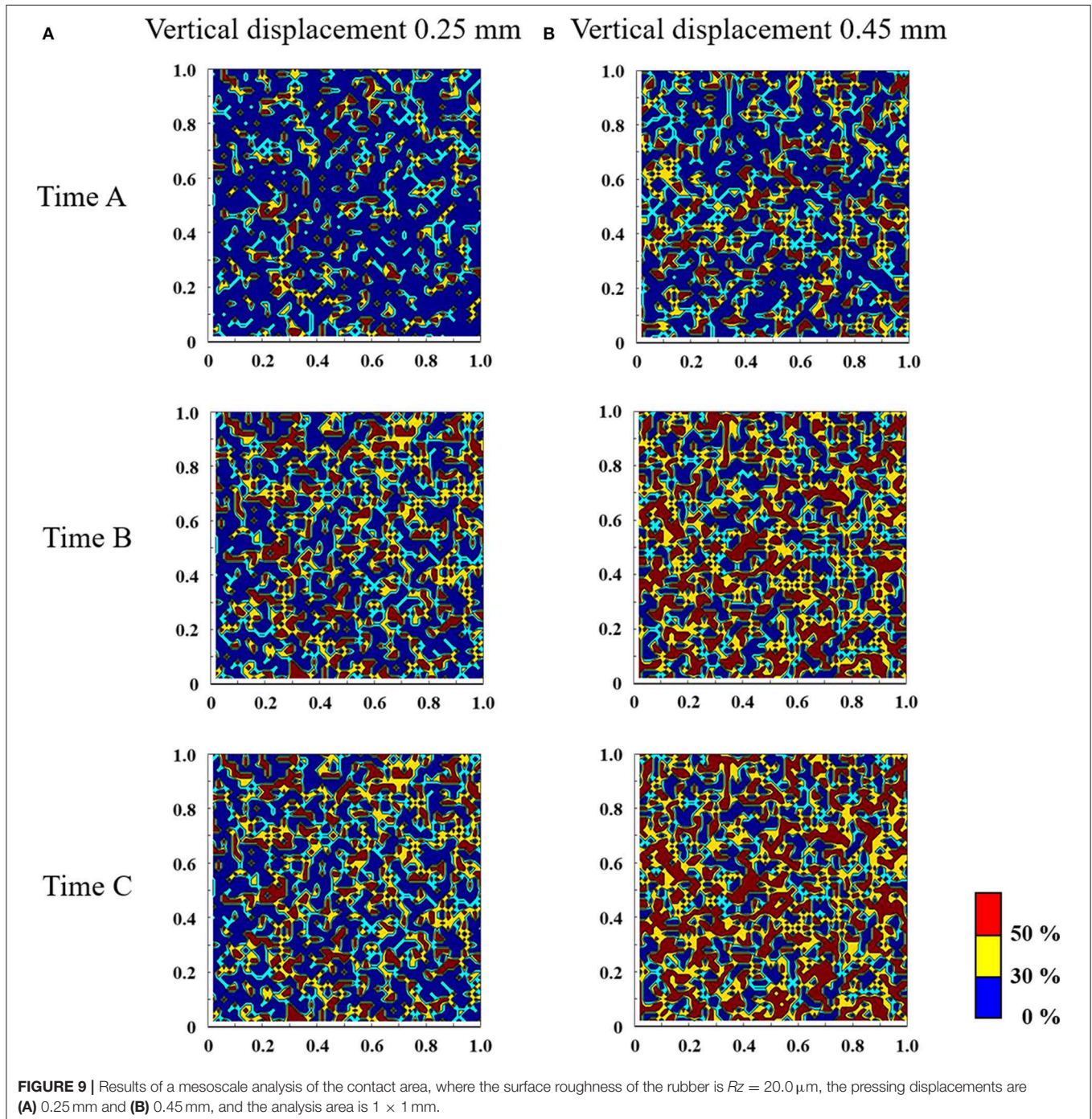


FIGURE 8 | Results of a meso-macro coupled analysis of the effect of vertical pressing displacement, where the surface roughness of the rubber was $R_z = 20.0 \mu\text{m}$: (A) History of the normal stress and sliding velocity obtained by the finite element analysis; (B) Localization analysis result of friction stress vs. sliding displacement.

and the parameters δ_i and w_i were calculated using Equations (1) and (4), respectively. Then, the distance d was repeatedly increased by a small amount until the calculated normal stress reached the target value. Thereafter, the sliding displacement was given and the tangential stress (friction stress) was calculated by Equation (6) under a constant normal stress. **Figure 2** shows examples of the asperity height distributions, where the 1×1 mm analysis area corresponds to the apparent contact area A_a , and is the representative contact area. Note that because

random numbers were used, the distributions of asperities shown in **Figure 2** were different for each analysis, and 10 tests were performed for each condition.

Figure 3 shows an example of the numerical test results for the relationship between frictional stress and sliding displacement under eight levels of normal stress. Here, R_z is $20.0 \mu\text{m}$. The averaged results of 10 tests are shown for each normal stress condition. As can be confirmed from the figure, the friction stress increases with sliding (the tangential movement of the



smooth flat plate) and reaches a steady state. In addition, because the real contact area in the mesoscale model varies depending on the normal stress, the magnitude of friction stress also varies.

Figure 4 shows the results of the pressure dependency of the friction coefficient (as a ratio of stress f_t/f_n). Plots in the graph are average values during the steady state of the 10 tests. It is confirmed that the friction coefficient decreases as the normal stress increases. The reason for the tendency shown in **Figure 4** is that the distance between the real contact points decreases as the normal stress increases, and the mutual interference between these contact points slows down the increase in the real contact area (Persson et al., 2002; Yang and Persson, 2008; Maegawa et al., 2015). Furthermore, the smaller the surface roughness, the larger the friction coefficient. These tendencies have also been confirmed in the actual friction tests simulated by these analyses (Maegawa et al., 2015).

Parameter Fitting of Macroscale Friction Model

In this section, the determination of the friction model parameters used for the macroscale finite element analysis is described. The following parameters were determined: the elastic contact stiffness moduli α_n and α_t ; r , which prescribes the smoothness of the stick-to-slip transition; and b , which prescribes the pressure dependency. Additionally, the shear strength shown in **Table 1** was adopted.

In the mesoscale model adopted in this study, the initial slope of the friction stress–sliding displacement relationship differs depending on the normal stress because the contact radius of asperity is used for the tangential stiffness [see Equation (5)]. Since the pressure dependencies of the elastic moduli were not considered in the macroscale friction model, the parameter fitting was performed by focusing on the result $f_n = 0.25$ MPa. The

determined parameters are accordingly listed below:

$$R_z = 20.0\mu\text{m} : \alpha_n = \alpha_t = 65\text{MPa/mm}, r = 1400\text{mm}^{-1}, \\ b = 2.03\text{MPa}^{-1}$$

$$R_z = 56.5\mu\text{m} : \alpha_t = \alpha_n = 35\text{MPa/mm}, r = 1200\text{mm}^{-1}, \\ b = 1.23\text{MPa}^{-1}$$

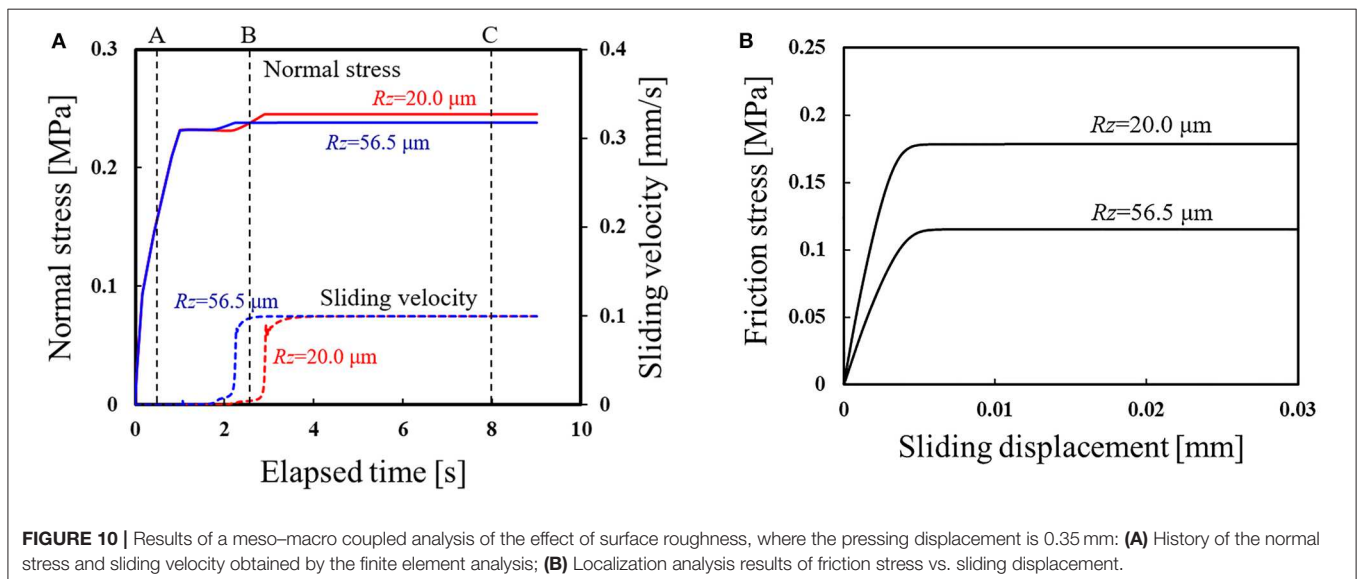
Figures 3, 4 also show an example of the fitting results by the macroscale friction model (the open circles). By using an appropriate macroscale friction model, the results of the mesoscale analysis can indeed be obtained at a low calculation cost.

Finite Element Analysis

In this section, the finite element analysis of the frictional contact between a rough rubber hemisphere and a smooth plate is described. The commercial software package MSC Marc (2020) and its related user subroutine for the implementation of the friction model was used in this analysis.

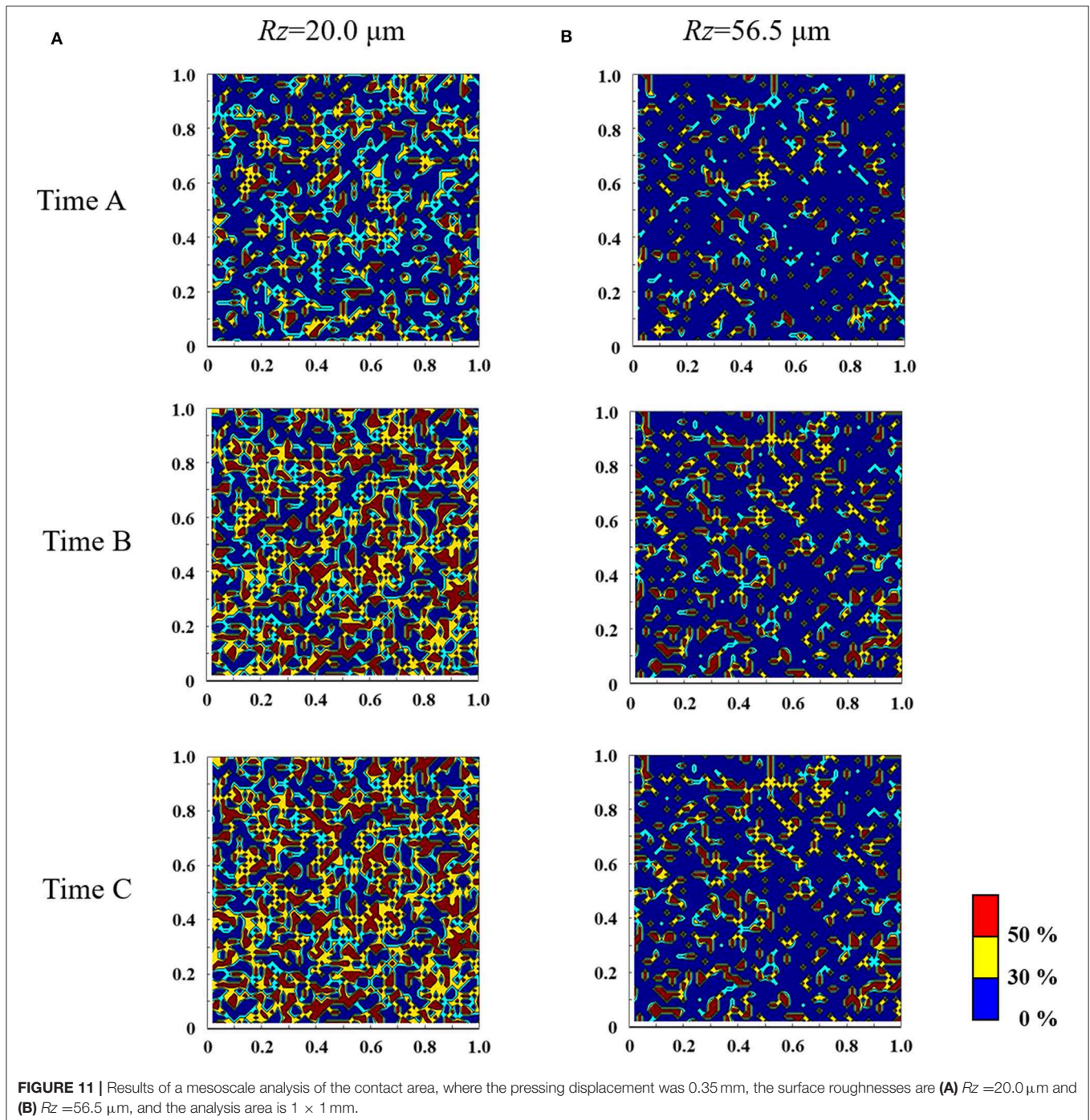
Figure 5 shows the finite element model and boundary conditions. The rubber hemisphere and the flat plate were discretized by eight-node solid elements. Note that a fine mesh was adopted for the rubber contact surface. The rubber hemisphere was a linear elastic body, and its Young's modulus and Poisson's ratio were set to $E = 1.5$ MPa and $\nu = 0.49$, respectively, while the flat PMMA plate was considered a rigid body. The radius of the rubber hemisphere was 11 mm. In the analysis, the upper surface of the rubber hemisphere was fixed in the in-plane direction and subjected to a prescribed forced displacement in the vertical direction (normal to the plate) in order to press it against the plate. Then, a forced velocity (0.1 mm/s) was applied to the plate in the x-direction to slide over one another. Note that the displacement of the plate was fixed in all directions except the x-direction.

Figure 6 shows an example of the macroscale analysis results, in which the vertical pressing displacement of the rubber



hemisphere is 0.25 mm. This figure shows the variation in friction force on the rubber hemisphere over time. From the figure, it can be seen that the frictional force gradually increases with the movement of the flat plate before reaching a steady state. In addition, the magnitude of the frictional force in the steady state reflects the roughness in the mesoscale model, i.e., the friction force decreases as the roughness increases. This tendency has also been confirmed in previous experiments (Maegawa et al., 2015).

Figure 7 shows the relationship between the friction coefficient and normal stress at the bottom of the rubber hemisphere. Here, the friction coefficient was evaluated using the contact stresses at the center contact node. As can be seen from the figure, the pressure dependence of the friction coefficient is expressed even in the finite element model via the macroscale friction model, thereby reflecting the mesoscale condition.



Localization Analysis

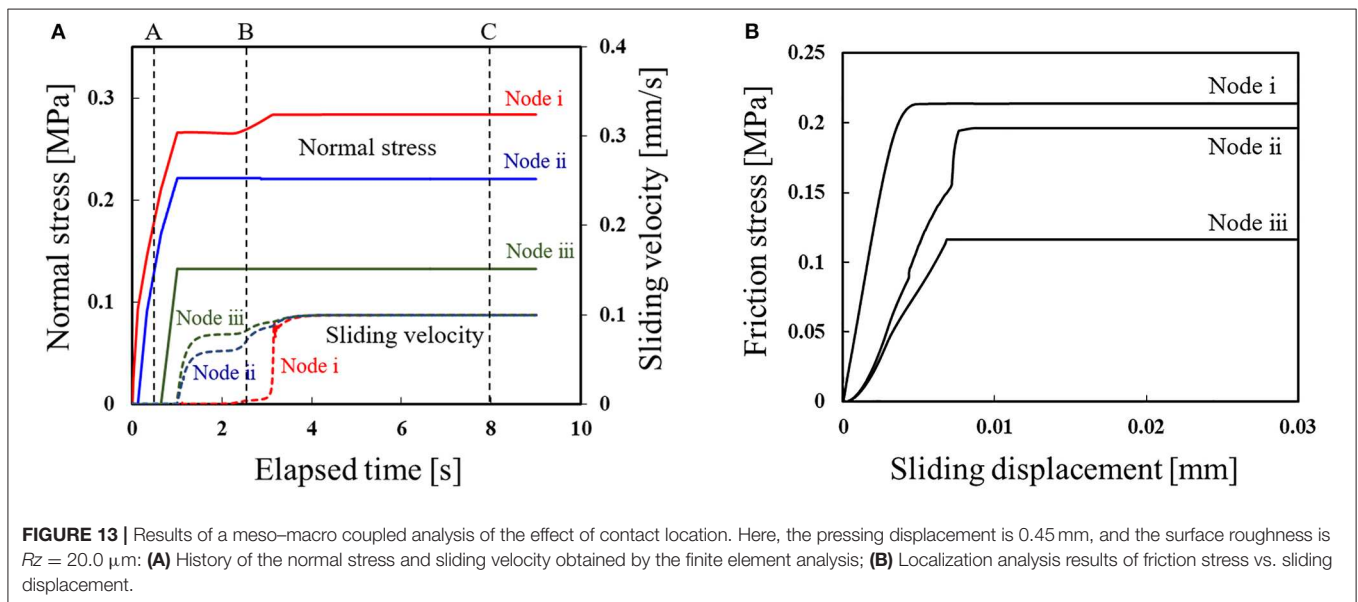
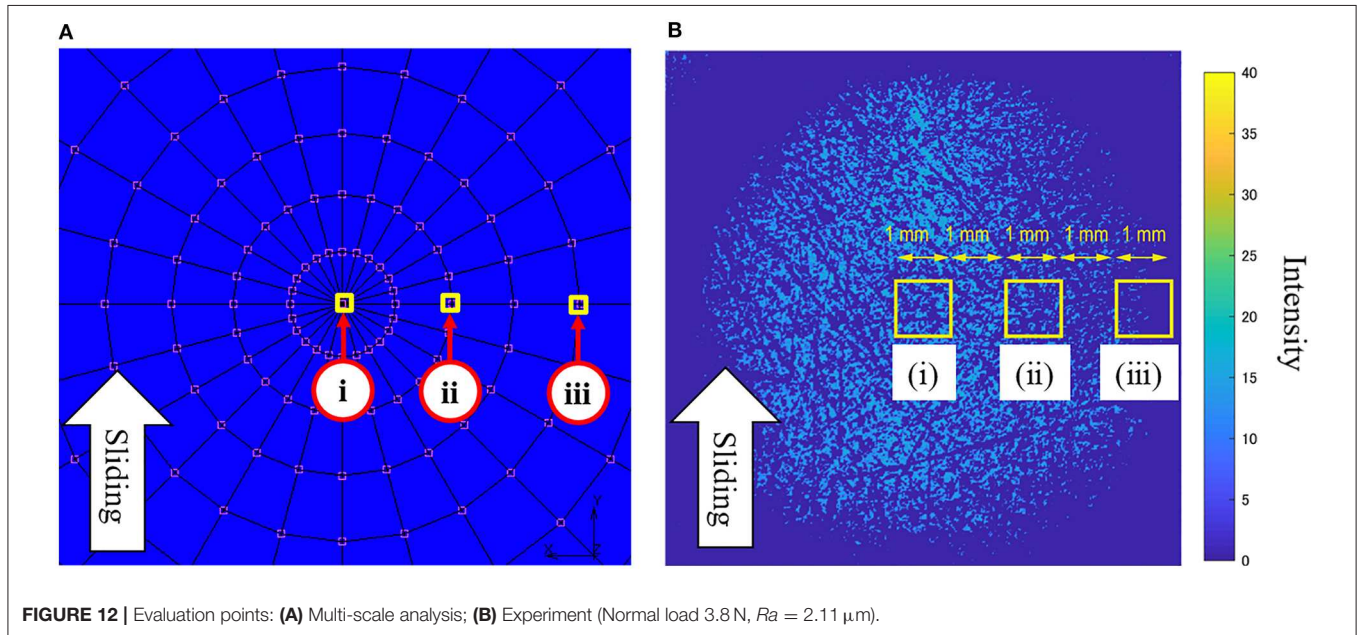
In this section, the results of the mesoscale localization analysis are discussed using the histories of the normal stress and sliding velocity of arbitrary contact nodes obtained by the macroscale finite element analysis as boundary conditions. The model used here is the same as that used in the numerical friction test described in Section Numerical Friction Test.

Effect of Vertical Pressing Displacement

In this section, the effect of the vertical pressing displacement of the rubber hemisphere is described. Here, the surface roughness of the rubber was $R_z = 20.0 \mu\text{m}$, and the vertical pressing displacement was evaluated at two levels. The node at the bottom

center of the hemisphere was used as the evaluation point. **Figure 8A** shows the variation in the normal stress and sliding velocity over time obtained by the finite element analysis. It is confirmed that the transition to gross sliding is delayed when the pressing displacement increases. Further, **Figure 8B** shows the results of the mesoscale analysis using the data in **Figure 8A** as input conditions; the graph shows the relationship between the frictional stress and the sliding displacement.

Times A, B, and C in **Figure 8A** represent the characteristic states corresponding to the vertical pressing process, the transient process between stick and slip, and the steady state, respectively. **Figures 9A,B** show the results of the mesoscale analysis of the contact area when the pressing displacement is 0.25 and



0.45 mm, respectively. Here, the contour range shows the ratio of contact area when the area $\pi\beta^2$ is set to 1.0. From the result at time A, it is observed that the magnitude of the normal stress affects the distribution of the real contact area. In addition, when the pressing displacement is smaller, the distribution of the real contact area at time B already shows the same area as under the steady state at time C. Meanwhile, when the pressing displacement is larger, sliding has hardly occurred at time B. Thereafter, gross sliding occurs and the normal pressure increases, and at time C, the real contact area increases. Thus, by performing the localization analysis, the microscopic state can be examined, such as the variation in friction stress and real contact area corresponding to the macroscale finite element analysis.

Effect of Surface Roughness

Next, the effect of surface roughness is discussed. Here, the pressing displacement was 0.35 mm, and two levels of rubber surface roughness were adopted ($R_z = 20.0\ \mu\text{m}$ and $56.5\ \mu\text{m}$). The node at the bottom center of the hemisphere was used as the evaluation point. **Figure 10A** shows the variation in the normal stress and sliding velocity with elapsed time obtained by the finite element analysis. **Figure 10B** shows the results of the mesoscale analysis using the data in **Figure 10A** as input conditions. The graph shows the relationship between the frictional stress and sliding displacement. From these figures, it is confirmed that the surface roughness influences the stress and sliding.

Figures 11A,B show the results of the mesoscale analysis of the contact area under conditions of $R_z = 20.0\ \mu\text{m}$ and $56.5\ \mu\text{m}$,

respectively. Here, the contour range shows the ratio of contact area when the area $\pi\beta^2$ is set to be 1.0. The focused times are shown in **Figure 10A**. From time A to time B, it can be seen that the real contact area increases, and its rate of increase is affected by the surface roughness. In the steady state (time C), the smaller the surface roughness, the larger the real contact area for the same levels of normal stress. As a result, as shown in **Figure 10B**, it is assumed that the variation in friction stress with elapsed time is affected.

Effect of Contact Location

This section describes the effect of the location of evaluation points in the contact zone on the determined friction behavior. **Figure 12A** shows the finite element model of the rubber hemisphere viewed from directly below. In the localization analysis, nodes (i)–(iii) shown in the figure were used as evaluation points.

Figure 13A shows the variation in the normal stress and sliding velocity with elapsed time obtained by finite element analysis. **Figure 13B** shows the results of the mesoscale analysis using data in **Figure 13A** as input conditions. Here, the pressing displacement is 0.45 mm and the surface roughness is $R_z = 20.0\ \mu\text{m}$. It can be seen in the figure that the normal stress at the bottom center [node (i)] is the largest, and the occurrence of gross sliding is the most delayed. In addition, it is confirmed that the normal stress decreases toward the outside of the contact area. The relationship between the friction stress and sliding displacement at the mesoscale also reflects the results of the

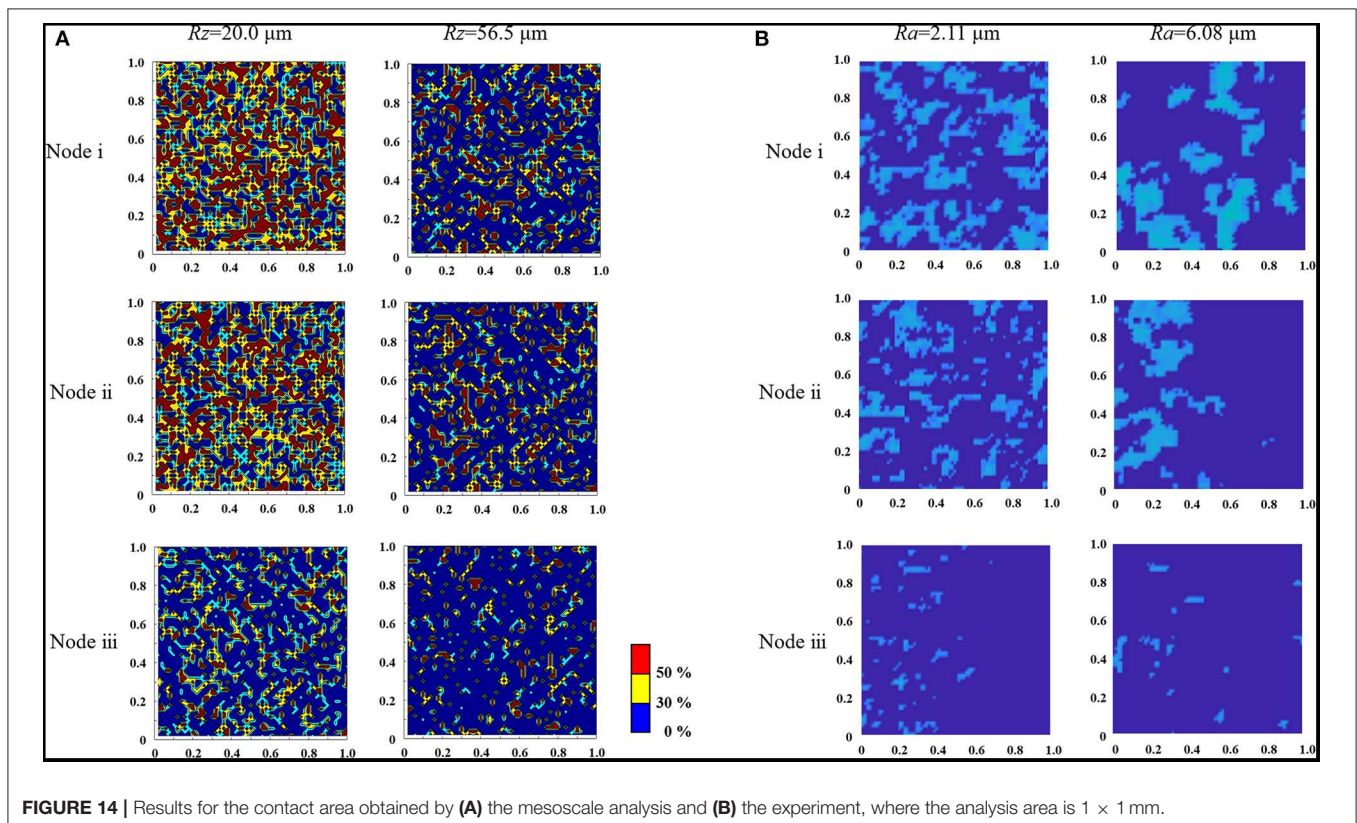


FIGURE 14 | Results for the contact area obtained by (A) the mesoscale analysis and (B) the experiment, where the analysis area is $1 \times 1\ \text{mm}$.

finite element analysis and its boundary conditions. Thus, by using the proposed meso–macro coupled analysis method, it is possible to study the microscopic behavior according to the geometric conditions and contact conditions with a macroscale finite element analysis.

Finally, the distribution characteristics of the real contact area obtained by the proposed meso–macro coupled analysis are qualitatively compared with the results of an experiment. **Figure 12B** shows the real contact area using a method for visualizing real contact regions formed in the apparent contact zone (Maegawa et al., 2015). Note that the rubbing materials used in the referenced study were the same as those used in this study, but their roughness conditions were different. The contour map in the figure shows the reflected light intensity after image processing, in which the bright portions correspond to the areas of real contact.

Figures 14A,B show a visualization of the real contact area results in steady state sliding obtained by the proposed multiscale analysis and previous experiment, respectively, corresponding to regions (i), (ii), and (iii) in **Figures 12A,B**, respectively. Although a quantitative comparison is difficult because of the different roughness profiles, it is confirmed that the relationship between the contact position and roughness determined by analysis and experiment reasonably agree. In other words, the ratio of the real contact area to apparent contact area decreases as the surface becomes rougher and closer to the outside of the contact area. However, the results of the mesoscale localization analysis are averaged over the representative contact area, so the normal stress distribution is homogeneous. Therefore, please note that the gradient of the results according to the stress distribution within the representative contact area is not represented as shown for the experiment results.

CONCLUSION

This study proposed a loop-type coupled analysis scheme to bridge the mesoscale and macroscale domains of friction analysis. Specifically, the mesoscale multipoint contact model was linked with the macroscale finite element analysis model

via the rate-, state-, and pressure-dependent friction model previously proposed by the authors (Ozaki et al., 2020). Next, the proposed method was applied to a contact problem between a rough rubber hemisphere and a smooth rigid plate. Then, systematic analyses were conducted, including a qualitative comparison with the results of a previous experiment, and the effectiveness of the proposed method in providing multiscale understanding of frictional sliding phenomena was examined. The proposed method was observed to enable a finite element analysis that reflects the characteristics of roughness at the mesoscale. Moreover, the analysis of elementary mesoscale behavior corresponding to the finite element analysis under arbitrary geometric and contact conditions can be performed in converse. Note that although other mesoscale and macroscale friction models (laws) can be used in the proposed method, the ability of the macroscale model to describe the complex frictional sliding phenomena model is essential. Additionally, attention must be paid to how the representative contact area is determined in the proposed method, as the contact area must be of a size for which the stress distribution in the macroscale analysis can be considered sufficiently uniform. Please note that a full quantitative comparison with experimental results still remains necessary. Thus, the authors intend to adopt a more advanced mesoscale contact model and use an actual profile of contact roughness in this model to perform a quantitative verification of the proposed method in a future study.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

SO and KM performed the numerical analysis and wrote the paper. SM conceived and designed the experiments and provided their detailed data. KN performed the conceptualization of mesoscale model.

REFERENCES

- Binnig, G., Quate, C. F., and Gerber, C. (1986). Atomic force microscope. *Phys. Rev. Lett.* 56, 930–933. doi: 10.1103/PhysRevLett.56.930
- Dieterich, J. H. (1972). Time-dependent friction in rocks. *J. Geophys. Res.* 77, 3690–3697. doi: 10.1029/JB077i020p03690
- Dieterich, J. H. (1979). Modeling of rock friction: 1. Experimental results and constitutive equations. *J. Geophys. Res.* 84, 2161–2168. doi: 10.1029/JB084iB05p02161
- Fuller, K. N. G., and Tabor, D. (1975). The effect of surface roughness on the adhesion of elastic solids. *Proc. R. Soc. Lond. A* 345, 327–342. doi: 10.1098/rspa.1975.0138
- Greenwood, J. A., and Williamson, J. B. P. (1966). Contact of nominally flat surface. *Proc. R. Soc. Lond. A* 295, 300–319. doi: 10.1098/rspa.1966.0242
- Hashiguchi, K., and Ozaki, S. (2008). Constitutive equation for friction with transition from static to kinetic friction and recovery of static friction. *Int. J. Plasticity*. 24, 2102–2124. doi: 10.1016/j.jiplas.2008.03.004
- Hyun, S., Pei, L., Molinari, J.-F., and Robbins, M. O. (2004). Finite-element analysis of contact between elastic self-affine surfaces. *Phys. Rev. E* 70:026117. doi: 10.1103/PhysRevE.70.026117
- Johnson, K. L., Greenwood, J. A., and Higginson, J. G. (1985). The contact of elastic regular wavy surfaces. *Int. J. Mech. Sci.* 27, 383–396. doi: 10.1016/0020-7403(85)90029-3
- Kikuchi, N., and Oden, J. T. (1988). *Contact Problem in Elasticity: A Study of Variational Inequalities and Finite Element Methods*. Philadelphia, PA: SIAM.
- Laursen, T. A. (2001). *Computational Contact and Impact Mechanics: Fundamentals of Modeling International Phenomena in Nonlinear Finite Element Analysis*. Berlin; Heidelberg: Springer-Verlag.
- Maegawa, S., Itoigawa, F., and Nakamura, T. (2015). Effect of normal load on friction coefficient for sliding contact between rough rubber surface and rigid smooth plane. *Tribol. Int.* 92, 335–343. doi: 10.1016/j.triboint.2015.07.014
- Manners, W. (2000). Pressure required to flatten an elastic random rough profile. *Int. J. Mech. Sci.* 42, 2321–2336. doi: 10.1016/S0020-7403(99)00098-3
- Marc, M. S. C. (2020). *User's Manual*. MSC Software.

- Ozaki, S., and Hashiguchi, K. (2010). Numerical analysis of stick-slip instability by a rate-dependent elastoplastic formulation for friction. *Tribol. Int.* 43, 2120–2133. doi: 10.1016/j.triboint.2010.06.007
- Ozaki, S., Hikida, K., and Hashiguchi, K. (2012). Elastoplastic formulation for friction with orthotropic anisotropy and rotational hardening. *Int. J. Solids Struct.* 49, 648–657. doi: 10.1016/j.ijsolstr.2011.11.010
- Ozaki, S., Ito, C., and Hashiguchi, K. (2013). Experimental verification of rate-dependent elastoplastic analogy friction model and its application to FE analysis. *Tribol. Int.* 64, 164–177. doi: 10.1016/j.triboint.2013.03.016
- Ozaki, S., Matsuura, T., and Maegawa, S. (2020). Rate-, state-, and pressure-dependent friction model based on the elastoplastic theory. *Friction* 1–16. doi: 10.1007/s40544-019-0321-3
- Persson, B. N. J. (2001). Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115, 3840–3861. doi: 10.1063/1.1388626
- Persson, B. N. J., Bucher, F., and Chiaia, B. (2002). Elastic contact between randomly rough surfaces: comparison of theory with numerical results. *Phys. Rev. B* 65:184106. doi: 10.1103/PhysRevB.65.184106
- Persson, B. N. J., and Volokitin, A. I. (2006). Rubber friction on smooth surfaces. *Eur. Phys. J. E* 21, 69–80. doi: 10.1140/epje/i2006-10045-9
- Roberts, A. D. (1992). A guide to estimating the friction of rubber. *Rubber Chem. Technol.* 65, 673–686. doi: 10.5254/1.3538633
- Roberts, A. D., and Thomas, A. G. (1975). The adhesion and friction of smooth surfaces. *Wear* 33, 45–64. doi: 10.1016/0043-1648(75)90223-9
- Ruina, A. L. (1983). Slip instability and state variable friction laws. *J. Geophys. Res.* 88, 359–370. doi: 10.1029/JB088iB12p10359
- Schallamach, A. (1971). How does rubber slide?. *Wear* 17, 301–312. doi: 10.1016/0043-1648(71)90033-0
- Tabor, D. (1960). Hysteresis losses in the friction of lubricated rubber. *Rubber Chem. Technol.* 33, 142–150. doi: 10.5254/1.3542120
- Watanabe, I., and Terada, K. (2010). A method of predicting macroscopic yield strength of polycrystalline metals subjected to plastic forming by micro-macro de-coupling scheme. *Int. J. Mech. Sci.* 52, 343–355. doi: 10.1016/j.ijmecsci.2009.10.006
- Wriggers, P. (2003). *Computational Contact Mechanics*. Chichester: John Wiley.
- Yang, C., and Persson, B. N. J. (2008). Contact mechanics: contact area and interfacial separation from small contact to full contact. *J. Phys. Condens. Matter.* 20:215214, doi: 10.1088/0953-8984/20/21/215214

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ozaki, Mieda, Maegawa and Nakano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Transient and Steady Sliding Friction of Elastomers: Impact of Vertical Lift

Ken Nakano* and Masaharu Kono

Faculty of Environment and Information Sciences, Yokohama National University, Yokohama, Japan

OPEN ACCESS

Edited by:

Mitjan Kalin,
University of Ljubljana, Slovenia

Reviewed by:

Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine
Andrey Dimaki,
Institute of Strength Physics and
Materials Science (ISPMS SB
RAS), Russia

*Correspondence:

Ken Nakano
nakano@ynu.ac.jp

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 16 March 2020

Accepted: 12 May 2020

Published: 28 July 2020

Citation:

Nakano K and Kono M (2020)
Transient and Steady Sliding Friction
of Elastomers: Impact of Vertical Lift.
Front. Mech. Eng. 6:38.
doi: 10.3389/fmech.2020.00038

The sliding friction of elastomers has been investigated numerically and theoretically for the line contact between a cylindrical rigid indenter and a “frictionless” Kelvin–Voigt foundation. The onset of sliding under an abrupt increase in the drive velocity has been simulated with different boundary conditions of the rigid indenter. When the rigid indenter is not allowed to move in any direction, just an abrupt change in the friction force appears, which is not accompanied by any transient processes. However, when the rigid indenter is able to move in the vertical direction, the transient sliding friction including three different time constants appears, resembling the typical transition from the static friction to the kinetic friction, in spite of no static friction considered in the simulation. The aforementioned drastic difference is caused by the “vertical lift” of the rigid indenter, which is induced by the damping of the Kelvin–Voigt foundation. In addition, the vertical lift strongly affects the characteristics of the steady sliding friction, which is explained well by using the critical velocity determined from the asymptotes in the master curve of friction coefficient.

Keywords: tribology, dynamics, rheology, elastomers, Kelvin–Voigt foundation

INTRODUCTION

It has been well-known that sliding contact of elastomers involves various types of dynamics due to friction. Physical phenomena such as wave propagations (Schallamach, 1971; Barquins, 1985; Rubinstein et al., 2004; Maegawa and Nakano, 2010), wear pattern formations (Schallamach, 1957; Fukahori and Yamazaki, 1994), and nonlinear vibrations (Nakano and Maegawa, 2009; Yamaguchi et al., 2011; Nakano et al., 2019) have attracted the interest of many scientific and engineering researchers. Among various elastomers, rubber is the material that has received the most attention in various practical applications (e.g., tires, seals, and shoes). Since the study done by Grosch (1963) showing the master curve of friction coefficient according to Williams–Landel–Ferry (WLF) theory (Williams et al., 1955), the importance of viscoelasticity has been recognized, and the dependence of the friction coefficient on temperature and velocity has been investigated [e.g., Popov et al. (2018)]. More recently, several swollen polymers showing high elasticity and ultra-low friction (e.g., hydrogels and polymer brushes) have been found (Gong et al., 2001; Nomura et al., 2011), some of which are being developed toward practical applications (Belin et al., 2018; Tadokoro et al., 2020), and their tribological properties have also been discussed in relation to their viscoelastic properties (Mizukami et al., 2019).

To understand the sliding friction of elastomers, various types of modeling methods have been proposed. Among them, those with “viscoelastic foundations” are known to have strong advantages not only of avoiding the difficulties of elastic contact stress theory but also of providing intuitive pictures of how the energy dissipation occurs inside the contact. The first was the extension of the elastic foundation (i.e., the Winkler foundation) from stationary contact problems of thin

films to steady sliding contact problems of elastomers (May et al., 1959; Johnson, 1985). Then recently, in the novel framework of the “Method of Dimensionality Reduction” invented by Popov and his colleagues, the modeling method has progressed, enabling us to analyze the true three-dimensional contact with high accuracy (Popov, 2010; Popov and Heß, 2015; Kusche, 2017). For example, Li et al. (2015) numerically and theoretically studied the kinetics of friction coefficient for the sliding contact between a flat elastomer (modeled by a Kelvin–Voigt foundation) and a rough rigid indenter (modeled by a self-affine fractal) under abrupt change in the drive velocity. As a result, they found interesting temporal changes (i.e., jumps and relaxations) of the friction coefficient depending on the drive velocity and the Hurst exponent of the self-affine fractal, under the consideration of quasi-static processes.

Based on the foregoing background, the aim of this study was to find the answer to the following question: What is the minimum requirement for modeling with viscoelastic foundations to describe the transient friction appearing in the onset of sliding between an elastomer and an indenter? In this article, to provide a possible answer in a minimal situation, we consider a simple line contact between a flat elastomer and a cylindrical rigid indenter, focusing on the boundary condition of the rigid indenter in the vertical direction, under the restriction that the elastomer is modeled by the conventional Kelvin–Voigt foundation. Two types of models with different boundary conditions reveal that the vertical dynamics of the rigid indenter strongly affects not only the occurrence of the transient sliding friction but also the characteristics of the steady sliding friction.

Note that the aforementioned idea on the boundary condition in the vertical direction originates from the typical structure of sliding systems under the fluid film lubrication, since when the fluid film lubrication is considered theoretically, it is natural to assume that the “slider” (corresponding to the “indenter” in this study) shifts vertically to find the position to balance with the applied normal load. Also note that in general, it has been known that the friction of elastomers arises by two different mechanisms. One is the “adhesion,” the energy dissipation of which occurs at the contact interface between the elastomer and indenter [i.e., the adhesive friction (e.g., Moore and Geyer (1972))], and the other is the “viscoelasticity,” the energy dissipation of which occurs inside the elastomers [i.e., the hysteresis friction (e.g., Moore and Geyer (1974))]. Although both mechanisms are important and should be interrelated, this study ignores the former and focuses on the latter.

MODELS

Structures

Figure 1 shows the two types of models for sliding friction of elastomers considered in this study. The left is termed the “fixed indenter (FI) model,” and the right is termed the “movable indenter (MI) model.” They are models describing the two-dimensional sliding contact between a rigid indenter and a viscoelastic foundation in the xz plane, where the x and z axes are taken in the horizontal and vertical directions, respectively.

The rigid indenter (mass per unit width: M) has a cylindrical shape (curvature radius: R), the bottom surface of which contacts with the viscoelastic foundation. In the FI model, the rigid indenter is mounted on the rigid walls, not to be able to move in any direction. In the MI model, on the other hand, the rigid indenter is supported by a frictionless linear bearing, to be able to move only in the vertical direction. Note that the boundary condition of the rigid indenter is the single difference between them.

The viscoelastic foundation consists of an infinite number of viscoelastic elements mounted on a rigid base at regular intervals in the horizontal direction. Every viscoelastic element is a one-degree-of-freedom Kelvin–Voigt element consisting of a vertical spring (stiffness per unit width: k) and a vertical damper (damping coefficient per unit width: c) with the same natural length; as a result, its non-disturbed surface is the horizontal plane $z = 0$. Note that by using the stiffness per unit area (K) and damping coefficient per unit area (C) of the viscoelastic foundation, k and c are given by $k = K/N$ and $c = C/N$, respectively, where N is the number of viscoelastic elements per unit length. Also note that here we assume that the upper end of every viscoelastic element makes contact with the rigid indenter surface with no adhesion and with no friction.

Owing to the aforementioned structures, the normal load per unit width (W) of the FI model is determined by the indentation depth δ [i.e., $W = W(\delta)$, where δ is constant], while that of the MI model is determined by the gravity [i.e., $W = Mg$, where g is the gravity constant]. In addition, in both models, the rigid base is driven horizontally at the drive velocity V . Therefore, we can say that the FI model represents a “constant-gap sliding system,” while the MI model represents a “dead-weight sliding system.” Or, we can also say that the FI and MI models represent the two limiting cases corresponding to a “very stiff apparatus” and a “very soft apparatus,” respectively.

Governing Equations

When the rigid indenter penetrates the viscoelastic foundation by $\delta > 0$, the indenter surfaces in the FI and MI models are given by

$$z = h(x) - \delta \text{ (FI model)} \quad (1)$$

$$z = h(x) - \delta(t) \text{ (MI model)} \quad (2)$$

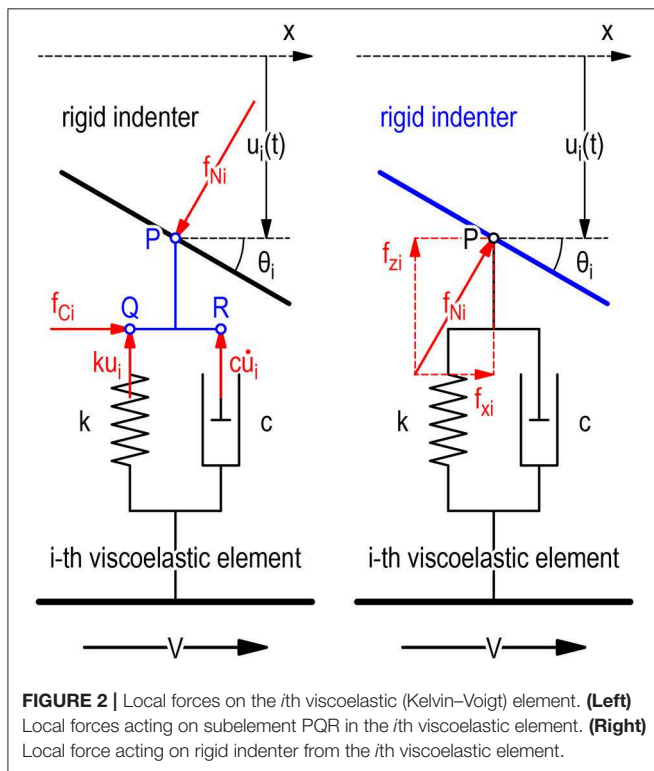
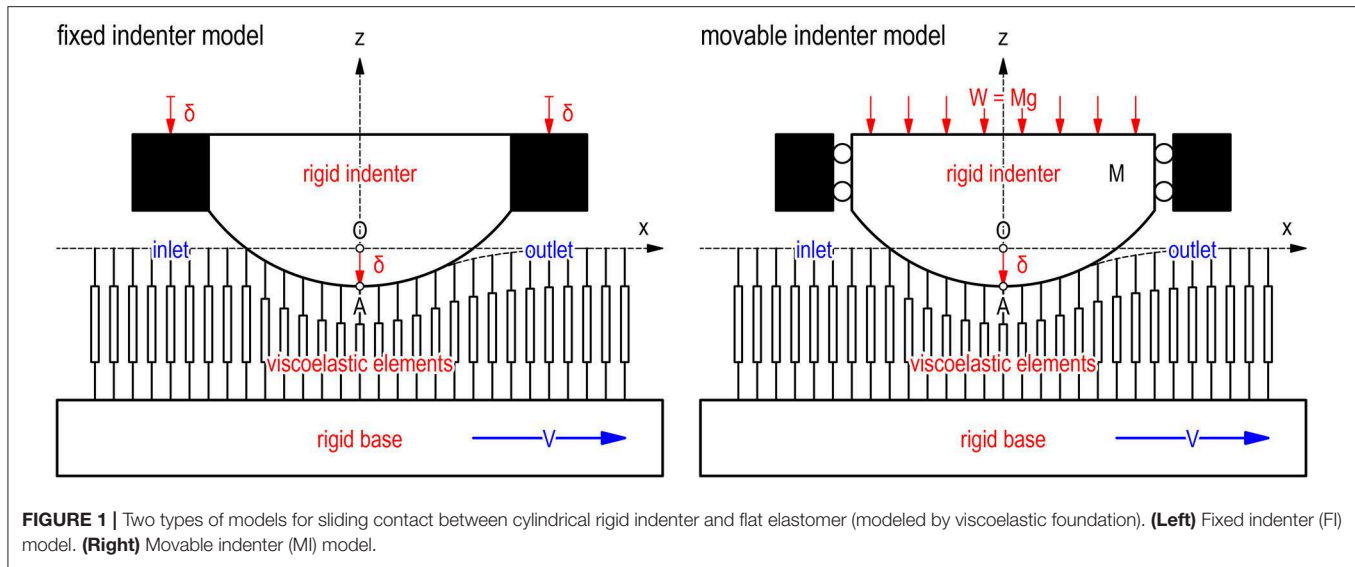
respectively, where t is the time, and $h(x)$ is the indenter shape function:

$$h(x) = \frac{x^2}{2R} \quad (3)$$

Let us assume that the i th viscoelastic element at $x = x_i(t)$ makes contact with the indenter surface (see **Figure 2**). The compression and compression rate of the i th viscoelastic element in the FI model are given by

$$u_i(t) = -h(x_i) + \delta \text{ (FI model)} \quad (4)$$

$$\dot{u}_i(t) = -h'(x_i) V \text{ (FI model)} \quad (5)$$



respectively, while in the MI model,

$$u_i(t) = -h(x_i) + \delta(t) \quad (\text{MI model}) \quad (6)$$

$$\dot{u}_i(t) = -h'(x_i)V + \dot{\delta}(t) \quad (\text{MI model}) \quad (7)$$

respectively, where (\cdot) and $(\dot{\cdot})$ are the derivatives with respect to t and x , respectively. Then, as shown in the left of **Figure 2**, there are four types of local forces acting on the subelement PQR: the normal force f_{Ni} at P (from the rigid indenter), the restoring force

ku_i at Q (from the spring), the damping force $c\dot{u}_i$ at R (from the damper), and the horizontal constraint force f_{Ci} (from the spring and damper). Considering that the subelement PQR is massless, we obtain the following force balance equations in the vertical and horizontal directions:

$$ku_i + c\dot{u}_i = f_{Ni} \cos \theta_i \quad (8)$$

$$f_{Ci} = f_{Ni} \sin \theta_i \quad (9)$$

respectively, where

$$\tan \theta_i = -h'(x_i) \quad (10)$$

under $-\pi/2 < \theta_i < \pi/2$, and from Equation (8),

$$f_{Ni} = \frac{ku_i + c\dot{u}_i}{\cos \theta_i} \quad (11)$$

Now, as shown in the right of **Figure 2**, the normal force acts on the rigid indenter at P from the *i*-th viscoelastic element, which is the reaction force of the normal force f_{Ni} on the left of **Figure 2**. Therefore, the vertical and horizontal components of the normal force are given by

$$f_{zi} = f_{Ni} \cos \theta_i = ku_i + c\dot{u}_i \quad (12)$$

$$f_{xi} = f_{Ni} \sin \theta_i = -h'(x_i)f_{zi} \quad (13)$$

respectively. Note that from Equations (9) and (13), the horizontal constraint force is determined as $f_{Ci} = f_{xi}$. Consequently, the total vertical force F_z and total horizontal force F_x acting on the rigid indenter from the viscoelastic foundation are given by

$$F_z = \sum_i f_{zi} \quad (14)$$

$$F_x = \sum_i f_{xi} \quad (15)$$

respectively. Finally, the vertical position of the indenter bottom A in the FI model is given by

$$z_A = -\delta \text{ (FI model)} \quad (16)$$

while in the MI model, it is determined by the equation of motion:

$$M\ddot{z}_A = F_z - Mg \text{ (MI model)} \quad (17)$$

which determines δ and $\dot{\delta}$ in Equations (6) and (7) as follows:

$$\delta(t) = -z_A(t) \text{ (MI model)} \quad (18)$$

$$\dot{\delta}(t) = -\dot{z}_A(t) \text{ (MI model)} \quad (19)$$

Note that when the i th viscoelastic element is not making contact with the indenter surface, it is obvious that $f_{Ni} = 0$ and therefore $f_{zi} = f_{xi} = 0$. Then, the compression rate of the non-contacting element is given by

$$\dot{u}_i = -\frac{u_i}{\tau} \text{ (non-contact)} \quad (20)$$

where τ is the retardation time of the viscoelastic element, defined as

$$\tau = \frac{c}{k} \quad (21)$$

or, by using the macroscopic properties of the viscoelastic foundation,

$$\tau = \frac{C}{K} \quad (22)$$

Therefore, when the contacting element violates the following condition:

$$u_i \geq h'(x_i) V \tau \text{ (FI model)} \quad (23)$$

$$u_i \geq h'(x_i) V \tau + \dot{z}_A \tau \text{ (MI model)} \quad (24)$$

it detaches from the rigid indenter surface. Note that the first-order differential equation (20) for the non-contacting element has the following solution:

$$u_i(t) = u_i(t_0) \exp\left(-\frac{t-t_0}{\tau}\right) \text{ (non-contact)} \quad (25)$$

where t_0 is the constant.

METHODS

The sets of the governing equations in the previous section were solved numerically for the FI and MI models. To solve the second-order ordinary differential equation (17) for the MI model, the Runge–Kutta method was used, where the time discretization was determined by $\Delta t = 2\Delta x/V = 2/NV$.

Standard parameter values for the numerical simulations are listed in **Table 1**. Note that both models include six independent parameters (R , K , τ , N , δ , and V in the FI model; R , M , K , τ , N , and V in the MI model). However, since $N (= 1/\Delta x)$ is the parameter for the space discretization, the number of

TABLE 1 | Standard parameter values for numerical simulations.

Rigid indenter	
Curvature radius R	10 mm
Mass per unit width M (for MI model)	1 kg/mm
Viscoelastic foundation	
Stiffness per unit area K	10 GN/m ³
Retardation time $\tau (= C/K)$	100 ms
Number of elements per unit length N	200 mm ⁻¹
Operation	
Indentation depth δ (for FI model)	300 μ m
Drive velocity V	100 mm/s

essential parameters is five for each model. Also note that if we imagine a “thin” elastic sheet of thickness $h = 1$ mm as the viscoelastic foundation, it is possible to say that the stiffness per unit area $K = 10$ GN/m³ corresponds to the effective elastic modulus $E \sim Kh = 10$ MPa, although it depends strongly on the boundary condition of mounting the elastomer on the rigid base [e.g., Popov (2010)].

RESULTS

Figure 3 shows the numerical results of temporal changes in representative variables for the FI model (left column) and MI model (right column) under the standard conditions (**Table 1**). Note that as shown by the blue lines in the top row, the drive velocity was abruptly increased from $V = 0$ to 100 mm/s at $t = 0$.

First, from the results for the FI model, we find that the indenter position provided the initial condition ($z_A = -0.3$ mm, $F_z \sim 10$ N/mm, and $F_x = 0$). Then, F_z and F_x were immediately increased at $t = 0$ and maintained the increased values ($z_A = -0.3$ mm, $F_z \sim 35$ N/mm, and $F_x \sim 5$ N/mm).

Then, from the results for the MI model, we find that the gravity force provided the same initial condition as that of the previous model. However, the responses were completely different, showing distinct “transient sliding” toward “steady sliding.” The rigid indenter started to move “upward” at $t = 0$ and then gradually approached $z_A \sim -0.1$ mm with a time constant in the order of “10 ms” (second row). The response of F_z was spiky: it immediately increased to $F_z \sim 30$ N/mm at $t = 0$, then rapidly decreased to $F_z \sim 9$ N/mm with a time constant in the order of “1 ms,” and gradually returned to $F_z \sim 10$ N/mm (third row). The response of F_x was also spiky: it immediately increased to $F_x \sim 5$ N/mm at $t = 0$, then rapidly decreased to $F_x \sim 2$ N/mm with the time constant “1 ms,” and gradually approached $F_x \sim 1$ N/mm with the time constant “10 ms” (bottom row). It should be noted that the response of F_x resembles the typical transition from the static friction to the kinetic friction, although no static friction was considered. This transient behavior will be discussed in the next section.

Figure 4 shows the numerical results of spatiotemporal changes in the contact pressure $p(x,t) = Nf_N(x,t)$ under the same conditions as those of the previous figure, where the abscissa

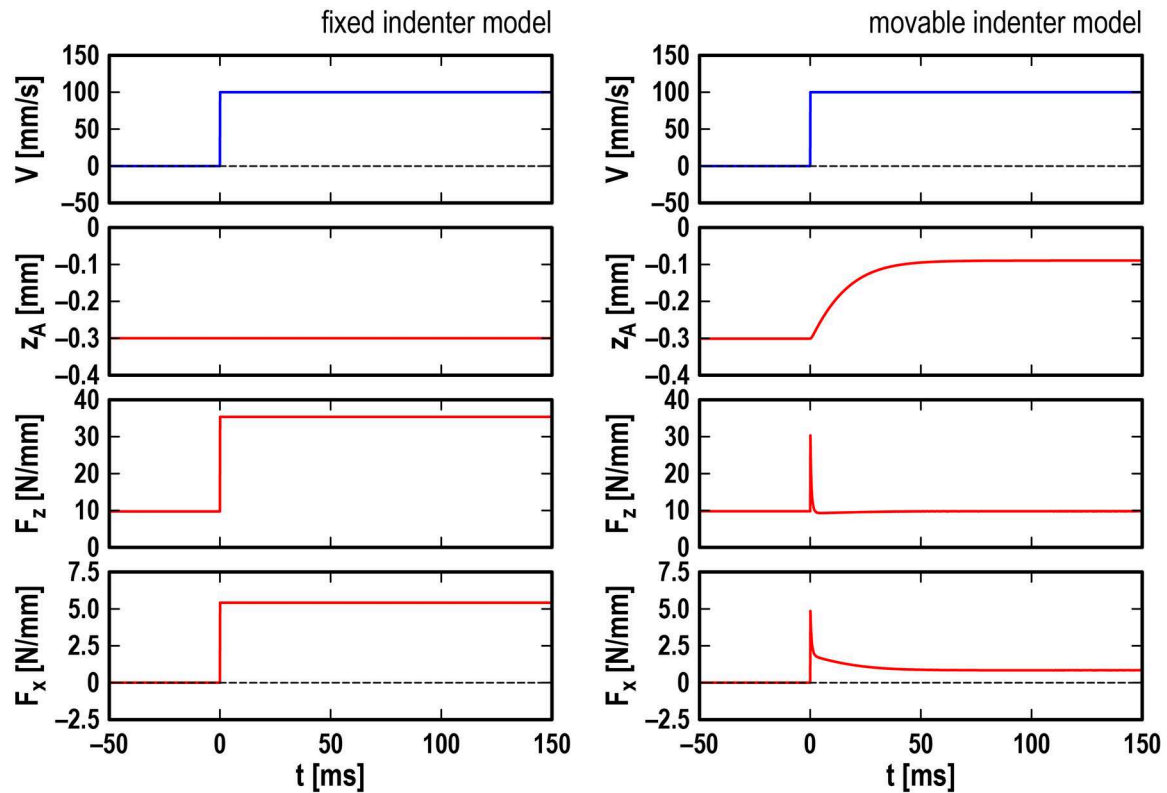


FIGURE 3 | Numerical results for fixed indenter model (Left) and movable indenter model (Right): temporal changes in drive velocity V (top row), indenter position z_A (second row), total vertical force F_z (third row), and total horizontal force F_x (bottom row) under standard conditions (see Table 1 for parameter values).

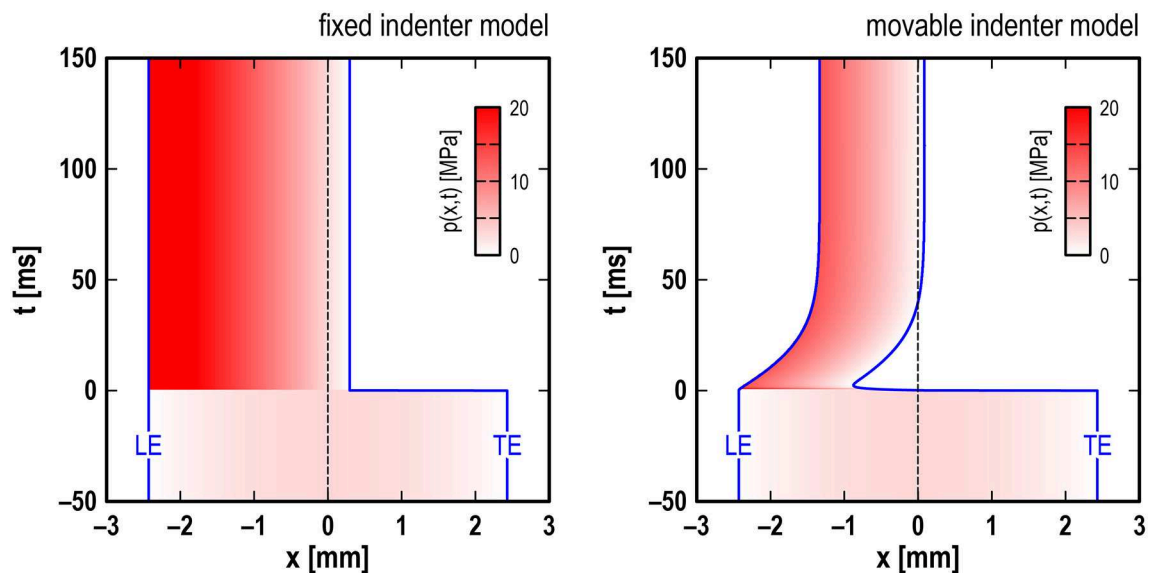


FIGURE 4 | Numerical results for fixed indenter model (Left) and movable indenter model (Right): spatiotemporal changes in contact pressure $p(x,t)$ under standard conditions (see Table 1 for parameter values). LE, leading edge of contact; TE, trailing edge of contact.

is the horizontal position x , the ordinate is the time t , and the magnitude of p is represented by the shade of the red color. The blue lines show temporal changes in the horizontal positions of the leading and trailing edges of the contact area, denoted by $x_{LE}(t)$ and $x_{TE}(t)$, respectively.

First, from the results for the FI model, we again find that the response was not accompanied by any transient processes. Regarding the edges, they were located at $x_{LE} = -2.4$ mm and $x_{TE} = 2.4$ mm as the initial state. After $t = 0$, the leading edge maintained the position, while the trailing edge immediately moved to the left to $x_{TE} = 0.3$ mm at $t = 0$ and maintained the position for $t > 0$ (which is observed as the peeling of the contact area in the outlet side). Regarding the contact pressure, it was symmetric for $t < 0$, showing the maximum $p_{max} \sim 3$ MPa at the contact center and the minimum $p_{min} = 0$ at both edges. After $t = 0$, it showed the maximum $p_{max} > 20$ MPa at the leading edge and the minimum $p_{min} = 0$ at the trailing edge.

Then, from the results for the MI model, we again find that the response was accompanied by distinct transient processes. Regarding the edges, they were located at the same positions as those of the previous model. However, after $t = 0$, not only the trailing edge but also the leading edge moved (which is observed as the simultaneous peeling of the contact area in both sides). The leading edge started to move to the right at $t = 0$ and then gradually approached $x_{LE} = -1.3$ mm with the time constant “10 ms.” Meanwhile, the trailing edge immediately moved to the left to $x_{TE} \sim 0$ at $t = 0$, then continued to move to the left to $x_{TE} \sim -1$ mm with the time constant “1 ms,” and then gradually approached a limiting value of $x_{TE} = 0.1$ mm with the time constant “10 ms.” Regarding the contact pressure, starting from the same initial state as that of the previous model, after $t = 0$, it showed the maximum at the leading edge and the minimum at the trailing edge. For example, in the steady sliding, $p_{max} \sim 14$ MPa at the leading edge and $p_{min} = 0$ at the trailing edge.

DISCUSSION

Transient Sliding Friction

Through direct comparison of the numerical results for the two types of models (Figure 1), we have found that the boundary condition of the rigid indenter is critical for the occurrence of transient sliding. Letting the rigid indenter be vertically movable causes its upward motion (termed the “vertical lift” of the rigid indenter) when the drive velocity is applied (Figure 3), leading to a drastic spatiotemporal change in the area and pressure of the contact (Figure 4).

First, in the FI model, the horizontal positions of the leading and trailing edges are given by

$$x_{LE} = -a \text{ (FI model)} \quad (26)$$

$$x_{TE} = \sqrt{a^2 + l_r^2} - l_r \text{ (FI model)} \quad (27)$$

respectively, where a is the half length of the stationary contact, and l_r is the retardation length, defined as

$$a = \sqrt{2R\delta} \quad (28)$$

$$l_r = V\tau \quad (29)$$

respectively. Note that x_{LE} is constant, while x_{TE} is a function of l_r . From Equations (26) and (27), we find that there are two limiting cases:

$$(x_{LE}, x_{TE}) = (-a, a) \text{ for } l_r \ll a \text{ (FI model)} \quad (30)$$

$$(x_{LE}, x_{TE}) = (-a, 0) \text{ for } l_r \gg a \text{ (FI model)} \quad (31)$$

where the former means the “non-peeling” of the entire contact area, and the latter means the “complete peeling” of the contact area in the outlet side. Therefore, for example, as the drive velocity is increased, the contact area changes from “symmetric” [Equation (30)] to “asymmetric” [Equation (31)]. Note that x_{TE} given by Equation (27) is immediately determined when the drive velocity is applied, which is the reason why the response of the FI model is not accompanied by any transient processes.

Then, in the MI model, the horizontal positions of the leading and trailing edges are given by

$$x_{LE} = -a \text{ (MI model)} \quad (32)$$

$$x_{TE} = \sqrt{a^2 \left(1 + \tau \frac{\dot{\delta}}{\delta}\right)} + l_r^2 - l_r \text{ (MI model)} \quad (33)$$

respectively, where $\delta = \delta(t)$ and therefore $a = a(\delta) = a(t)$, which is the reason why the MI model creates transient processes. In Equations (32) and (33), we find two sources creating temporal changes. One is $a \sim \delta^{1/2}$, which shrinks the contact area symmetrically by the vertical lift of the rigid indenter. From the temporal change in x_{LE} in the right of Figure 4, we can say that under the standard condition, the time constant “10 ms” was caused by this effect. (In addition, the temporal change in z_A on the right of Figure 3 supports this conclusion.) The other is $\dot{\delta}/\delta$ located under the square-root sign in Equation (33), working only for the trailing edge, which deforms the contact area asymmetrically. From the temporal change in x_{TE} on the right of Figure 4, we can say that under the standard condition, the time constant “1 ms” was caused by this effect.

Based on the foregoing, let us consider the spiky response of F_x on the right of Figure 3. As we saw in the previous section, the temporal change in F_x in the MI model is quite similar to the typical transition from the static friction to the kinetic friction, in spite of no static friction considered in the simulation. If we saw this type of response in experiments, we probably believed that this was caused by the typical adhesive friction consisting of two types of friction. In fact, the spiky response consists of the following three parts. The first is the immediate increase responding to the abrupt increase in the drive velocity at $t = 0$. This is obviously caused by the damping C of the Kelvin-Voigt foundation, which is essentially identical to the response observed in the FI model at $t = 0$. The second is the rapid decrease with the time constant “1 ms.” Considering the discussion in the previous paragraph, we can conclude that the rapid decrease is caused by the rapid motion of the trailing edge to reduce the contact area. The third is the gradual decrease with the time constant “10 ms.” Considering the discussion in the previous paragraph, it is natural to say that the gradual decrease in F_x is caused by the gradual motion of both edges. However,

also considering that the contact pressure takes the maximum p_{\max} at the leading edge and the minimum $p_{\min} = 0$ at the trailing edge (Figure 4), we can conclude that the gradual decrease in F_x is mainly caused by the gradual motion of the leading edge. Again note that the second and third parts of F_x in the MI model never appear in the FI model, which tells us that the vertical lift of the rigid indenter is essential to the spiky response.

Steady Sliding Friction

Through the numerical simulations to examine the response to the abrupt increase in the drive velocity, we have found that the response of the FI model is not accompanied by any transient processes (which means that the steady sliding friction appears immediately), while the response of the MI model is accompanied by distinct transient processes followed by the steady sliding friction. In this section, we focus on the steady sliding friction.

The situation of steady sliding is given by

$$\dot{\delta} = 0 \text{ (steady sliding)} \quad (34)$$

which makes the governing equations for the MI model reduce to those for the FI model. For example,

$$u_i(t) = -h(x_i) + \delta \text{ (steady sliding)} \quad (35)$$

$$\dot{u}_i(t) = -h'(x_i) V \text{ (steady sliding)} \quad (36)$$

However, as seen in the numerical results for $t > 100$ ms in Figure 3, the normal load $W (= F_z \text{ in steady sliding})$ and friction force $F (= F_x \text{ in steady sliding})$ for the MI model are considerably different from those for the FI model (i.e., $W \sim 35$ N/mm and $F \sim 5$ N/mm for the FI model, while $W \sim 10$ N/mm and $F \sim 1$ N/mm for the MI model). Note that the difference is caused by the boundary condition of the rigid indenter: the FI model represents a “constant-gap sliding system” in which the indentation depth δ is controlled, while the MI model represents a “dead-weight sliding system” in which the normal load $W (= Mg)$ is controlled. Therefore, when one tries to measure the sliding friction of elastomers, it is mandatory to pay much attention to the boundary condition of the counter surface: otherwise, measured values could lose their meaning.

The foregoing discussion is supported by Figure 5, which shows numerical results in the steady sliding for the FI model (left column) and the MI model (right column). The dependences of δ (top row), W (second row), F (third row), and μ (bottom row) on V in the steady sliding are summarized, where μ is the friction coefficient in steady sliding, defined as

$$\mu = \frac{F}{W} \text{ (steady sliding)} \quad (37)$$

The τ -values are 10^{-2} s (black), 10^{-1} s (blue), and 10^0 s (red), and the other parameter values are the same as those of the standard condition (Table 1). Owing to the difference of boundary conditions, the velocity dependences of μ for the FI and MI models are different from each other: under high- V conditions, the FI model shows the limiting value $\mu \sim 0.2$,

while the MI model shows a negative dependence of μ on V with a slope of -0.5 . In addition, the three curves in every graph in Figure 5 are found to be located at regular intervals, which means that the product of V and τ (i.e., $l_r = V\tau$) is an essential parameter for both models.

Through a series of numerical simulations under various sets of parameters, master curves on μ in the steady sliding for the two types of models were obtained (see Figure 6). The red curve in the left graph is the master curve for the FI model, the ordinate and abscissa of which are $\mu(R/\delta)^{1/2}$ and $V\tau(R\delta)^{-1/2}$, respectively, while the red curve in the right graph is the master curve for the MI model, the ordinate and abscissa of which are $\mu(KR^2/W)^{1/3}$ and $V\tau(K/RW)^{1/3}$, respectively. It should be stressed that every quantity assigned to the axes in Figure 6 is dimensionless.

To examine the asymptotes of the master curves, we again consider the two limiting cases:

$$(x_{LE}, x_{TE}) = (-a, a) \text{ for } l_r \ll a \text{ (steady sliding)} \quad (38)$$

$$(x_{LE}, x_{TE}) = (-a, 0) \text{ for } l_r \gg a \text{ (steady sliding)} \quad (39)$$

where again, the former means the “non-peeling” of the entire contact area, and the latter means the “complete peeling” of the contact area in the outlet side. First, when $l_r \ll a$, W and F can be estimated by

$$W \sim \int_{-a}^a Ku(x) dx = 2\alpha R^{1/2} K \delta^{3/2} \quad (40)$$

$$F \sim \int_{-a}^a -h'(x) C \dot{u}(x) dx = 2\alpha R^{-1/2} CV \delta^{3/2} \quad (41)$$

respectively, where $\alpha = 2\sqrt{2}/3$. Therefore, μ for the first limiting case [Equation (38)] is given by

$$\mu = \frac{CV}{RK} \text{ (asymptotes L}_{1\text{-FI}} \text{ and L}_{1\text{-MI}}) \quad (42)$$

Then, when $l_r \gg a$, W and F can be estimated by

$$W \sim \int_{-a}^0 C \dot{u}(x) dx = CV\delta \quad (43)$$

$$F \sim \int_{-a}^0 -h'(x) C \dot{u}(x) dx = \alpha R^{-1/2} CV \delta^{3/2} \quad (44)$$

respectively. Therefore, μ for the second limiting case [Equation (39)] is given by

$$\mu = \alpha \sqrt{\frac{\delta}{R}} \text{ (asymptote L}_{2\text{-FI}}) \quad (45)$$

or, by using Equation (43),

$$\mu = \alpha \sqrt{\frac{W}{RCV}} \text{ (asymptote L}_{2\text{-MI}}) \quad (46)$$

The black broken lines in Figure 6 are the above asymptotes. Now we find that excellent agreement of the asymptotes with the master curves. Note that Equations (42) and (46) are the same as those shown by Popov (2010).

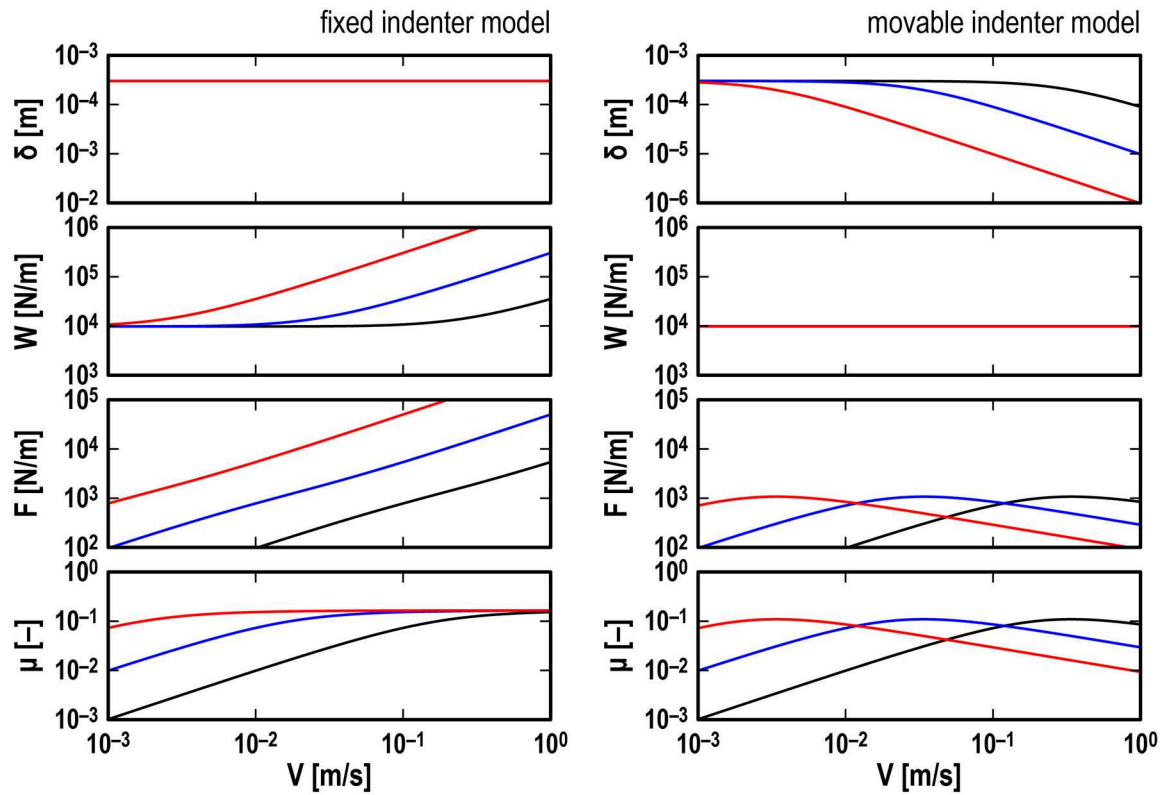


FIGURE 5 | Numerical results in steady sliding for fixed indenter model (**Left**) and movable indenter model (**Right**): effects of retardation time τ on velocity dependences of indentation depth δ (**Top**), normal load W (second row), friction force F (third row), and friction coefficient μ (bottom row). Black lines, $\tau = 10^{-2}$ s; blue lines, $\tau = 10^{-1}$ s; red lines, $\tau = 10^0$ s (see **Table 1** for other parameter values).

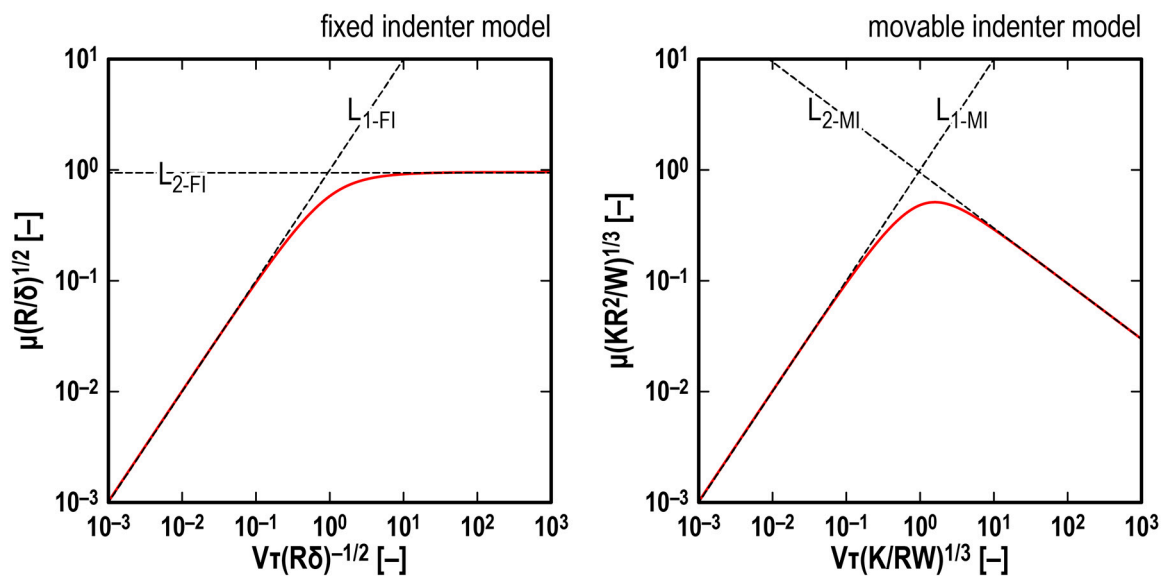


FIGURE 6 | Master curves on friction coefficient μ in steady sliding for fixed indenter model (**Left**) and movable indenter model (**Right**). Red solid lines: master curves obtained numerically, black broken lines: asymptotes obtained theoretically.

TABLE 2 | Velocity dependences of friction coefficient in steady sliding ($\alpha = 2\sqrt{2}/3$).

	FI model	MI model
Critical velocity V^*	$V^* \sim \frac{K\sqrt{R\delta}}{C}$	$V^* \sim \frac{\sqrt[3]{RK^2W}}{C}$
Friction coefficient μ for $V \ll V^*$	$\mu = \frac{CV}{RK} \sim V^1$	$\mu = \frac{CV}{RK} \sim V^1$
Friction coefficient μ for $V \gg V^*$	$\mu = \alpha\sqrt{\frac{\delta}{R}} \sim V^0$	$\mu = \alpha\sqrt{\frac{W}{RCV}} \sim V^{-1/2}$

Velocity Dependences of Friction Coefficient

In this section, based on **Figure 6**, let us consider the velocity dependences of the friction coefficient. In **Figure 6**, the drive velocity V is included only on the abscissa. Therefore, from the intersection of the two asymptotes, we introduce the critical velocity V^* defined as

$$V^* \sim \frac{K\sqrt{R\delta}}{C} \quad (\text{FI model}) \quad (47)$$

$$V^* \sim \frac{\sqrt[3]{RK^2W}}{C} \quad (\text{MI model}) \quad (48)$$

By using V^* , the formulas for estimating friction coefficient are summarized in **Table 2**.

First, we consider the case of $V \ll V^*$. From the third row of **Table 2**, we find that the formulas for both models are the same, which means that when $V \ll V^*$, the vertical lift effect is negligible, although the rigid indenter could move vertically in the MI model. In addition, considering the assumption given by Equation (38), we find that when $V \ll V^*$, the peeling of the contact area in the outlet side is also negligible in both models. The velocity dependence of μ is $\mu \sim V^1$, which means that it is caused by the damping of the Kelvin–Voigt foundation.

Then, we consider the case of $V \gg V^*$. From the bottom row of **Table 2**, we find that the formulas for the two types of models are different, which means that when $V \gg V^*$, the vertical lift effect strongly appears in the MI model. From Equation (43), we find that the vertical lift effect appears according to

$$\delta \sim V^{-1} \quad \text{for } V \gg V^* \quad (\text{MI model}) \quad (49)$$

which is confirmed by the numerical results shown in the upper right graph of **Figure 5**, where a decrease in δ means an increase in $z_A (= -\delta)$: that is, the vertical lift of the rigid indenter. In addition, considering the assumption given by Equation (39), we find that when $V \gg V^*$, the complete peeling of the contact area in the outlet side occurs in both models. Regarding the velocity dependence of μ in the FI model, it is $\mu \sim V^0$ (i.e., μ is constant). This is because when $V \gg V^*$, the damping becomes dominant, and therefore the restoring becomes negligible, which leads to the situation that both of W and F are proportional to V , as shown in Equations (43) and (44). Regarding the velocity dependence of μ in the MI model, on the other hand, it shows the negative dependence $\mu \sim V^{-1/2}$, which is caused by the vertical lift of the rigid indenter.

As a result, we find that in the MI model, the function $\mu = \mu(V)$ has a local maximum at $V \sim V^*$. Note that for many decades, this type of velocity dependence has been observed and discussed by a number of researchers for the sliding friction of elastomers, which seems to be basically understood as the frequency dependence of viscoelasticity (Persson, 2001; Momozono et al., 2010; Carbone and Putignano, 2013). However, the MI model produces a qualitatively similar dependence, although the frequency dependence of the Kelvin–Voigt foundation has no local maximum, where it is caused by the vertical lift of the rigid indenter.

Also note that the negative dependence $\mu \sim V^{-1/2}$ means

$$\mu \rightarrow 0 \quad \text{for } \frac{V}{V^*} \rightarrow \infty \quad (\text{MI model}) \quad (50)$$

which gives us an idea of the ultra-low friction. If we try to embody the situation of Equation (50) in real systems, not only increasing V but also decreasing V^* is promising, the method of which is shown by Equation (48). Probably, the most effective parameter is the damping C (or the retardation time $\tau = C/K$). According to the equation, increasing C leads to decreasing V^* , which leads to approaching the situation of Equation (50). It should be noted that the concept can never be embodied in the FI model, because increasing C under a constant δ just linearly increases F , as shown in Equation (44). Therefore, the key is the vertical lift effect arising with the movable boundary condition of the rigid indenter.

Recently, several swollen polymers showing ultra-low friction (e.g., hydrogels and polymer brushes) have attracted attention, with inspirations from natural tribosystems in human bodies (e.g., eyes and joints) (Klein et al., 1994; Lee and Spencer, 2008). An important aspect is obviously the low-adhesive properties of surfaces caused by their “microscopic” structures. However, as another important aspect, proper “macroscopic” structures are also needed to utilize them properly under various conditions. At least, recalling the results on the transient sliding (**Figure 3**), we can say that in the lubricated sliding of elastomers, macroscopic structures utilizing the vertical lift effect by viscoelasticity seem to have strong advantages for smooth transition to the fluid film lubrication regime, especially at the onset of sliding.

CONCLUSIONS

In this study, the sliding friction of elastomers was investigated numerically and theoretically for the line contact between a cylindrical rigid indenter and a “frictionless” Kelvin–Voigt foundation. The onset of sliding under an abrupt increase in the drive velocity was simulated with different boundary conditions of the rigid indenter. The main conclusions are as follows:

1. When the rigid indenter is not allowed to move in any direction, just an abrupt change in the friction force appears, which is not accompanied by any transient processes. However, when the rigid indenter is able to move in the vertical direction, the transient sliding friction including three different time constants appears, resembling the typical

transition from the static friction to the kinetic friction, in spite of no static friction considered in the simulation. The aforementioned drastic difference is caused by the “vertical lift” of the rigid indenter induced by the damping of the Kelvin–Voigt foundation.

2. When the drive velocity is sufficiently low, the vertical lift effect is negligible, where the restoring is more dominant than the damping, which leads to little peeling of the entire contact area and the friction coefficient proportional to the drive velocity. On the other hand, when the drive velocity is sufficiently high, the vertical lift effect becomes strong, where the damping is more dominant than the restoring, leading to the complete peeling of the contact area in the outlet side. The vertical lift of the rigid indenter strongly affects the characteristics of the steady sliding friction, which is explained well by using the critical velocity determined from the asymptotes in the master curve of friction coefficient.

At the end, it is noted again that the foregoing conclusions are obtained for the Kelvin–Voigt foundation against a cylindrical rigid indenter. In general, since behaviors are greatly affected by the rheology of the viscoelastic foundation or the shape of the rigid indenter (Popov and Heß, 2015), further investigations are needed.

REFERENCES

- Barquins, M. (1985). Sliding friction of rubber and Schallamach waves: a review. *Mater. Sci. Eng.* 73, 45–63. doi: 10.1016/0025-5416(85)90295-2
- Belin, M., Arafune, H., Kamijo, T., Perret-Liaudet, J., Morinaga, T., Honma, S., et al. (2018). Low friction, lubricity, and durability of polymer brush coatings, characterized using the relaxation tribometer technique. *Lubricants* 6:52. doi: 10.3390/lubricants6020052
- Carbone, G., and Putignano, C. (2013). A novel methodology to predict sliding and rolling friction of viscoelastic materials: theory and experiments. *J. Mech. Phys. Solids* 61, 1822–1834. doi: 10.1016/j.jmps.2013.03.005
- Fukahori, Y., and Yamazaki, H. (1994). Mechanism of rubber abrasion: part I: abrasion pattern formation in natural rubber vulcanizate. *Wear* 171, 195–202. doi: 10.1016/0043-1648(94)90362-X
- Gong, J. P., Kurokawa, T., Narita, T., Kagata, G., Osada, Y., Nishimura, G., et al. (2001). Synthesis of hydrogels with extremely low surface friction. *J. Am. Chem. Soc.* 123, 5582–5583. doi: 10.1021/ja003794q
- Grosch, K. A. (1963). The relation between the friction and visco-elastic properties of rubber. *Proc. Roy. Soc. Ser. A* 274, 21–39. doi: 10.1098/rspa.1963.0112
- Johnson, K. L. (1985). *Contact Mechanics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139171731
- Klein, J., Kumacheva, E., Mahalu, D., Perahia, D., and Fetters, L. J. (1994). Reduction of frictional forces between solid surfaces bearing polymer brushes. *Nature* 370, 634–636. doi: 10.1038/370634a0
- Kusche, K. (2017). Frictional force between a rotationally symmetric indenter and a viscoelastic half-space. *Z. Angew. Math. Mech.* 97, 226–239. doi: 10.1002/zamm.201500169
- Lee, S., and Spencer, N. D. (2008). Sweet, hairy, soft, and slippery. *Science* 319, 575–576. doi: 10.1126/science.1153273
- Li, Q., Dimaki, A., Popov, M., Psakhie, S. G., and Popov, V. L. (2015). Kinetics of the coefficient of friction of elastomers. *Sci. Rep.* 4:5795. doi: 10.1038/srep05795
- Maegawa, S., and Nakano, K. (2010). Mechanism of stick-slip associated with Schallamach waves. *Wear* 268, 924–930. doi: 10.1016/j.wear.2009.12.018
- May, W. D., Morris, E. L., and Atack, D. (1959). Rolling friction of a hard cylinder over a viscoelastic material. *J. Appl. Phys.* 30, 1713–1724. doi: 10.1063/1.1735042

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

KN designed the study and wrote the manuscript. MK conducted the simulations. KN and MK analyzed the results. All authors contributed to the article and approved the submitted version.

FUNDING

This study was partly supported by ACCEL (Accelerated Innovation Research Initiative Turning Top Science and Ideas into High-Impact Values) under Grant No. JPMJAC1503 sponsored by Japan Science and Technology Agency.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Professor Valentin L. Popov and Professor Hiroshi Watanabe for fruitful and stimulating discussions.

- Mizukami, M., Gen, M., Hsu, S. Y., Tsujii, Y., and Kurihara, K. (2019). Dynamics of lubricious, concentrated PMMA brush layers studied by surface forces and resonance shear measurements. *Soft Matter*. 15, 7765–7776. doi: 10.1039/C9SM01133A
- Momozono, S., Nakayama, K., and Kyogoku, K. (2010). Theoretical model for adhesive friction between elastomers and rough solid surfaces. *J. Chem. Phys.* 132:114105. doi: 10.1063/1.3356220
- Moore, D. F., and Geyer, W. A. (1972). A review of adhesion theories for elastomers. *Wear* 22, 113–141. doi: 10.1016/0043-1648(72)90271-2
- Moore, D. F., and Geyer, W. A. (1974). A review of hysteresis theories for elastomers. *Wear* 30, 1–34. doi: 10.1016/0043-1648(74)90055-6
- Nakano, K., Kawaguchi, K., Takeshima, K., Shiraishi, Y., Forsbach, F., Benad, J., et al. (2019). Investigation on dynamic response of rubber in frictional contact. *Front. Mech. Eng.* 5:9. doi: 10.3389/fmech.2019.00009
- Nakano, K., and Maegawa, S. (2009). Stick-slip in sliding systems with tangential contact compliance. *Tribol. Int.* 42, 1771–1780. doi: 10.1016/j.triboint.2009.04.039
- Nomura, A., Okayasu, K., Ohno, K., Fukuda, T., and Tsujii, Y. (2011). Lubrication mechanism of concentrated polymer brushes in solvents: effect of solvent quality and thereby swelling state. *Macromolecules* 44, 5013–5019. doi: 10.1021/ma200340d
- Persson, B. N. J. (2001). Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115, 3840–3861. doi: 10.1063/1.1388626
- Popov, V. L. (2010). *Contact Mechanics and Friction: Physical Principles and Applications*. Berlin: Springer. doi: 10.1007/978-3-642-10803-7
- Popov, V. L., and Heß, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin: Springer. doi: 10.1007/978-3-642-53876-6
- Popov, V. L., Voll, L., Kusche, S., Li, Q., and Rozhkova, S. V. (2018). Generalized master curve procedure for elastomer friction taken into account dependencies on velocity, temperature and normal force. *Tribol. Int.* 120, 376–380. doi: 10.1016/j.triboint.2017.12.047
- Rubinstein, S. M., Cohen, G., and Fineberg, J. (2004). Detachment fronts and the onset of dynamic friction. *Nature* 430, 1005–1009. doi: 10.1038/nature02830
- Schallamach, A. (1957). Friction and abrasion of rubber. *Wear* 1, 384–417. doi: 10.1016/0043-1648(58)90113-3
- Schallamach, A. (1971). How does rubber slide? *Wear* 17, 301–312. doi: 10.1016/0043-1648(71)90033-0

- Tadokoro, T., Sato, K., Nagamine, T., Nakano, K., Sasaki, S., Sato, T., et al. (2020). Concentrated polymer brush as reciprocating seal material for low leakage and low friction. *Tribol. Trans.* 63, 20–27. doi: 10.1080/10402004.2019.1650213
- Williams, M. L., Landel, R. F., and Ferry, J. D. (1955). The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids. *J. Am. Chem. Soc.* 77, 3701–3707. doi: 10.1021/ja01619a008
- Yamaguchi, T., Morishita, M., Doi, M., Hori, T., Sakaguchi, H., and Ampuero, J. P. (2011). Gutenberg-Richter's law in sliding friction of gels. *J. Geophys. Res.* 116, B12306. doi: 10.1029/2011JB008415

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Nakano and Kono. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Influence of Vibration on Friction: A Contact-Mechanical Perspective

Mikhail Popov^{1,2*}

¹ Department of Continuum Mechanics and Constitutive Theory, Technische Universität Berlin, Berlin, Germany, ² Tomsk State University, Tomsk, Russia

A unified model for active control of static and sliding friction by normal, tangential, and transverse oscillations is discussed, building on a series of past publications. The model in question is quasi-static, uses Amontons friction and takes into account *contact stiffness* in both normal and tangential directions. This makes the model fully macroscopic, which stands in contrast to Prandtl-Tomlinson-derived microscopic models that seem to be the currently preferred explanation for the influence of vibration on friction. While many technical details and numerical simulations based on our model have already appeared in a series of publications, here we attempt to give a high-level overview and discuss the main properties of friction under oscillation as generally as possible, while making a minimum of assumptions.

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Shingo Ozaki,
Yokohama National University, Japan
Varvara Romanova,
Institute of Strength Physics and
Materials Science (ISPMS SB RAS),
Russia
Vladislav Aleshin,
UMR8520 Institut d'électronique, de
Microélectronique et de
Nanotechnologie (IEMN), France

*Correspondence:

Mikhail Popov
mpopov@fastmail.fm

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 29 May 2020

Accepted: 14 July 2020

Published: 20 August 2020

Citation:

Popov M (2020) The Influence of
Vibration on Friction: A
Contact-Mechanical Perspective.
Front. Mech. Eng. 6:69.
doi: 10.3389/fmech.2020.00069

Keywords: vibration, friction, contact mechanics, active control, stick-slip actuation

1. INTRODUCTION

The fact that vibration can be used to significantly reduce the force of friction has been known since at least the 1950s (Fridman and Levesque, 1959). Since then, the effect has been studied extensively and exploited in many practical applications. Classical examples are to be found in wire drawing (Murakawa and Jin, 2001; Siegert and Ulmer, 2001), press forming (Eaves et al., 1975; Ashida and Aoyama, 2007), cutting (Thoe et al., 1998; Eggers et al., 2004), and other machining processes. Also well-known is the use of vibration for stabilization of system dynamics, e.g., suppression of brake squeal (Müller and Ostermeyer, 2007) and cornering noise (Heckl and Huang, 2000).

There are also a number of advanced applications that move beyond simple reduction of sliding or static friction, and involve vibration-driven directed transport or exact positioning (Popov, 2017). The most famous example of this are traveling wave motors (Schmidt et al., 1996; Storck et al., 2002), which are used to adjust focus in camera lenses, among many other applications. Similar principles are employed in high-precision linear actuators and positioning systems (Socoliuc et al., 2006), vibrational conveyors (Gaberson, 1971, 1972), and other types stick-slip drives.

The above examples are only a small sample of technical applications at the intersection of friction and vibration. Correspondingly, there is a large body of existing research in this field (see e.g., Pohlman and Lehfeldt, 1966; Godfrey, 1967; Storck et al., 2002; Chowdhury and Helali, 2008). Most of it is practical in nature, even though several well-known theoretical models have been proposed as well (De Wit et al., 1995). However, it is the contention of the author that an important factor is missing from currently popular models: the compliance of the contact and its interaction with the applied oscillation. The currently prevailing tendency is to ascribe the reduction of friction by vibration mostly to processes at the micro-scale (Popov et al., 2010). However, here we will argue that the *primary* (but not necessarily exclusive) mechanism is to be found on the macro-scale, in ordinary contact mechanics. It should be noted that this does not automatically invalidate previous work. In fact, it seems likely that a truly accurate model will be multiscale, combining both macroscopic dynamics and microscopic processes.

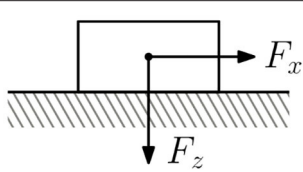


FIGURE 1 | Static friction model. The intrinsic coefficient of friction in the contact is μ_0 . The y-axis points out of the plane.

The primary advantage of our model is its simplicity. It relies only on macroscopic contact mechanics and introduces no new physics. In fact, it is likely to be the *simplest possible* model that is rich enough to describe almost the full range of behaviors exhibited by friction under the influence of external vibration. For this reason, the present paper can be seen as an exercise in minimalism, attempting to cover as much phenomenological ground as possible with a minimum of assumptions and variables.

1.1. Contributions

This work draws heavily on results recently published in a series of papers with participation of the present author (Mao et al., 2017; Popov et al., 2017; Benad et al., 2018a,b; Popov and Li, 2018). While there is substantial overlap with these papers, the present work is organized differently, seeking to present a “big picture” view without getting bogged down in details. Several results have been generalized from previous publications, while the discussion of the influence of oscillation waveforms has, to the best knowledge of the author, not previously appeared in the literature.

2. STATIC FRICTION

To warm up, we consider static friction. This case is much simpler than the sliding case and leads to some satisfyingly general results. The system under consideration consists of a body resting on a plane (Figure 1). The body is pressed into the plane with a force F_z and pulled sideways with a force F_x . The coefficient of friction between the body and the plane is assumed to be constant and equal to μ_0 . The body remains at rest while

$$|F_x| < \mu_0 F_z \quad (1)$$

where $\mu_0 F_z$ is the critical force at which the body just begins to slide. The static coefficient of friction is defined as the ratio of this critical force to the normal load. In the absence of oscillation, it is equal to μ_0 :

$$\mu_s = \mu_0 \quad (2)$$

Things get slightly more interesting when we add an oscillatory force component. If the force oscillation acts normal to the plane, we denote it by $A_z g(t)$, where A_z is the amplitude. The stick condition in that case needs to be amended to:

$$|F_x| < \mu_0 (F_z + A_z g(t))^+ \quad (3)$$

The $(\cdot)^+$ notation denotes the ramp function, which clips negative values to zero. It is necessary because the normal force does not turn negative when contact is lost.

It is easy to see that the critical force is reduced relative to the non-oscillatory case, since the above inequality must hold at all times, including the times when the normal force drops below its mean value F_z . In other words, static friction is limited by the *minimum* of normal force encountered during the oscillation. For the coefficient of static friction under normal oscillation we thus obtain:

$$\mu_{s,z} = \mu_0 (1 - A_z/F_z)^+ \quad (4)$$

In a similar fashion, we can add an oscillatory component $A_x g(t)$ that is aligned with the tangential force F_x . This results in the stick condition

$$|F_x + A_x g(t)| < \mu_0 F_z \quad (5)$$

Note that this inequality is only satisfiable when $A_x < \mu_0 F_z$. Otherwise the body starts to slide in place and the contact loses its ability to statically sustain a lateral force. Thus, μ_s can be expressed as:

$$\mu_{s,x} = (\mu_0 - A_x/F_z)^+ \quad (6)$$

Notice the slight difference between this result and Equation (4). In particular, note that a tangential oscillation will reduce μ_s by a larger amount than a normal oscillation of the same amplitude if $\mu_0 < 1$, and by a smaller amount otherwise.

Transverse oscillations are also able to reduce static friction. This case is qualitatively similar to that of tangential oscillation, with the difference that we need to use the vector norm of the in-plane forces instead of adding them directly:

$$F_x^2 + (A_y g(t))^2 < (\mu_0 F_z)^2 \quad (7)$$

Once again, stick is impossible if $A_y \geq \mu_0 F_z$, and for the static coefficient of friction we obtain:

$$\mu_{s,y} = \sqrt{(\mu_0^2 - A_y^2/F_z^2)^+} \quad (8)$$

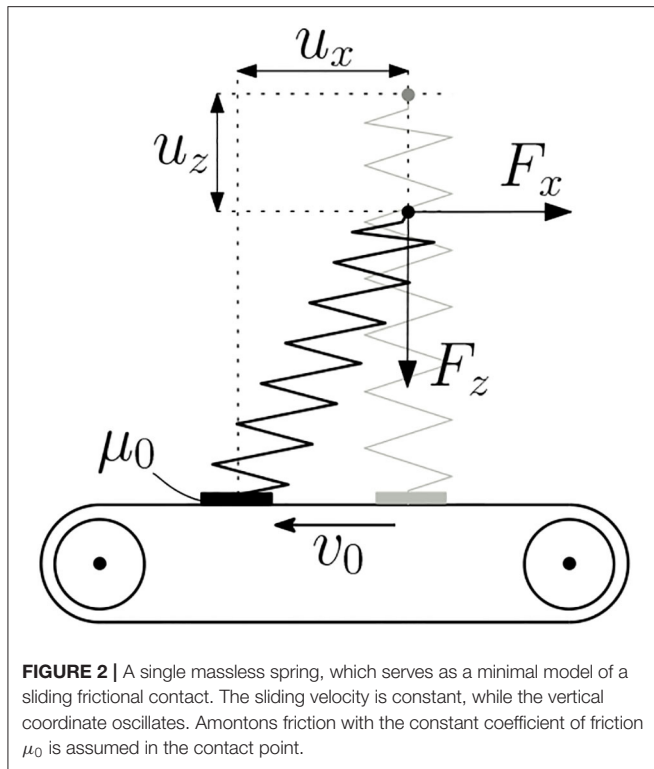
One particularly useful thing about these results is that they are quite general, and in particular independent of contact geometry, frequency of oscillation, and the shape of the oscillation waveform.

2.1. Static Friction Under Superimposed Oscillation

Things become considerably less transparent when we consider simultaneous oscillation in multiple directions. The stick condition itself does not change much, and in the most general case can be expressed as:

$$(F_x + A_x g_x(t))^2 + (A_y g_y(t))^2 < \mu_0^2 (F_z + A_z g_z(t))^2 \quad (9)$$

Unfortunately, actually finding the maximal F_x that still satisfies this inequality at all times quickly becomes unwieldy, leading to



a large number of case distinctions—if a closed-form solution is possible at all. In addition, when the compliance of the contact is taken into account, the static coefficient of friction may become *negative*, in the sense that a constant force needs to be applied to *prevent* the contact from sliding. This effect is what frictional drives and actuators are based on. For an analysis of this case the reader is referred to Popov and Li (2018). In this paper, however, we ignore superimposed oscillation.

3. SLIDING FRICTION UNDER NORMAL OSCILLATION

The key feature of the model that we use to describe dynamic friction is that the *compliance* of the contact is taken into account. In the initial formulation, the contact is modeled as a single Hookean spring that has an associated normal stiffness k_z and a lateral stiffness k_x (Figure 2). This is a reasonable approximation of a flat-ended cylinder in contact with a plane. The model can also be extended to cover arbitrary curved contacts with the help of the Method of Dimensionality Reduction (Popov and Heß, 2016). However, for a general analysis, a single spring is quite sufficient.

The model considered here is displacement-controlled and quasi-static. A force-controlled and/or inertial model can be formulated within the same framework, which, however, leads to certain complications (e.g., resonances) that are outside the focus of the present paper. For an analysis of such a model, the reader is referred to Mao et al. (2017).

The kinematics of the model is as follows: The contact spring is pulled over a flat plane with a constant velocity v_0 , although for convenience we consider the spring to be stationary, while the substrate slides underneath it. The normal displacement u_z of the spring is measured relative to the state of unstressed first contact with the substrate. $u_z(t)$ represents the externally applied oscillation and is thus given explicitly. The lateral displacement u_x , on the other hand, depends on the current state of the system and is the only unknown variable.

We assume that Amontons' law of friction (with a constant coefficient of friction μ_0 that is the same for both static and sliding friction) holds in the contact point. In general, this may be an unrealistic assumption. However, the use of a constant coefficient of friction not only simplifies calculations, but also eliminates all possible micro-scale influences from the model. Since one of the primary aims of this paper is to advertise the feasibility of a purely macroscopic theory of friction under oscillation, making μ_0 constant is actually a prerequisite.

The *effective* coefficient of friction $\bar{\mu}$, which is to be determined in the sequel, is defined as the average tangential force exerted by the spring divided by the average normal force:

$$\bar{\mu} = \frac{\langle F_x(t) \rangle}{\langle F_z(t) \rangle} \quad (10)$$

where $\langle \dots \rangle$ denotes averaging over one period of oscillation.

Previous publications on the topic assumed that the imposed normal oscillation is harmonic, so as to simplify analysis. However, this turned out to be an unnecessary restriction, so here we will work with a general periodic function that is parameterized as follows:

$$u_z(t) = \bar{u}_z + A_z w(ft) \quad (11)$$

Here \bar{u}_z is the mean indentation, A_z is the amplitude and f the frequency of the oscillation. $w(\varphi)$ is a dimensionless function describing the “shape” of the oscillation, with $\varphi = ft$. The waveform w is normalized such that it is zero-mean, with a period of 1 and a minimum value of -1 . Note however, that the *maximum* of w is left unconstrained.

3.1. Pure Sliding

While the behavior of a frictional couple under oscillation has its complexities in general, there are two extreme cases that lend themselves to easy and precise analysis: One of them, static friction, was already discussed above. The second, pure sliding, is briefly discussed here. The most important thing about pure sliding is that oscillations do not influence the coefficient of friction in that mode. This can be easily seen from the fact that the instantaneous tangential force is uniquely defined during slip ($F_x = \mu_0 F_z$), from which the effective coefficient of friction is immediately obtained:

$$\bar{\mu}_{\text{slip}} = \frac{\langle F_x(t) \rangle}{\langle F_z(t) \rangle} = \frac{\langle \mu_0 F_z(t) \rangle}{\langle F_z(t) \rangle} = \mu_0 \quad (12)$$

Irrespective of how complex the dependence $F_z(t)$ may be, it always cancels out—by linearity of sliding friction. While

this result may seem unimpressive by itself, it establishes an important “boundary condition” for the more general case of friction with stick-slip. Also, as in the static case, the coefficient of friction in pure slip has the important property of not being dependent on contact geometry and oscillation parameters. The result $\bar{\mu}_{\text{slip}} = \mu_0$ is also valid for tangential and combined normal/tangential oscillations. It can also be shown to be valid in the inertial case (Mao et al., 2017). However, *transverse* oscillations do not, strictly speaking, have this limiting case, although the deviation becomes negligible at high velocities. This will be discussed in more detail later.

3.2. Stick-Slip

Two extreme points have now been established: pure stick (static friction) and pure slip (plateau). Reason suggests that there is also something in between. It would be physically implausible for the coefficient of friction under oscillation to “snap” from near zero back to μ_0 due to arbitrarily slow sliding. And this is in fact not observed experimentally: At a given amplitude and frequency, the static coefficient of friction is lowest, and then smoothly increases with the sliding velocity until reaching a plateau of sorts. Fortunately, the transition region can also be described in our model. Unsurprisingly, it is dominated by stick-slip.

Let us now consider this phenomenon in more detail. During sliding, we have $F_x = \mu_0 F_z$, which can also be written as $k_x u_x(t) = \mu_0 k_z u_z(t)$. Substituting u_z from Equation (11) and rearranging gives us the lateral displacement and velocity of the contact point:

$$u_x(t) = \mu_0 \frac{k_z}{k_x} (\bar{u}_z + A_z w(ft)) \quad (13)$$

$$\dot{u}_x(t) = \mu_0 \frac{k_z}{k_x} A_z f w'(ft) \quad (14)$$

A transition from slip to stick happens when the relative motion between the substrate and the contact point vanishes, i.e., when $\dot{u}_x(t) = v_0$. From this condition, the point of stick onset can be determined:

$$\varphi_1 = ft_1 = (w')^{-1} \left(\frac{k_x v_0}{\mu_0 k_z A_z f} \right) \quad (15)$$

It becomes obvious that φ_1 is a function of a single compound variable, which combines all parameters of the system, *except* \bar{u}_z . To simplify further calculations, we introduce some dimensionless variables, α (corresponding to amplitude), β (corresponding to velocity), and φ (phase):

$$\alpha = \frac{A_z}{\bar{u}_z} \quad (16)$$

$$\beta = \frac{k_x v_0}{\mu_0 k_z A_z f} \quad (17)$$

$$\varphi = ft \quad (18)$$

Using these variables, the static coefficient of friction (Equation 4) can be expressed as $\mu_{s,x} = \mu_0(1 - \alpha)^+$, while Equation (15) can be written as

$$\varphi_1 = (w')^{-1}(\beta) \quad (19)$$

Noting that β is a positive quantity and assuming that w is differentiable (but not necessarily invertible—there can be multiple stick events), it can be seen that the above equation has solutions if

$$\beta < \max_{\varphi} w'(\varphi) = \beta_c \quad (20)$$

where β_c denotes the critical value that separates the stick-slip region from the continuous sliding region. A harmonic oscillation, for example, has $\beta_c = 2\pi$, while a right-leaning sawtooth function has $\beta_c = 2$, which is in fact the smallest possible value. The larger β_c , the more effective the waveform is at reducing friction at high velocities, but more on that later.

Once stick is initiated, the contact point is dragged along by the substrate with velocity v_0 , so that the tangential displacement and force increase linearly with time:

$$F_{\text{stick}}(t) = \mu_0 F_z(t_1) + k_x v_0 (t - t_1) \quad (21)$$

This continues while the condition for static friction holds:

$$F_{\text{stick}}(t) < \mu_0 F_z(t) \quad (22)$$

Trivial as it is, this inequality lies at the core of reduction of friction in our model. It serves as the sole source of nonlinearity that allows the system to break free of the trivial solution exemplified by Equation (12). With pure slip, the spring force is always equal to $\mu_0 F_z(t)$, while in stick-slip it is sometimes lower, which leads to lower average force and coefficient of friction (see also **Figure 3**). Another way of looking at it is that the contact point stands still when the normal force is highest, and covers more distance when the normal load diminishes. This leads to lower energy dissipation over the same distance. The whole process is somewhat similar to walking, where one leg carries the load without dissipation, while the other is lifted and advanced to the next position. Something analogous happens in our model, only there is just one “leg” and it is not necessarily lifted all the way.

The stick phase ends at time t_2 when the condition $F_{\text{stick}}(t_2) = \mu_0 F_z(t_2)$ is met. Expanding this condition yields

$$\mu_0 k_z u_z(t_1) + k_x v_0 (t_2 - t_1) = \mu_0 k_z u_z(t_2) \quad (23)$$

or, more conveniently,

$$\frac{v_0 k_x}{\mu_0 k_z} (t_2 - t_1) = u_z(t_2) - u_z(t_1) \quad (24)$$

Substituting u_z and $t = \varphi/f$, this can be rewritten as:

$$\beta (\varphi_2 - \varphi_1) = w(\varphi_2) - w(\varphi_1) \quad (25)$$

Once again \bar{u}_z cancels out, leaving us with a function of only β . Unfortunately, the equation is implicit and cannot be solved symbolically for φ_2 except in the simplest cases (sawtooth, square wave, etc). In the case of a harmonic oscillation, for example, Equation (25) takes the form $(\cos x = a + bx)$, which does not have a closed-form solution in terms of standard functions. Numerical solution is required in most cases.

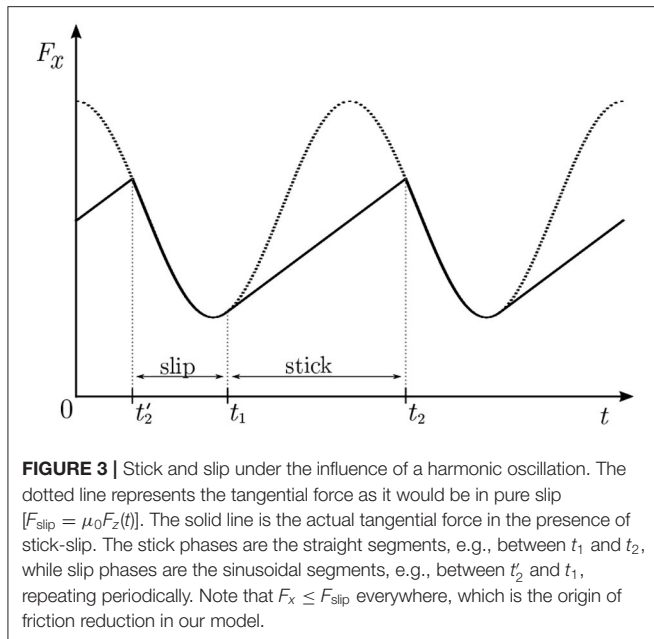


FIGURE 3 | Stick and slip under the influence of a harmonic oscillation. The dotted line represents the tangential force as it would be in pure slip [$F_{\text{slip}} = \mu_0 F_z(t)$]. The solid line is the actual tangential force in the presence of stick-slip. The stick phases are the straight segments, e.g., between t_1 and t_2 , while slip phases are the sinusoidal segments, e.g., between t'_2 and t_1 , repeating periodically. Note that $F_x \leq F_{\text{slip}}$ everywhere, which is the origin of friction reduction in our model.

3.3. Effective Coefficient of Friction

We define the “macroscopic” or “effective” force of friction simply as the tangential force averaged over one period $T = 1/f$:

$$\langle F_x \rangle = \frac{1}{T} \int_0^T F_x(t) dt \quad (26)$$

However, it will become clear in a moment that it is more convenient to consider the difference or *reduction* of the force of friction relative to the state of continuous sliding:

$$\Delta F_x = \langle F_{\text{slip}} \rangle - \langle F_x \rangle = \frac{1}{T} \int_0^T (F_{\text{slip}}(t) - F_x(t)) dt \quad (27)$$

Since F_x only differs from F_{slip} during the stick phase, we can tighten the integration bounds:

$$\Delta F_x = \frac{1}{T} \int_{t_1}^{t_2} (\mu_0 F_z(t) - F_{\text{stick}}(t)) dt \quad (28)$$

This form is convenient for numerical solution. However, some additional properties can be gleaned by expanding F_{stick} and $u_z(t)$ and making the substitution $dt = T d\varphi$:

$$\begin{aligned} \Delta F_x &= \frac{1}{T} \int_{t_1}^{t_2} (\mu_0 F_z(t) - \mu_0 F_z(t_1) - k_x v_0 (t - t_1)) dt \\ &= \frac{1}{T} \int_{t_1}^{t_2} \mu_0 k_z (\bar{u}_z + A_z w(f t) - \bar{u}_z - A_z w(f t_1) \\ &\quad - \frac{k_x v_0}{\mu_0 k_z} (t - t_1)) dt \\ &= \mu_0 k_z A_z \int_{\varphi_1}^{\varphi_2} (w(\varphi) - w(\varphi_1) - \beta (\varphi - \varphi_1)) d\varphi \end{aligned} \quad (29)$$

It becomes apparent that the expression for ΔF_x can be split into the dimensional factor $\mu_0 k_z A_z$ and a dimensionless function Ψ_w of a single variable:

$$\Delta F_x = \mu_0 k_z A_z \Psi_w(\beta) \quad (30)$$

where

$$\Psi_w(\beta) = \int_{\varphi_1}^{\varphi_2} (w(\varphi) - w(\varphi_1) - \beta (\varphi - \varphi_1)) d\varphi \quad (31)$$

We refrain from integrating this expression, since a closed-form solution is precluded by the lack of an explicit formula for φ_2 . We merely draw attention to the fact that ΔF_x is invariant with respect to mean indentation. The same is not true for the *coefficient of friction*:

$$\bar{\mu} = \frac{\langle \mu_0 F_z \rangle - \Delta F_x}{\langle F_z \rangle} = \mu_0 - \frac{\Delta F_x}{k_z \bar{u}_z} \quad (32)$$

However, the dependence on \bar{u}_z is incidental, merely reflecting the fact that ΔF_x is subtracted from different baselines of friction force. Using our dimensionless variables, the above can also be written in the following compact form:

$$\bar{\mu} = \mu_0 (1 - \alpha \Psi_w(\beta)) \quad (33)$$

Further, it can be shown that Ψ_w is a fairly well-behaved function that has unit range and is monotonously decreasing and convex for all waveforms and any number of stick events per cycle of oscillation. However, space considerations prevent us from including a formal proof of these properties.

3.4. Oscillation Waveforms

The functional dependence (33) presented in the previous section permits an interesting observation: the overall strength of the friction reduction effect is primarily governed by the *amplitude* of the oscillation and not by the frequency. In principle, the effective coefficient of friction can be reduced to very low values, but that requires a force amplitude that is comparable to the mean normal force. Thus, the technique is not very useful for reducing friction in highly loaded contacts, e.g., rail-car or truck wheels.

Furthermore, a higher frequency cannot be used to compensate for small amplitude. However, frequency is still an important parameter, since it determines the “velocity-resistance” of the effect: As has been pointed out before, the largest reduction is always seen in the static case, and becomes lower with increasing sliding velocity. The frequency determines the scaling of this decline, and a strong reduction can be achieved even at high sliding velocities if the frequency of the applied oscillation is high enough. However, frequency is not the only factor that determines this “velocity-resistance.” The waveform of the oscillation is also quite important, which is why we briefly discuss it here.

By far the most important property of a waveform w is the maximal positive value of its first derivative, or β_c . A *right-leaning* sawtooth function, for example, has $\beta_c = 2$; a harmonic

oscillation has $\beta_c = 2\pi$, which is slightly better; however, a *left-leaning* sawtooth function has $\beta_c = \infty$, which is ideal. An infinite value of β_c implies that the oscillation will provide some measure of friction reduction at arbitrarily high velocities. While this is not really possible in practice, the general rule for waveform selection is nevertheless that the load should increase as fast as possible and then relax slowly. Thus, approximations of left-leaning sawtooth or the square wave are preferable to smooth and symmetric functions like the harmonic oscillation. Naturally, this recommendation is subordinate to practical technological constraints. For example, a high-amplitude harmonic oscillation could be generated by exciting a natural vibrational mode of the system, while a square wave would likely require more sophisticated equipment.

We conclude this section by giving Ψ_w for a few common waveforms explicitly. For both sawtooth variants and the square wave Ψ_w can be calculated in closed form. However, for most oscillations, including sinusoidal ones, this is not possible. Nonetheless, the function can easily be computed numerically for arbitrary waveforms, and so we include two empirical approximations for the harmonic oscillation, which were first obtained in Popov et al. (2017). The first approximation is slightly more accurate.

$$\Psi_{\text{str}}(\beta) = 1 - \frac{\beta}{2} \quad (34)$$

$$\Psi_{\text{stl}}(\beta) = \frac{2}{2 + \beta} \quad (35)$$

$$\Psi_{\text{sqr}}(\beta) = \begin{cases} 1 - \beta/8, & \text{for } \beta < 4 \\ 2/\beta, & \text{for } \beta > 4 \end{cases} \quad (36)$$

$$\Psi_{\text{sin}}(\beta) \approx \frac{3}{4} (1 - \beta/\beta_c)^2 + \frac{1}{4} (1 - \beta/\beta_c)^4 \quad (37)$$

$$\approx (1 - \beta/\beta_c)^{2.4} \quad (38)$$

For a visual comparison, the dependence of the coefficient of friction on β is plotted in **Figure 4** for all four of the above waveforms. To keep things simple, only the case of *maximal* friction reduction is shown $\alpha = 1$, in which case Equation (33) reduces to $\bar{\mu} = \mu_0[1 - \Psi_w(\beta)]$. This is why all curves show zero static friction. For other values of α the shapes of the curves would remain the same, but they would start at non-zero values of μ_s and would be scaled accordingly.

As a final remark, we note that there is a unique optimal waveform with regards to reduction of friction. It is given by the periodic extension of $\delta(\varphi) - 1$, where δ is the Dirac delta function. This “impulse wave” is -1 everywhere, except for very short positive spikes (impulses) that occur with a period of 1 and each integrate to 1, so that the average of the function is zero. With this degenerate waveform, the system slides most of the time, with only an infinitesimal stick phase at each spike, which implies that Ψ is very close to 1 for all β :

$$\Psi_{\text{imp}}(\beta) \rightarrow 1 \quad (39)$$

Thus, we conclude that friction can be reduced, in principle, to an arbitrary degree even at high sliding velocities, by effectively

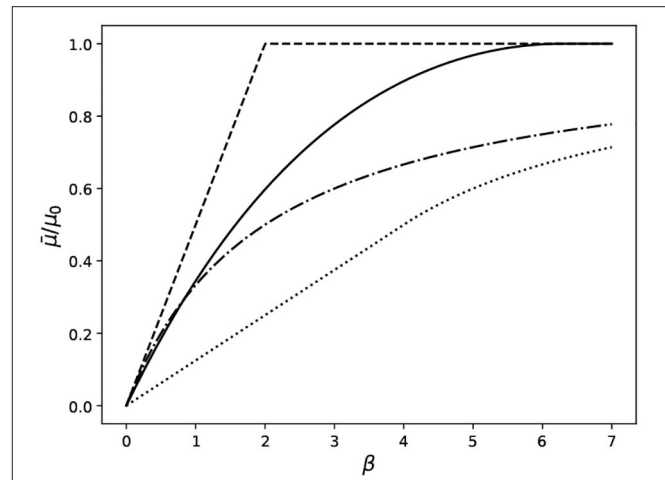


FIGURE 4 | Coefficient of friction under normal vibration with different waveforms, computed using Equation (33) and the individual influence factors Equations (34)–(37). Note that for all curves $\alpha = 1$, which corresponds to *maximal* friction reduction. Legend: dashed line—right-leaning sawtooth function (Equation 34); solid line—harmonic oscillation (Equation 37); dash-dotted line—left-leaning sawtooth function (Equation 35); dotted line—square wave (Equation 36).

hopping over the surface. In practice, this approach will be limited by plastic deformation, radiation of elastic waves and the sheer difficulty of generating such an oscillation.

4. TANGENTIAL AND TRANSVERSE OSCILLATIONS

Most of this paper was devoted to reduction of friction by normal oscillations. This focus is explained partly by the fact that the normal case is easiest to analyze, and partly because normal oscillations are generally the most efficient way to reduce friction, out of the three possible directions. Nevertheless, both tangential (in the direction of sliding) and transverse (in-plane, but orthogonal to sliding) vibration can reduce friction. For detailed analysis of the tangential case the reader is referred to Popov and Li (2018) and for the transverse case to Benad et al. (2018a). Here we only present some highlights and point out the major differences between normal oscillations and the other two modes.

In the tangential oscillation case the normal indentation is kept constant while an oscillatory component is added to the base of the spring. Sliding friction under such conditions can proceed in three modes: (I) pure sliding, in which the effective coefficient of friction is equal to μ_0 , as argued previously. (II) simple stick-slip, which occurs for obvious reasons when the velocity amplitude is greater than the mean sliding velocity ($A_{\text{xf}}w'(ft) > v_0$). (III) multiple stick-slip, which occurs when the velocity amplitude is much larger than v_0 , so that the contact point slides back-and-forth in each cycle, going through two stick and slip phases each.

The most important difference between friction reduction by normal and tangential oscillations is that normal oscillations actually reduce the total dissipated energy through a walking-like mechanism, while tangential oscillations do not. The author is not aware of a good analogy to visualize the mechanism in the tangential case. But it is clear that, since the normal load (and therefore the force of sliding friction) is constant, the dissipated energy is simply friction force times distance (in mode II). Although the effective coefficient of friction (i.e., average spring force) may be lowered, the missing energy must be supplied by the oscillator. In mode III, when the amplitude is large enough to cause in-place sliding, the total sliding distance actually increases, and the total energy expenditure becomes *larger* than without oscillations, even though the effective coefficient of friction will still appear lower than μ_0 .

Friction reduction by transverse oscillation always operates in something like mode III of tangential oscillation: it causes additional sliding in the direction orthogonal to the main sliding motion, thereby increasing the total path and therefore energy expenditure. However, the apparent coefficient of friction is reduced, because the magnitude of the local friction force is still limited to $\mu_0 F_z$, but now shared between the force components parallel and orthogonal to the main sliding direction. Thus, transverse oscillations are effectively “stealing” the friction vector from the slider, but at considerable expense of energy by the oscillator. This also accounts for the fact, mentioned previously, that the system never formally reaches the “invariant plateau” ($\bar{\mu}_{\text{slip}} = \mu_0$) even at high sliding velocities, because the projection of the local friction force onto the sliding direction is always less than its total magnitude, so long as the transverse amplitude is non-zero. However, for sufficiently large sliding velocities this difference becomes very small, so for all practical purposes the plateau exists in the transverse case as well.

To summarize, normal oscillations are most effective at reducing dynamic friction and should be used in preference to the other directions. Not only do they actually reduce the total dissipated energy, but normal oscillations also act at right angles (by definition) to the sliding motion. Thus, they technically do not require energy to keep going. Of course, this is never quite the case in practice, but by exciting a resonant frequency the power needed to drive the oscillator can usually

be made quite small. Compared to that, tangential oscillation requires a powerful oscillator (except in the static case), while transverse oscillation is even more energetically expensive, and also less effective overall. There are cases, however, where energy expenditure is not a primary concern (e.g., stabilization of system dynamics) and normal oscillations cannot be easily applied due to technological constraints. In such cases, tangential and even transverse oscillations are viable alternatives.

5. CONCLUSION

The present paper summarizes and generalizes a series of recent works that aim to establish a simple macroscopic contact model as a viable explanation for active control of friction by externally applied vibration. Despite its apparent simplicity, the model not only captures the full range of experimentally observed effects, but is also very flexible, being able to adapt to static and dynamic friction, oscillations in normal, tangential and transverse directions, contacts of curved bodies, etc. Apart from straight-forward reduction-of-friction settings, the model can also be applied to the study of frictional drives and actuators under complicated loading scenarios. A similar approach was also highly successful in modeling positioning systems without using any modified friction laws such as the elastoplastic model (see e.g., Teidelt et al., 2012; Grzempa et al., 2014; Teidelt, 2015).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

MP conceived and conducted the research and prepared the manuscript.

ACKNOWLEDGMENTS

This work was supported in part by the *Tomsk State University competitiveness improvement programme*, which the author gratefully acknowledges.

REFERENCES

- Ashida, Y., and Aoyama, H. (2007). Press forming using ultrasonic vibration. *J. Mater. Process. Technol.* 187, 118–122. doi: 10.1016/j.jmatprotec.2006.11.174
- Benad, J., Nakano, K., Popov, V. L., and Popov, M. (2018a). Active control of friction by transverse oscillations. *Friction* 7, 1–12. doi: 10.1007/s40544-018-0202-1
- Benad, J., Popov, M., Nakano, K., and Popov, V. L. (2018b). Stiff and soft active control of friction by vibrations and their energy efficiency. *Forsch. Ingenieurw.* 82, 331–339. doi: 10.1007/s10010-018-0281-1
- Chowdhury, M. A., and Helali, M. (2008). The effect of amplitude of vibration on the coefficient of friction for different materials. *Tribol. Int.* 41, 307–314. doi: 10.1016/j.triboint.2007.08.005
- De Wit, C. C., Olsson, H., Astrom, K. J., and Lischinsky, P. (1995). A new model for control of systems with friction. *IEEE Trans. Autom. Control* 40, 419–425. doi: 10.1109/9.376053
- Eaves, A., Smith, A., Waterhouse, W., and Sansome, D. (1975). Review of the application of ultrasonic vibrations to deforming metals. *Ultrasonics* 13, 162–170. doi: 10.1016/0041-624X(75)90085-2
- Eggers, G., Klein, J., Blank, J., and Hassfeld, S. (2004). Piezosurgery®: an ultrasound device for cutting bone and its use and limitations in maxillofacial surgery. *Br. J. Oral Maxillof. Surg.* 42, 451–453. doi: 10.1016/j.bjoms.2004.04.006
- Fridman, H. D., and Levesque, P. (1959). Reduction of static friction by sonic vibrations. *J. Appl. Phys.* 30, 1572–1575. doi: 10.1063/1.1735002
- Gaberson, H. A. (1971). Particle motion on oscillating conveyors: the equations of motion and the rules for predicting motion form transitions. *J. Eng. Ind.* 94, 50–56. doi: 10.1115/1.3428154

- Gaberson, H. A. (1972). Particle motion on oscillating conveyors-part 2. practical solutions to the equations of motion and the extension of the theory to beds of granular material. *J. Eng. Ind.* 94, 57–63. doi: 10.1115/1.3428155
- Godfrey, D. (1967). Vibration reduces metal to metal contact and causes an apparent reduction in friction. *ASLE Trans.* 10, 183–192. doi: 10.1080/05698196708972178
- Grzempa, B., Pohrt, R., Teidelt, E., and Popov, V. L. (2014). Maximum micro-slip in tangential contact of randomly rough self-affine surfaces. *Wear* 309, 256–258. doi: 10.1016/j.wear.2013.11.050
- Heckl, M. A., and Huang, X. (2000). Curve squeal of train wheels, part 3: active control. *J. Sound Vibrat.* 229, 709–735. doi: 10.1006/jsvi.1999.2512
- Mao, X., Popov, V. L., Starcevic, J., and Popov, M. (2017). Reduction of friction by normal oscillations. ii. in-plane system dynamics. *Friction* 5, 194–206. doi: 10.1007/s40544-017-0146-x
- Müller, M., and Ostermeyer, G. (2007). Cellular automata method for macroscopic surface and friction dynamics in brake systems. *Tribol. Int.* 40, 942–952. doi: 10.1016/j.triboint.2006.02.045
- Murakawa, M., and Jin, M. (2001). The utility of radially and ultrasonically vibrated dies in the wire drawing process. *J. Mater. Process. Technol.* 113, 81–86. doi: 10.1016/S0924-0136(01)00635-5
- Pohlman, R., and Lehfeldt, E. (1966). Influence of ultrasonic vibration on metallic friction. *Ultrasonics* 4, 178–185. doi: 10.1016/0041-624X(66)90244-7
- Popov, M., and Li, Q. (2018). Multimode active control of friction, dynamic ratchets and actuators. *Phys. Mesomech.* 21, 24–31. doi: 10.1134/S1029959918010046
- Popov, M., Popov, V. L., and Popov, N. V. (2017). Reduction of friction by normal oscillations. I. influence of contact stiffness. *Friction* 5, 45–55. doi: 10.1007/s40544-016-0136-4
- Popov, V. L. (2017). Oscillation-based methods for actuation and manipulation of nano-objects. *AIP Conf. Proc.* 1882:020056. doi: 10.1063/1.5001635
- Popov, V. L., and Heß, M. (2016). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin: Springer. doi: 10.1007/978-3-642-53876-6
- Popov, V. L., Starcevic, J., and Filippov, A. (2010). Influence of ultrasonic in-plane oscillations on static and sliding friction and intrinsic length scale of dry friction processes. *Tribol. Lett.* 39, 25–30. doi: 10.1007/s11249-009-9531-6
- Schmidt, J. P., Hagedorn, P., and Bingqi, M. (1996). A note on the contact problem in an ultrasonic travelling wave motor. *Int. J. Nonlin. Mech.* 31, 915–924. doi: 10.1016/S0020-7462(96)00112-6
- Siegert, K., and Ulmer, J. (2001). Superimposing ultrasonic waves on the dies in tube and wire drawing. *J. Eng. Mater. Technol.* 123, 517–523. doi: 10.1115/1.1397779
- Socoliuc, A., Gnecco, E., Maier, S., Pfeiffer, O., Baratoff, A., Bennewitz, R., and Meyer, E. (2006). Atomic-scale control of friction by actuation of nanometer-sized contacts. *Science* 313, 207–210. doi: 10.1126/science.1125874
- Storck, H., Littmann, W., Wallaschek, J., and Mracek, M. (2002). The effect of friction reduction in presence of ultrasonic vibrations and its relevance to travelling wave ultrasonic motors. *Ultrasonics* 40, 379–383. doi: 10.1016/S0041-624X(02)00126-9
- Teidelt, E. (2015). *Oscillating contacts: friction induced motion and control of friction* (Ph.D. thesis). TU Berlin, Berlin, Germany.
- Teidelt, E., Willert, E., Filippov, A., and Popov, V. L. (2012). Modeling of the dynamic contact in stick-slip microdrives using the method of reduction of dimensionality. *Phys. Mesomech.* 15, 287–292. doi: 10.1134/S1029959912030071
- Thoe, T., Aspinwall, D., and Wise, M. (1998). Review on ultrasonic machining. *Int. J. Mach. Tools Manuf.* 38, 239–255. doi: 10.1016/S0890-6955(97)00036-9

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Popov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Friction Influenced by Vibrations: A Refined Contact-Mechanics View on Lateral and Rotational Oscillations

Roman Pohrt*

Department of System Dynamics and Friction Physics, Institute of Applied Mechanics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Carminé Putignano,
Politecnico di Bari, Italy
Ken Nakano,
Yokohama National University, Japan

*Correspondence:

Roman Pohrt
roman.pohrt@tu-berlin.de

Specialty section:

This article was submitted to
Tribology, a section of the journal
Frontiers in Mechanical Engineering

Received: 28 May 2020

Accepted: 14 September 2020

Published: 05 October 2020

Citation:

Pohrt R (2020) Friction Influenced by
Vibrations: A Refined Contact-
Mechanics View on Lateral and
Rotational Oscillations.
Front. Mech. Eng. 6:566440.
doi: 10.3389/fmech.2020.566440

It is known that superposed movements can lower the friction felt at the macroscale. This is well documented for in-plane and normal translatory oscillations. Contact mechanics are a suitable approach to model this effect but so far have not gone beyond single-slider dynamics. In this study, we make use of 3D Boundary Elements Simulations to study the macroscopic friction reduction. This approach allows us to take into account also partial sliding of the contact zone. We first revisit the case of transversal in-plane translatory oscillations. Here, we argue that the behavior at small velocities can best be described when partial slip is indeed taken into account. Next, we investigate the frictional response of a Hertzian indenter when the lateral movement is superposed with a rotational bore movement. An analytical approximation is given for the steady state solution with constant angular velocity. The third case under investigation is an oscillating bore rotation. We present numerical results for the reduction of the macroscopic friction. In two limiting cases, an analytical prediction is given, following the lines used in the translatory case. For extremely large amplitudes, it is based on the idea that a rotational steady state is assumed at every instant. For small velocities we adapt our new approach including partial sliding. We find these predictions to be good but not perfect, slightly underestimating the reduction that rotational oscillations can provide.

Keywords: friction, vibrations, microslip, fretting, shakedown

INTRODUCTION

When frictional contacts are subjected to external oscillations, the friction felt at the macroscale is generally reduced. Since this constitutes a relatively simple way to control friction, the effect is used in countless manufacturing applications (Siegert and Ulmer, 2001a; Siegert and Ulmer, 2001b; Murakawa, 2001; Egashira and Mizutani, 2002; Ashida and Aoyama, 2007) as well as in noise control (Thomsen, 1999). On the negative side, it can lead to an error in measurements of the coefficient of friction (Kado et al., 2014).

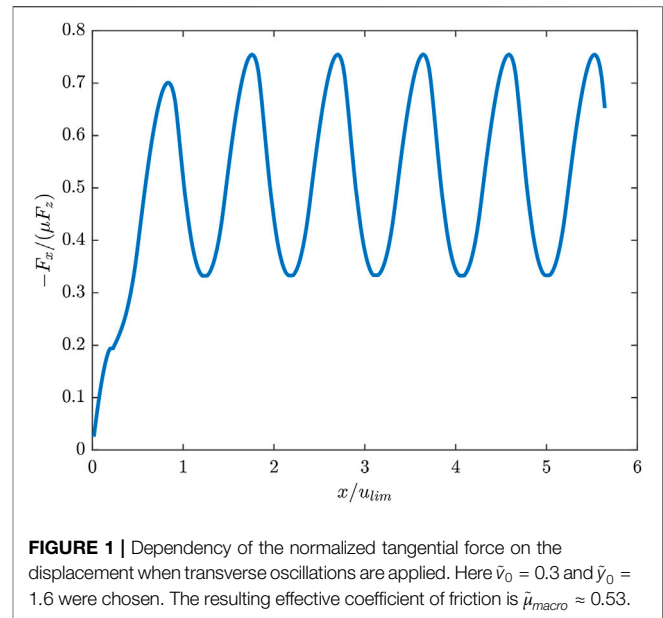
The effect has been described experimentally in the middle of the last century (Fridman and Levesque, 1959; Pohlman and Lehfeldt, 1966; Godfrey, 1967; Lenkiewicz, 1969) but modeling was not attempted. In 2002, Storck (2002) investigated translatory in-plane oscillations experimentally and analytically. They argued that the interplay of forward friction and the superposed vibrations can be modeled using the perspective of contact mechanics with the simple Coulomb law of friction of a single slider. Kumar and Hutchings (2004) use a comparable modeling to interpret their experimental data. Tsai and Tseng (2005) employed the Dahl friction

model to estimate the reduction for in-plane oscillations and found that rigid slider models overestimate the friction reduction. Teidelt, (2012) reported an impressive series of experimental data on the reduction for oscillations in the two in-plane directions as well as the out-of-plane direction and interpreted the results using a model of rigid sliders. Starcevic and Filippov also worked on this dataset. They used the Method of Dimensionality Reduction to study static friction under the influence of in-plane parallel oscillations. Their setup consisted of two coupled parabolic sliders and included micro-slip (Starcevic and Filippov, 2012). The same method was used by Teidelt et al. (2012) to study the performance of Microdrive actuators. The simultaneous acting of normal and in-plane-parallel oscillation was studied by Popov and Li (2018) using an elastic slider model. The case of in-plane oscillation transversal to the macroscopic forward motion was modeled in a recent study by Benad et al., (2019). They used a coulomb-type frictional slider attached to a linear spring. In this work, results are given as a function of normalized input quantities and explicit dependencies are given for two limiting cases, corresponding to very low forward velocities (static friction) on the one hand and very large oscillation amplitudes on the other hand. Their work can be considered the starting point of the current paper. “*Transverse Oscillations*” section of the current paper will build on their results but will introduce explicit partial slip.

The setup under investigation is the following: We focus solely on superposed in-plane motion, leaving the normal contact unchanged. Unless stated otherwise, the indenter is a sphere with radius R , approximated by a parabola shape. It is compressed elastically against a non-deformable flat counterbody. The resulting contact zone and normal stresses are of the Hertzian type and are assumed to be unaffected by tangential stresses. The elastic sphere and the rigid counterbody interact locally with a constant coefficient of friction μ , such that the effective tangential stress inside the contact zone is always less or equal to the normal stress at that point. While the indentation depth in normal direction (z) is held constant, the sphere is subjected to controlled in-plane displacements. “*Transverse Oscillations*” section will start with a constant velocity in forward (x) direction and harmonic oscillation in y -direction. In “*Continuous Rotation (Bore)*” section, the sphere will also move with constant velocity but will also rotate (bore) around its central z -axis at constant angular velocity. “*Oscillating Rotation (Bore)*” section will cover the case of oscillatory bore rotation. In all cases, the macroscopic forward friction will be decreased. The greatest reduction is generally associated with larger amplitude, frequency or velocity, respectively, of the superposed motion.

TRANSVERSE OSCILLATIONS

The case of transverse oscillations consists of a frictional indenter which is moved with constant velocity in x -direction while its position in y -direction is given by a sine-function



$$\begin{aligned} x(t) &= v_0 \cdot t \\ y(t) &= y_0 \sin(\omega_{transv} t) \end{aligned} \quad (1)$$

The simplest approach is to assume that **Eq. 1** describe not only the macroscopic motion of the indenter but also the exact motion of the *contact spot*. This approach neglects the lateral elasticity of the indenter. In contrast, the model of Benad et al. (2019) consists of a single slider but connected to the coordinates given in **Eq. 1** by a linear spring. They showed that the stiffness has a considerable influence on the system behavior, in particular for small oscillation amplitudes. We agree with this assessment. Because the system is quasistatic, it makes sense to formulate in problem in a way that eliminates the time. Similar to Benad et al. (2019), we employ a set of two dimensionless variables for such system including elasticity. The amplitude of the oscillation is described as

$$\tilde{y}_0 = \frac{y_0}{u_{lim}} \quad (2)$$

where $u_{lim} = \mu d \frac{E^*}{G^*}$ is the maximum unidirectional tangential displacement of the indenter before full sliding occurs, see (Mindlin, 1949; Popov, 2019). Here d is the indentation depth in normal direction, E^* , G^* are the reduced moduli of shear and elasticity **Eq. 1**, ν is Poisson ratio.

$$E^* = \frac{E}{1 - \nu^2}, \quad G^* = \frac{4G}{2 - \nu} \quad (3)$$

The forward speed is normalized with respect to the speed of the oscillation and is expressed as the dimensionless variable.

$$\tilde{v}_0 = \frac{v_0}{\omega_{transv} \cdot u_{lim}}. \quad (4)$$

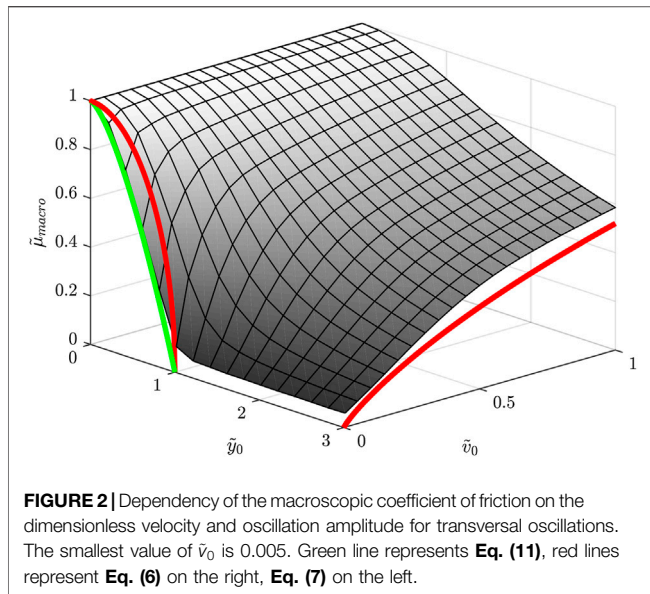


FIGURE 2 | Dependency of the macroscopic coefficient of friction on the dimensionless velocity and oscillation amplitude for transversal oscillations. The smallest value of \tilde{v}_0 is 0.005. Green line represents Eq. (11), red lines represent Eq. (6) on the right, Eq. (7) on the left.

The definitions \tilde{v}_0 and \tilde{y}_0 coincide with those from Benad et al. (2019) when l_0 of their paper equals u_{lim} , which is the most reasonable interpretation since “ l_0 is the elongation at which sliding starts.”

The dependent variable is the apparent coefficient of friction $\tilde{\mu}_{\text{macro}}$ which is determined from the tangential force F_x in x -direction and the normal force F_z

$$\tilde{\mu}_{\text{macro}} = \frac{\langle F_x(t) \rangle}{\mu F_z}. \quad (5)$$

Here $\langle \dots \rangle$ means averaging over one period of oscillation, after the system has tuned in.

Results

Figure 1 shows a typical evolution of the resistive tangential force in x -direction. Such curves were obtained for a variety of combinations for parameters \tilde{v}_0 and \tilde{y}_0 . The resulting macroscopic coefficient of friction is depicted in Figure 2 and mostly reaffirms known results (Tsai and Tseng, 2005). In Benad et al. (2019), an almost identical graph is not given in the paper but can be found on the cover page of the journal's issue.

Limiting Solution for Large Oscillation Amplitudes

An estimate for the apparent coefficient of friction at large oscillation amplitudes is given as

$$\tilde{\mu}_{\text{macro, large}} = \frac{2}{\pi} \int_0^{\pi/2} \frac{d\tau}{\sqrt{1 + \left(\frac{\tilde{y}_0}{\tilde{v}_0} \sin \tau\right)^2}}. \quad (6)$$

In the range covered in this current work, we find the same as (Benad et al., 2019): We basically confirm (6) and find it slightly overestimates the reduction in μ_{macro} for finite \tilde{y}_0 , see the red curve in Figure 2. This is to be expected, since the turning points in the oscillation delay the reducing effect.

Limiting Solution for Small Driving Velocities

The limit of small driving velocities can be considered the case of static friction. From a macroscopic point of view, this describes simply the value of the tangential force below which the indenter does not advance. Looking at the situation within an oscillation cycle, the limit translates into the requirement that at least during one infinitesimal instant of the cycle, a forward motion is possible.

In the elastic-single-slider model of (Benad et al., 2019), these instants coincide with the extremal values of the oscillations and a value for the tangential force can be associated. For μ_{macro} a closed form solution at $\tilde{v}_0 \rightarrow 0$ is deduced

$$\tilde{\mu}_{\text{macro, static, Benad}} = \sqrt{1 - \tilde{y}_0^2} \quad (7)$$

This solution is applicable for $\tilde{y}_0 \leq 1$ and is shown in Figure 2. For any larger values, static macroscopic friction vanishes.

Our numerical simulations show that Eq. 7 overestimates the macroscopic friction. In other words, the reduction resulting from the oscillation is underestimated. Here we'd like to argue that this is due to the exclusion of partial slip in the model. Partial slip does not occur in a single slider element but requires a finite contact zone and distribution of normal stress.

For a Hertzian indenter in the absence of a forward motion, the oscillation with amplitude $\tilde{y}_0 \leq 1$ gives rise to the formation of a ring-shaped slip zone and an inner circle of stick. According to Mindlin theory, the radius of the stick region c depends on \tilde{y}_0 according to

$$c = a \sqrt{1 - \tilde{y}_0^2}. \quad (8)$$

where a is the contact radius. From this state, what condition must be met in order to achieve a forward motion? Since $\tilde{v}_0 \rightarrow 0$, the oscillations are at high frequency in the outer ring of slip and are very effective in reducing, even suppressing the friction there. Therefore it is reasonable to assume that the driving forward motion only has to overcome the friction inside the remaining stick zone. With the total normal force F_z and the normal force only acting inside the stick zone $F_{z,c}$, we can conjecture

$$\tilde{\mu}_{\text{macro, static}} = \frac{F_{z,c}}{F_z} = \frac{\int_0^c p(r) r dr}{\int_0^a p(r) r dr}. \quad (9)$$

Here $p(r)$ is the distribution of normal contact stress. In the particular case of a Hertzian stress distribution, Eq. 9 reduces to

$$\tilde{\mu}_{\text{macro, static}} = 1 - \left(1 - \frac{c^2}{a^2}\right)^{3/2} \quad (10)$$

Furthermore in the particular case of translatory oscillations using Eq. 8 it reduces to

$$\tilde{\mu}_{\text{macro, static}} = 1 - \tilde{y}_0^{3/2} \quad (11)$$

which is also plotted in Figure 2 and coincides much better with direct simulation results. Thus we see reason that unless the result is zero or one, $\tilde{\mu}_{\text{macro, static}}$ cannot be predicted by a single slider approach.

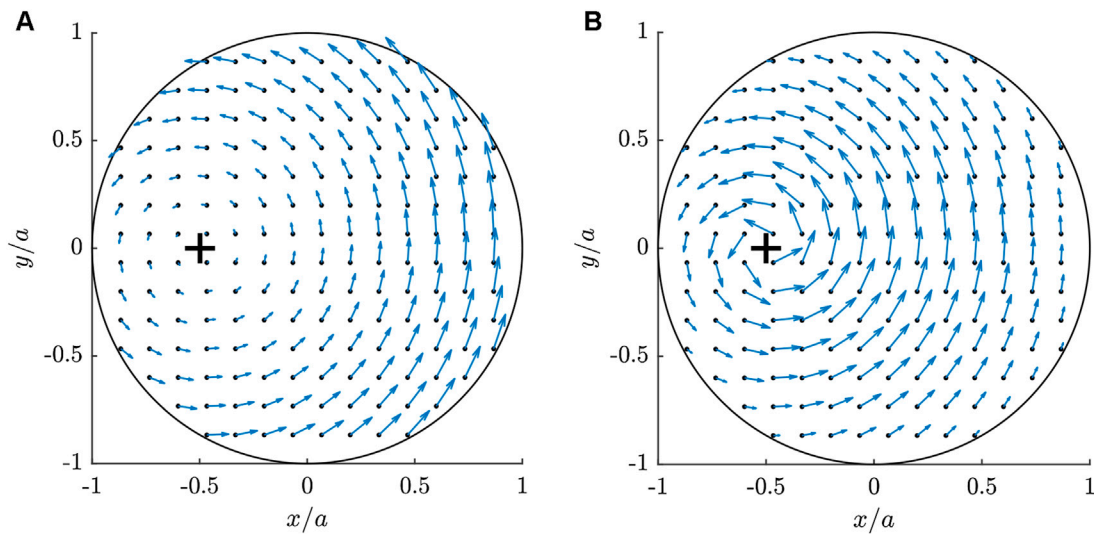


FIGURE 3 | Hertzian indenter slipped in positive y -direction and rotated. Normalized forward speed $\tilde{v} = 0.5$. Quiver plots of **(A)** local velocity field, **(B)** tangential surface stresses. The stress vector field also rotates around $x/a = -0.5$ but its amplitude is related to the contact pressure.

CONTINUOUS ROTATION (BORE)

We now investigate a rotationally symmetric indenter which is rotated around the z -axis while sliding in the x - y -plane.

$$\begin{aligned} x(t) &= 0 \\ y(t) &= v_0 t \\ \varphi(t) &= \omega t \end{aligned} \quad (12)$$

Since the problem is quasistatic, it can be reduced to a single system parameter and dimensionless time

$$\tilde{v} = \frac{v_0}{a\omega}, \quad \tilde{t} = \frac{t}{v_0 a} \quad (13)$$

In this configuration, the instantaneous center of rotation in the undeformed remote parts of the body is situated at

$$x_{\text{center}} = -\tilde{v}a, \quad y_{\text{center}} = v_0 t. \quad (14)$$

See **Figure 3A**. The steady state can be analyzed using the following simplifications.

- The velocities of all spots on the surface are assumed to be the same as their corresponding spots in the remote body
- For the calculation of frictional torque, surface stresses are assumed to act at the undeformed spot.

In particular, the above simplifications are valid for small μ . The amplitudes of the frictional stresses τ are given by $|\tau(x, y)| = \mu p(x, y)$ and they are directed opposite to the velocity in each surface point. This can be expressed as

$$\begin{aligned} \tau_x(x, y) &= -\frac{y}{\sqrt{y^2 + (x - x_{\text{center}})^2}} \mu p(x, y) \\ \tau_y(x, y) &= -\frac{(x - x_{\text{center}})}{\sqrt{y^2 + (x - x_{\text{center}})^2}} \mu p(x, y) \end{aligned} \quad (15)$$

See **Figure 3B** for an example. The resulting tangential forces and torque can then be calculated as

$$F_x = 0 \quad (16)$$

$$F_y = \int_0^a \tau_y dA = \mu \int_0^a p(r) r Q_Y(r/\tilde{v}a) dr \quad (17)$$

$$M_z = \int_0^a (\tau_y x - \tau_x y) dA = \mu \int_0^a p(r) r^2 Q_M(r/\tilde{v}a) dr \quad (18)$$

wherein $\int_0^a dA$ means integrating over the contact area and we introduced

$$Q_Y(\tilde{r}) = \int_0^{2\pi} \frac{\tilde{r} \cos \varphi + 1}{\sqrt{\tilde{r}^2 + 2\tilde{r} \cos \varphi + 1}} d\varphi \quad (19)$$

$$Q_M(\tilde{r}) = \int_0^{2\pi} \frac{\tilde{r} + \cos \varphi}{\sqrt{\tilde{r}^2 + 2\tilde{r} \cos \varphi + 1}} d\varphi \quad (20)$$

The additional rotational movement eases the forward movement and effectively reduces the apparent coefficient of friction. Vice versa, the forward motion eases the rotation. Therefore, the required torque M_z also depends on \tilde{v} . A macroscopic coefficient of friction can be formulated for both the lateral movement and the rotation.

$$\tilde{\mu}_y(\tilde{v}) = F_y / \mu F_z = \frac{\int_0^1 Q_Y(\tilde{r}/\tilde{v}) p(\tilde{r}a) \tilde{r} d\tilde{r}}{2\pi \int_0^1 p(\tilde{r}a) \tilde{r} d\tilde{r}} \quad (21)$$

$$\tilde{\mu}_M(\tilde{v}) = M_z / \mu M_{z,0} = \frac{\int_0^1 Q_M(\tilde{r}/\tilde{v}) p(\tilde{r}a) \tilde{r}^2 d\tilde{r}}{2\pi \int_0^1 p(\tilde{r}a) \tilde{r}^2 d\tilde{r}} \quad (22)$$

where $M_{z,0}$ is the steady state torque at zero forward velocity.

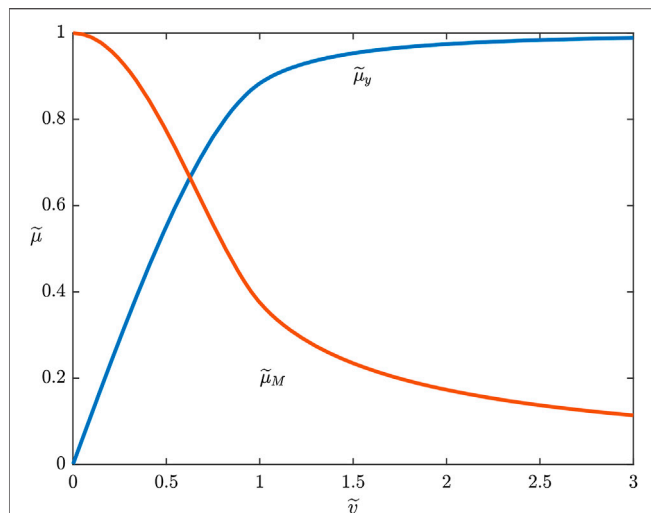


FIGURE 4 | Theoretical prediction of the dependency of the macroscopic coefficient of friction on the dimensionless velocity for added continuous rotational (bore) movement to a Hertzian Indenter according to **Eqs 21** and **22**. The frictional resistances on both the forward motion and on the rotational torque are reduced.

For a Hertzian Indenter the dependencies of **Eqs 21** and **22** on \tilde{v} are shown in **Figure 4**. In our transient numerical simulations, these values are approached after the influence of the initial

conditions has vanished. Typical evolutions of the tangential forces and torque are shown in **Figure 5**. It can be seen that the macroscopic forces and torque approach a steady state which is indeed well predicted by **Eqs 21** and **22**.

Interestingly, our simulations never showed zones of stick when the steady state was reached, even when instantaneous center of rotation is located inside the contact region.

OSCILLATING ROTATION (BORE)

In this section, we will set a continuous forward motion and the additional bore rotation will be oscillating. Let us impose a displacement of the indenter according to

$$\begin{aligned} x(t) &= 0 \\ y(t) &= v_0 \cdot t \\ \varphi(t) &= \varphi_0 \sin(\Omega t), \quad \omega = \dot{\varphi} = \varphi_0 \Omega \cos(\Omega t) \end{aligned} \quad (23)$$

Some normalization is again required for the oscillation amplitude. In “*Transverse Oscillations*” section, it was divided by the maximum tangential deflection before the start of macroscopic slip, which is when the inner stick region fully vanishes. In the case of pure rotation of a Hertzian indenter however, there is no such maximum angle at which the stick region vanishes. According to Lubkin (1951) (indexed “lk”) (Popov, 2019), the stick radius c and the torsion angle φ are related as

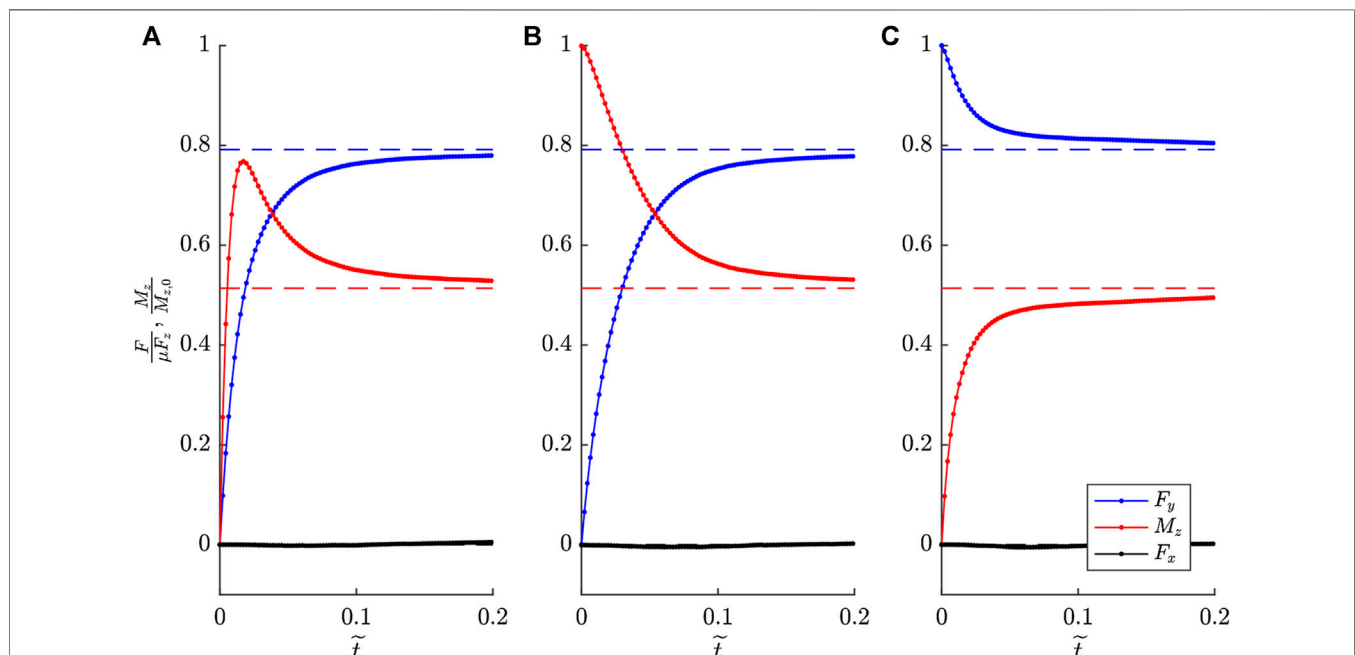


FIGURE 5 | Coefficient of friction regarding Forces and torque of a Hertzian Indenter which is displaced and rotated simultaneously at $\tilde{v} = 0.8$. Dashed lines represent theoretical prediction according to **Eqs 21** and **22**. Initial conditions were (A) full stick without tangential stress, (B) full rotational slip without lateral displacement and (C) full lateral slip, no rotation.

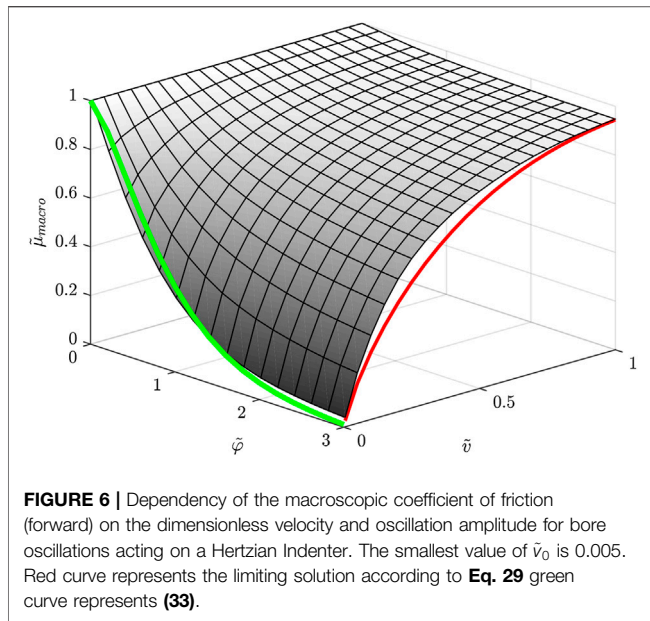


FIGURE 6 | Dependency of the macroscopic coefficient of friction (forward) on the dimensionless velocity and oscillation amplitude for bore oscillations acting on a Hertzian Indenter. The smallest value of \tilde{v}_0 is 0.005. Red curve represents the limiting solution according to **Eq. 29** green curve represents **(33)**.

$$\begin{aligned}\varphi_{lk}(c) &= \frac{\mu E^* a}{\pi GR} \left[K(\sqrt{1 - c^2/a^2}) - E(\sqrt{1 - c^2/a^2}) \right] \\ &= \frac{u_{lim}}{a} \frac{4}{\pi(2 - \nu)} D(c/a)\end{aligned}\quad (24)$$

where we introduced

$$D(m) = K(\sqrt{1 - m^2}) - E(\sqrt{1 - m^2}) \quad (25)$$

With the complete elliptical integrals of the first and second kind

$$\begin{aligned}K(k) &= \int_0^{\pi/2} (1 - k^2 \sin^2 \varphi)^{-1/2} d\varphi, \\ E(k) &= \int_0^{\pi/2} (1 - k^2 \sin^2 \varphi)^{1/2} d\varphi\end{aligned}\quad (26)$$

Indeed **Eq. 24** tends to infinity for small c , meaning that the very center of the contact will always be in stick state. Normalization of the rotational amplitude will be instead be done with respect to the angle necessary to achieve a stick radius of $c = a/2$ in the absence of lateral displacement.

$$\tilde{\varphi} = \frac{\varphi_0}{\varphi_{lk}(a/2)} = \varphi_0 \frac{a}{u_{lim}} \frac{\pi(2 - \nu)}{4D(1/2)} \quad (27)$$

The normalization of the forward velocity is done with respect to the maximum speed of the contact zone edge at $r = a$

$$\tilde{v} = \frac{v_0}{\Omega \varphi_0 a} \quad (28)$$

With these definitions, simulations and evaluation can be done in analogy to “Transverse Oscillations” section.

Results

Figure 6 shows the results of μ_{macro} for a range of parameters \tilde{v} and $\tilde{\varphi}$. At first sight, the general appearance resembles the dependencies shown in **Figure 2** of “Transverse Oscillations”

section. For $\tilde{\varphi} = 0$, no reduction is found and thus $\tilde{\mu}_{macro} = 1$. For large amplitudes of the rotational amplitude, friction is greatly reduced, especially for small forward velocities \tilde{v} . Please note however, that even in the limit of $\tilde{v} \rightarrow 0$, the macroscopic friction is not reduced to zero.

Limiting Solution for Large Oscillation Amplitudes

For the limiting case of large amplitudes, a similar expression to **Eq. 6** can be found, when it is assumed that the contact is always in a rotational-transversal steady state such as “Continuous Rotation (Bore)” section has introduced. Then, the results found in **Eq. 21** can be averaged over one period with momentary values of $\omega(t) = \varphi_0 \Omega \cos(\Omega t)$ yielding

$$\begin{aligned}\tilde{\mu}_{macro, largerot} &= \frac{2}{\pi} \int_0^{\pi/2} \mu_y \left(\frac{v_0}{a \varphi_0 \Omega \cos(\tau)} \right) d\tau \\ &= \frac{2}{\pi} \int_0^{\pi/2} \mu_y \left(\frac{\tilde{v}}{\cos(\tau)} \right) d\tau\end{aligned}\quad (29)$$

This finding is shown in **Figure 6** with the red line. Please note the resemblance to **Eq. 6** in **Figure 2**.

Limiting Solution for Small Driving Velocities

Following the approach laid out in “Limiting Solution for Small Driving Velocities,” it is possible to formulate an approximation for the static friction in the limit of $\tilde{v} \rightarrow 0$. Let us take the exact same steps for the rotational problem.

In a situation without forward motion, the rotational oscillation will give rise to a zone of constant stick. The oscillation amplitude φ_0 and the static stick radius c are related by **Eq. 24**

$$\varphi_0 = \varphi_{lk}(c) \quad (30)$$

which we can now write in normalized form, introducing $\tilde{c} = c/a$

$$\tilde{\varphi}_{lk}(\tilde{c}) = \frac{\varphi_{lk}(c)}{\varphi_{lk}(a/2)} = \frac{D(\tilde{c})}{D(1/2)} \quad (31)$$

This is a monotonous dependency with $D(1/2) \approx 0.9455$. Let $\tilde{c}_{lk}(\tilde{\varphi})$ be the inverse function

$$\tilde{c}_{lk}(\tilde{\varphi}) = D^{-1}[D(1/2)\tilde{\varphi}] \quad (32)$$

We can then use this stick radius with **Eq. 10** to find

$$\tilde{\mu}_{macro, static} = 1 - (1 - \tilde{c}_{lk}(\tilde{\varphi}_0))^2)^{3/2} \quad (33)$$

Which is shown in **Figure 6** with the green line. It can be seen that **Eq. 33** constitutes a reasonably good prediction of static friction, but not an exact one. It appears to accurately predict the static friction very well for $\tilde{\varphi}_0$ greater than approximately 2. At this point one should note that our judgment on the accuracy is limited, because in order to reach a steady state, the velocities used in the simulation must be strictly positive and smaller values require more computation time. Therefore, the minimum value represented in **Figure 6** is $\tilde{v}_0 = 0.005$. For $\tilde{\varphi}_0 < 2$ however, **Eq. 33** clearly underestimates the reduction of friction.

This means that the forward force does not need to overcome the entire normal stress inside the oscillatory stick zone multiplied with μ , but less. This in turn means that during the forward motion of the indenter, not the entire inner circle slides forward. Indeed, the time series of these simulations indicate that at no point in time there is full sliding in the contact zone. Instead, patches of stick alternate between the left and the right hand side of the contact zone.

CONCLUSIONS

In this paper, we investigate friction under the influence of superposed motion using explicit contact mechanics. In all cases, the superposed motion reduces friction. The stronger the motion (faster, higher amplitude) the greater the reduction. Under the influence of any given superposed motion, the forward friction increases with the forward velocity, stabilizing the frictional system.

Based on the findings presented here for transversal oscillations, the spring model given in Benad et al. (2019) provides satisfactory results, except for very small forward velocities, so-to-speak for static friction when the oscillations provoke some partial slip, which should be taken into account. We postulate that in order to start a forward motion, only the friction inside this stick region must be overcome.

For friction under the influence of additional continuous bore movement, an approximation for the steady state is identified and given in “Continuous Rotation (Bore)” section. Both the friction opposing forward motion and opposing the rotation are reduced.

When the bore movement oscillates, a reduction of the macroscopic forward friction is again found and the dependency on forward velocity and oscillation amplitude resembles the transversal case (Compare **Figures 2, 6**). Here, two limiting cases are again identified and analyzed. At large amplitudes, assuming steady state in every instant leads to a good approximation. For small velocities near static friction, a new estimate with good agreement is suggested. It is again based on the stick zone caused by the oscillations alone. However, simulations show that the reduction of forward friction is even greater than predicted, hinting to a more elaborate mechanism.

Rotational bore oscillations are a promising way to control static friction or micro-propulsion due to two favorable properties over transversal oscillations. First, $\tilde{\mu}_{macro,static}$ does not sharply saturate at zero for increasing amplitudes but is instead very well controllable in this region. Second, the reduction in $\tilde{\mu}_{macro,static}$ is very effective even at small amplitudes.

REFERENCES

- Ashida, Y., and Aoyama, H. (2007). Press forming using ultrasonic vibration. *J. Mater. Process. Technol.* 187–188, 118–122. doi:10.1016/j.jmatprotec.2006.11.174

METHOD

We employ the Boundary Elements Method for contact problems (Pohrt and Li, 2014). The normal contact is solved following standard procedures. Interaction of normal and in-plane deformations are not considered. Interactions between the two in-plane directions are considered. For the tangential contact, an iterative scheme is employed that ensures each point satisfies the following conditions in each time step.

- Tangential stress is below the frictional threshold $|\tau| < \mu p$ (“stick state”) or,
- Tangential stress $|\tau| = \mu p$ and the difference in point coordinates including deformation, the local velocity, is directed opposite of tangential stress (“slip state”)

where τ is the tangential stress, having x and y components, p is the local pressure and μ is the coefficient of friction. For the stress inequality we typically allowed an error of 0.1% and for the alignment of local stresses and velocities we typically allowed an error of 4°. Achieving convergence in the iterative scheme proved be time-consuming. Furthermore, the investigations of low forward speeds required the use of up to 8,000 time steps. As a consequence we limited the spatial resolution of the contact region to 32 times 32 points in bulk simulations such as those shown in **Figures 2, 6**. We used $\nu = 0$ for the Poisson ratio and $\mu = 0.2$.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RP developed the BEM code, performed the simulations and created the text and figures.

FUNDING

This research was funded by regular financing of researchers at Technische Universität Berlin.

ACKNOWLEDGMENTS

Discussions with colleagues from TU Berlin were helpful in efforts of the author.

- Benad, J., Nakano, K., Popov, V. L., and Popov, M. (2019). Active control of friction by transverse oscillations. *Friction* 7 (1), 74–85. doi:10.1007/s40544-018-0202-1
- Egashira, K., Mizutani, K., and Nagao, T. (2002). Ultrasonic vibration drilling of microholes in glass. *CIRP Ann.* 51 (1), 339–342. doi:10.1016/S0007-8506(07)61531-5

- Fridman, H. D., and Levesque, P. (1959). Reduction of static friction by sonic vibrations. *J. Appl. Phys.* 30 (10), 1572–1575. doi:10.1063/1.1735002
- Godfrey, D. (1967). Vibration reduces metal to metal contact and causes an apparent reduction in friction. *Tribol. Trans.* 10, 183–192. doi:10.1080/05698196708972178
- Kado, N., Tadokoro, C., and Nakano, K. (2014). Kinetic friction coefficient measured in tribotesting: influence of frictional vibration. *Tribol. Online.* 9 (2), 63–70. doi:10.2474/trol.9.63
- Kumar, V. C., and Hutchings, I. M. (2004). Reduction of the sliding friction of metals by the application of longitudinal or transverse ultrasonic vibration. *Tribol. Int.* 37 (10), 833–840. doi:10.1016/j.triboint.2004.05.003
- Lenkiewicz, W., (1969). The sliding friction process: effect of external vibrations. *Wear* 13, 99–108. doi:10.1016/0043-1648(69)90505-5
- Lubkin, J. L. (1951). The torsion of elastic spheres in contact. *J. Appl. Mech.* 73, 183–187.
- Mindlin, R. D. (1949). Compliance of elastic bodies in contact. *J. Appl. Mech.* 16, 259–268.
- Murakawa, M. (2001). The utility of radially and ultrasonically vibrated dies in the wire drawing process. *J. Mater. Process. Technol.* 113 (1–3), 81–86. doi:10.1016/S0924-0136(01)00635-5
- Pohlman, R., and Lehfeldt, E. (1966). Influence of ultrasonic vibration on metallic friction. *Ultrasonics* 4 (4), 178–185. doi:10.1016/0041-624X(66)90244-7
- Pohrt, R., and Li, Q. (2014). Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17 (4), 334–340. doi:10.1134/S1029959914040109
- Popov, M., and Li, Q. (2018). Multimode active control of friction, dynamic ratchets and actuators. *Phys. Mesomech.* 21 (1), 24–31. doi:10.1134/s1029959918010046
- Popov, V. L., Heß, M., and Willert, E. (2019). *Handbook of contact mechanics. Exact solutions of axisymmetric contact problems.* 1st Edn. Berlin: Springer-Verlag, 347.
- Siegert, K., and Ulmer, J. (2001a). Influencing the friction in metal forming process by superimposing ultrasonic waves. *CIRP Ann.* 50 (1), 195–200. doi:10.1016/S0007-8506(07)62103-9
- Siegert, K., and Ulmer, J. (2001b). Superimposing ultrasonic waves on the dies in tube and wire drawing. *J. Eng. Mater. Technol.* 123 (4), 517–523. doi:10.1115/1.1397779
- Starcevic, J., and Filippov, A. E. (2012). Simulation of the influence of ultrasonic in-plane oscillations on dry friction accounting for stick and creep. *Phys. Mesomech. Mesomech.* 15 (5–6), 330–332. doi:10.1134/S1029959912030150
- Storck, H. (2002). The effect of friction reduction in presence of ultrasonic vibrations and its relevance to travelling wave ultrasonic motors. *Ultrasonics* 40, 1–8. doi:10.1016/s0041-624x(02)00126-9
- Teidelt, E. (2012). Influence of ultrasonic oscillation on static and sliding friction. *Tribol. Lett.* 48, 51–62. doi:10.1007/s11249-012-9937-4
- Teidelt, E., Willert, E., Filippov, A. E., and Popov, V. L. (2012). Modeling of the dynamic contact in stick-slip microdrives using the method of reduction of dimensionality. *Phys. Mesomech.* 15 (5–6), 287–292. doi:10.1134/S1029959912030071
- Thomsen, J. (1999). Using fast vibrations to quench friction-induced oscillations. *J. Sound Vib.* 228, 1079–1102. doi:10.1006/jsvi.1999.2460
- Tsai, C. C., and Tseng, C. H. (2005). The effect of friction reduction in the presence of in-plane vibrations. *Arch. Appl. Mech.* 75, 164–176. doi:10.1007/s00419-005-0427-0

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pohrt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Metal Transfer and Wear

James A. Greenwood*

Department of Engineering, University of Cambridge, Cambridge, United Kingdom

Keywords: transfer layer, wear fragments, riders, prowls, radioactivity

INTRODUCTION

These thoughts are offered as a reminder that Tribology is not all about the normal contact of fractal surfaces, and indeed, not all about elastic contact of rubber and polymers, or even about dry contact. Machines do still contain metal surfaces sliding past each other, hopefully separated by an oil film; and sometimes, when tolerances have been pushed too far, or running with starved lubrication when the oil or grease supply is inadequate, with some metal to metal contact. Fortunately this is not always disastrous: surfaces do often run-in, so that after running with contact and a contribution of dry-contact friction, there is steady wear and contacts no longer occur. The traditional design criterion for gears and ball races was, and still is, the Λ – ratio: the ratio of the predicted film thickness for smooth surfaces to the rms roughness. Certainly a Λ – ratio of 3 or more¹ usually leads to full-film lubrication: but to anyone with the slightest background in surface roughness this is an absurd rule. Assuming, as is usually done, that the predicted smooth-surface film thickness refers to the distance between the mean planes of the roughness, the rms roughness says nothing about the how much contact there will be. And if running-in is successful, and the high points of the surface wear away, the rms (and the Λ – ratio) may hardly change, but there will be successful operation.

But when will running-in be successful? What determines when instead of running-in there will be scuffing, and disaster?

The traditional picture of the “mixed friction” regime is that when the local film thickness falls to zero, additives (or perhaps happy accidents) provide a boundary lubricant in the oil: some form of long chain polymer, which has a reactive end which attaches itself to the metal, and carries the load on its free ends: with low friction but, more importantly, preventing metal to metal contact. The Blok scuffing criterion was that the maximum surface temperature must be below a specific value: and there was the problem, what should it be? In Bowden and Tabor’s laboratory experiments, using a known, pure, organic compound, clear links with the known properties could be found; but in engineering practice perhaps all that can be done is to ensure that the calculated maximum temperature in a new application is no more than in an existing application: the ISO guide concentrates on the temperature calculation, not on the temperature found.

But what happens when boundary lubrication fails? Fortunately it seems that we do not move completely into the dry wear scenario. The failure will usually be local, and the dry wear process interrupted. An earlier work (Sakmann et al., 1944) reported that in a pin on disc experiment, flooding the surface with a plain mineral oil halved the transfer at a light load, but produced only a small reduction at a heavier load. But flooding with oleic acid largely eliminated transfer.

Here it seems desirable to review what has been learnt about dry wear, and perhaps, forgotten.

TRANSFER AND WEAR

The obvious starting point is the “Archard” wear equation. This was predicted by Holm in 1938 [Holm (1938)], by postulating that for every encounter of a pair of atoms (within the contact area found as W/p_m) there was a fixed probability of one being pulled out of its parent surface. Detailed

¹The “traditional” requirement is $\Lambda > 3$. But see Cann et al. (1994), or Greenwood (2020) for alternatives.

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Evgeny Viktorovich Shilko,
Institute of Strength Physics and
Materials Science (ISPMS SB
RAS), Russia

*Correspondence:

James A. Greenwood
jag5@cam.ac.uk

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 13 May 2020

Accepted: 30 June 2020

Published: 11 August 2020

Citation:

Greenwood JA (2020) Metal Transfer
and Wear. *Front. Mech. Eng.* 6:62.
doi: 10.3389/fmech.2020.00062

experimental confirmation was provided by Burwell and Strang (1952), but from electron micrographs of transfer particles they argue that the unit event is the encounter of two asperities. Both models predict that the volume of wear V is proportional to the distance slid L and the load W , and inversely proportional to the hardness p_m : $V = k \cdot L \cdot (W/p_m)$. Archard's contribution was to show that it is not necessary to assume that the average size of the contact areas or wear particles is constant, and to calculate the probabilities k implied by the results of all the available experimental wear combinations ... and to go on to contribute to the great wear research program of Hirst's group at AEI (Associated Electrical Industries) Aldermaston (see Archard, 1953; Kerridge, 1955, etc.).

WEAR OF SOFT STEEL AGAINST HARD STEEL

The natural meaning of the term "wear" is the weight, or volume, lost from the device concerned: and early researchers merely noted that this could become either transfer particles attached to the "wrong" partner, or loose wear debris. The important distinction between transfer and wear was first made when Kerridge (1955) found that when a (radioactive) steel pin was loaded against a rotating hard steel ring ("hollow drum" perhaps conveys the picture), a radioactive transfer layer built up on the ring, but the radioactivity (and therefore the amount of transfer) then became constant. When the active pin was replaced by an inactive one, the activity reduced, mirroring the path of the increase: it was not that the transfer layer had a maximum size, and could build up no further, but that a steady state had been reached where the transfer to the ring equaled the rate of loss from the ring. At this point the pin wear rate fell to the steady rate required by the wear law. The wear fragments were carefully collected and monitored, little radioactivity being found at first, but ultimately matching the wear rate of the pin: and consisting of relatively large, *oxidized*, particles. The detachment of the transfer layer, and so presumably its oxidation rate, was the rate-determining process. Experiments in air at 10^{-3} mm mercury found the wear rate reduced to a tenth (or lower at low loads) of the atmospheric value, confirming this.

Thus, for this combination, wear is a multi-stage process. First, metallic transfer from pin to ring as submicroscopic particles. These are then smeared out to form a (harder) transfer layer. The transfer layer then oxidizes, helped by the temperature produced by frictional heating, weakening its attachment to the ring, and finally the oxidized layer is rubbed off as large oxide particles to form the wear debris.

WEAR OF BRASS AGAINST HARD STEEL (SEVERE WEAR)

Kerridge and Lancaster (1956) followed this by replacing the soft steel pin by an irradiated (60/40) brass pin. Initially both wear of the pin and transfer increased exponentially at almost the same rate, with nothing left over to form wear fragments. When the experiment was repeated with a load $5\times$ higher but lubricated

with cetane, the same was true, but the amounts reduced to one quarter. The significant feature was that as the transfer film reached its limiting value, the wear rate fell to its uniform rate. As in Kerridge's experiments, when the active pin was replaced by an inactive one, the decrease from the limiting activity rate to zero mirrored the growth from zero to limiting activity. And when an inactive pin was replaced by an active one, there was at first no activity in the wear debris: *wear particles are not produced directly from the pin*. Convincing evidence was given that individual wear particles at a given time after a pin replacement all had the same activity, i.e., each had the same mix of "new" and "old" transfer fragments.

The authors emphasize that the term "transfer film" is misleading; the transfer layer is composed of identifiable fragments. These grow by accretion of further fragments to become incipient wear particles, perhaps $50\times$ larger than the transfer fragments, before becoming detached as wear particles. But the fragments are themselves much larger than the likely size of asperity contacts, more approximating to "the total real area of contact as estimated from the ratio of the load to the flow pressure of 60/40 brass," so it seems (although the authors do not quite say this) that incipient transfer fragments form *on the pin* by accretion from individual intermetallic junctions.

Thus, for this combination, wear is again a multi-stage process: first, transfer fragments are assembled on the pin, which then transfer from pin to ring, and then become larger by accretion: these are smeared out to become harder, thicker, flakes, which finally become detached as wear particles.

The authors, rather as an aside, report that both the transfer rate and the wear rate are proportional to the load, and that the same applies to the *sizes* of transfer fragments and wear particles: concluding that "*the number of individual events occurring ... is the same at all loads.*" This is a flat contradiction of the Archard (and Greenwood!) belief, that the typical event is the same at all loads, and that only the number of events is proportional to load.

The authors also make the very perceptive remark (in view of the work of Cocks and Antler to be described below), "Once a fragment has been transferred to the ring, the load will be concentrated on that fragment during subsequent passages beneath the pin."

WEAR OF HARD STEELS AGAINST HARD STEEL (MILD WEAR)

One might expect mild wear to be the more straightforward process, but this was not what Archard and Hirst (1957) found. When rubbing hardened tool steel pins against a disk of the same material, the initial behavior was as described above (but on a scale two orders of magnitude less): the pin wore by transfer to the disk, the transfer fragments then aggregated into larger (metallic) fragments which became detached as wear particles. But as the conformity between pin and disk improved, many of the transferred fragments were worn away by a much smaller scale process. And then in a final stage the wear rate increased by a factor of five, and seems to have occurred as abrasion by large (oxidized) wear particles. At this stage the transfer "layer"

consisted of areas with no transfer and patches very much thicker than the average (50–100 Å) thickness.

WEDGES (PROWS)

Tribologists have long regretted the impossibility of directly examining the area of contact between sliding metals: but in 1962 two researchers reported the results of looking sideways at the contact²... which “of course” could reveal nothing. But it did!

When Cocks (1962) loaded a hemispherically ended copper rider against the surface of a rotating large cylindrical copper drum (load 700 g, speed 0.038 cm/s) he found: “*This immediately revealed some unexpected phenomena. Most of the time during sliding, the drum and rider were separated by a wedge of displaced metal which emerged from the drum surface and became trapped between them.*”

The same was found for other metal pairs. Rider and drum were polished up to grade 0000 paper, and then cleaned to be wettable by water and yielding a high coefficient of friction (1.1–1.4). The rider was given a slight transverse movement to avoid repeating the same path on the drum.

A wedge of metal displaced from the drum becomes trapped between the two sliding members and holds them apart. Thus, the drum surface slides against the wedge instead of the rider itself. As the sliding proceeds, the wedge continuously receives more metal from the drum. At the same time fragments of metal from the wedge are intermittently redeposited back on the drum, and can be seen emerging from the contact on the drum surface. From time to time the wedge itself breaks away from the rider, and is carried away on the drum surface. However, when this happens a new wedge immediately begins to form and the surfaces are soon visibly separated again.

Neither the hardness, nor the relative hardnesses of rider and drum appeared to matter.

The whole assembly of rider, wedge and part of the drum could be mounted in a cold-setting plastic without ever separating them, and then sectioned, polished and etched. **Figure 1** shows a result. The wedge is initiated by transfer from the drum, and remains bonded to the rider. It is “a relatively compact mass of metal, not an aggregate of individual fragments ...” The wedge grows by accretion, at the front, where “*the material of the wedge is continuous with that of the drum*”: so the relative motion must cause shearing of the metal, presumably rather similar to what happens in the formation of a chip in the machining of a ductile metal.

The life of a wedge is very variable: as short as 20 s (sliding distance 0.8 cm); often 10–20 min (sliding distance 25–45 cm) and sometimes longer. After perhaps 1 min it seems an equilibrium size is reached where continued accretion at the front is balanced by fragments breaking off from the rear (and sometimes the entire wedge breaks away).

Cocks (1964) showed that the wedge formation and consequent separation of the sliding bodies occurs in the same

way between two flat surfaces, usually with several wedges in action at once. The wedge may be formed by accretion from either disk, and occasionally (see one beautiful micrograph of a wedge formed by a pair of SAE1020 steel disks at 5 kg load) by accretion at both ends. These later experiments showed that the wedge mechanism seemed able to continue indefinitely (copper on copper), to reduce to a much smaller scale (steel on steel), or apparently to cease completely (nickel on nickel).

Antler (1962), working with metals suited to electrical contacts, mostly using lighter loads (100 g) and longer runs (40 m sliding distance) found similar behavior. He emphasizes the irrelevance of the relative hardnesses of rider and flat, and how the prow (his term, but also “wedge-shaped prow”) becomes very much harder than either, making its resemblance to a machining tool clearer. But he gives many examples of the complexity of the process: one example being for a gold rider sliding on a palladium flat:

- (a) *First pass: gold rider deposits adherent particles on palladium flat with little palladium transfer to gold;*
- (b) *Second and third passes: rider removes gold from flat;*
- (c) *Subsequent passes: a severely work hardened gold prow gouges [the] flat, producing palladium particles that adhere to the rider;*
- (d) *From this point, sliding is identical to the all-palladium system, regardless of length of run. Practically all debris is palladium.*

Note the hardnesses: Gold, 79 kg/mm²; Palladium, 142 kg/rnm²: but it is the soft gold which wears away the hard palladium!

Antler (1964) notes that *wear by the prow-formation mechanism tends to change to the rider-wear severe regime on prolonged sliding in the same track. This was observed with a variety of metals, including aluminum, copper, gold, palladium, and silver sliding on themselves.* This differs from what Cocks reports for his pin on disc experiments, but he never ran on the same track.

RABINOWICZ'S CONTRIBUTION

The first use of radioactivity to study transfer was when Sakmann et al. (1944) slid an inactive slider over an activated block, and measured the transfer with a Geiger counter. Gregory (1946) slid a radioactive lead slider³ over an inactivated flat surface: this permitted the use of autoradiography to study the transfer. He noted average film thicknesses between 10 and 100 lead atoms thick. Rabinowicz and Tabor (1951) modestly state:

This paper describes an extension of Gregory's autoradiographic method to a study of the friction and pick-up occurring between stationary and sliding metal surfaces in the absence and in the presence of lubricant films.

carrying understatement too far! They established the proportionality between pick-up and load: that for like metals

²Cocks (1958) reported preliminary results in 1958 (cited by Antler, 1962), so undoubtedly has priority.

³A lead slider containing a radioactive isotope (of radon?); by implication all that was available in Australia at the time.

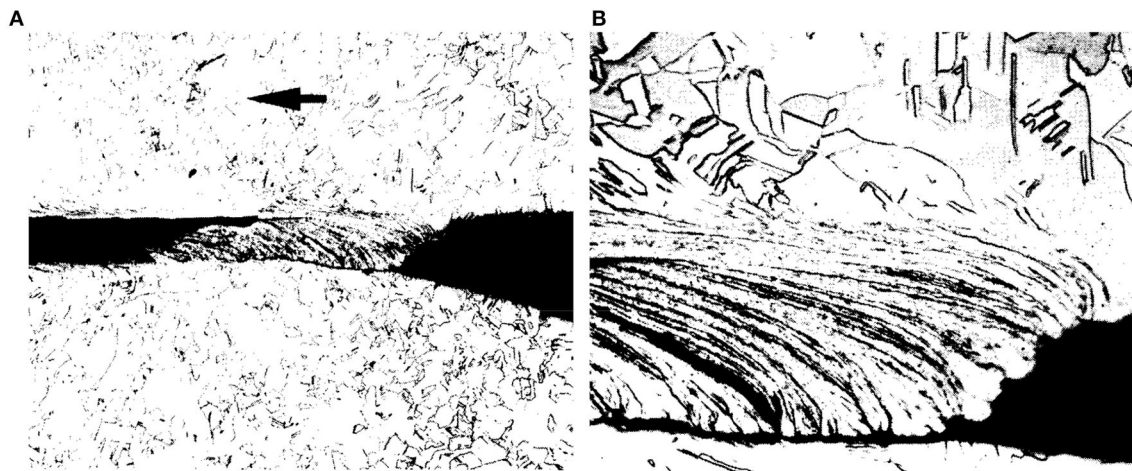


FIGURE 1 | Section through the wedge linking drum and rider, both of oxygen-free high-conductivity copper, after sliding at a speed of 0.038 cm/s with a load of 710 g. **(A)** Complete wedge. Magnification 100 \times . **(B)** Right-hand portion of the section in **(A)**. Magnification 340 \times [Reproduced from Cocks (1962), with the permission of AIP Publishing].

there was no correlation between pick-up and hardness: that for dissimilar metals the pick-up was typically less by a factor 50, mainly due to the smaller size of the transfer fragments, typically 10^{-8} g compared to 10^{-6} g. They conclude that *metallic transfer does not occur as a uniform smear but as a relatively small number of discrete fragments*.

which seems to be the first clear statement of this fact⁴. But sadly for the present purpose, the main thrust of the paper is to study the effect of lubrication, certainly on both friction and transfer, but really hoping to get some clue to the origin of friction.

However, Rabinowicz and Tabor make the important discovery that transfer takes place under purely normal loading, although on a much reduced scale, with transfer fragments typically 10^{-11} g: still enormous by atomic standards (10^{10} atoms). Rabinowicz (1952) followed this up more carefully, finding the total transfer (copper to copper: load 4 kg) of 3×10^{-8} to 3×10^{-9} g, but falling to 10^{-10} g when great care was taken to avoid lateral motion.

Are such transfer fragments consistent with an atom by atom transfer process?

We note with interest Rabinowicz (1953) “A quantitative study of the wear process” –which is entirely about transfer fragments, with no mention of wear fragments! However, following Kerridge (1955) and Kerridge and Lancaster (1956), Rabinowicz (1958) recognized the distinction between transfer and wear, and propounded a criterion for the size of wear fragments (The AEI publications contain estimates of the sizes of transfer and wear fragments, but make no attempt to predict them). Rabinowicz starts from the basic Bowden and Tabor idea of a lump of metal torn out of the base and crushed against the slider so that it adheres and becomes a transfer fragment. When the crushing load is removed, the fragment relaxes, but remains stretched because it adheres to the slider, and residual stresses

remain. A fragment of volume V with mean residual stresses $\bar{\sigma}_r$ will have a strain energy $(\bar{\sigma}_r^2/2E) \cdot V$. Rabinowicz explain that to separate the transfer particle from the slider involves the creation of new free surface, so for an area A and a work of adhesion w_{ab} for the pair of materials will require an energy $A(w_{ab})$: the first introduction of surface energy into Tribology⁵. Thus, the minimum size of loose wear fragments will be when $(\bar{\sigma}_r^2/2E) \cdot V = A(w_{ab})$. Accordingly, for a hemispherical fragment of radius r , we must have $r \geq (3E/\bar{\sigma}_r^2) \cdot (w_{ab})$.

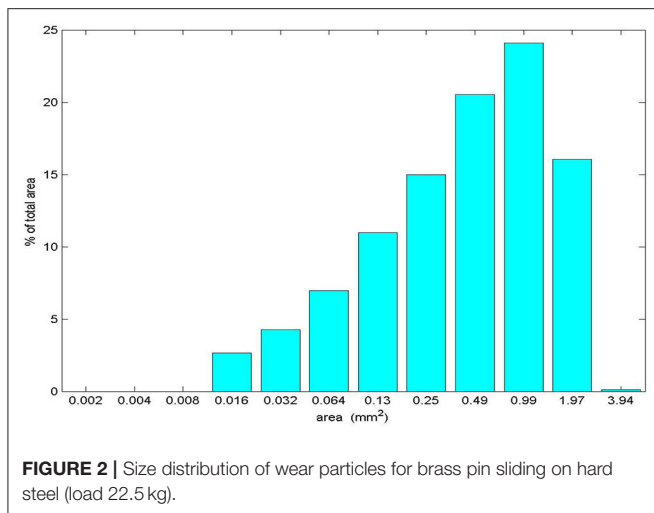
The maximum possible strain energy density is $Y^2/2E$ where Y is the yield stress in tension, and clearly the mean residual stress after unloading will be less. Rabinowicz (1958) suggests that the strain energy might be perhaps 10% of the maximum possible [In his 1961 paper, Rabinowicz (1961) suggests residual stresses are often found to be $\sigma_r = \nu Y$, which when Poisson’s ratio $\nu = 0.3$ gives 9%]⁶. He goes on to argue that for most metals the elastic strain at yield may be taken as $3 \cdot 10^{-3}$, so sets $Y/E = 3 \cdot 10^{-3}$, and takes $Y = p_m/3$ where p_m is the hardness, so a simplified, practical criterion is $r \geq 30,000 \cdot (w_{ab})/p_m$. But in wear tests to confirm this, he switches from the predicted *minimum* fragment size to the “average” size (total weight of larger particles equals total weight of smaller ones): despite some scatter, the agreement is reasonable. But what then do we deduce about the origin of the multitude of smaller fragments contributing half the total weight? The size distribution of the wear fragments is not given, but the Kerridge and Lancaster paper described above, and **Figure 2** shows their results.

There is no evidence *in these experiments* of a minimum size for wear particles. And it should be noted that in experiments in which active and inactive pins were interchanged, individual wear fragments had different specific activities: i.e., each particle contained a mixture of active and inactive atoms. No particle has a life history like that postulated in the Rabinowicz model.

⁵Or should one count Derjaguin’s (1934) estimate of the force of adhesion?

⁶I find these estimates hard to accept: stresses of order Y may well be found near the interface, but will these not be very local?

⁴But sliding was not completely steady, with some stick-slip. Did this matter?



DISCUSSION

We all know that the underlying problem in studying wear is the multiplicity of wear mechanisms. Even when considering a given materials pair, as the mild wear of hard steel against itself as studied by Archard and Hirst (above) or gold against platinum as described by Antler, there was a sequence of different mechanisms. Challen and Oxley (1979) used a slip-line field solution to model the steady motion when a hard slider traversed a rigid/perfectly plastic half-space, and found three modes of deformation; (1) plastic flow of the surface by a wave pushed ahead of the slider, referred to by Challen and Oxley as “rubbing mode”; (2) a deformed “prow” which becomes detached from the surface; (3) a cutting mode in which a “chip” is continuously cut from the surface. Kayaba et al. (1986) refer to all three as abrasive wear: the rubbing mode, rather inappropriately, as plowing (though they do sometimes find a gentle trace). They experimented with a hard steel pin against brass, mild steel and stainless steel. All three combinations, when unlubricated, show each of the three modes, as the parameter $\sqrt{W/R^2H_V}$ (where R is the pin radius, H_V the hardness, and W the load) increases. They show scanning electron micrographs of prow formation (with no reference to Cocks or Antler), and similar pictures for steel against brass, where the prow material passes under the slider and becomes a chip. When lubricated with a smear of silicone grease, with the two steels all three modes are found: however lubricated brass gave just two modes: the rubbing mode at higher loads but changing at light loads to *flaking*: the flakes then forming wear debris. This appears to be the delamination wear mode studied by Suh (1973), and caused by incremental plastic flow as the Challen and Oxley plastic wave moves over the surface and produces below the surface an *almost* reversible plastic strain cycle. I cannot resist including his conclusion (Suh, 1977) “(2) *The wear rate of metals may be predicted in the near future, based on first principles and fundamental material properties.*”

And of course these are just a few wear mechanisms: one should certainly add metallographic phase transformations [Welsh (1965) and his demonstration that increasing the load or speed not only changes wear from mild to severe, but, because of the frictional temperature rise, there are “three transitions: (1) a change from mild to severe wear at relatively light loads (T_1): (2) a change from severe wear back to mild wear at higher loads, (T_2): (3) a perturbation in the mild-wear rate at even higher loads, (T_3) with the wear rates of the pin and ring diverging”].

Others emphasize the development of subsurface cracks, or tribochemistry, or simple corrosion.

I have confined my attention to the largely forgotten classic contributions.

For they are largely forgotten. Exceptionally, deRoos and Schipper (2001) and deRoos et al. (2013) know of Cocks and Antler’s work, so are able to build on it in their investigation of how hard transfer fragments build up on a deep-drawing tool and ruin the finish. In contrast, a recent admirable paper (Tarasov et al., 2017) studies the sliding of a hard steel pin on an aluminum disk. Their interest is in Friction Stir Welding, so they contribute new information about the effect of temperature on the iron/aluminum combination: but they rediscover prow formation, the formation of a transfer layer, and back transfer to the disk—all without noticing that it has been discovered already 60 years ago. But no radiography, so they do not learn that transfer particles are formed by accretion over time, not torn out bodily as lumps.

Is it all right to forget all this, because prows and riders are just examples of Godet’s “Third Body,” so need not be treated separately? There seems no doubt that Antler’s “riders” are indeed third bodies. It is less clear that Cocks’ prows are: certainly sometimes they are the very opposite: they are the link which makes the first and second bodies into one! There are not two bodies sliding past each other: there is a single body shearing along a neck. It seems clear that the formation of transfer layers is not a three body process: and the transfer layer is not itself a third body any more than the oxide film can usefully be treated as such: perhaps there has been insufficient study of the five-body problem, where debris particles on their way out of the conjunction roll about between two oxide films?

Yes, study the behavior of trapped particles: but if all transfer and wear is attributed to third bodies, the term has become too wide to be useful. But at least stop studying the purely normal contact of rough surfaces.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research has received no funding. I hope, and trust, that the University of Cambridge will provide the open access publication fee.

REFERENCES

- Antler, M. (1962). Wear, friction, and electrical noise phenomena in severe sliding systems. *ASLE Trans.* 5, 297–307. doi: 10.1080/05698196208972473
- Antler, M. (1964). Processes of metal transfer and wear. *Wear* 7, 181–203. doi: 10.1016/0043-1648(64)90053-5
- Archard, J. F. (1953). Contact and rubbing of flat surfaces. *J. Appl. Phys.* 24, 981–988. doi: 10.1063/1.1721448
- Archard, J. F., and Hirst, W. (1957). An examination of a mild wear process. *Proc. R. Soc. A* 238, 515–528. doi: 10.1098/rspa.1957.0015
- Burwell, J. T., and Strang, D. C. (1952). On the empirical law of adhesive wear. *J. Appl. Phys.* 23, 18–28. doi: 10.1063/1.1701970
- Cann, P., Ioannides, E., Jacobson, B., and Lubrecht, A. A. (1994). The lambda ratio—a critical re-examination. *Wear* 175, 177–188. doi: 10.1016/0043-1648(94)90181-3
- Challen, J. M., and Oxley, P. L. B. (1979). An explanation of different regimes of friction and wear using asperity deformation models. *Wear* 53, 229–243. doi: 10.1016/0043-1648(79)90080-2
- Cocks, M. (1958). Wear debris in contact between sliding metals. *J. Appl. Phys.* 29, 1609–1610. doi: 10.1063/1.1723007
- Cocks, M. (1962). Interaction of sliding metal surfaces. *J. Appl. Phys.* 33, 2152–2161. doi: 10.1063/1.1728920
- Cocks, M. (1964). Role of displaced metal in the sliding of flat metal surfaces. *J. Appl. Phys.* 35, 1807–1814. doi: 10.1063/1.1713746
- Derjaguin, B. V. (1934). Theorie des Anhaftens kleiner teilchen. *Kolloid Zeitschrift*. 69, 155–164. doi: 10.1007/BF01433225
- deRooij, M. B., and Schipper, D. J. (2001). Analysis of material transfer from a soft workpiece to a hard tool: part I—lump growth model: part II: experimental verification: *ASME J. Tribol.* 123, 474–478. doi: 10.1115/1.1308023
- deRooij, M. B., van der Linde, G., and Schipper, D. J. (2013). Modelling material transfer on a single asperity scale. *Wear* 307, 198–208. doi: 10.1016/j.wear.2013.09.006
- Greenwood, J. A. (2020). “Elastohydrodynamic Lubrication,” in *Encyclopedia of Tribology*, eds Q. J. Wang and Y. W. Chung (Boston, MA: Springer), 51–80.
- Gregory, J. N. (1946). Radioactive tracers in the study of friction and lubrication. *Nature* 157, 443–444. doi: 10.1038/157443b0
- Holm, R. (1938). Wiss. Veroffent. Siemens-Werken 17:43.
- Kayaba, T., Hokkirigawa, K., and Kato, K. (1986). Analysis of the abrasive wear mechanism by successive observations of wear processes in a scanning electron microscope. *Wear* 110, 419–430. doi: 10.1016/0043-1648(86)90115-8
- Kerridge, M. (1955). Metal transfer and the wear process. *Proc. Phys. Soc. B* 68, 400–407. doi: 10.1088/0370-1301/68/7/302
- Kerridge, M., and Lancaster, J. K. (1956). The stages in a process of severe metallic wear. *Proc. R. Soc. A* 236, 250–264. doi: 10.1098/rspa.1956.0133
- Rabinowicz, E. (1952). Metal transfer during static loading and impacting. *Proc. Phys. Soc. B*. 65:630.
- Rabinowicz, E. (1953). On the looseness of wear fragments. *J. Appl. Phys.* 24:367. doi: 10.1063/1.1721285
- Rabinowicz, E. (1958). The effect of size on the looseness of wear fragments. *Wear* 2, 4–8. doi: 10.1016/0043-1648(58)90335-1
- Rabinowicz, E. (1961). Influence of surface energy on friction and wear phenomena. *J. Appl. Phys.* 32, 1440–1444. doi: 10.1063/1.1728375
- Rabinowicz, E., and Tabor, D. (1951). Metallic transfer between sliding metals: an autoradiographic study. *Proc. R. Soc. Lond. A* 208, 455–475. doi: 10.1098/rspa.1951.0174
- Sakmann, B. W., Burwell, J. T., and Irvine, W. J. (1944). Measurements of the adhesion component in friction by means of radioactive indicators. *J. Appl. Phys.* 15, 459–473. doi: 10.1063/1.1707457
- Suh, N. P. (1973). The delamination theory of wear. *Wear* 25, 111–124.
- Suh, N. P. (1977). An overview of the delamination theory of wear. *Wear* 44, 1–16. doi: 10.1016/0043-1648(77)90081-3
- Tarasov, S. Y., Filippov, A. V., Kolubaev, E. A., and Kalashnikova, T. A. (2017). Adhesion transfer in sliding a steel ball against an aluminum alloy. *Tribol. Int.* 115, 191–198. doi: 10.1016/j.triboint.2017.05.039
- Welsh, N. C. (1965). The dry wear of steels part I. The general pattern of behaviour. *Phil. Trans. R. Soc. A* 257, 31–70. doi: 10.1098/rsta.1965.0001

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Greenwood. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comparative Analysis of Error Sources in the Determination of Wear Volumes of Oscillating Ball-on-Plane Tests

Manuel Reichelt and Brunero Cappella*

Federal Institute for Material Research and Testing (BAM), Berlin, Germany

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Andrey I. Dmitriev,
Institute of Strength Physics and
Materials Science (ISPM SB
RAS), Russia
Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine

*Correspondence:

Brunero Cappella
brunero.cappella@bam.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 25 February 2020

Accepted: 15 April 2020

Published: 08 May 2020

Citation:

Reichelt M and Cappella B (2020)
Comparative Analysis of Error Sources
in the Determination of Wear Volumes
of Oscillating Ball-on-Plane Tests.
Front. Mech. Eng. 6:25.
doi: 10.3389/fmech.2020.00025

The accurate determination of wear volumes is a prerequisite for the study of numerous tribological phenomena. Wear volumes can be measured with different techniques (profilometry, confocal microscopy, white light interferometry, atomic force microscopy) or else be calculated starting from some quantities (usually the width and the planimetric wear) measured from the wear scar. Advantages and drawbacks of the mentioned measuring techniques are shown by means of wear scars and calottes resulting from ball-on-plane tests with 100Cr6 specimens. When measuring wear volumes, white light interferometry results to be one of the most suitable techniques, since it offers high accuracy and is not as time consuming as atomic force microscopy. When wear volumes are calculated, errors result mainly from two sources: (1) the arbitrary choice of one or few line profiles for the determination of the width and of the planimetric wear, and (2) approximations in the calculation, which are even necessary when values of the wear volumes of the single tribological partners, i.e., ball and plane, and not only the total volume, are of interest. The effect of both the statistical distribution of values of the width and of the planimetric wear and the propagation of errors due to approximations on the accuracy in the determination of wear volumes is characterized and elucidated by examples. It is found that errors due to approximations are negligible when compared to errors due to the arbitrary choice of one line profile.

Keywords: error sources analysis, AFM, white light interferometry, statistical analysis, planimetric wear, volumetric wear, oscillating ball-on-disc test

INTRODUCTION

The variation in volumetric wear data from tribological tests is quite large. Volumetric wear is a value influenced by the whole tribological system (Meng and Ludema, 1995) and its variation affects seriously the repeatability and reproducibility of tribological measurements and hence the determination of wear coefficients k . Many properties of the system are unknown and therefore not predictable. The variation of the material of the specimens due to production process, grinding, polishing or heat treatment, and properties of tribometers such as stiffness of the mechanical elements or the accuracy of sensors are only few examples of parameters influencing the volumetric wear, the wear coefficient and/or their determination. Through additional sources of error in the determination of the volumetric wear, the chance to identify error sources in the system and to

recognize correlations with experimental parameters is further reduced (Reichelt and Cappella, 2020). Therefore, it is essential to determine and to minimize all possible sources of error in the method for the determination of the volumetric wear.

The methods for the determination of the volumetric wear have changed due to novel capabilities of microscopy tools and technologies. In the past, tactile devices were employed to measure individual lines perpendicular to the sliding direction, which in turn are used to calculate the wear volume. Rarely the entire friction track was recorded using many profile lines. Rather, the volume was calculated analytically from one or some of these lines. Today, white light interferometers (WLI), laser scanning systems, confocal microscopes, atomic force microscopes (AFM) etc. offer together with their analysis software a variety of possibilities to determine the wear volume more precisely within a reasonable measuring time.

Tactile profilometers (Yost, 1983; Wehbi et al., 1986; Kalin and Vizentin, 2000) for measuring roughness have been on the market since the 1930s. A diamond tip serves as a scanning probe. It is brought into contact with the surface and moved over the surface under contact forces that do not damage or even deform the sample. The contact force is kept constant by controlling the deflection of the tip. The electrical voltage required for the deflection of the tip is recorded and converted into the corresponding height profile. Distances of a few centimeters can be recorded with height differences of a few hundred micrometers. 3D measurements are extremely time-consuming due to the low scanning speed, so that usually just profile lines are recorded.

Non-contact optical measuring instruments, which have been available only since the 1980s and have been providing scientifically usable results since the 1990s through standardization, including calibration, are more suitable for fast surface measurements. White light interferometer microscopes (WLI) (Guilemany et al., 2001; Cox, 2006) are widely used in tribology. In these microscopes the interference of broadband white light is used to measure the topography of surfaces. The resolution of this optical method is better than that of profilometers (some hundreds of nanometers instead of 1 μm). However, at maximum resolution, the measuring field is limited to a few hundred micrometers. This increases the measuring time and requires stitching (the combination of multiple images to form an overall topography image), which is subject to errors. WLI is an indirect measurement of the height.

Atomic force microscopy (AFM) (Bhushan, 2001, 2003; Yu et al., 2009) is the method with the best resolution and the lowest susceptibility to errors, since it is a direct measurement of heights. A deflectable cantilever with a tip at its end serves as a scanning probe. The radius of the tip is usually in the range 5–20 nm. A laser beam is focused on the cantilever and the reflected beam changes its angle when the cantilever deflects. The deflection is thus detected by a four-quadrant photodiode. The position of the sample in the z -direction is controlled by means of piezos to ensure a constant deflection of the cantilever. The piezo voltage required for this purpose is converted to the corresponding height z . AFM has a sub-nanometer resolution, both laterally and vertically. The major disadvantage of this measuring method

is the very small measuring range both laterally (some tens of microns) and in height (some microns). By time-consuming stitching these ranges can be theoretically many times larger. However, the gained accuracy does not justify in most cases the very long measuring time.

In the present article, the method for the calculation of volumetric wear in oscillating sliding tests with a ball-on-plane configuration is described and errors arising from approximations are analyzed. Furthermore, this method is exemplarily applied to a tribological test result without significant anomalies on the wear track and compared with WLI measurements. Also, the values of volumes determined through WLI is compared with AFM measurements. In this way, both errors due to measurement techniques and to calculation methods can be estimated.

RESULTS

Calculation of Volumetric Wear

The total wear volume W_v is the sum of the wear volume of the ball, $W_{v,ball}$, and that of the plane, $W_{v,flat}$:

$$W_v = W_{v,flat} + W_{v,ball}. \quad (1)$$

When the direct measurement of the volumes is not possible, they must be calculated using the ball radius R , the planimetric wear W_q , and the track width d_{\perp} , determined through profile measurements. This method (W_q method) is exposed in the following. The analysis of the procedure for the calculation of the wear volumes is important also for comparison of experiments in which volumes have partly been measured and partly calculated, as it is usually the case for old experiments.

In the middle of the wear track of the plane, a profile line perpendicular to the sliding direction is plotted, from which the track width d_{\perp} and the planimetric wear W_q can be determined (Figure 1). Both d_{\perp} and W_q depend strongly on the roughness of the unworn surface and on the roughness of the scar, which influence significantly the accuracy in the determination of the zero value of the height and hence the identification of scar borders. Errors resulting from this issue may vary considerably due to automated analysis or to individual discretion.

In the following, Figure 2 is used to show how R' of a wear scar is calculated.

From the geometry shown in Figure 2 follows:

$$W_q = R'^2 \arcsin\left(\frac{d_{\perp}}{2R'}\right) - \frac{1}{2}R'd_{\perp}\sqrt{1 - \left(\frac{d_{\perp}}{2R'}\right)^2}. \quad (2)$$

If $d_{\perp} \ll R'$, the following approximations result from Taylor developments:

$$\arcsin\left(\frac{d_{\perp}}{2R'}\right) \cong \frac{d_{\perp}}{2R'} + \frac{1}{6}\left(\frac{d_{\perp}}{2R'}\right)^3 \quad (3)$$

and

$$\sqrt{1 - \left(\frac{d_{\perp}}{2R'}\right)^2} \cong 1 - \frac{1}{2}\left(\frac{d_{\perp}}{2R'}\right)^2. \quad (4)$$

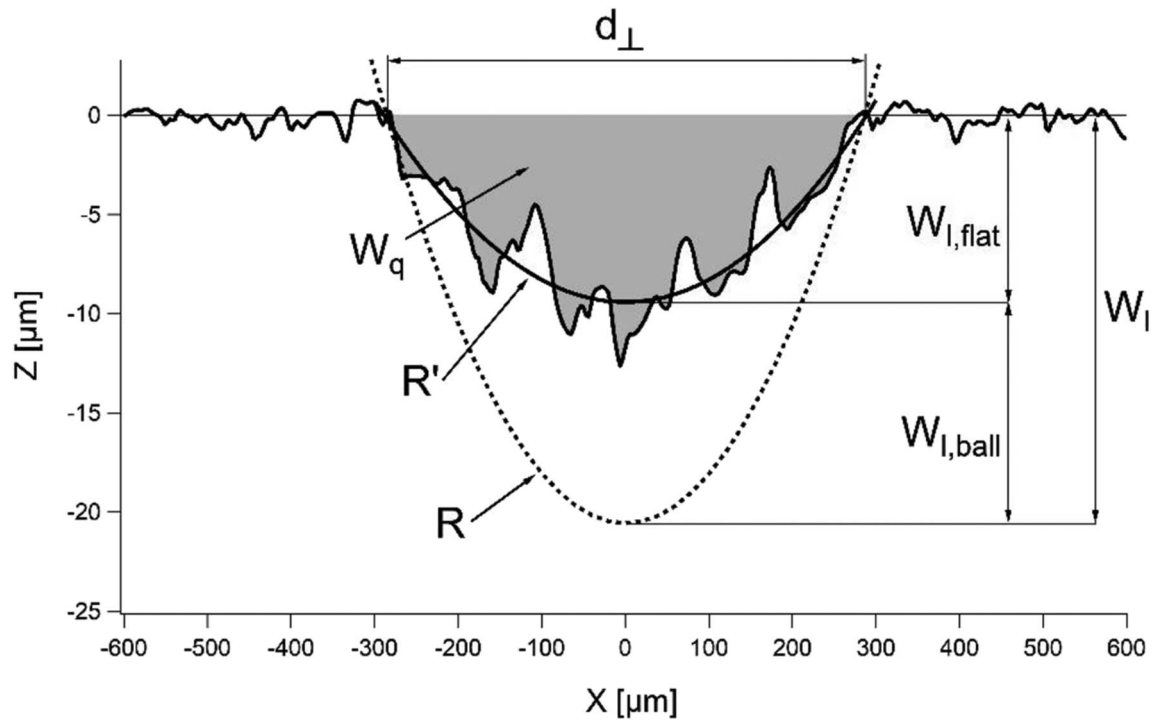


FIGURE 1 | Line profile of a wear track on a 100Cr6 plane taken from a WLI measurement. The picture shows the track width d_{\perp} , the planimetric wear W_q (cross-sectional area of the wear track, filled in gray), the fitted circumference with radius of curvature R' , the circumference with known sphere radius R , the total linear wear W_l , and the linear wear of ball and plane, $W_{l,ball}$ and $W_{l,flat}$.

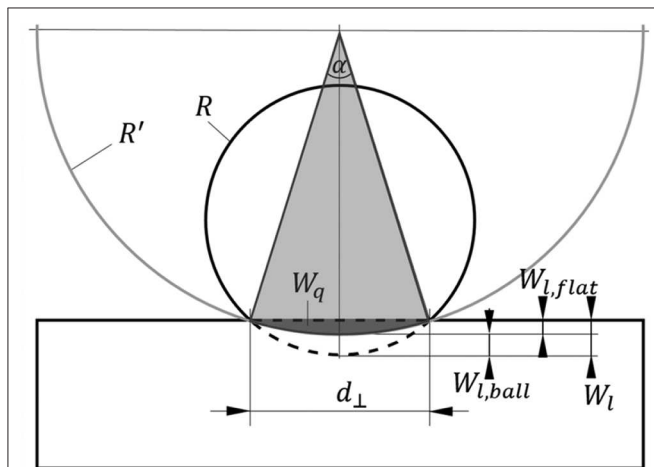


FIGURE 2 | Schematic representation of the geometry of the worn sphere and plane normal to the sliding direction. The figure shows the quantities used to calculate the radius of the curved worn surface R' in the friction test: radius of the ball R , wear track width d_{\perp} , angle of the circular sector α , planimetric wear W_q , total linear wear W_l , and linear wear of ball and plane, $W_{l,ball}$ and $W_{l,flat}$. The original surface profile lines of the sphere and plane are shown as dashed lines.

Substituting in Equation (2), we get:

$$W_q \cong \frac{d_{\perp}^3}{12R'} \quad (5)$$

If the track width d_{\perp} and the planimetric wear W_q are known, the radius R' of the wear track can be calculated:

$$R' \cong R_{app}' = \frac{d_{\perp}^3}{12W_q} \quad (6)$$

Hence, the wear volume of the plane is given by:

$$\begin{aligned} W_{v,flat} &= \frac{1}{3} \pi W_{l,flat}^2 (3R' - W_{l,flat}) + W_q \Delta x \\ &\cong \frac{\pi d_{\perp}^4}{64R'} + W_q \Delta x, \end{aligned} \quad (7)$$

where $W_{l,flat}$ is the linear wear of the plane and Δx the stroke. Using the total linear wear W_l , the wear volume of the ball can be written as:

$$\begin{aligned} W_{v,ball} &= \frac{1}{3} \pi W_l^2 (3R - W_l) - \frac{1}{3} \pi W_{l,flat}^2 (3R' - W_{l,flat}) \\ &= \frac{1}{3} \pi \left(\frac{d_{\perp}^2}{8R} \right)^2 \left(3R - \frac{d_{\perp}^2}{8R} \right) - \frac{1}{3} \pi \left(\frac{d_{\perp}^2}{8R'} \right)^2 \\ &\quad \left(3R' - \frac{d_{\perp}^2}{8R'} \right) \cong \pi \frac{d_{\perp}^4}{64} \left(\frac{1}{R} - \frac{1}{R'} \right), \end{aligned} \quad (8)$$

where the higher terms proportional to d_{\perp}^6 have been neglected both in Equations (7) and (8).

The total wear volume W_v can be calculated without approximations and is given by:

$$W_v = \pi R^3 \left(1 - \sqrt{1 - \left(\frac{d_{\perp}}{2R} \right)^2} \right)^2 - \frac{1}{3} \pi R^3 \left(1 - \sqrt{1 - \left(\frac{d_{\perp}}{2R} \right)^2} \right)^3 + W_q \Delta x. \quad (9)$$

The approximation $\left(1 - \sqrt{1 - \left(\frac{d_{\perp}}{2R} \right)^2} \right)^3 \cong 0$ yields a slightly larger value of W_v :

$$W_v \cong \pi R^3 \left(1 - \sqrt{1 - \left(\frac{d_{\perp}}{2R} \right)^2} \right)^2 + W_q \Delta x. \quad (10)$$

On the contrary, the approximation $\sqrt{1 - \left(\frac{d_{\perp}}{2R} \right)^2} \cong 1 - \frac{1}{2} \left(\frac{d_{\perp}}{2R} \right)^2$ reduces W_v :

$$W_v \cong \frac{1}{4} \pi R^3 \left(\frac{d_{\perp}}{2R} \right)^4 \left(1 - \frac{1}{6} \left(\frac{d_{\perp}}{2R} \right)^2 \right) + W_q \Delta x. \quad (11)$$

Using both approximations, we get:

$$W_v \cong \pi \frac{d_{\perp}^4}{64R} + W_q \Delta x. \quad (12)$$

In all cases in which the length of the calotte is different from its width ($d_{\parallel} \neq d_{\perp}$), d_{\perp} must be replaced by $\sqrt{d_{\parallel} d_{\perp}}$ [e.g., in Equations (7), (8), and (12)].

It is important to remember that the error in the determination of R' does not affect the calculation of the total wear volume W_v [see Equation (9)]. Nevertheless, it affects indeed the calculation of the wear volumes of the single partners [see Equations (7) and (8)]. Therefore, it is important to analyze the error done through the approximations in Equations (3) and (4).

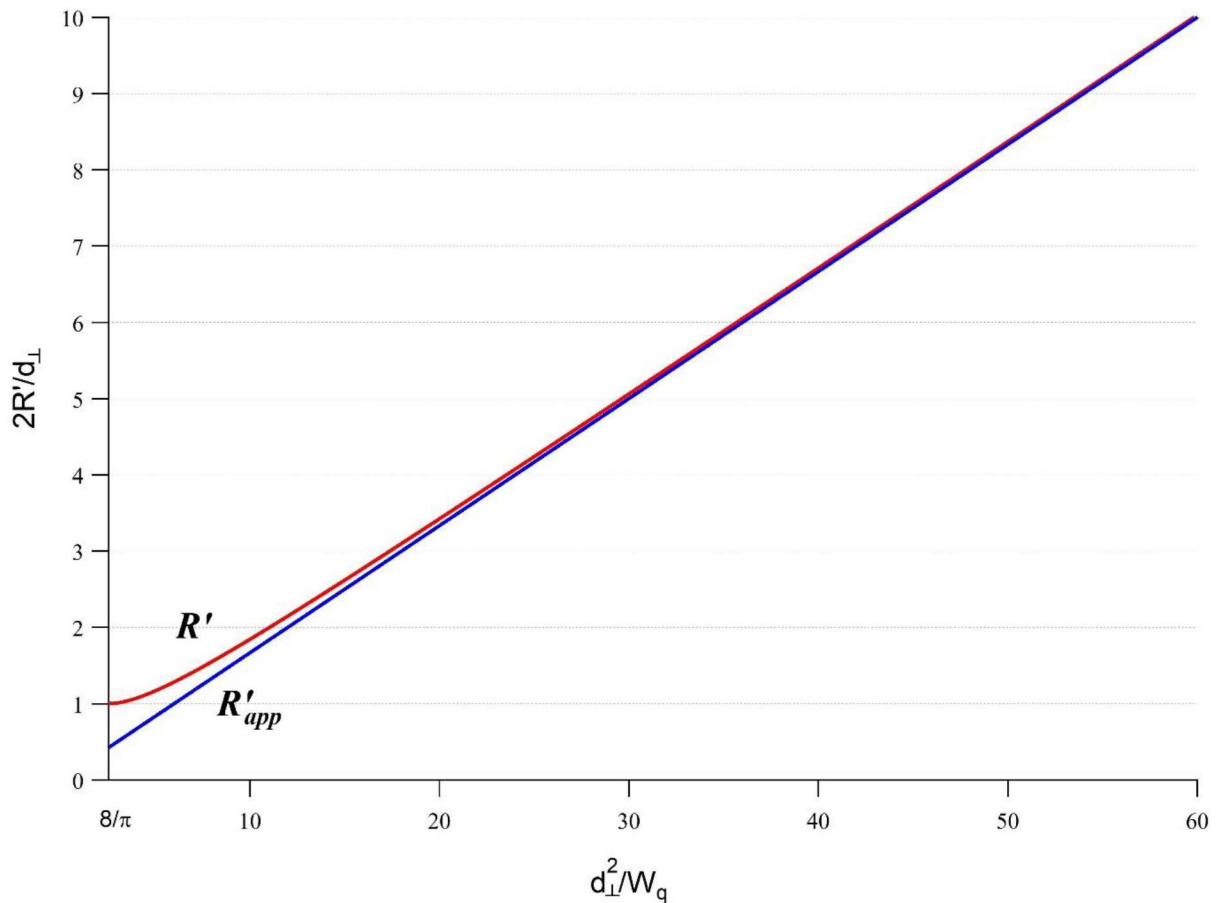


FIGURE 3 | Comparison of the exact numerical solution of Equation (2) (upper curve, red) with the approximated solution (lower curve, blue) of the calculation of R' from Equation (6). The diagram shows the ratio of twice the radius R' to track width d_{\perp} over the ratio of the square of the track width d_{\perp}^2 to the planimetric wear W_q .

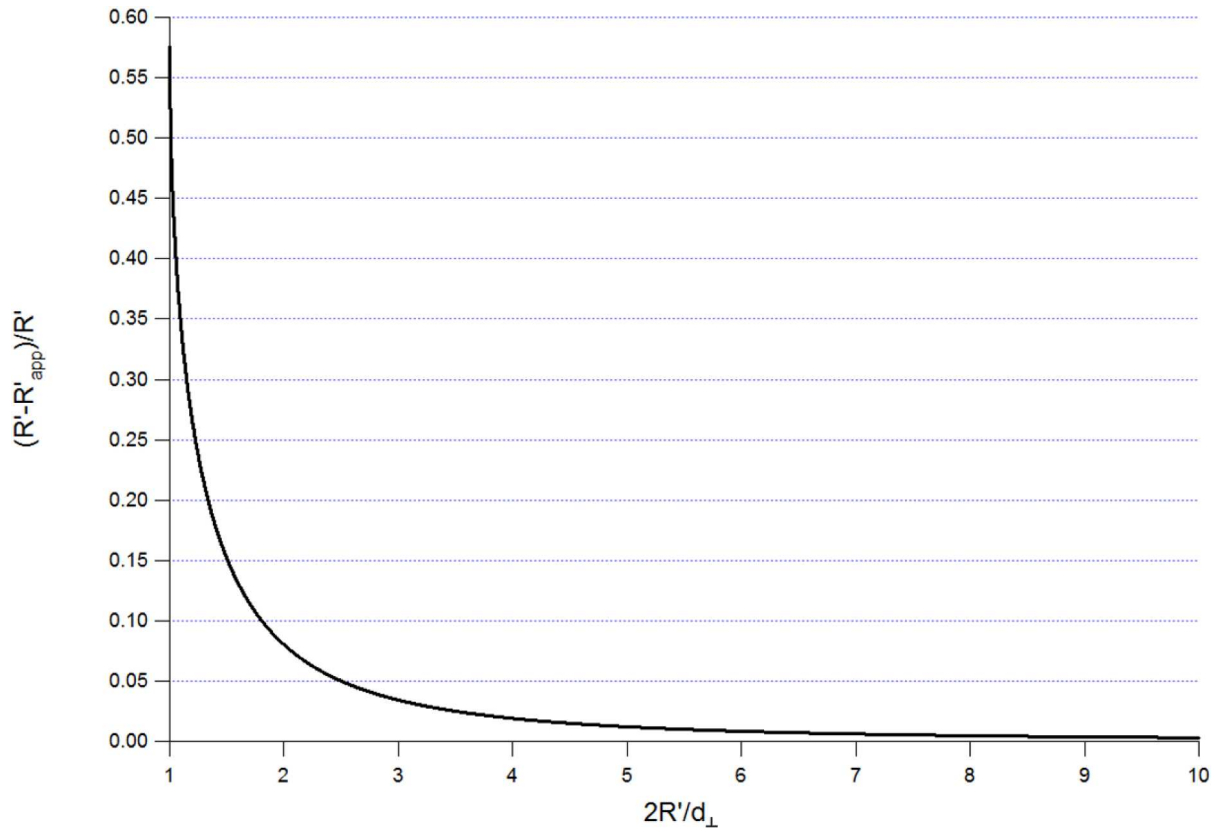


FIGURE 4 | Relative error in the calculation of R' by using the approximation in Equation (6) over the ratio of twice the radius of curvature R' to the track width d_{\perp} .

The error due to both approximations is shown in **Figure 3** by comparing R'_{app} [Equation (6)] and the numerical solution of Equation (2) for R' . The approximation always leads to an underestimation of R' ($R'_{app} < R'$).

The error of the approximation of R' reaches a maximum when the ball does not wear and the plane wears maximally with $R' = R = W_l = d_{\perp}/2$. The area W_q in this case is half the area of a circle with diameter d_{\perp} , $\pi d_{\perp}^2/8$, and the approximation in Equation (6) yields $R'_{app} \cong 2d_{\perp}/3\pi = 4R/3\pi = 0.424R$. The maximum deviation $(R' - R'_{app})/R'$ in this extreme case amounts to $1 - 4/3\pi \cong 57.6\%$ and can also be seen in **Figure 4**, which shows the relative error of the approximate equation. One can see that the error is $<10\%$ if $2R'/d_{\perp} > 1.8$ (i.e., $d_{\perp} < 1.1R'$); for $2R'/d_{\perp} > 2.485$ (i.e., $d_{\perp} < 0.8R'$) the error is even $<5\%$.

As concerns the propagation of the error on R' in the functions expressing the volumes, it is easy to show that $\partial W_{v,flat}(\partial R') = -\partial W_{v,ball}(\partial R')$, since W_v does not depend on R' .

Further, by neglecting the terms proportional to d_{\perp}^6 as in Equations (7) and (8), we get:

$$\frac{\partial (W_{v,flat} - W_q \Delta x)}{W_{v,flat} - W_q \Delta x} = \left(1 - \frac{2}{\sqrt{1 - \left(\frac{d_{\perp}}{2R'}\right)^2}} \right) \frac{\partial R'}{R'}. \quad (13)$$

Hence, the error on $W_{v,flat}$ due to the error on R' is always negative and is between $-\partial R'/R'$ and $-2\partial R'/R'$ for $0 < d_{\perp} < 1.5R'$.

Comparison of AFM and WLI Measurements

In order to investigate the accuracy of the WLI images, comparative measurements were carried out after a tribological test by means of stitched AFM contact measurements on the surfaces of the 100Cr6 plane and the 100Cr6 sphere ($R = 2$ mm). WLI images were acquired with a NewView 5022 (Zygo, Middlefield, Connecticut). For AFM measurements, a Ciphcr (Asylum Research, Santa Barbara, California) was used with a maximum lateral scanning range of $30 \mu\text{m}$ and a vertical range of $3 \mu\text{m}$. The AFM tip had a radius of $\sim 15\text{--}20$ nm.

The WLI image of the wear track of the plane is shown in **Figure 5**.

A series of AFM topographies was acquired on the sphere and on the plane transverse to the sliding direction (26 and 28 images, respectively), and the images were then combined into two single images. These images, together with the profile lines extracted from them, are shown in **Figure 6**. They cover only a small part of the WLI images. Since the topography of the calotte perpendicular to the sliding direction is quite uniform, one line profile is representative of the whole scanned area, whereas five

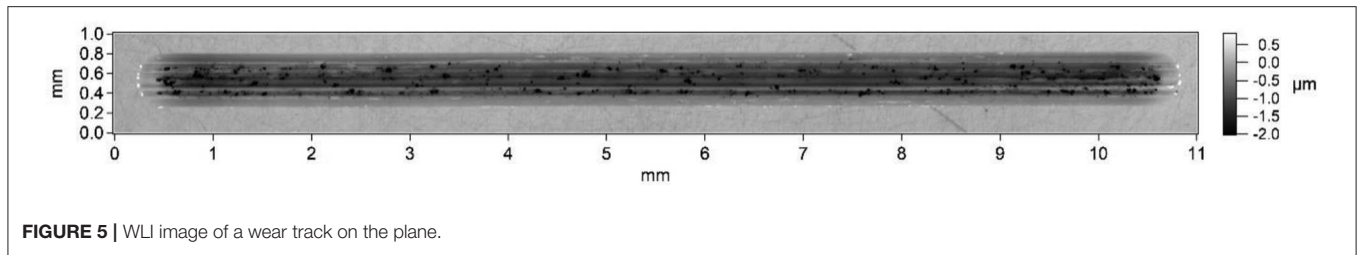


FIGURE 5 | WLI image of a wear track on the plane.

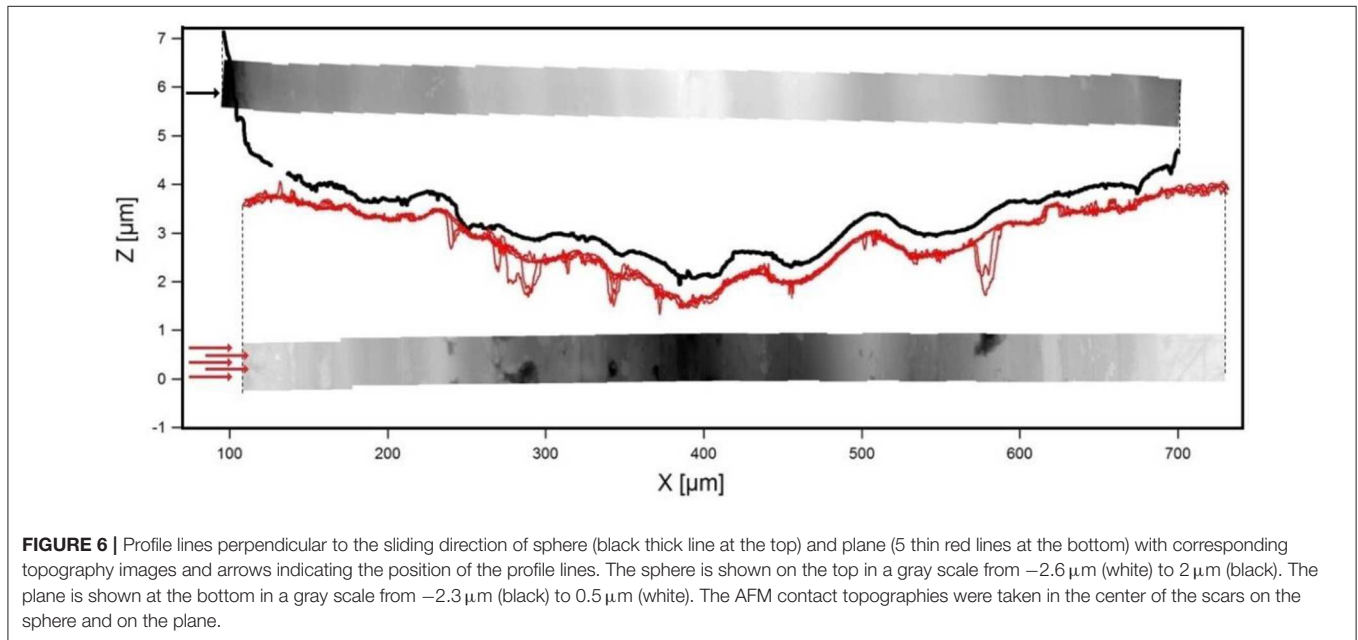


FIGURE 6 | Profile lines perpendicular to the sliding direction of sphere (black thick line at the top) and plane (5 thin red lines at the bottom) with corresponding topography images and arrows indicating the position of the profile lines. The sphere is shown on the top in a gray scale from $-2.6 \mu\text{m}$ (white) to $2 \mu\text{m}$ (black). The plane is shown at the bottom in a gray scale from $-2.3 \mu\text{m}$ (black) to $0.5 \mu\text{m}$ (white). The AFM contact topographies were taken in the center of the scars on the sphere and on the plane.

lines are needed for the plane. For clarity, the profile lines of the cross-sections of the sphere and plane are shown with a vertical offset of $\sim 400 \text{ nm}$. The profiles match very well. Apart from a few defects on the plane and small deviations on the spherical calotte, the curves deviate very little from each other along the x -axis. This shows that AFM should be used for the characterization of the worn surfaces whenever the fine structure of the tracks and calottes is relevant (Wäsche et al., 2014).

A comparison of the white light measurement data with those of the AFM is shown using the example of the calotte on a sphere in **Figure 7**. The region scanned with AFM is shown in copper color scale. By superimposing it to the gray scale WLI image, a calibration error of the white light interferometer was found, which made it necessary to rescale the lateral dimensions of the WLI image by ca. 1%. After this rescaling, the volume measured through WLI differs by $<0.5\%$ from the volume determined by AFM. This shows that the WLI measurement data have a high accuracy. Similar results could already be shown by Wäsche et al. (2014).

It can be concluded that the determination of wear volumes by white light interferometry is recommendable. On the contrary, the use of AFM stitching is usually too time consuming. Yet, as shown in previous works (Wäsche et al., 2014; Cappella et al., 2015), AFM measurements are necessary when wear volumes are

very small, i.e., linear wear is in the range of roughness asperities, or when the fine structure of the tracks is of interest.

Comparison of Measured and Calculated Wear

The statistical distributions of W_q and d_{\perp} are useful to assess the influence of a randomly selected profile line of the wear track of a plane on the wear indicators calculated from it. The WLI measurement in **Figure 5** is used as an example for this in the following. The analysis of all 3350 profile lines along Δx excluding the edge regions results in average values of $559 \pm 16 \mu\text{m}$ for d_{\perp} , $452 \pm 36 \mu\text{m}^2$ for W_q and $1135 \pm 65 \text{ nm}$ for $W_{l,flat}$. As already stated at the beginning of section Calculation of volumetric wear, these values are strongly influenced by the roughness of the specimens; yet this influence in the determination of e.g. d_{\perp} cannot be quantified. Using the maxima and minima of these distributions, the resulting mean values as well as the deviations for the exact solution of W_v according to Equation (9) and those for the approximated solutions according to Equations (10), (11), and (12) were calculated. The same procedure was used to calculate the ranges for $W_{v,flat}$ and $W_{v,ball}$ according to Equations (7) and (8) and for the linear wear values. $W_{l,flat}$ and $W_{l,ball}$ were determined

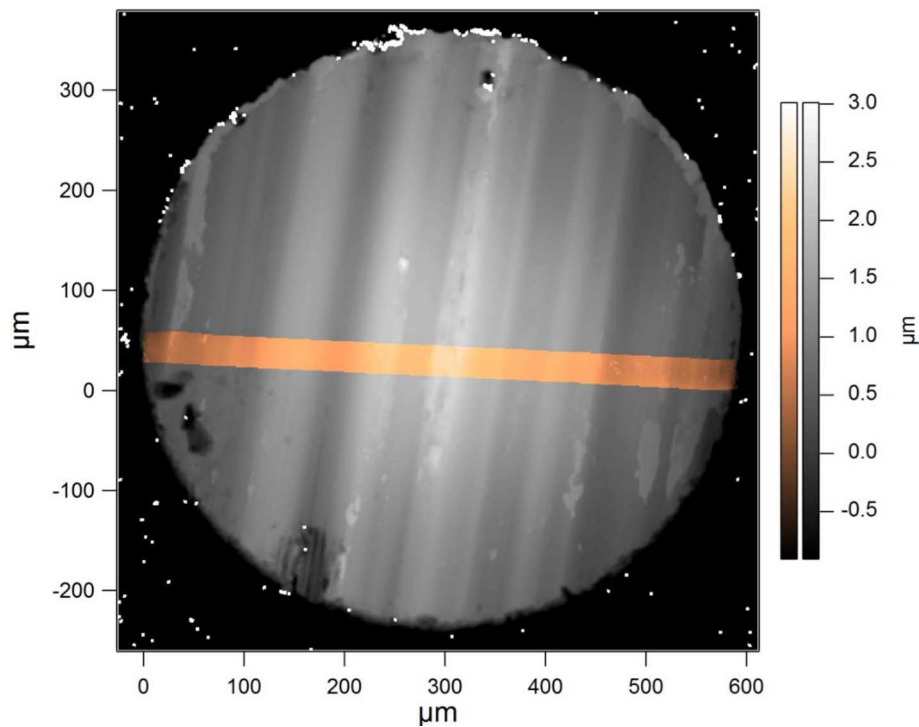


FIGURE 7 | Superposition of two topographic images of a sphere wear calotte ($R = 2$ mm): total image in gray color scale acquired with a white light interferometer, section in copper color scale stitched from 26 AFM contact images.

through the following equations:

$$W_{l,flat} \cong \frac{3W_q}{2d_{\perp}}, \quad (14)$$

$$W_{l,ball} = R - \sqrt{R^2 - \left(\frac{d_{\perp}}{2}\right)^2} - W_{l,flat}. \quad (15)$$

All values are listed in **Table 1**. Since the WLI results have been verified through AFM measurements, the uncertainty of the measured values is well below 10^{-6} mm^3 .

The mean values of the results of the three approximated equations for W_v deviate by a maximum of $24 \cdot 10^{-6} \text{ mm}^3$ from the exact solution given by Equation (9), corresponding to 0.3% of the calculated volume. The standard deviation of the exact solution is $637 \cdot 10^{-6} \text{ mm}^3$, which, with a factor of ~ 27 , is considerably higher than the error due to approximations.

As shown in this example, the error caused by the approximations in Equations (10), (11), and (12) can be neglected compared to the error caused by the random selection of the profile line used for calculations. By using the mean values of d_{\perp} and W_q , the volume is only 2.9% smaller than the measured value. Yet, even if a profile line is selected whose d_{\perp} and W_q deviate less than σ from the mean values, the error in volume can be as high as 12%. The maximum error caused by the random selection of a profile line is 25% in this example. As concerns the volumes $W_{v,flat}$ and $W_{v,ball}$, the error when choosing a profile line with d_{\perp} and W_q deviating less than σ from the mean values is 14 and 27%, respectively. In agreement with the analysis

of the error propagation, the errors are in opposite directions ($W_{v,ball}$ is underestimated and $W_{v,flat}$ is overestimated) and they partially cancel out in the calculation of the total volume. The error done in the calculation of the linear wear quantities is about 10%.

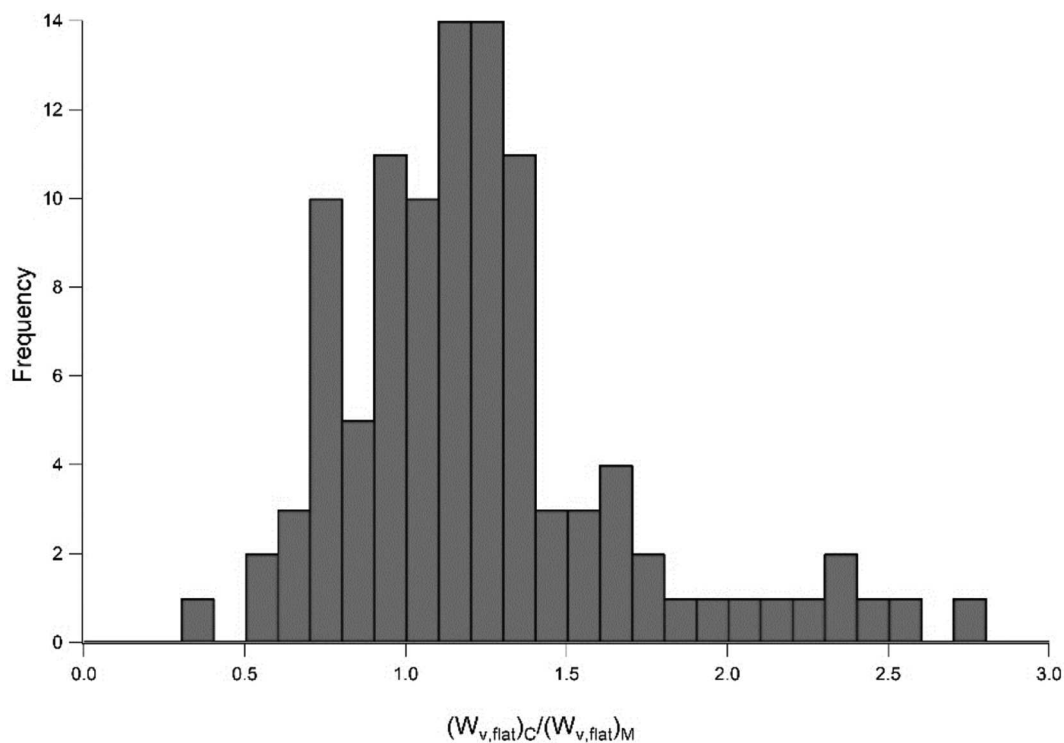
The track chosen for this example has a quite regular shape. Other scars may present significant anomalies. These are described in the following section. In these cases, significantly larger standard deviations are to be expected, making the error due to the approximations even less significant.

With help of experiments performed with 100Cr6 balls on 100Cr6 planes, in which the wear volumes of the planes have been measured with WLI, the comparison just shown for one example can be extended to 102 cases. **Figure 8** shows the histogram of the ratios of the calculated wear volumes of the planes, $(W_{v,flat})_C$, to the measured values, $(W_{v,flat})_M$. The calculated values were determined using a randomly selected profile line with the W_q method according to Equation (7).

The histogram is not symmetric; hence, the data are not normally distributed. As in the example previously shown, the calculated volumes of the planes are usually larger than the values determined by WLI. The maximum frequency of the ratios occurs for ~ 1.2 ; the mean value is 1.13. Hence, the calculation of the volumes engenders an average error of 20%; in some cases, the error is even larger than 200%. The standard deviation (0.27) is rather large. A correlation between the deviation of the ratios from the value one and test parameters

TABLE 1 | Comparison between measured and calculated wear quantities.

Quantity and method	Symbol	Sphere	Plane	Total	Unit
Track width from WLI	d_{\perp}		559 ± 16		μm
Planimetric wear from WLI	W_q		452 ± 36		μm^2
Wear volumes from WLI	$W_{v,ball}$ $W_{v,flat}$ W_v	2716	4435	7151	10^{-6} mm^3
Wear volume calculated with Equation (9)	W_v			6944 ± 637	10^{-6} mm^3
Wear volume calculated with approximate Equation (10)	W_v			6952 ± 638	10^{-6} mm^3
Wear volume calculated with approximate Equation (11)	W_v			6920 ± 633	10^{-6} mm^3
Wear volumes calculated with approximate Equations (7), (8), and (12)	$W_{v,ball}$ $W_{v,flat}$ W_v	2260 ± 282	4674 ± 402	6928 ± 635	10^{-6} mm^3
Linear wear from WLI	$W_{l,ball}$ $W_{l,flat}$ W_l	19.3 ± 0.5	1.135 ± 0.065	20.5 ± 0.5	μm
Linear wear calculated with approximate Equations (14) and (15)	$W_{l,ball}$ $W_{l,flat}$ W_l	18.3 ± 1.2	1.2 ± 0.3	19.5 ± 1.1	μm

**FIGURE 8** | Histogram of the ratios of the calculated wear volumes of the planes $(W_{v,flat})_C$ to the measured wear volumes of the planes $(W_{v,flat})_M$.

or topographic parameters, such as track width, could not be established.

In case of the ratio of the calculated total wear volumes to the measured, a similar comparison can be done only for 41 tests, since wear volumes of the spheres could not always be measured with WLI. Hence, the significance of the distribution (with a mean value of 0.964 and a standard deviation of 0.5) is not as good as for the planes. Nevertheless, it can be noticed that the error on $W_{v,flat}$ is partially compensated by the error on $W_{v,ball}$.

Anomalies in the Wear Track

Anomalies in the wear track are strong deviations from the model of a uniform wear track on the plane with regular shape. Therefore, for wear tracks with anomalies, it is not possible to calculate any R' . Anomalies can occur either randomly or due to varying loading parameters, material inhomogeneities or irregular movements of the sample. In these cases, 3D measurements of the wear volume are of great advantage compared to calculations based on a few profile lines, as these randomly selected profile lines are very unlikely to be representative of the mean cross section of the track and can lead to very incorrect results. The most important anomalies are described below:

1. Irregularities of the wear calottes in the sliding direction can occur due to material breakouts or transfers, e.g., as a result of (micro) welding. Material transfer in general is actually no wear and cannot be identified, even by 3D measurements. The material adhering locally in the friction track of the plane can lead to considerable deviations when calculating volumes with the W_q method. In some cases, especially with isolated material transfers, such anomalies can be detected and taken into account.
2. Pile-ups at the edge of the wear track (e.g., due to plastic deformation or adhering wear particles) can make a track appear wider. In this case, the determination of the position of the edges, which has an important influence on the wear volume calculation, is particularly difficult.
3. Atypical wear of the ball at the edges, but not or only very slightly in the middle of the track, can make the friction track appear narrower on the plane than on the ball.
4. In case of the so-called W-shaped profile, the wear in the center of the track on the plane is very small. This leads to a different geometry of the track than the one in **Figure 5** and consequently to larger errors in the calculation of $W_{v,ball}$ and $W_{v,flat}$.
5. A very small wear in the range of the roughness asperities can be considered as an anomaly, too. In this case, by defining the zero value of the height in the cross-sectional profile, roughness depressions are also assigned to the worn volume and contribute to its calculated value. A better way to determine the wear volume in this case is to subtract the 3D data of an AFM measurement of the worn sample from those of the unstressed sample.
6. A last kind of anomaly is due to the fact that in some cases the linear wear $W_{l,flat}$ depends on the position along the sliding direction. This may occur due to the wear dependence on the sliding speed or when the ball is moved not parallel to the plane, e.g., along a circular arc.

CONCLUSIONS

The current method for the analytical determination of the volumetric wear (W_v , $W_{v,ball}$, and $W_{v,flat}$) has been described in detail. The error due to inevitable approximations has been analyzed and the error propagation has been calculated, too. This analytical method (W_q method) is based on the arbitrary choice of a profile line of the wear track, usually measured with tactile techniques.

The errors due to the use of a WLI for the measurement of the profile line, to the approximations in the equations of the W_q method and to the arbitrary choice of a profile line have been assessed with help of an example. To this aim, the samples, namely a 100Cr6 ball and a 100Cr6 plane worn through an unlubricated oscillating sliding test, have been additionally measured with an AFM.

The error on the volume due to the use of a WLI resulted to be smaller than 0.5%. The error resulting from the approximations is smaller than 0.3% and can be neglected.

The arbitrary choice of a profile line for the W_q method turned out to generate the largest errors. By using the mean values of d_{\perp} and W_q , the error on the volume is 2.9%. The choice of a profile line with d_{\perp} and W_q differing less than σ from the mean values causes an error, which can be as high as 12%. Higher error values must be expected for scars with important anomalies, which have been listed at the end of the article.

As a conclusion, due to the large errors engendered by the arbitrary choice of a profile line for the W_q method, it is recommended to measure the volumetric wears of ball and plane and not to calculate them. For all cases in which the linear wear is comparable with the roughness of the samples, the quite time-consuming AFM stitching is necessary. Otherwise the accuracy of WLI measurements is sufficient.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

Authors have contributed in equal measure to each part of the work and of the article.

REFERENCES

- Bhushan, B. (2001). Nano-to micro scale wear and mechanical characterization using scanning probe microscopy. *Wear* 251, 1105–1123. doi: 10.1016/S0043-1648(01)00804-3
- Bhushan, B. (2003). *Handbook of Nanotechnology*. Berlin: Springer.
- Cappella, B., Hartelt, M., and Wäsche, R. (2015). High resolution imaging of macroscopic wear scars in the initial stage. *Wear* 338–339, 372–378. doi: 10.1016/j.wear.2015.07.013
- Cox, G. (2006). *Optical Imaging Techniques in Cell Biology*. Boca Raton, FL: CRC Press.
- Guilemany, J. M., Miguel, J. M., Armada, S., Vizcaino, S., and Climent, F. (2001). Use of scanning white light interferometry in the characterization of wear mechanisms in thermal-sprayed coatings. *Mater. Charact.* 47, 307–314. doi: 10.1016/S1044-5803(02)00180-8
- Kalin, M., and Vizingin, J. (2000). Use of equations for wear volume determination in fretting experiments. *Wear* 237, 39–48. doi: 10.1016/S0043-1648(99)00322-1
- Meng, H. C., and Ludema, K. C. (1995). Wear models and predictive equations: their form and content. *Wear* 181, 443–457. doi: 10.1016/0043-1648(95)90158-2
- Reichelt, M., and Cappella, B. (2020). Influence of relative humidity on wear of self-mated 100Cr6 steel. *Wear* 446–447:203202. doi: 10.1016/j.wear.2020.203239
- Wäsche, R., Hartelt, M., and Cappella, B. (2014). The use of AFM for high resolution imaging of macroscopic wear scars. *Wear* 309, 120–125. doi: 10.1016/j.wear.2013.11.009
- Wehbi, D., Clerc, M. A., Roques-Carmes, C., and Barquins, M. (1986). Three-dimensional quantification of wear tracks on amorphous NiB coatings. *Wear* 107, 263–278. doi: 10.1016/0043-1648(86)90229-2
- Yost, F. G. (1983). Two profilometric measurements of wear. *Wear* 92, 135–142. doi: 10.1016/0043-1648(83)90013-3
- Yu, J. X., Qian, L. M., Yu, B. J., and Zhou, Z. R. (2009). Nano-fretting behaviors of mono crystalline silicon (100) against diamond tips in atmosphere and vacuum. *Wear* 267, 322–329. doi: 10.1016/j.wear.2008.11.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Reichelt and Cappella. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Contact Geometry Adaptation in Fretting Wear: A Constructive Review

Ivan Argatov^{1*} and Young Suck Chai²

¹ Institut für Mechanik, Technische Universität Berlin, Berlin, Germany, ² School of Mechanical Engineering, Yeungnam University, Gyeongsan, South Korea

Fretting is a special type of wear, which appears at the contact interface between two solids subjected to constant normal load and periodic tangential forces. Although most studies on fretting have been executed experimentally, some approaches for simulating fretting wear have also been introduced during the last decades. In particular, fretting wear analysis is concerned with the evolution of the surface profiles of the contacting bodies due to wear, and its modeling was executed using numerical, finite-element, semi-analytical, and analytical methods, including the method of dimensionality reduction. In the present review we discuss recent analytical results on fretting wear contact geometry adaptation.

Keywords: fretting wear, elastic contact, limiting contact profile, steady state, wearing-in period, wear accumulation

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Stanislaw Stupkiewicz,
Polish Academy of Sciences, Poland
Yoshitaka Nakanishi,
Kumamoto University, Japan

*Correspondence:

Ivan Argatov
ivan.argatov@campus.tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 13 March 2020

Accepted: 05 June 2020

Published: 03 July 2020

Citation:

Argatov I and Chai YS (2020) Contact
Geometry Adaptation in Fretting Wear:
A Constructive Review.
Front. Mech. Eng. 6:51.
doi: 10.3389/fmech.2020.00051

1. INTRODUCTION

Contact of two machine parts established under external compressive loads and subjected to oscillating shear forces is often accompanied by friction and wear (Ciavarella and Demelio, 2001). In the case of oscillatory tangential motion of small amplitude, the particular type of wear which occurs at the sliding interface is called fretting (Vingsbo and Söderberg, 1988). Due to surface wear and damage, the shapes of the contacting bodies change and this process is called the contact geometry adaptation. Fretting phenomena, including fretting wear and accompanied variation of contact geometry, are encountered in many industrial applications, where contact parts experience oscillating small relative movements. For instance, the fretting wear characteristics of Inconel 690 U-tubes strongly influences the structural integrity of steam generators in nuclear power plants (Chai et al., 2005; Lee et al., 2009).

The problem of contact geometry adaptation in fretting wear can be formulated as a spatial-temporal contact problem with a variable contact geometry. In particular, fretting wear analysis is concerned with the evaluation of the surface profiles of the contacting bodies due to wear. Mathematical methods of solving elastic contact problems with wear were reviewed in a number of review papers by Aleksandrov and Kovalenko (1984) and Kovalenko (2001). Numerical simulation aspects of wear modeling were recently discussed in detail by Huajie and Hongzhao (2018), using the integration size as a principal characteristic (which is absent in analytical models). Recent studies on fretting wear damage in coated systems were reviewed by Ma et al. (2019). Fretting wear mechanisms and modeling were considered by Yue and Wahab (2019) and Meng et al. (2020) with a particular focus on models of debris and third-body fretting wear.

In the present review paper, we discuss different approaches to modeling fretting wear with the emphasis on analytical and semi-analytical methods, including the method of dimensionality reduction. It is to note here that though the present review is somewhat biased with the focus on the recent work of the authors, an up-to-date account of relevant studies is given as well. The aim of this review paper is to explore the theoretical ideas, analytical models and results relating to the

concept of contact geometry adaptation in fretting wear to facilitate their further development (for instance, by extending the solutions of two-dimensional problems to the three-dimensional case). For this reason, the term “constructive review” is emphasized.

One of the things we shall concentrate on is the so-called wearing-in period in gross-slip fretting wear, when the initial contact state progresses into a kind of steady-state, in which the applied contact load is redistributed along the contact area in accordance with the wear equation. Under partial-slip fretting wear conditions, no such steady state exists, and, theoretically speaking, the initial contact state is expected to evolve into a kind of steady state (called limiting state), characterized by transferring the contact load primarily through the stick zone, where no wear occurs. In both cases our particular interest is focused on estimating the time needed to achieve the steady state or the limiting state.

1.1. Archard Wear Equation and Its Generalizations

The local wear is usually characterized by the linear wear rate, \dot{w} , where a dot denotes the derivative with respect to time. According to Archard's equation of wear (Archard, 1953), which is adopted in the majority of studies on fretting wear published to date, we have

$$\dot{w} = k_w p v, \quad (1)$$

where p is the contact pressure, v is the absolute value of the relative sliding velocity, and k_w is the coefficient of wear.

A mathematically straightforward generalization of Equation (1) leads to the Archard–Kragelsky wear equation

$$\dot{w} = K_w p^\alpha v^\beta, \quad (2)$$

which was considered in a number of studies (Kragelsky, 1965; Meng and Ludema, 1995; Kragelsky et al., 2013). Recently, Argatov and Chai (2019a) put the Archard–Kragelsky equation into an ANN (artificial neural networks) framework, which allows to account for the dependence of the wear coefficient K_w on material parameters and operational conditions.

1.2. Reciprocal Sliding Wear

Let ΔT denote the period of tangential oscillations. Then, according to Equation (1), the linear wear resulting from one cycle will be

$$\Delta w(x, t) = k_w \int_t^{t+\Delta T} p(x, \bar{t}) v(x, \bar{t}) d\bar{t}. \quad (3)$$

One simplification of the wear relation (3) is that, under certain conditions (e.g., under constant normal load and relatively small wear coefficient), the contact pressure may be assumed to not change appreciably during one cycle. In this way, Equation (3) simplifies as

$$\Delta w(x, t) = k_w p(x, t) \int_t^{t+\Delta T} v(x, \bar{t}) d\bar{t}. \quad (4)$$

Another simplification is admissible in the gross slip regime, when the relative sliding velocity is supposed to become independent of the position of the point x on the contact interface, and thus, Equation (4) simplifies further $\Delta w(x, t) = k_w p(x, t) \bar{v} \Delta T$, where \bar{v} is the average absolute value of sliding velocity, that is

$$\bar{v} = \frac{1}{\Delta T} \int_t^{t+\Delta T} v(\bar{t}) d\bar{t}. \quad (5)$$

Let now Δx denote the stroke of tangential oscillations. Then, Equation (5) can be rewritten as

$$\bar{v} = \frac{2\Delta x}{\Delta T}. \quad (6)$$

In reciprocating sliding, it is convenient to operate both with the number of cycles, N , and the effective time variable, t , such that $N = \lfloor t/\Delta T \rfloor$, where $\lfloor x \rfloor$ denotes the floor function (that is the largest integer less than or equal to x).

1.3. Energy Wear Equation

By introducing the sliding distance, s , such that $ds = v dt$, Equation (1) can be represented in the differential form as $dw = k_w p ds$. Moreover, let μ denote the coefficient of friction. Then, introducing the frictional shear stress is $q = \mu p$, the wear increment can be further rewritten as

$$dw = \alpha_V q ds, \quad (7)$$

where α_V is the energy wear coefficient, such that

$$k_w = \mu \alpha_V. \quad (8)$$

In fretting wear, by combining Equations (3) and (7), (8), we arrive at the following wear equation (Mróz and Stupkiewicz, 1994; Fouvry et al., 1996, 2003):

$$\Delta w(x, t) = \alpha_V \Delta E_d(x, t). \quad (9)$$

Here, $\Delta E_d(x, t)$ is the frictional dissipated energy during one fretting cycle, that is

$$\Delta E_d(x, t) = \int_s^{s+\Delta s} q(x, t) ds. \quad (10)$$

While, in view of (8), the wear equations (3) and (10) are equivalent, the energy wear equation incorporates the friction mechanism and allows to account for variable coefficient of friction (Cheikh et al., 2007).

2. FRETTING WEAR IN GROSS SLIP REGIME

2.1. Formulation of the Model Wear Contact Problem

In the present review, we consider both two-dimensional and three-dimensional settings, highlighting their similarities and differences. The analysis of contact deformations is limited to

the framework of linear elasticity and small strain analysis. While the Hertzian half-plane or half-space approximation can be employed in modeling the local stress-displacement field, in many cases, the developed analytical approach can be directly generalized to the case of layered elastic bodies with planar contact interface by utilizing the corresponding surface influence function. However, the effect of finite geometry on the wear scar profiles requires a special consideration. In what follows, the effect of local contact geometry is accounted for by means of the local gap function.

Let $\varphi_0(x)$ denote the undeformed gap between the contacting surfaces as a function of Cartesian coordinate x . Then, according to Equation (1), the worn gap under plane-deformation conditions will be given by the time-continuous equation

$$\varphi(x, t) = \varphi_0(x) + k_w \int_0^t p(x, \bar{t}) v(x, \bar{t}) d\bar{t}, \quad (11)$$

or by the time-increment equation $\Delta\varphi(x, t_{i+1}) = k_w p(x, t_i) v(x, t_i) \Delta t$, where $\Delta t = t_{i+1} - t_i$ is the time increment. Note that, by the definition, the contact pressure $p(x, t)$ is positive, and, therefore, the value of the integral in (11) monotonically increases with time, provided $v(x, t) \neq 0$.

Another simplification which is implicitly or explicitly present in many studies on contact with wear, is to replace the problem for two contacting elastic bodies with the wear contact problem for one equivalent elastic body (whose surface influence function is composed from the surface influence functions of the given bodies) and a rigid punch, whose shape function is determined by the local gap function. Though, this is a usual approach in studying contact problems (e.g., for elastic bodies with rough surfaces), the formulation of the wear contact problem requires to consider the partition of the linear wear between the two wearing bodies, when the resulting wear scar profiles are determined.

A wear contact problem in gross slip fretting regime can be formulated as follows (Goryacheva, 1998):

$$\frac{2}{\pi E^*} \int_{-a(t)}^{a(t)} K(x - \bar{x}) p(\bar{x}, t) d\bar{x} = \delta_0(t) - \varphi(x, t). \quad (12)$$

Here, $K(x)$ is the normalized surface influence function, E^* is the reduced elastic modulus, $\delta_0(t)$ is the normal approach between the contacting bodies, and $a(t)$ is the variable half-width of the contact interval.

Yet another simplifying assumption is incorporated into the governing integral equation (12), which implicitly states that the wear, while changing the contact geometry, does not significantly affect the surface influence function. In other words, small changes of the contact shape due to wear are assumed, and their influence on the contact pressure distribution is neglected. The geometry dependence of the Green functions was accounted for by Peigney (2004) using a first-order perturbation approach.

In light of (11) and (6), we have

$$\varphi(x, t) = \varphi_0(x) + \frac{2k\Delta x}{\Delta T} \int_0^t p(x, \bar{t}) d\bar{t}. \quad (13)$$

Finally, in unilateral contact, the extent of the contact zone is determined by the condition of vanishing contact pressure at $x = \pm a(t)$.

It is to note here that the gross-slip regime assumes that the condition of sliding occurs at the entire contact interface. It is clear that the simple equation of wear (13) is violated at the turning points of the wear, since the sliding velocity goes through zero. This aspect, as well as the relative value of the displacement stroke, should be taken into account when considering the application of the analytical model to the analysis of experimental results.

The integral equation kernel $K(x)$ depends on both the global shapes of the contacting bodies and the boundary conditions. Using the half-plane approximation, one obtains $K(x - \bar{x}) = -\ln[|x - \bar{x}|/H] - d_0$, where H is the characteristic length of the contact pair, and d_0 is the asymptotic constant (Aleksandrov et al., 1978; Argatov, 2001). It was Galin (1976) who first considered a two-dimensional wear contact problem with a constant area and applied the method of variables separation. Galin's method was further developed by Aleksandrov et al. (1978) and Komogortsev (1985) and extended to three-dimensional wear contact problems (Galin and Goriacheva, 1977; Kovalenko, 1985).

Numerical methods for solving the fretting wear contact problem of the type (12), (13) were developed in a number of studies, of which we refer especially to the papers by McColl et al. (2004), Chai et al. (2005), Mary and Fouvry (2007), and Bae et al. (2009) on finite-element simulations and (Serre et al., 2001; Sfantos and Aliabadi, 2006) on boundary-element simulations. Alternatives to finite and boundary element methods were proposed by Lee et al. (2009), using the influence function method, and (Nowell, 2010), using a quadratic programming technique. A general method for the analysis of plane contact problems for layered elastic structures in the presence of sliding wear was developed by Aleksandrov and Kovalenko (1980). Recent modeling results for coated systems, including functionally graded material (FGM) coatings, are discussed elsewhere (Ma et al., 2019).

2.2. Force-Controlled Steady-State Regime

In gross slip fretting wear, the following natural assumption makes sense (Galin, 1976; Komogortsev, 1985):

$$p(x, t) = p_\infty + q(x, t), \quad p_\infty = \frac{P}{2a}, \quad (14)$$

where $q(x, t) \rightarrow 0$ as $t \rightarrow \infty$. In other words, the contact pressure is equalized during wear, i.e., $p(x, t) \rightarrow p_\infty$ as $t \rightarrow \infty$. Of course, formula (14) holds true, provided the following assumption is fulfilled: both the contact force P and the contact half-width a are kept constant during the wear process.

While formula (14) has been proved to hold for a constant area of contact (Aleksandrov et al., 1978; Komogortsev, 1985), Argatov and Tato (2012) extended its applicability to the case of constant contact load P and variable contact half-width

$a(t)$, when

$$P = \int_{-a(t)}^{a(t)} p(\bar{x}, t) d\bar{x}, \quad (15)$$

by including into $q(x, t)$ the boundary layers appearing near the ends of the contact interval. However, the problem of explicit constructing these boundary layers still remains open.

In the steady-state regime, the variation of contact zone depends on the contact geometry. For a paraboloidal gap $\varphi_0(x) = x^2/(2R)$, the following differential equation holds (Argatov et al., 2011):

$$a^2 \frac{da}{dt} = \frac{1}{2} k_w \bar{v} R P. \quad (16)$$

Observe that Equation (16) states that in quasi steady state the contact zone evolves being governed by the wear coefficient k_w , the average sliding velocity \bar{v} , and the undeformed contact geometry, which is characterized by the curvature radius R . This fact was utilized by Lengiewicz and Stupkiewicz (2013) in their model of evolution of contact zone and wear accumulation, without referring to the underlying elasticity problem. The effect of elastic deformations on the wear scar profiles was accounted for by Argatov and Tato (2012). The model of fretting wear in quasi-steady-state gross-slip regime was generalized by Argatov et al. (2011) for the Archard–Kragelsky wear equation (2), which assumes a power law dependence of the wear rate on the contact pressure.

In contrast to the force-controlled loading, when the left-hand side of Equation (15) is suggested to be known, in the displacement-controlled loading, the variation of the contact approach will be specified. In such a case, due to the dissipative nature of wear accompanied by material removal, the total contact load should gradually vanish to the end of the fretting process. Such situation was analyzed by Peigney (2004), who determined the asymptotically stabilized state reached by an elastic body subjected to wear contact with a rigid indenter in the displacement-controlled cyclic loading.

2.3. Steady-State Contact Profile

Apparently, the problem of determining the worn shape in the operating state, that is the contact shape function $\varphi_\infty(x)$ that produces a uniform contact pressure under a constant load was first solved by Dundurs and Comninou (1980) in the case of an elastic half-space. In a more general case (12), the steady-state profile is given by

$$\varphi_\infty(x) = \frac{P}{\pi E^* a} \left\{ \int_{-a}^a K(\bar{x}) d\bar{x} - \int_{-a}^a K(x - \bar{x}) d\bar{x} \right\}. \quad (17)$$

We note that $\varphi_\infty(0) = 0$, and this explains the first term in the curly braces in (17). The steady-state solution in the case of Archard–Kragelsky model of wear (2) was obtained by Goryacheva (1998), who also studied its asymptotic stability.

Interestingly, the steady-state profile (17) is shown to be optimal, if the wear coefficient is assumed to be constant (Banichuk et al., 2010). The optimal shapes generated by wear process were evaluated by exploiting the dissipative nature of the

wear process, which can be characterized, e.g., by minimization of friction dissipation power (Páczelt and Mróz, 2007). Based on the Hertzian half-space approximation, a three-dimensional computational method for determining the optimum contact geometry in fretting under the gross slip regime was developed by Gallego et al. (2006). Recently, Argatov and Chai (2019c) considered a practically important question of approximating the ideal profile (17) with a symmetric smooth profile composed of three parabolic arcs, which was introduced by Vázquez et al. (2010), exploiting the idea of compound curvature. The effect of friction is shown to result in the profile asymmetry, which depends on the direction of sliding (Argatov and Chai, 2020b).

2.4. Wearing-In Period

Any fretting wear test starts from the initial contact state, which is fully characterized by the initial contact geometry and loading conditions. The initial time interval, during which the contact pressure evolves from the initial one to the steady state pattern, is called the wearing-in period. Based on the Galin type analysis of the wear contact problem with a fixed contact zone, Argatov and Fadin (2011) have estimated the duration of the wearing-in period, T_{in} , as follows:

$$T_{in} \sim \frac{a}{\lambda_{min} k_w \bar{v} E^*}. \quad (18)$$

Here, λ_{min} is the minimum characteristic value of the corresponding integral eigenvalue problem.

It is to note that the right-hand side of the relation (18) does not depend on the loading level. It is interesting that the wearing-in period in the displacement-controlled regime is about five times greater than that under the force-controlled loading (Argatov and Fadin, 2011). It is also shown (Argatov and Chai, 2020b) that the effect of friction extends the wearing-in period.

2.5. Wear of Functionally-Graded Wear-Resisting Materials

A range of wear contact profiles with variable wear resistance of a sliding punch was considered by Goryacheva (1998) in the case of the Archard–Kragelsky wear model (2) with a particular focus on the steady-state solutions. The transient wear contact problems for composite materials were considered recently using different approaches, including the method of dimensionality reduction with application to an axisymmetric heterogeneous annular cylindrical punch (Li et al., 2018) and a level-set based shape and topology optimization method with application to a Pasternak elastic foundation model (Feppon et al., 2017). By using an appropriate symmetrization of the integral equation kernel, Argatov and Chai (2019b) extended the Galin method for analyzing the transient contact pressure distribution and derived an upper estimate for the wearing-in period. Also, the effective wear coefficient was represented as

$$K_{eff}(t) = K_{eff}^\infty + \frac{1}{P} \int_{-a}^a k_w(x) [p(x, t) - p_\infty(x)] dx, \quad (19)$$

where $k_w(x)$ is a variable wear coefficient, $p_\infty(x)$ is the steady-state contact pressure distribution, and K_{eff}^∞ is the steady-state

value of the effective wear coefficient given by

$$K_{\text{eff}}^{\infty} = 2a \left(\int_{-a}^a \frac{dx}{k_w(x)} \right)^{-1}. \quad (20)$$

Moreover, it was shown that the second term on the right-hand side of Equation (19) decreases exponentially with time during the wearing-in period, for which the following estimate was established (Argatov and Chai, 2019b):

$$T_{\text{in}} \leq \frac{2a}{\pi k_w^{\text{max}} \bar{v} E^*} \sqrt{\int_{-1}^1 \int_{-1}^1 |\mathcal{L}_x(\xi, \bar{\xi})|^2 d\bar{\xi} d\xi}. \quad (21)$$

Here, k_w^{max} is the maximum wear coefficient, and $\mathcal{L}_x(\xi, \bar{\xi})$ is the normalized integral kernel.

Interestingly, Equation (20) represents a generalization of the mixture rule of Khrushov (1974) for functionally graded wear-resisting composite materials (Friedrich, 1993; Yen and Dharan, 1996). Observe also (Argatov and Chai, 2019b) that while K_{eff}^{∞} is independent of the distribution of the wear resistance, that is independent of the distribution of the phases in a wearable multiphase material the duration of the wearing-in period T_{in} , in contrast, is sensitive to the phase fraction distribution.

2.6. Three-Dimensional Fretting Wear Contact Problems

The main difference between the 2D and 3D cases is a higher variability of the sliding velocity orientation that can occur in practice, so that the concept of anisotropic wear (Mróz and Stupkiewicz, 1994; Zmitrowicz, 2006) can be introduced in the spatial case. Otherwise, many features of the gross slip fretting wear are similar in the two cases. In particular, in the axisymmetric case with a constant contact radius a , Equation (14) still applies with $p_{\infty} = P/(\pi a^2)$. Also, formula (17) for the steady-state profile can be simply generalized by the appropriate choice of the integral equation kernel (which is determined by the corresponding surface influence function) and, of course, by extending the integration to the whole contact area. For the Hertzian type contact geometry with the initial gap $\varphi_0(x) = (x_1^2 + x_2^2)/(2R)$ and a variable contact area of radius $a(t)$, the analog of Equation (16) reads as follows (Argatov, 2011):

$$a^3 \frac{da}{dt} = \frac{1}{\pi} k_w \bar{v} R P. \quad (22)$$

The axisymmetric model of fretting wear based on the Archard–Kragelsky wear model (2) was developed by Argatov et al. (2011), who have observed the phenomenon of the decrease of the wear coefficient due to the increase of the contact area followed by decrease of the contact pressure. Apparently, such an effect depends on the exponent α in Equation (2). In the mentioned study, it was evaluated to be greater than one.

The general Hertzian type contact with a variable elliptical contact area was considered by Argatov (2011) in application to local interwire contact under reciprocal sliding. A special consideration is required in the case of torsional fretting wear

with an annular contact area, when the relative sliding velocity varies proportionally to the distance from the axis of symmetry. The analogous rotational contact problem with a sliding wear was analyzed by Galin and Goriacheva (1977), and Kovalenko (1985). The finite element method study of torsional fretting for a ball-on-flat configuration was conducted by Liu et al. (2014) under the assumption of variable coefficient of friction, whose variation due to the abrasive wear degradation is governed by the local contact history and the accumulated slip distance.

3. FRETTING WEAR IN PARTIAL SLIP REGIME

3.1. Stick Zone

Recall that the local contact of two elastically similar bodies during cyclic loading-unloading by a normal force only (i.e., at zero tangential force) is not accompanied by the relative tangential displacements at the contact interface, whereas some fretting can occur at the cyclic normal contact of dissimilar bodies. However, the tangential displacements mismatch appears for elastically similar bodies even in the case of cyclic tangential loading with constant normal load.

In a certain regime of fretting, called partial slip regime, the contact area, $\omega(t)$, which may vary in time, contains inside a stick zone, ω_* , where the contacting surfaces stick one to another, so that the contact geometry inside the stick zone remains untouched by wear, whereas wear occurs in a slip zone, where the contacting surfaces experience relative tangential movement. The theory of tangential contact with partial slip for the Hertzian geometry was developed by Cattaneo (1938) and Mindlin (1949). The 2D theory of tangential contact for elastically similar semi-infinite solids was developed by Jäger (1998) and Ciavarella (1998). The 3D Cattaneo–Mindlin model was extended by Jäger (1996) for the case of stepwise oblique loading. Further progress in its development is associated with devising the method of memory diagrams (Aleshin and Van Den Abeele, 2013; Aleshin et al., 2015). Recently, the 3D Cattaneo–Mindlin model was outlined for transversely isotropic materials by Argatov et al. (2018).

In the axisymmetric case, Jäger (1995) generalized the model of local tangential contact for two elastically similar bodies with arbitrary gap function. In the general non-axisymmetric three-dimensional case, Ciavarella (1998) introduced a simplified version of the Cattaneo–Mindlin theory (without Poisson's effect), which was recently extended for the case of transversely isotropic materials by Chai and Argatov (2018), who also applied the self-similarity solutions by Borodich (1983, 1989) to derive explicit tangential force-displacement relations in the case of self-similar gap between the contacting surfaces.

Further, Hills and Sosa (1999) reviewed analytical solutions for general elastic contact problems with partial slip, which can be used, for instance, in estimating frictional energy and the local wear rate in the initial stage of fretting. Analytical aspects of fretting fatigue damage were considered by Ciavarella and Demelio (2001) with application to dovetail joints. Numerical method for partial-slip frictional contact problems have been

developed in a number of studies (Chen and Wang, 2008; Wang et al., 2013).

Let $\varphi(x_1, x_2)$ be the gap between the surfaces of two elastically similar transversely isotropic semi-infinite bodies that do not exhibit the effect of coupling between the distributions of shearing contact tractions (in the case of identical materials this assumption implies the zero Poisson's ratio, see, e.g., Ciavarella, 1998). In many cases it may be assumed that the gap is described by a homogeneous function of degree d , such that

$$\varphi(cx_1, cx_2) = c^d \varphi(x_1, x_2), \quad (23)$$

where c is an arbitrary positive constant.

Let also l and l^* be the characteristic sizes of the contact area ω and the stick zone ω_* , respectively. Then, the following relations hold between the relative size of the stick zone, the ratio of the tangential to the normal contact force, and the ratio of the tangential to the normal displacement (Chai and Argatov, 2018):

$$\frac{l^*}{l} = \left(1 - \frac{F_1}{\mu F_3}\right)^{1/(d+1)}, \quad \frac{l^*}{l} = \left(1 - \frac{M_1}{M_3} \frac{\delta_1}{\mu \delta_3}\right)^{1/d}. \quad (24)$$

Here, M_1 and M_3 are effective elastic moduli. The Cattaneo-Mindlin model is recovered from Equation (24) for $d = 2$.

3.2. MDR-Based Approach

It is well-known (De Mul et al., 1986) that for a wide class of contact geometries, the Hertzian half-space analysis of local contact gives reliable results. The method of dimensionality reduction (MDR) developed by Popov and Heß (2015), and their collaborators, combines the axisymmetric Hertz theory of normal contact and the axisymmetric Cattaneo-Mindlin theory of tangential contact into a unified modeling framework by transforming a given 3D contact problem for two elastically similar solids into one-dimensional contact problem for an equivalent rigid punch and the Popov foundation, which is a linearly-elastic spring-like foundation that possesses both normal and tangential stiffnesses. The relation between the gap function $\varphi(r, t)$ and the equivalent 1D profile $g(x, t)$ represents the direct mapping rule from the original 3D contact problem into the 1D equivalent contact problem. Since the Popov foundation is spring-like, its normal reaction is given by $q_z(x, t) = E^*(\delta(t) - g(x, t))$, where $|x| \leq a(t)$ and $a(t)$ is a root of the equation $g(a, t) = \delta(t)$. The integration of the 1D normal contact reaction over the contact interval $(-a(t), a(t))$ yields the total normal force, F_3 (Note that in the displacement controlled mode, the contact approach $\delta(t)$ is assumed to be known).

The Popov foundation can be discretized by introducing a discretization step Δx , so that the normal and tangential stiffnesses of every individual spring element will be $\Delta k_z = E^* \Delta x$ and $\Delta k_x = G^* \Delta x$, respectively, where G^* is the effective shear modulus. Correspondingly, if an individual elastic spring element with a coordinate x receives normal, $u_z(x, t)$, and tangential, $u_x(x, t)$, displacements, the values of normal and tangential reaction forces, respectively, will be $\Delta f_N(x, t) = u_z(x, t) \Delta k_z$ and $\Delta f_T(x, t) = u_x(x, t) \Delta k_x$ (with compression-positive sign convention taken into account).

In a stick zone, the tangential displacement $u_x(x, t)$ is determined by the punch's tangential displacement, $u_x^{(0)}(t)$, whereas in a slip zone $\Delta f_T(x, t) = \pm \mu \Delta f_N(x, t)$, as it is prescribed by Coulomb's law, where μ is the coefficient of friction. It is suggested (Dimaki et al., 2014, 2016) to consider the fretting wear as an incremental process, such that $\Delta u_x^{(0)}(t) = u_x^{(0)}(t + \Delta t) - u_x^{(0)}(t)$ is the tangential displacement increment of the 1D equivalent punch, which exactly corresponds to the increment of the relative tangential displacement of the contacting solids. According to the Archard wear equation (1), the linear change of the 3D profile is given as follows (Dimaki et al., 2016):

$$\Delta \varphi(r, t) = k_w p(r, t) (\Delta u_x^{(0)}(t) - \Delta u_x^{(3D)}(r, t)). \quad (25)$$

Here, $u_x^{(3D)}(r, t)$ and $\Delta u_x^{(3D)}(r, t)$ are the relative tangential displacement at the contact interface and its increment.

Thus, the numerical implementation of the MDR-based approach using Equation (25) will require the application of the inverse mapping (from the 1D contact problem to the 3D contact problem) for evaluating $u_x^{(3D)}(r, t)$ and the normal contact pressure $p(r, t)$. The corresponding numerical procedures have been developed for both the gross-slip (Dimaki et al., 2016) and partial-slip (Dimaki et al., 2014) regimes.

3.3. Limiting Profile

Evidently, in the partial slip fretting wear, by the definition, there is a stick zone which remains untouched by wear for the entire periodic loading process. For instance, in the force-controlled mode with a constant normal contact load and a constant amplitude sinusoidal tangential force, the effect of wear on the contacting surfaces will exhibit itself in an increase of the contact approach as well as in an increase of the contact area. However, as it was observed by Ciavarella and Hills (1999), this process eventually leads to some limiting contact geometry that is characterized by the absence of wear outside the stick zone. In the case of the Archard wear model, the latter means that the contact pressure vanishes in the final slip zone.

In the axisymmetric case under displacement-controlled loading (when $\delta = \text{const}$), the solution for the limiting profile can be easily obtained in terms of the limiting profile for the equivalent punch as follows (Popov, 2014):

$$g_\infty(x) = \begin{cases} g_0(x), & |x| \leq c, \\ \delta, & c < |x| \leq a_\infty. \end{cases} \quad (26)$$

Here, $g_0(x)$ is the initial equivalent profile, c is the radius of the stick zone, which depends on the tangential displacement amplitude $u_x^{(0)}$, δ is the contact approach, a_∞ is the limiting radius of the contact area, which is determined by the equation $\varphi_\infty(a) = \varphi_0(a)$. Finally, the function $\varphi_\infty(r)$ itself is determined by the inverse transform applied to $g_\infty(x)$.

Of course, formula (26) can be easily extended to the case of force-controlled loading by replacing δ with δ_∞ , where δ_∞ is the limiting contact approach, and expressing c in terms of the normal and tangential contact loads F_3 and F_1 , e.g., using Equation (24). However, it is instructive to distinguish the two cases, especially since it makes sense in the 3D case.

Table 1 shows the state of the art of the analysis of the final (limiting or asymptotic) contact geometry. We note that the solutions of Heß (2019) and Willert et al. (2019) were obtained for the case of power-law graded materials with Young's modulus varying with depth as $E(z) = E_0(z/z_0)^k$, where $k \in (-1, 1)$.

The MDR-based approach (Popov, 2014) was generalized by Chai and Popov (2016) for fretting wear in an adhesive contact in the Dugdale approximation when the adhesive (attractive) stress outside the contact area is assumed to be constant up to some critical distance and vanishing beyond this range. Recently, the analytical approach developed by Popov (2014) was extended by Dmitriev et al. (2016) and Mao et al. (2016) (see also Li, 2016) to a dual-mode fretting under the influence of superimposed normal oscillations (with amplitude Δu_z^0 and frequency ω_z) and tangential oscillations (with amplitude Δu_x^0 and frequency ω_x). It is interesting that Dmitriev et al. (2016) also provided experimental evidence for the correctness of the theoretically predicted limiting shape. We note also that recent experimental aspects of fretting wear were discussed in an extensive review by Meng et al. (2020).

3.4. Wear Accumulation

A number of numerical methods have been devised for simulating the contact geometry adaptation during fretting wear in the partial slip regime (Gallego and Nelias, 2007; Dimaki et al., 2014; Wang et al., 2015; Cardoso et al., 2019), when both the contact profiles (outside the stick zone) and the contact pressure distribution evolve in relation to each other.

Interestingly, a special focus on modeling of the evolution of the worn volume, V , was not shown until recently (Kasarekar et al., 2007). A simple mathematical model of wear accumulation in the case of initial Hertzian contact was developed by Chai and Argatov (2019) based on the dissipation energy model for the volume wear rate (Fouvry et al., 2003). The following one fitting parameter formula was suggested for the non-monotonic variation of the volume wear rate:

$$\frac{dV}{dN} = w_0 \left(1 + \beta_1 \frac{N}{N_1} \right) \exp \left(-\frac{N}{N_1} \right). \quad (27)$$

TABLE 1 | Limiting shapes of profiles in fretting wear.

	Initial profile	Displacement-controlled regime	Force-controlled regime
2D case	Hertzian		Goryacheva et al., 2001; Hills et al., 2009
	Arbitrary		Goryacheva and Goryachev, 2006
3D axisymmetric	Hertzian	Popov, 2014	Dini et al., 2008
	Arbitrary	Popov, 2014	Argatov and Chai, 2018
3D non-axisymmetric	Hertzian		Argatov et al., 2018
3D axisymmetric FGM	Hertzian	Willert et al., 2019	Heß, 2019
	Arbitrary	Willert et al., 2019	

Here, w_0 is the initial volume wear rate, which can be evaluated using the Cattaneo–Mindlin theory (Johnson, 1955), N is the number of cycles, N_1 is an auxiliary parameter, which is related to the total worn volume $V_\infty = w_0(1 + \beta_1)N_1$, and β_1 is the only fitting parameter.

The analytical model (27) implies that the wearing-in period is proportional to N_1 , which, in turn, is estimated as

$$N_1 \sim \frac{G^*}{(E^*)^{5/3}} \frac{R^{2/3}}{\mu k_w P^{1/3}} n_1(\chi), \quad (28)$$

where $\chi = c/a_0$ is the relative stick-zone radius, and a_0 is the initial (Hertzian) contact radius. In contrast to Equation (18), formula (28) shows that the duration of the wearing-in period (measured in number of cycles) in partial-slip fretting wear depends on the load level.

4. DISCUSSION AND CONCLUSIONS

The obtained results for wearing-in period (18), (21), and (28) merit comment. First, as we might intuitively expect, the larger the initial contact zone, the larger the wearing-in period. Second, as it could be foreseen from the physical dimension of the wear coefficient, the duration of the wearing-in period is inversely proportional to the characteristic value of the coefficient of wear. Third, in the gross-slip and partial-slip regimes, the wearing-in period is inversely proportional to E^* and $(E^*)^{2/3}$, respectively, so that the elasticity effect weakens in the second case.

4.1. Limitations of the Analytical Approach

It goes without saying that analytical methods are not so flexible as numerical ones, especially in applications to specific engineering problems. In contact mechanics, the success of analytical approach is critically dependent on the possibility to approximate the surface influence function, which, in turn, strongly depends on the contact geometry. The phenomenon of wear manifests itself in the variation of the latter, and, generally speaking, the main limitation of the current state-of-the-art analytical techniques is in their inability to effectively deal with varying geometry due to wear loss.

Perhaps, another reason for the slow progress in solving transient wear contact problems is the difficulty which may arise with the introduction of non-linear equations of wear. However, a further rapid advance can be achieved, for instance, in estimating the duration of wearing-in period, provided the problem formulation admits the existence a steady-state regime.

4.2. Open Problems

First, observe that **Table 1** has a few empty cells, which indicate a number of still unsolved problems on limiting shapes of profiles in fretting wear. The limiting profile problem formulation assumes that in the limiting state the contact pressure vanishes outside of the stick zone. It makes sense to investigate whether a threshold model of wear, which assumes no wear below certain level of contact pressures, is suitable for describing the limiting state in practical fretting problems. Second, we point out that the analytical solutions (16) and (22) were obtained

for the Hertzian local gap, though the method with which they were derived allows such generalizations. By the way, it is still interesting to analyze the boundary-layer problem in the wear contact problem with variable contact zone. Third, it is to note here that the effective wear coefficient (20) for functionally-graded wear-resisting materials was evaluated under the simplifying assumption of homogeneous elastic properties, which is implicitly or explicitly present in a majority of studies of wear contact problems for composite materials. Further, as it was already mentioned in section 1.3, the energy wear equation allows to account for variable coefficient of friction, which can be done in a straightforward manner in the gross slip regime.

It should be noted that the wearing-in period in gross-slip fretting wear is associated with the redistribution of the macroscopic contact pressure from the initial pattern, which is caused by the intact contact geometry (e.g., the Hertzian contact pattern in ball-on-plate contact), to an approximately uniform distribution, which reproduces the steady-state shape profile. In the case of non-conforming contact, there is still uncertainty regarding the experimental relation between the wearing-in period and the running-in period, which is primarily associated with the evolution of surface topography and shows the exponential evolution of the wear rate as well (Zhang et al., 2018). It is to note here that the recently developed artificial neural network modeling framework (Argatov and Chai, 2020a) can be utilized in analysis of experimental studies of wearing-in period and the so-called true wear coefficient. It is also apparently an open question as to whether the duration of the wearing-in period in periodic contact is longer than that in the wear contact problem for a single contact.

It was shown (Argatov and Chai, 2020b) that the eigenvalue problem associated with the wear contact problem with friction, generally speaking, may possess the complex-valued spectrum, and therefore, the contact pressure is predicted to approach the steady state exponentially decaying and oscillating. Such an oscillating behavior of the contact pressure during the wearing-in period, if exists, would be experimentally observed for large values of the friction coefficient.

4.3. Directions for Future Research

Based on the above review, one can suggest that further progress is expected in studying transient wear contact problems

for functionally-graded wear-resisting materials. Also, further attention needs to be paid to the optimization problems in fretting wear, while combining different strategies for optimizing contacting parts, including geometrical (shape optimization) and material grading.

Observe that Archard's equation of wear or the work rate model are usually adopted in the majority of analytical studies to date. Further progress is needed in extending the results obtained for these linear wear models to other models, including non-linear, like the Archard–Kragelsky equation, or models which account for the effect of debris formation and third-body fretting wear.

Further, it is well-known and practically important that the Archard equation (1) as well as the Archard–Kragelsky equation (2) treat wear as a local process. In this respect, it would of interest to investigate what role the non-local nature of wear damage plays in the partial lip regime of fretting, especially near the boundary of the stick zone.

Lastly, caution is urged when the simple analytical models are applied for analyzing practical problems. For instance, one may argue that the analytical solutions for limiting profiles outlined in **Table 1** do not predict the final state that is practically achieved in partial-slip fretting, since they assume complete removal of worn material. Nevertheless, the limiting shape profiles are useful, as they allow to upper estimate the differences in contact behavior that can be observed both in practice and experiment (Dini et al., 2008). An important asset of the developed analytical solutions is their explicit dependence on the model parameters, which can be effectively used for solving design and optimization problems.

AUTHOR CONTRIBUTIONS

IA: methodology, formal analysis, and writing—original draft preparation. YC: conceptualization, methodology, and writing—reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2017M2B2A9072449).

REFERENCES

- Aleksandrov, V. M., Galin, L. A., and Piriev, N. P. (1978). A plane contact problem for an elastic layer of considerable thickness in the presence of wear [in Russian]. *Mech. Solids* 4, 60–67.
- Aleksandrov, V. M., and Kovalenko, E. V. (1980). Plane contact problems of the theory of elasticity for nonclassical regions in the presence of wear. *J. Appl. Mech. Tech. Phys.* 21, 421–427. doi: 10.1007/BF00920786
- Aleksandrov, V. M., and Kovalenko, E. V. (1984). "Mathematical methods in contact problems with wear," in *Nonlinear Models and Problems of Mechanics of Deformable Solids*, ed K. V. Frolov (Moscow: Nauka), 77–89.
- Aleshin, V., Matar, O. B., and Van Den Abeele, K. (2015). Method of memory diagrams for mechanical frictional contacts subject to arbitrary 2D loading. *Int. J. Solids Struct.* 60, 84–95. doi: 10.1016/j.ijsolstr.2015.02.016
- Aleshin, V., and Van Den Abeele, K. (2013). General solution to the Hertz–Mindlin problem via Preisach formalism. *Int. J. Nonlinear Mech.* 49, 15–30. doi: 10.1016/j.ijnonlinmec.2012.09.003
- Archard, J. F. (1953). Contact and rubbing of flat surfaces. *J. Appl. Phys.* 24, 981–988. doi: 10.1063/1.1721448
- Argatov, I., and Tato, W. (2012). Asymptotic modeling of reciprocating sliding wear—comparison with finite-element simulations. *Eur. J. Mech. A Solids* 34, 1–11. doi: 10.1016/j.euromechsol.2011.11.008
- Argatov, I. I. (2001). Solution of the plane Hertz problem. *J. Appl. Mech. Tech. Phys.* 42, 1064–1072. doi: 10.1023/A:1012534432055
- Argatov, I. I. (2011). Asymptotic modeling of reciprocating sliding wear with application to local interwire contact. *Wear* 271, 1147–1155. doi: 10.1016/j.wear.2011.05.028

- Argatov, I. I., Bae, J. W., and Chai, Y. S. (2018). The limiting shape of the transversely isotropic elastically similar solids in fretting. *Int. J. Appl. Mech.* 10:1850089. doi: 10.1142/S1758825118500898
- Argatov, I. I., and Chai, Y. S. (2018). Limiting shape of profiles in fretting wear. *Tribol. Int.* 125, 95–99. doi: 10.1016/j.triboint.2018.04.026
- Argatov, I. I., and Chai, Y. S. (2019a). An artificial neural network supported regression model for wear rate. *Tribol. Int.* 138, 211–214. doi: 10.1016/j.triboint.2019.05.040
- Argatov, I. I., and Chai, Y. S. (2019b). Effective wear coefficient and wearing-in period for a functionally graded wear-resisting punch. *Acta Mech.* 230, 2295–2307. doi: 10.1007/s00707-019-2366-9
- Argatov, I. I., and Chai, Y. S. (2019c). A note on optimal design of contact geometry in fretting wear. *Int. J. Mech. Mater. Design* 16, 415–422. doi: 10.1007/s10999-019-09467-9
- Argatov, I. I., and Chai, Y. S. (2020a). Artificial neural network modeling of sliding wear. *J. Proc. Instit. Mech. Eng. J.* doi: 10.1177/1350650120925582. [Epub ahead of print].
- Argatov, I. I., and Chai, Y. S. (2020b). Wear contact problem with friction: Steady-state regime and wearing-in period. *Int. J. Solids Struct.* 193–194:213–221. doi: 10.1016/j.ijsolstr.2020.02.019
- Argatov, I. I., and Fadin, Y. A. (2011). A macro-scale approximation for the running-in period. *Tribol. Lett.* 42, 311–317. doi: 10.1007/s11249-011-9775-9
- Argatov, I. I., Gómez, X., Tato, W., and Urchegui, M. A. (2011). Wear evolution in a stranded rope under cyclic bending: Implications to fatigue life estimation. *Wear* 271, 2857–2867. doi: 10.1016/j.wear.2011.05.045
- Bae, J. W., Lee, C. Y., and Chai, Y. S. (2009). Three dimensional fretting wear analysis by finite element substructure method. *Int. J. Precis. Eng. Manufact.* 10, 63–69. doi: 10.1007/s12541-009-0072-6
- Banichuk, N. V., Ragnedda, F., and Serra, M. (2010). Some optimization problems for bodies in quasi-steady state wear. *Mech. Based Design Struct. Mach.* 38, 430–439. doi: 10.1080/15397734.2010.483574
- Borodich, F. M. (1983). Similarity in the problem of contact between elastic bodies. *J. Appl. Math. Mech.* 47, 440–442. doi: 10.1016/0021-8928(83)90077-1
- Borodich, F. M. (1989). Hertz contact problems for an anisotropic physically nonlinear elastic medium. *Strength Mater.* 21, 1668–1676. doi: 10.1007/BF01533408
- Cardoso, R. A., Doca, T., Néron, D., Pommier, S., and Araújo, J. A. (2019). Wear numerical assessment for partial slip fretting fatigue conditions. *Tribol. Int.* 136, 508–523. doi: 10.1016/j.triboint.2019.03.074
- Cattaneo, C. (1938). Sul contatto di due corpi elastici: distribuzione locale degli sforzi. *Rendiconti dell'Accademia nazionale dei Lincei* 27, 342–348, 434–436, 474–478.
- Chai, Y., Lee, C., Bae, J., Lee, S., and Hwang, J. (2005). Finite element analysis of fretting wear problems in consideration of frictional contact. *Key Eng. Mater.* 297–300:1406–1411. doi: 10.4028/www.scientific.net/KEM.297-300.1406
- Chai, Y. S., and Argatov, I. I. (2018). Local tangential contact of elastically similar, transversely isotropic elastic bodies. *Meccanica* 53, 3137–3143. doi: 10.1007/s11012-018-0870-y
- Chai, Y. S., and Argatov, I. I. (2019). Fretting wear accumulation in partial-slip circular Hertzian contact. *Mech. Res. Commun.* 96, 45–48. doi: 10.1016/j.mechrescom.2019.02.005
- Chai, Y. S., and Popov, V. L. (2016). Limiting shape due to fretting wear in an adhesive contact in Dugdale approximation. *Phys. Mesomech.* 19, 378–381. doi: 10.1134/S1029959916040044
- Cheikh, M., Quilici, S., and Caillaud, G. (2007). Presentation of KI-COF, a phenomenological model of variable friction in fretting contact. *Wear* 262, 914–924. doi: 10.1016/j.wear.2006.10.001
- Chen, W. W., and Wang, Q. J. (2008). A numerical model for the point contact of dissimilar materials considering tangential tractions. *Mech. Mater.* 40, 936–948. doi: 10.1016/j.mechmat.2008.06.002
- Ciavarella, M. (1998). Tangential loading of general 3D contacts. *ASME J. Appl. Mech.* 65, 998–1003. doi: 10.1115/1.2791944
- Ciavarella, M., and Demelio, G. (2001). A review of analytical aspects of fretting fatigue, with extension to damage parameters, and application to dovetail joints. *Int. J. Solids Struct.* 38, 1791–1811. doi: 10.1016/S0020-7683(00)00136-0
- Ciavarella, M., and Hills, D. A. (1999). Brief note: Some observations on oscillating tangential forces and wear in general plane contacts. *Eur. J. Mech. A Solids* 18, 491–497. doi: 10.1016/S0997-7538(99)00117-5
- De Mul, J. M., Kalker, J. J., and Fredriksson, B. (1986). The contact between arbitrarily curved bodies of finite dimensions. *J. Tribol.* 108, 140–148. doi: 10.1115/1.3261134
- Dimaki, A. V., Dmitriev, A. I., Chai, Y. S., and Popov, V. L. (2014). Rapid simulation procedure for fretting wear on the basis of the method of dimensionality reduction. *Int. J. Solids Struct.* 51, 4215–4220. doi: 10.1016/j.ijsolstr.2014.08.003
- Dimaki, A. V., Dmitriev, A. I., Menga, N., Papangelo, A., Ciavarella, M., and Popov, V. L. (2016). Fast high-resolution simulation of the gross slip wear of axially symmetric contacts. *Tribol. Trans.* 59, 189–194. doi: 10.1080/10402004.2015.1065529
- Dini, D., Sackfield, A., and Hills, D. A. (2008). An axi-symmetric Hertzian contact subject to cyclic shear and severe wear. *Wear* 265, 1918–1922. doi: 10.1016/j.wear.2008.04.031
- Dmitriev, A. I., Voll, L. B., Psakhie, S. G., and Popov, V. L. (2016). Universal limiting shape of worn profile under multiple-mode fretting conditions: theory and experimental evidence. *Sci. Rep.* 6:23231. doi: 10.1038/srep23231
- Dundurs, J., and Comninou, M. (1980). Shape of a worn slider. *Wear* 62, 419–424. doi: 10.1016/0043-1648(80)90183-0
- Feppon, F., Michailidis, G., Sidebottom, M. A., Allaire, G., Krick, B. A., and Vermaak, N. (2017). Introducing a level-set based shape and topology optimization method for the wear of composite materials with geometric constraints. *Struct. Multidiscipl. Optimizat.* 55, 547–568. doi: 10.1007/s00158-016-1512-4
- Fouvry, S., Kapsa, P., and Vincent, L. (1996). Quantification of fretting damage. *Wear* 200, 186–205. doi: 10.1016/S0043-1648(96)07306-1
- Fouvry, S., Liskiewicz, T., Kapsa, P., Hannel, S., and Sauger, E. (2003). An energy description of wear mechanisms and its applications to oscillating sliding contacts. *Wear* 255, 287–298. doi: 10.1016/S0043-1648(03)00117-0
- Friedrich, K. (1993). “Wear models for multiphase materials and synergistic effects in polymeric hybrid composites,” in *Advances in Composite Tribology, Vol. 8 of Composite Materials Series*, ed K. Friedrich (Amsterdam: Elsevier), 209–273. doi: 10.1016/B978-0-444-89079-5.50010-6
- Galin, L. A. (1976). Contact problems of the theory of elasticity in the presence of wear. *J. Appl. Math. Mech.* 40, 931–936. doi: 10.1016/0021-8928(76)90132-5
- Galin, L. A., and Goriacheva, I. G. (1977). Axisymmetric contact problem of the theory of elasticity in the presence of wear. *J. Appl. Math. Mech.* 41, 826–831. doi: 10.1016/0021-8928(77)90164-2
- Gallego, L., and Nelias, D. (2007). Modeling of fretting wear under gross slip and partial slip conditions. *J. Tribol.* 129, 528–535. doi: 10.1115/1.2736436
- Gallego, L., Nelias, D., and Jacq, C. (2006). A comprehensive method to predict wear and to define the optimum geometry of fretting surfaces. *J. Tribol.* 128, 476–485. doi: 10.1115/1.2194917
- Goryacheva, I. G. (1998). *Contact Mechanics in Tribology*. Dordrecht: Kluwer. doi: 10.1007/978-94-015-9048-8
- Goryacheva, I. G., and Goryachev, A. P. (2006). The wear contact problem with partial slippage. *J. Appl. Math. Mech.* 70, 934–944. doi: 10.1016/j.jappmathmech.2007.01.010
- Goryacheva, I. G., Rajeev, R. T., and Farris, T. N. (2001). Wear in partial slip contact. *J. Tribol.* 123, 848–856. doi: 10.1115/1.1338476
- Heß, M. (2019). A study on gross slip and fretting wear of contacts involving a power-law graded elastic half-space. *Fact. Univers. Ser. Mech. Eng.* 17, 47–64. doi: 10.22190/FUME190121010H
- Hills, D. A., Sackfield, A., and Paynter, R. J. H. (2009). Simulation of fretting wear in halfplane geometries: part I-the solution for long term wear. *J. Tribol.* 131:031401. doi: 10.1115/1.3118785
- Hills, D. A., and Sosa, G. U. (1999). Origins of partial slip in fretting-a review of known and potential solutions. *J. Strain Anal. Eng. Design* 34, 175–181. doi: 10.1243/0309324991513731
- Huajie, L., and Hongzhao, L. (2018). “An overview of the numerical simulation wear models in mechanical systems,” in *2018 IEEE 9th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)* (IEEE), 105–109. doi: 10.1109/ICMIMT.2018.8340430
- Jäger, J. (1995). Axi-symmetric bodies of equal material in contact under torsion or shift. *Arch. Appl. Mech.* 65, 478–487. doi: 10.1007/BF00835661
- Jäger, J. (1996). Stepwise loading of half-spaces in elliptical contact. *ASME J. Appl. Mech.* 63, 766–773. doi: 10.1115/1.2823361

- Jäger, J. (1998). A new principle in contact mechanics. *J. Tribol.* 120, 677–684. doi: 10.1115/1.2833765
- Johnson, K. (1955). Surface interaction between elastically loaded bodies under tangential forces. *Proc. R. Soc. Lond. Ser. A* 230, 531–548. doi: 10.1098/rspa.1955.0149
- Kasarekar, A. T., Bolander, N. W., Sadeghi, F., and Tseregounis, S. (2007). Modeling of fretting wear evolution in rough circular contacts in partial slip. *Int. J. Mech. Sci.* 49, 690–703. doi: 10.1016/j.ijmecsci.2006.08.021
- Khrushchov, M. M. (1974). Principles of abrasive wear. *Wear* 28, 69–88. doi: 10.1016/0043-1648(74)90102-1
- Komogortsev, V. F. (1985). Contact between a moving stamp and an elastic half-plane when there is wear. *J. Appl. Math. Mech.* 49, 243–246. doi: 10.1016/0021-8928(85)90110-8
- Kovalenko, E. V. (1985). Study of the axisymmetric contact problem of the wear of a pair consisting of an annular stamp and a rough half-space. *J. Appl. Math. Mech.* 49, 641–647. doi: 10.1016/0021-8928(85)90085-1
- Kovalenko, E. V. (2001). “Contact problems for coated solids [in Russian],” in *Mechanics of Contact Interactions*, eds I. I. Vorovich and V. M. Aleksandrov (Moscow: Fizmatli), 459–475.
- Kragelsky, I. V. (1965). *Friction and Wear*. London: Butterworths.
- Kragelsky, I. V., Dobychin, M. N., and Komalov, V. S. (2013). *Friction and Wear: Calculation Methods*. Oxford: Pergamon Press.
- Lee, C. Y., Tian, L. S., Bae, J. W., and Chai, Y. S. (2009). Application of influence function method on the fretting wear of tube-to-plate contact. *Tribol. Int.* 42, 951–957. doi: 10.1016/j.triboint.2009.01.005
- Lengiewicz, J., and Stupkiewicz, S. (2013). Efficient model of evolution of wear in quasi-steady-state sliding contacts. *Wear* 303, 611–621. doi: 10.1016/j.wear.2013.03.051
- Li, Q. (2016). Limiting profile of axisymmetric indenter due to the initially displaced dual-motion fretting wear. *Fact. Univers. Ser. Mech. Eng.* 14, 55–61. doi: 10.22190/FUME1601055L
- Li, Q., Forsbach, F., Schuster, M., Pielsticker, D., and Popov, V. L. (2018). Wear analysis of a heterogeneous annular cylinder. *Lubricants* 6:28. doi: 10.3390/lubricants6010028
- Liu, J., Shen, H. M., and Yang, Y. R. (2014). Finite element implementation of a varied friction model applied to torsional fretting wear. *Wear* 314, 220–227. doi: 10.1016/j.wear.2014.01.006
- Ma, L., Eom, K., Geringer, J., Jun, T.-S., and Kim, K. (2019). Literature review on fretting wear and contact mechanics of tribological coatings. *Coatings* 9:501. doi: 10.3390/coatings9080501
- Mao, X., Liu, W., Ni, Y., and Popov, V. L. (2016). Limiting shape of profile due to dual-mode fretting wear in contact with an elastomer. *Proc. Instit. Mech. Eng. C J. Mech. Eng. Sci.* 230, 1417–1423. doi: 10.1177/0954406215619450
- Mary, C., and Fouvry, S. (2007). Numerical prediction of fretting contact durability using energy wear approach: optimisation of finite-element model. *Wear* 263, 444–450. doi: 10.1016/j.wear.2007.01.116
- McColl, I. R., Ding, J., and Leen, S. B. (2004). Finite element simulation and experimental validation of fretting wear. *Wear* 256, 1114–1127. doi: 10.1016/j.wear.2003.07.001
- Meng, H. C., and Ludema, K. C. (1995). Wear models and predictive equations: their form and content. *Wear* 181, 443–457. doi: 10.1016/0043-1648(95)90158-2
- Meng, Y., Xu, J., Jin, Z., Prakash, B., and Hu, Y. (2020). A review of recent advances in tribology. *Friction* 8, 221–300. doi: 10.1007/s40544-020-0367-2
- Mindlin, R. D. (1949). Compliance of elastic bodies in contact. *ASME J. Appl. Mech.* 16, 259–268.
- Mróz, Z., and Stupkiewicz, S. (1994). An anisotropic friction and wear model. *Int. J. Solids Struct.* 31, 1113–1131. doi: 10.1016/0020-7683(94)90167-8
- Nowell, D. (2010). Simulation of fretting wear in half-plane geometries-Part II: analysis of the transient wear problem using quadratic programming. *J. Tribol.* 132:021402. doi: 10.1115/1.4000733
- Páczelt, I., and Mróz, Z. (2007). Optimal shapes of contact interfaces due to sliding wear in the steady relative motion. *Int. J. Solids Struct.* 44, 895–925. doi: 10.1016/j.ijsolstr.2006.05.027
- Peigney, M. (2004). Simulating wear under cyclic loading by a minimization approach. *Int. J. Solids Struct.* 41, 6783–6799. doi: 10.1016/j.ijsolstr.2004.05.022
- Popov, V. L. (2014). Analytic solution for the limiting shape of profiles due to fretting wear. *Sci. Rep.* 4:3749. doi: 10.1038/srep03749
- Popov, V. L., and Heß, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Heidelberg: Springer. doi: 10.1007/978-3-642-53876-6
- Serre, I., Bonnet, M., and Pradeilles-Duval, R.-M. (2001). Modelling an abrasive wear experiment by the boundary element method. *Comptes Rendus de l'Académie des Sciences Series IIB Mechanics* 329, 803–808. doi: 10.1016/S1620-7742(01)01402-7
- Sfantos, G. K., and Aliabadi, M. H. (2006). Application of BEM and optimization technique to wear problems. *Int. J. Solids Struct.* 43, 3626–3642. doi: 10.1016/j.ijsolstr.2005.09.004
- Vázquez, J., Sackfield, A., Hills, D., and Domínguez, J. (2010). The mechanical behaviour of a symmetrical punch with compound curvature. *J. Strain Anal. Eng. Design* 45, 209–222. doi: 10.1243/03093247JSA633
- Vingsbo, O., and Söderberg, S. (1988). On fretting maps. *Wear* 126, 131–147. doi: 10.1016/0043-1648(88)90134-2
- Wang, Z., Jin, X., Keer, L. M., and Wang, Q. (2013). Novel model for partial-slip contact involving a material with inhomogeneity. *J. Tribol.* 135:041401. doi: 10.1115/1.4024548
- Wang, Z., Yu, C., and Wang, Q. (2015). An efficient method for solving three-dimensional fretting contact problems involving multilayered or functionally graded materials. *Int. J. Solids Struct.* 66, 46–61. doi: 10.1016/j.ijsolstr.2015.04.010
- Willert, E., Dmitriev, A. I., Psakhie, S. G., and Popov, V. L. (2019). Effect of elastic grading on fretting wear. *Sci. Rep.* 9:7791. doi: 10.1038/s41598-019-44269-1
- Yen, B. K., and Dharan, C. K. H. (1996). A model for the abrasive wear of fiber-reinforced polymer composites. *Wear* 195, 123–127. doi: 10.1016/0043-1648(95)06804-X
- Yue, T., and Wahab, M. A. (2019). A review on fretting wear mechanisms, models and numerical analyses. *Comput. Mater. Contin.* 59, 405–432. doi: 10.32604/cmc.2019.04253
- Zhang, Y., Kovalev, A., and Meng, Y. (2018). Combined effect of boundary layer formation and surface smoothing on friction and wear rate of lubricated point contacts during normal running-in processes. *Friction* 6, 274–288. doi: 10.1007/s40544-018-0228-4
- Zmitrowicz, A. (2006). Wear patterns and laws of wear—a review. *J. Theor. Appl. Mech.* 44, 219–253.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Argatov and Chai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Theoretical Study of Fretting Wear Rate Evolution in Axi-Symmetrical Elastic Contact Using the Method of Dimensionality Reduction

Andrey V. Dimaki*

Laboratory of Computer-Aided Design of Materials, Institute of Strength Physics and Materials Science of Siberian Branch of Russian Academy of Sciences, Tomsk, Russia

OPEN ACCESS

Edited by:

Alessandro Ruggiero,
University of Salerno, Italy

Reviewed by:

Jiliang Mo,
Southwest Jiaotong University, China
Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine

*Correspondence:

Andrey V. Dimaki
dav@ispms.tsc.ru

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 28 May 2020

Accepted: 26 August 2020

Published: 09 October 2020

Citation:

Dimaki AV (2020) Theoretical Study of
Fretting Wear Rate Evolution in
Axi-Symmetrical Elastic Contact Using
the Method of Dimensionality
Reduction.
Front. Mech. Eng. 6:566703.
doi: 10.3389/fmech.2020.566703

We simulated wear in elastic tangential contact in partial-slip mode using the method of dimensionality reduction. The obtained numerical dependencies of wear rate on the number of loading cycles were approximated with existing analytical dependencies; at that, the estimated values of parameters of approximating equations are close to analytical estimates given before. The present results demonstrate the possibility of application of the method of dimensionality reduction to the theoretical study of fretting wear rate evolution.

Keywords: fretting, wear rate, elastic contact, method of dimensionality reduction, numerical simulation

INTRODUCTION

Fretting represents a specific kind of wear in contact zones appearing under multi-cycle tangential loading with relatively small amplitude providing no gross sliding to take place. The mentioned loading conditions are typical for a wide range of machine parts and joints subject to vibrations. Wear mode is determined by many parameters and conditions, including physical-mechanical properties of a material, sliding velocity, contact pressure, local, and ambient temperatures, etc. (Odfalk and Vingsbo, 1990; Goryacheva et al., 2001; Matikas and Nicolaou, 2009; Leonard et al., 2012). Variation of these parameters may lead to wear rate change by orders of magnitude (Lim and Ashby, 1987). This underlines the complexity of the problem as well as the actuality of the application of theoretical methods allowing its efficient parametric study that is difficult to carry out in full-scale experiments (Kasarekar et al., 2007).

The theoretical solution of a wear problem requires a calculation of normal and tangential stresses in contact area as well as the evolution of the shape of contacting bodies during the wear process (Dimaki et al., 2014). An equation, describing the local rate of irreversible contact shape change, represents a local wear law. This equation postulates a dependence of local wear rate on material parameters, stress-strain state, and loading conditions in a local area of contact (Popov, 2017).

The well-known assumption of wear rate proportionality to the ratio of dissipated energy to material hardness σ_0 was first proposed by Reye (1860). In the present paper, we use the wear law, given in Archard and Hirst (1956) and Rabinowicz (1995) and based on the hypothesis of direct proportionality between the wear rate and work of frictional forces.

The local form of this wear law for an axially symmetrical contact area reads

$$\Delta f(r) = \frac{k_{\text{wear}}}{\sigma_0} \tau(r) \left(\Delta u_x^{(0)} - \Delta u_x^{(3D)}(r) \right). \quad (1)$$

where $\Delta f(r)$ —local change of the contact profile $f(r)$; r —polar radius in the contact plane; $u_x^{(0)}$ —relative tangential displacement of the contacting bodies; σ_0 —hardness of a material; $u_x^{(3D)}(r)$ —a portion of relative displacement due to elastic deformation of the medium; $\tau(r)$ —shear stress in the contact plane; and Δ —indicates an increment of a corresponding value during a time step. Note that Equation (1) contains a non-dimensional wear coefficient k_{wear} that includes a combination of material parameters determining the local wear rate.

As a rule, the wear process described by Equation (1) is steady because contact pressure distribution tends to uniform in the course of wearing. This is because the wear rate is higher in contact patches subject to higher normal pressure. At that, it is necessary to explicitly take into account the evolution of contact area and size (Dimaki et al., 2016).

In the paper, we use a theoretical model of wear based on the method of dimensionality reduction (Popov and Hess, 2015) and wear law (1). The model allows obtaining an exact solution to the wear problem that agrees with the analytical solutions of Galin (1961) and (Sneddon, 1965).

As it is seen from Equation (1), wear intensity varies over contact radius and time. The latter is due to time variation (as the number of cycles increases) of distributions of shear stress and strain in the contact area. Theoretical and experimental studies show that wear rate changes non-monotonically—in the beginning of the wear process, the wear rate increases, then reaches a maximum and gradually decreases to zero (as a contact profile approaches so-called “limiting” profile (Popov, 2014). In the present paper, we obtained theoretical dependencies of wear rate on the loading cycle number for different profiles of contacting bodies of revolution. The results of the theoretical study are compared with known analytical estimations.

MODEL AND SETUP

Let us briefly describe the keystones of the method of dimensionality reduction (Popov and Hess, 2014, 2015). Consider the contact of a three-dimensional body of revolution having the profile $z = f(r)$ and an infinite elastic foundation. The given three-dimensional profile is transformed into a one-dimensional profile $g(x)$ based on the multifactor dimensionality reduction (MDR) rules (Popov and Hess, 2014, 2015):

$$g(x) = |x| \int_0^{|x|} \frac{f'(r) dr}{\sqrt{x^2 - r^2}} \quad (2)$$

A transformation of the one-dimensional profile back into a three-dimensional one reads

$$f(r) = \frac{2}{\pi} \int_0^r \frac{g(x)}{\sqrt{r^2 - x^2}} dx \quad (3)$$

The one-dimensional profile (2) is pressed to a certain depth d into an elastic foundation that represents a set of non-interacting springs having a spatial size Δx . The normal and tangential stiffness of the springs are given by Popov and Hess (2014):

$$\begin{aligned} k_z &= E^* \Delta x \\ k_x &= G^* \Delta x \end{aligned} \quad (4)$$

where E^* is the effective elastic modulus

$$\frac{1}{E^*} = \frac{1 - \nu_1^2}{E_1} + \frac{1 - \nu_2^2}{E_2} \quad (5)$$

and G^* is the effective shear modulus

$$\frac{1}{G^*} = \frac{(2 - \nu_1)}{4G_1} + \frac{(2 - \nu_2)}{4G_2} \quad (6)$$

To satisfy the rules of MDR, we assume the contacting materials satisfy the “elastic similarity” condition:

$$\frac{1 - 2\nu_1}{G_1} = \frac{1 - 2\nu_2}{G_2} \quad (7)$$

that provides an ability to solve the normal and tangential contact problems independently (Johnson, 1987). The vertical displacement of an individual spring in the contact area reads

$$u_z(x) = d - g(x) \quad (8)$$

and the resulting normal force in a spring is given by

$$F_z(x) = E^* \Delta x (d - g(x)). \quad (9)$$

The linear force density is therefore

$$q_z(x) = \frac{F_z(x)}{\Delta x} = E^* u_z(x) = E^* (d - g(x)). \quad (10)$$

The contact radius a can be estimated from the following equation:

$$g(a) = d. \quad (11)$$

Having the value of a , it is possible to calculate the total normal force over the contact area:

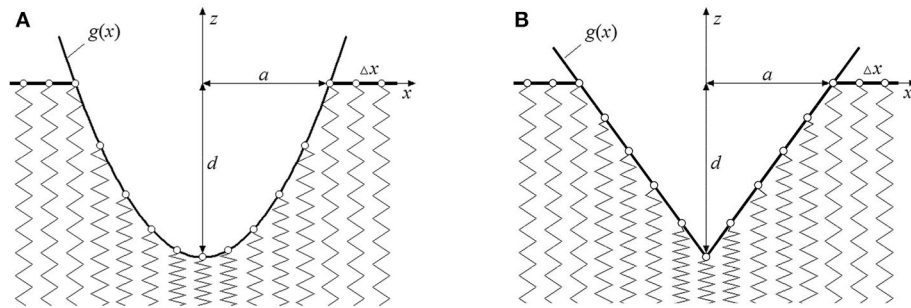


FIGURE 1 | Profiles of contacting rigid bodies and a flat elastic foundation: **(A)** parabolic indenter; **(B)** conical indenter.

TABLE 1 | Estimated values of the parameters of Equation (21), obtained based on the approximation of the numerical simulation results.

Indenter shape	β_1	N_1
Parabolic	3.05	0.34
Conical	2.73	0.21

$$F_z^{total} = 2E^* \int_0^a (d - g(x)) dx. \quad (12)$$

According to the MDR rules, the distribution of normal pressure p in the initial three-dimensional problem can be calculated using the following integral transformation (Popov and Hess, 2014):

$$p(r) = -\frac{1}{\pi} \int_r^\infty \frac{q_z'(x)}{\sqrt{x^2 - r^2}} dx = \frac{E^*}{\pi} \int_r^a \frac{g'(x)}{\sqrt{x^2 - r^2}} dx \quad (13)$$

Assume the indenter moves to displacement $u_x^{(0)}$ in a tangential direction. A one-dimensional spring of the half-space remains in a contact state with the indenter until a tangential force $f_x = k_x u_x^{(0)}$ in the spring reaches a critical value μf_z where μ denotes the friction coefficient. After that, the spring becomes in sliding state, and the tangential force equals to μf_z . This behavior is independent of an initial stress state of a spring and results in the following equations describing the relation between tangential deformation and reaction force of a spring:

$$\begin{aligned} \Delta u_x(x) &= \Delta u_x^{(0)}, \text{ if } |k_x u_x(x)| < \mu f_z \\ u_x(x) &= \pm \frac{\mu f_z(x)}{k_x}, \text{ in the sliding state} \end{aligned} \quad (14)$$

The sign in Equation (14) depends on the direction of the motion of the indenter. The sign “minus” corresponds to motion along the contact plane in a positive direction of axis oX ; the sign “plus” corresponds to motion in the opposite direction. Having a given time dependence of the indenter position, it is possible to

calculate its increment $\Delta u_x^{(0)}$ and thus to determine a tangential deformation of a one-dimensional spring in each point of a contact area. Obviously, the tangential reaction force can also be determined as follows:

$$f_x = k_x u_x(x) = G^* \Delta x \cdot u_x(x). \quad (15)$$

Three-dimensional radial distributions of tangential stresses $\tau(r)$ and displacements $u_x^{(3D)}(r)$ can be calculated by means of the integral transformations similar to (3):

$$u_x^{(3D)}(r) = \frac{2}{\pi} \int_0^r \frac{u_x(x) dx}{\sqrt{r^2 - x^2}}, \quad (16)$$

$$\tau(r) = -\frac{1}{\pi} \int_r^\infty \frac{q_x'(x) dx}{\sqrt{x^2 - r^2}} = -\frac{G^*}{\pi} \int_r^\infty \frac{u_x'(x) dx}{\sqrt{x^2 - r^2}}, \quad (17)$$

where $q_x(x)$ is a tangential force density:

$$q_x(x) = \frac{f_x}{\Delta x} = G^* u_x(x). \quad (18)$$

The distributions of stresses and displacements obtained above represent exact solutions of the corresponding three-dimensional problem, which is one of the main advantages of the MDR (Dimaki et al., 2014; Li et al., 2014). The transformation (2) represents a relationship between the full three-dimensional contact problem and a one-dimensional contact with an elastic foundation (Galin, 1961). Stresses and displacements in the three-dimensional contact problem with a linearly elastic foundation can be obtained using the corresponding integral transformations given earlier. The obtained solution is “exact” and can be applied to contact problems that reduce to a normal contact problem.

In this paper, we consider a contact of parabolic and conical rigid indenter with a flat elastic foundation (see **Figure 1**). The choice of these two shapes is conditioned

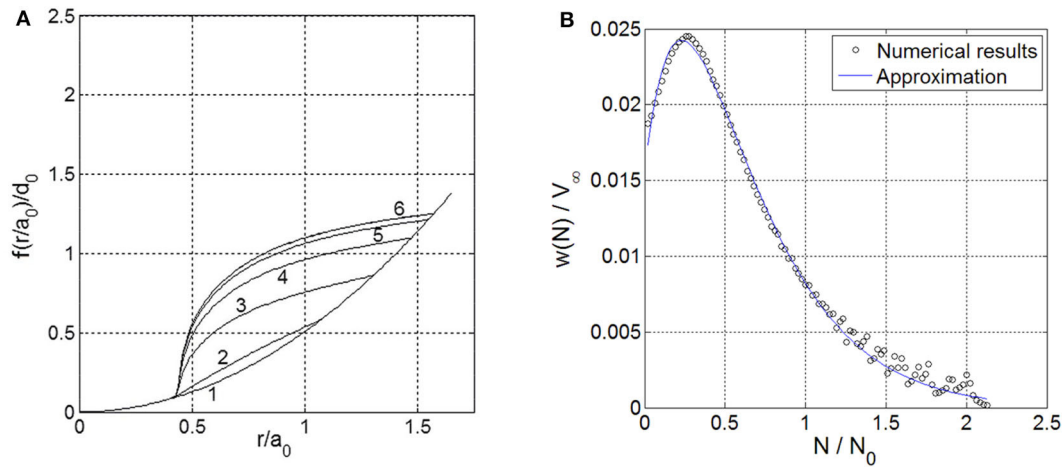


FIGURE 2 | (A) Development of the shape of an initially parabolic profile due to fretting wear. Curve “1” shows the initial profile; further curves correspond to the following normalized number of oscillation cycles: (2) $\tilde{N} = 0.04$; (3) $\tilde{N} = 0.16$; (4) $\tilde{N} = 0.36$; (5) $\tilde{N} = 0.64$; (6) $\tilde{N} = 1$. **(B)** Dependency of normalized wear rate on a dimensionless number of cycles for a parabolic indenter under constant normal load. Points indicate the results of numerical simulation, and solid line indicates the approximation with Equation (21).

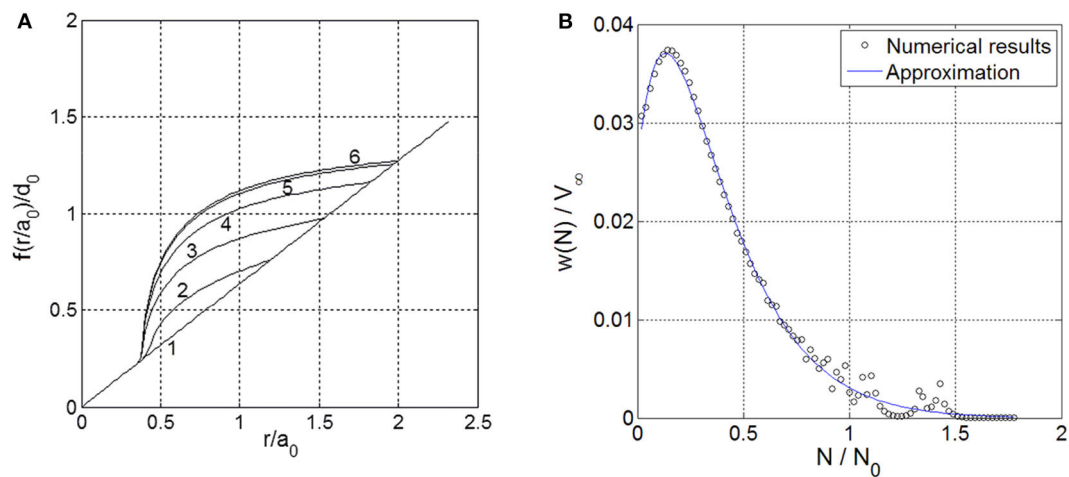


FIGURE 3 | (A) Development of the shape of an initially conical profile due to fretting wear. Curve “1” shows the initial profile; further curves correspond to the following normalized number of oscillation cycles: (2) $\tilde{N} = 0.1$; (3) $\tilde{N} = 0.4$; (4) $\tilde{N} = 0.9$; (5) $\tilde{N} = 1.6$; (6) $\tilde{N} = 2.5$. **(B)** Dependency of normalized wear rate on a dimensionless number of cycles for a conical indenter under a constant normal load. Points indicate the results of numerical simulation, and solid line indicates the approximation with Equation (21).

by the fact that many theoretical approaches use these shapes for describing an asperity shape in single-asperity and multi-asperity models of rough surfaces [see, e.g., Popov (2017)]. For a parabolic indenter with $f(r) = r^2/2R$, the corresponding one-dimensional profile is $g(x) = x^2/R$. For a three-dimensional conical indenter with an initial profile given by $f(r) = r \tan \theta$, the corresponding MDR-image is $g(x) = \frac{\pi}{2} |x| \tan \theta$.

If the indenter is subject to tangential oscillations with an amplitude $U^{(0)}$, the characteristic wear volume per one cycle of oscillation can be roughly estimated as follows: $\Delta f \approx \frac{k}{\sigma_0} \frac{\mu F_N}{\pi a_0^2} U^{(0)}$, where a_0 is the initial contact radius. At that, an estimation of

several cycles needed to reach the wear depth of the indenter of an order of magnitude of d_0 , which reads:

$$N_0 = \frac{d_0}{\Delta f} = \frac{\pi a_0^2 d_0 \sigma_0}{k \mu F_N U^{(0)}}. \quad (19)$$

In the results presented later, an actual number of cycles will be normalized to the characteristic value (19) as follows:

$$\tilde{N} = \frac{N}{N_0}. \quad (20)$$

Further, we operate with the normalized number of cycles (20) to provide universality of the obtained results.

SIMULATION RESULTS AND DISCUSSION

We carried out a theoretical study of the dependence of normalized volume wear rate on normalized cycle number \tilde{N} for parabolic and conical indenters in partial slip mode. We assume that both bodies in contact are elastic, but only the indenter is subject to wear. It is known that for oscillations with a magnitude $< \mu d_0$, the contact area separates into “stick” and “slip” regions (Jäger, 1995; Hills et al., 2009; Lengiewicz and Stupkiewicz, 2013). Wear takes place only outside the “stick” area. At that, in the absence of “gross slip” motion, the wear process finishes at the so-called “shakedown” state. After achieving the “shakedown” state (in other words, after achieving some “final” contact profile), the wear rate becomes equal to zero (Ciavarella and Hills, 1999) and after oscillations proceed in the “absence” of wear. This allows, in particular, to estimate a total wear volume for a given initial contact shape, material parameters, and loading conditions. It is necessary to note that in the mentioned conditions, wear occurs at lower spatial scales (Gnecco and Meyer, 2015).

The results of the performed numerical simulations (see **Figures 2, 3**) show that the wear rate initially grows, approaches a maximum value at a certain value of \tilde{N} , and then decreases to zero that corresponds to the “shakedown.” The similar character of the dependence of wear rate on cycle number $w(N) = \frac{dV}{dN}$ is obtained in experimental and theoretical studies (Kasarekar et al., 2007). In this connection, it is interesting to compare the quantitative estimations of parameters of the dependence $w(N)$, obtained in the developed MDR-based model, with analytic estimations.

Chai and Argatov (2019) give an analytic equation for wear rate against cycle number based on the results of numerical simulation by Kasarekar et al. (2007), which reads:

$$w(N) = w_0 \left[1 + \beta_1 \frac{N}{N_1} \right] \exp \left(-\frac{N}{N_1} \right), \quad (21)$$

where the initial wear rate w_0 is defined as follows (Chai and Argatov, 2019):

$$w_0 = \alpha_V \Delta \xi_0 \quad (22)$$

where $\alpha_V = k_{\text{wear}}/\mu$ and $\Delta \xi_0$ —energy dissipation per first wear cycle. The parameter N_1 can be estimated as follows (Chai and Argatov, 2019):

$$N_1 = \frac{V_\infty}{w_0(1 + \beta_1)}, \quad (23)$$

where V_∞ is a total worn volume (assuming that wear rate tends to zero with an increasing number of cycles).

Simulation results for a parabolic indenter are shown in **Figure 2**. The obtained numerical dependence of wear rate on

the number of cycles was approximated with Equation (21). The values of parameters w_0 , β_1 , and N_1 were estimated by means of the non-linear least-squares method (Marquardt, 1963). It is seen that Equation (21) represents a good approximation of the numerically obtained results. The estimated values of the parameters β_1 and N_1 are given in **Table 1**. In the paper by Chai and Argatov (2019), estimation of the parameter β_1 is given: $\beta_1 = 3.11$. In the present study, estimation of β_1 is about $\beta_1 \approx 3$ for a parabolic indenter, which is in good agreement with the given analytically obtained quantity earlier.

Simulation results for a conical indenter are presented in **Figure 2**. In this case, Equation (21) also allows performing a good approximation of the numerically obtained dependence of wear rate on cycle number. However, for a conical indenter, the estimated value of β_1 is $\beta_1 = 2.73$.

The shape and estimated values of the parameters of the dependence of wear rate on cycle numbers are in agreement with previously obtained analytic estimations (Chai and Argatov, 2019). In general, the results of performed simulations show that the method of dimensionality reduction allows us to adequately describe the dynamics of wear of arbitrary bodies of revolution.

CONCLUSIONS

We have performed numerical simulations of the evolution of wear rate of elastic bodies of revolution within a theoretical model based on the method of dimensionality reduction. It was shown that numerical dependencies of wear rate on a normalized number of cycles coincide with analytical equations for such dependencies. For a parabolic indenter, estimations of the parameter $\beta_1 \approx 3.05$ obtained in the present study and the paper (Chai and Argatov, 2019) agree. For a conical indenter, the estimation of β_1 is $\beta_1 \approx 2.73$. A detailed description of the indenter shape that influences on wear rate evolution requires further studies. Summarizing, we can confirm that the present results demonstrate the applicability of the method of dimensionality reduction for analysis of wear dynamics, including quantitative prediction of wear rate and a total volume of worn material.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

AD developed the model, conducted the simulations, and wrote the manuscript.

ACKNOWLEDGMENTS

The author cordially acknowledges Profs. Valentin Popov and Ivan Argatov for fruitful discussions. The work was performed according to the Government research assignments for ISPMS SB RAS, Project No. III.23.1.4.

REFERENCES

- Archard, J. F., and Hirst, W. (1956). The wear of metals under unlubricated conditions. *Proc. R. Soc. London A* 236, 397–410. doi: 10.1098/rspa.1956.0144
- Chai, Y. S., and Argatov, I. I. (2019). Fretting wear accumulation in partial-slip circular Hertzian contact. *Mech. Res. Comm.* 96, 45–48. doi: 10.1016/j.mechrescom.2019.02.005
- Ciavarella, M., and Hills, A. (1999). Brief Note: some observations on the oscillating tangential forces and wear in general plane contacts. *Eur. J. Mech.* 18, 491–497. doi: 10.1016/S0997-7538(99)00117-5
- Dimaki, A. V., Dmitriev, A. I., Chai, Y. S., and Popov, V. L. (2014). Rapid simulation procedure for fretting wear on the basis of the method of dimensionality reduction. *Int. J. Solids Struct.* 51, 4215–4220. doi: 10.1016/j.ijsolstr.2014.08.003
- Dimaki, A. V., Dmitriev, A. I., Menga, N., Papangelo, A., Ciavarella, M., and Popov, V. L. (2016). Fast high-resolution simulation of the gross slip wear of axially symmetric contacts. *Tribol. Trans.* 59, 189–194. doi: 10.1080/10402004.2015.1065529
- Galin, L. A. (1961). *Contact Problems in the Theory of Elasticity*. Raleigh, NC: North Carolina State College.
- Gnecco, E., and Meyer, E. (2015). *Fundamentals of Friction and Wear on Nanoscale*. Cham: Springer. doi: 10.1007/978-3-319-10560-4
- Goryacheva, I., Rajeev, R., and Farris, T. (2001). Wear in partial slip contact. *Int. J. Precis. Engng. Manuf.* 123, 848–856. doi: 10.1115/1.1338476
- Hills, D. A., Sackfield, A., and Paynter, R. J. H. (2009). Simulation of fretting wear in half-plane geometries: part I – the solution for long term wear. *J. Tribol.* 131:031401. doi: 10.1115/1.3118785
- Jäger, J. (1995). Axi-symmetric bodies of equal material in contact under torsion or shift. *Arch. Appl. Mech.* 65, 478–487. doi: 10.1007/BF00835661
- Johnson, K. L. (1987). *Contact Mechanics*. Cambridge: Cambridge University Press.
- Kasarekar, A., Bolander, N., Sadeghi, F., and Tseregounis, S. (2007). Modeling of fretting wear evolution in rough circular contacts in partial slip. *Int. J. Mech. Sci.* 49, 690–703. doi: 10.1016/j.ijmecsci.2006.08.021
- Lengiewicz, J., and Stupkiewicz, S. (2013). Efficient model of evolution of wear in quasi-steady-state sliding contacts. *Wear* 303, 611–621. doi: 10.1016/j.wear.2013.03.051
- Leonard, B. D., Sadeghi, F., Shinde, S., and Mittelbach, M. (2012). A numerical and experimental investigation of fretting wear and a new procedure for fretting wear maps. *Tribol. Trans.* 55, 313–324. doi: 10.1080/10402004.2012.654598
- Li, Q., Filippov, A. E., Dimaki, A. V., Chai, Y. S., and Popov, V. L. (2014). Simplified simulation of fretting wear using the method of dimensionality reduction. *Phys. Mesomech.* 17, 236–241. doi: 10.1134/S1029959914030102
- Lim, S. C., and Ashby, M. F. (1987). Overview no. 55. Wear-Mechanism maps. *Acta Metallurg.* 35, 1–24. doi: 10.1016/0001-6160(87)90209-4
- Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.* 11, 431–441. doi: 10.1137/0111030
- Matikas, T. E., and Nicolaou, P. D. (2009). Prediction of contact temperature distribution during fretting fatigue in titanium alloys. *Tribol. Trans.* 52, 346–353. doi: 10.1080/10402000802563117
- Odfalk, M., and Vingsbo, O. (1990). Influence of normal force and frequency in fretting. *Tribol. Trans.* 33, 604–610. doi: 10.1080/10402009008981995
- Popov, V. L. (2014). Analytic solution for the limiting shape of profiles due to fretting wear. *Sci. Rep.* 4:3749. doi: 10.1038/srep03749
- Popov, V. L. (2017). *Contact Mechanics and Friction. Physical principles and applications: 2nd Edn*. Berlin: Springer. doi: 10.1007/978-3-662-53081-8
- Popov, V. L., and Hess, M. (2014). Method of dimensionality reduction in contact mechanics and friction: a user's handbook. I. Axially-symmetric contacts. *Facta Univ. Mech. Engng.* 12, 1–14. Available online at: <http://casopisi.junis.ni.ac.rs/index.php/FUMechEng/article/view/155>
- Popov, V. L., and Hess, M. (2015). *Method of Dimensionality Reduction in Contact Mechanics and Friction*. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-53876-6
- Rabinowicz, E. (1995). *Friction and Wear of Materials*. New York, NY: John Wiley and Sons.
- Reye, T. (1860). Zur Theorie der Zapfenreibung. *Der Civilingenieur* 4, 235–255.
- Sneddon, I. N. (1965). The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *Int. J. Engng. Sci.* 3, 47–57. doi: 10.1016/0020-7225(65)90019-4

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Dimaki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Adhesion Between Rigid Indenter and Soft Rubber Layer: Influence of Roughness

Iakov A. Lyashenko^{1,2*} and Roman Pohrt¹

¹ Department of System Dynamics and Friction Physics, Technische Universität Berlin, Berlin, Germany, ² Department of Applied Mathematics and Complex Systems Modelling, Sumy State University, Sumy, Ukraine

We report a series of experiments on the indentation of steel indenters into a soft layer of transparent rubber with relatively high adhesion. The roughness properties of the steel indenters are varied by undergoing preparation using sandpaper with different grain sizes. Starting from a smooth surface, additional roughness increases the adhesive strength up to a critical roughness value, after which it significantly decreases. Furthermore, we look at the evolution of the contact area during slow indentation and detachment. It was found that, during indentation, the contact area changes more sharply compared to detachment (pull-off).

Keywords: surface energy, contact area, adhesive contact, indentation experiments, hysteresis, rough surfaces

OPEN ACCESS

Edited by:

Marco Paggi,
IMT School for Advanced Studies
Lucca, Italy

Reviewed by:

Michele Ciavarella,
Politecnico di Bari, Italy
Antonio Papangelo,
Politecnico di Bari, Italy

*Correspondence:

Iakov A. Lyashenko
i.lyashenko@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 09 March 2020

Accepted: 03 June 2020

Published: 29 July 2020

Citation:

Lyashenko IA and Pohrt R (2020)
Adhesion Between Rigid Indenter and
Soft Rubber Layer: Influence of
Roughness. *Front. Mech. Eng.* 6:49.
doi: 10.3389/fmech.2020.00049

INTRODUCTION

Roughness plays a significant role for the adhesive strength of contacts between materials. The roughness of many surfaces in nature is of the fractal type (Persson, 2014). When such surfaces approach one another, the highest asperities are first to come into contact. For hard materials, such as metals or rocks, the load is then carried by a few peaks alone. As a result, the real contact area is much smaller than the apparent contact area (Persson, 2006). Because adhesive interactions are extremely short-ranged, they effectively only act in the vicinity of real contact patches. For that reason, the observed adhesion force between hard objects is very small even at high values of surface energy. When at least one of the materials is very soft, like rubber, and the roughness is not too high, a much larger real contact area can be realized, with the rubber partially filling the gaps between roughness peaks. When compared to a perfectly smooth surface, one with mild roughness was shown to increase the adhesive toughness. This can be explained by the increased effective surface in contact or by the instability of the detachment process, when roughness peaks effectively hinder the advancing of the detachment front. In Guduru (2007), this was evidenced by choosing a special roughness in the form of axisymmetric waviness on the indenter surface. Indeed during detachment, multiple stable configurations of the contact succeed one another. The load–displacement curve is not monotonous and not continuous (Guduru, 2007; Jin et al., 2013). A similar effect can be observed when otherwise flat surfaces are equipped with shallow dimples. Introducing a moderate superposed roughness here can affect the total adhesive strength in both positive and negative ways (Papangelo and Ciavarella, 2017, 2018).

The existence of an adhesive enhancement for moderate roughness leads to the conclusion that there should be optimal surfaces with roughness properties, at which the adhesive strength of the contact has the highest value. Due to the importance of rough contacts for industry and daily life, there are many works done by different scientific groups with investigations of the particularities of contact between rough surfaces in the presence of adhesion. There are different experiments

(Benz et al., 2006; Tiwari et al., 2017; Liu et al., 2018; Sahli et al., 2019), theories (Johnson, 1995; Persson, 2006; Persson and Scaraggi, 2014; Ciavarella, 2015; Papangelo et al., 2019), as well as computer simulations (Pastewka and Robbins, 2014; Rey et al., 2017; Li et al., 2019). A short historical review of the development of adhesion theories can be found in Popova and Popov (2018). The majority of theories for rough contacts are based on JKR-type adhesion interaction model (Johnson et al., 1971). For smooth spherical contacts, the JKR model predicts the dependency of the normal force F_N on the indentation depth d to be a single universal curve, independently from whether the motion takes place in the direction of approach or detachment. In reality, an adhesive hysteresis (Deng and Kesari, 2019) is typically observed. This hysteresis leads to energy dissipation in any indentation–detachment cycle.

In the present paper, we experimentally investigate the behavior at indentation and at detachment of spherical surfaces with different parameters of roughness.

PROBE PREPARATION

In our experiments, we used spherical steel indenters with radius $R = 30, 40, 50$, and 100 mm. With each indenter, we conducted 12 experiments with different roughness values of its surface. The roughness was changed with sandpaper *Matador* with different grit numbers, which correspond to the different sizes of grains. For the first experiments, the surfaces were prepared using P2000 sandpaper and had almost mirror-like properties. After processing with sandpaper, the indenter was cleaned in the chemical solution EMAG EM-404 for 16 min, using the ultrasonic cleaner Codyson CD-4800, before going into the adhesion experiment.

After each experiment, we measured the 3D topography of the indenter surface using a KEYENCE VK-100 series microscope, at $\times 10$ magnification. We measured 200–300 single images separately, with size of $1,398 \text{ mm} \times 1,048 \text{ mm}$ each. These parts had resolution of $1,024 \times 768$, with fixed horizontal distance between points $\Delta x = \Delta y = 1,365 \text{ nm}$. To exclude the macroscopic curvature in our roughness characterization, we subtracted from each part a polynomial shape of power four, which we obtained by the least square method. **Figure 1** depicts a typical view of the measured inclined surface (left panel) and after subtraction of the polynomial shape (right panel).

After completion of the experiments for one roughness value, we used the next sandpaper with bigger corn size and repeated the same for all sandpapers listed in **Table 1**. After the last sandpaper P40 treatment, we conducted two additional experiments with even greater roughness. This was obtained by manual treatment with a hacksaw and the subsequent removal of some features. (“P0” refers to the hacksaw surface; “P0+180” is for additional processing of indenter P0 after experiment with sandpaper P180).

EXPERIMENTAL SETUP

The freshly prepared indenter is pushed into the rubber base to a maximal indentation depth of 0.4 mm (0.3 mm for $R = 100 \text{ mm}$). Then, it is pulled off until full detachment, when

complete vanishing of the contact is realized. The rubber base consists of a 5-mm-thick layer of optically smooth, clear rubber TARNAC CRG N3005, located on a surface of silicate glass. No special measures were taken to fixate the rubber to the glass due to its good adhesive properties. Throughout the experiment, the contact was observed from below with a digital camera [see **Figure 2A** (pos. 8)]. The indenters were moved in normal direction with linear actuator PI M-403.2DG (pos. 1 in **Figure 2A**), controlled by controller PI C-863 (pos. 7). The experiments were done with distance control, not force control, and a constant velocity of $v = 1 \text{ } \mu\text{m/s}$ was used in all the experiments. The normal force was measured by the force sensor ME K3D40 (pos. 3) with amplifier GSV-1A4 SubD37/2, which was connected to a PC with 16-bit NI USB-6211 analog to a digital converter. While the rubber layer can be tilted in two horizontal axis (with tilting mechanism on pos. 5) for use in flat-indenter experiments (Popov et al., 2017), such adjustments were not necessary for the current investigation with spheres. All measurements and saving of the obtained displacement, normal force, and contact area were done within a LabVIEW computer program.

The lighting of the contact area was realized using 80 LEDs, which illuminated the contact zone from all four sides (see **Figure 2B**). Because the light enters the rubber at a low angle, a good contrast between contact and non-contact zones is visible in the camera image. We employed a computer algorithm based on differences in pixel intensity to differentiate the two. Even though from the analysis of the digital pictures we technically obtained the *visible* contact area, we assume that it represents the *real* contact area with reasonable accuracy.

While the experimental equipment is capable of investigating contact phenomena in the presence of tangential motion (see 2nd linear stage in **Figure 2A**, pos. 2 and 6), this capacity was not used in the current work.

We investigated the situation with a very slow vertical indenter velocity $v = 1 \text{ } \mu\text{m/s}$. This was done to best achieve a quasi-static situation and allowed to use theories for static contacts (Hertz, JKR) to describe the contact properties. For faster motion, the velocity affects the obtained results critically, for instance, *via* viscoelastic properties. The effect of different velocities was investigated in Waters and Guduru (2010) for tangential contact and in Lyashenko and Popov (2019) for the setup covered in this paper.

EXPERIMENTAL RESULTS

Adhesive Strength

After preparation of the indenter for each roughness, we did three cycles of indentation in a series, without cleaning or changing the indenter properties between these cycles.

During each cycle, a series of digital images was recorded. In **Figure 3**, sample pictures of the contact area at different roughness of the indenter are shown.

For each experiment, the dependencies of the normal force F_N and contact area A on the indentation depth d followed the general behavior as shown in **Figure 4**. The blue curves with negative values of the indentation depth and normal

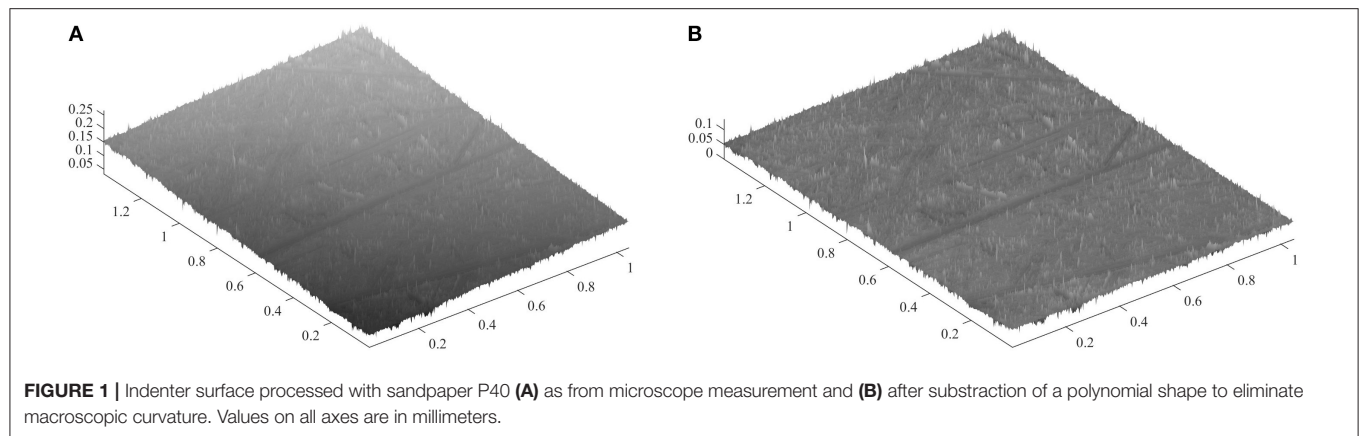
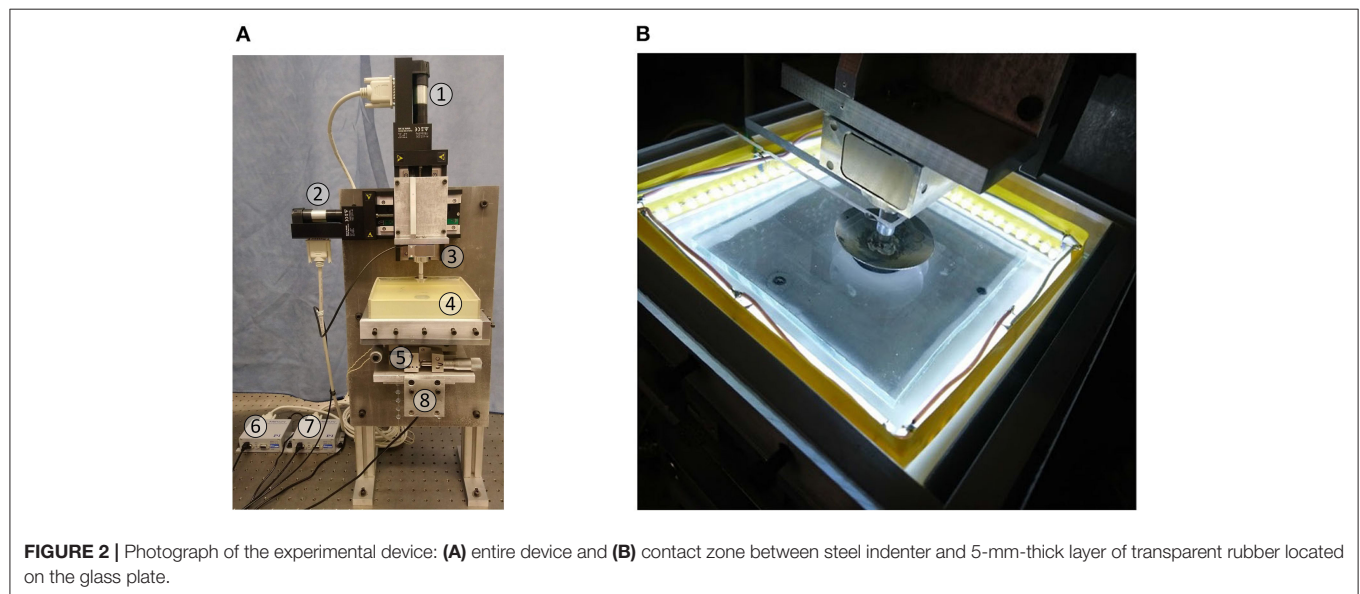


TABLE 1 | Average size of the grains of sandpaper according to standard ISO-6344.

No.	P2000	P1500	P1000	P800	P400	P320	P180	P80	P60	P40
Size (μm)	10.3	12.6	18.3	21.8	35	46.2	82	201	269	425



force correspond to the detachment process (pull-off). At this phase, adhesion plays a significant role; otherwise, $F_N < 0$ cannot be observed. The absolute value of minimal force F_N is called *adhesion strength* of the contact. In a force-controlled experiment, this value would have to be exceeded in order to separate the surfaces (Johnson et al., 1971).

During indentation (dashed red lines), the normal force F_N has practically only positive values because the influence of adhesion at indentation is very weak. This phase can thus be approximated by the Hertz contact theory (Hertz, 1881) and the normal force can thus be calculated with the well-known formula $F_N = 4E^*a^3/(3R)$. Here $a = (Rd)^{1/2}$ is the contact radius, R is the radius of spherical indenter, $E^* = E/(1-\nu^2)$ is the reduced value of elastic modulus, E is the elastic modulus,

and ν is the Poisson ratio. However, these formulas are correct only in half-space approximation. In our experiments, we used a rubber layer with thickness $h = 5$ mm. Therefore, the normal force is always higher compared to the half-space solution. It is necessary to do additional simulations to obtain the correct dependencies $F_N(d)$. We did such simulations in the framework of boundary elements method (BEM) for layered systems (Li et al., 2020) based on JKR-type interaction between contacted bodies. For our rubber material, we assumed Poisson ratio $\nu = 0.47$ (near ratio 0.5 for incompressible elastomers). From a comparison between the experimental results and the theory for different cylindrical and spherical indenters, we have found the approximate value of the rubber elastic modulus to be $E \approx 0.324 \cdot 10^6$ Pa (Lyashenko and Popov, 2019). Our experiments

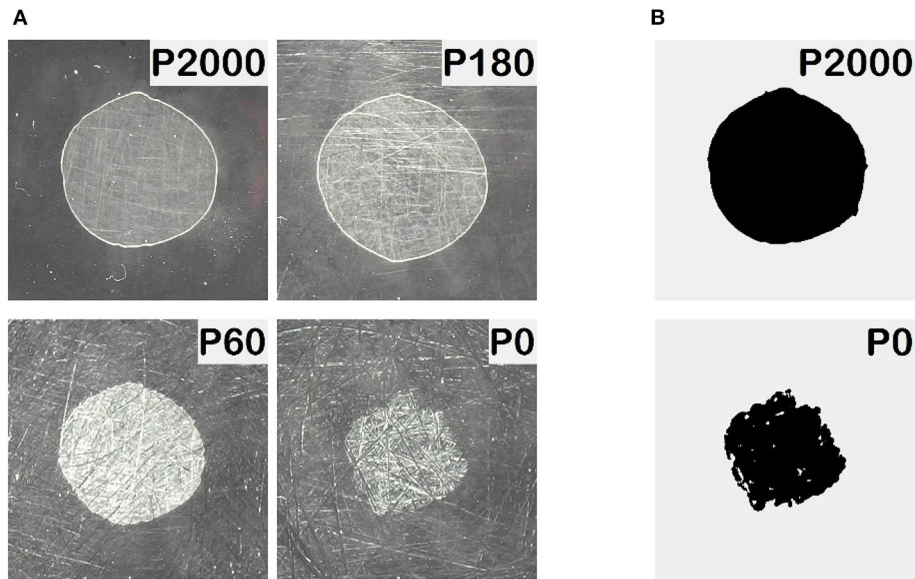


FIGURE 3 | (A) Pictures of the contact area at different roughness of indenter with radius $R = 100$ mm and zero indentation depth $d = 0$ mm during phase of detachment. **(B)** Visible contact area as determined by the computer program.

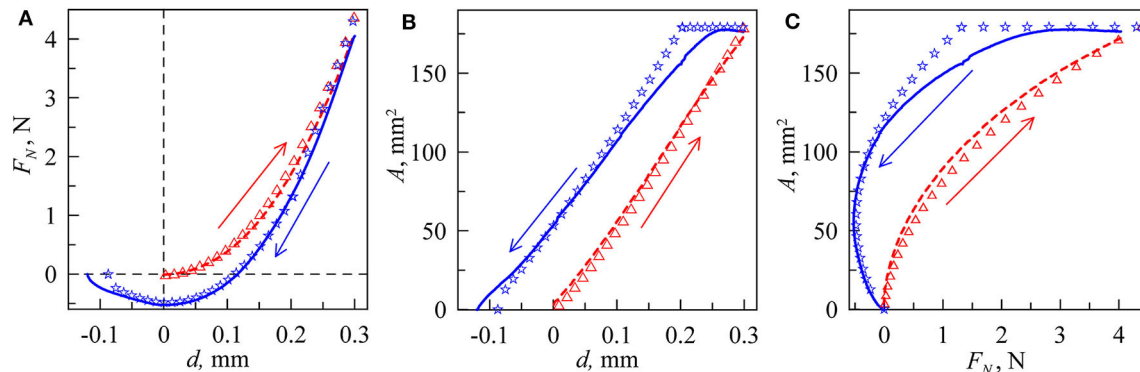


FIGURE 4 | General behavior of main macroscopic parameters during indentation (dashed red lines) and pull-off (solid blue lines): **(A)** dependence of normal force F_N vs. indentation depth d , **(B)** dependence of contact area A vs. indentation depth d , **(C)** dependence of contact area A vs. normal force F_N . These dependencies were obtained in real experiment in the case of P2000 for the indenter of the radius $R = 100$ mm. With symbols, the results of the computer simulation are shown: triangles for absence of adhesive interaction $\gamma_{12} = 0$ J/m² and stars for the value $\gamma_{12} = 0.9$ J/m².

show that, during the indentation phase, adhesion does not play a significant role, and in the point of the first contact, the contact area barely spreads. The normal force F_N shows small negative values near zero at the beginning of the indentation phase. In all plots in **Figure 4**, the results of BEM simulations are shown by triangles. In these simulations, the value of specific work of adhesion γ_{12} was set to zero during indentation, so the contact shown is effectively adhesionless. What remains is the Hertz contact, but generalized to the situation of a finite elastic layer. From this figure, it can be seen that a fairly good coincidence between simulations and experimental results is realized. In the phase of detachment, we approximated the experimental results with the curve, obtained in BEM simulations

with the value $\gamma_{12} = 0.9$ J/m². These results are shown in **Figure 4A** with stars. The situation of indentation without adhesion and detachment with adhesion is similar to the case presented in Figure 3 in Ciavarella et al. (2019). When the direction of motion is changed from indentation to pull-out, the contact area remains constant for some time, similar to what was observed in Waters and Guduru (2010). Such type of experimental behavior can be explained within the framework of the approach of “chemical heterogeneity” of the contact area (Popov, 2020).

In **Figure 5**, the dependencies of normal force F_N vs. indentation depth d for indenters with different roughness are shown (see also **Supplementary Video 1**). As can be seen, all

curves show a similar behavior in the indentation phase, but not in the detachment phase. Some curves at the indentation phase decline from universal behavior, for example, curves P0 and P0 + 180 for the case of indenter $R = 50$ mm. Since we set $d = 0$ at the point of the first contact, this corresponds only to the contact

of the highest roughness peaks at a large roughness amplitude. For the detachment phase, a simple trend in the dependency on the roughness is not immediately visible. In **Figure 6**, we show the adhesive strength (absolute value of minimal normal force at phase of detachment) for all indenters as a function of sandpaper

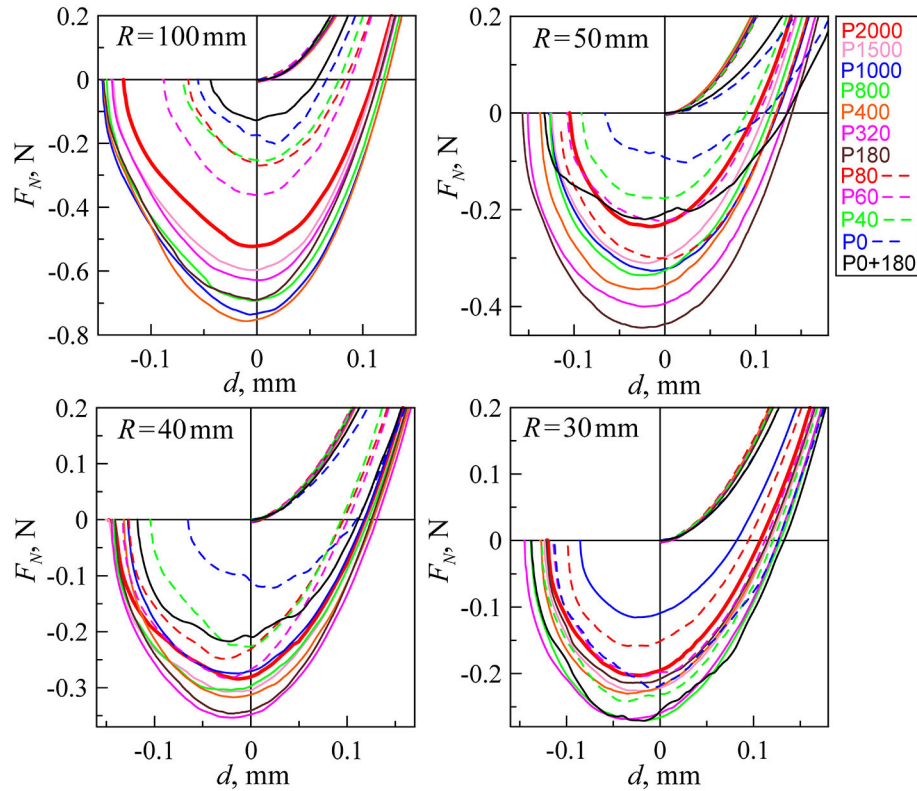


FIGURE 5 | Dependencies of normal force F_N vs. indentation depth d for all experiments (only first indentation cycles), with indenters of radii at 100, 50, 40, and 30 mm. Each plot has 12 dependencies for different roughness. The solid red lines in all figures correspond to the situation with smallest roughness P2000.

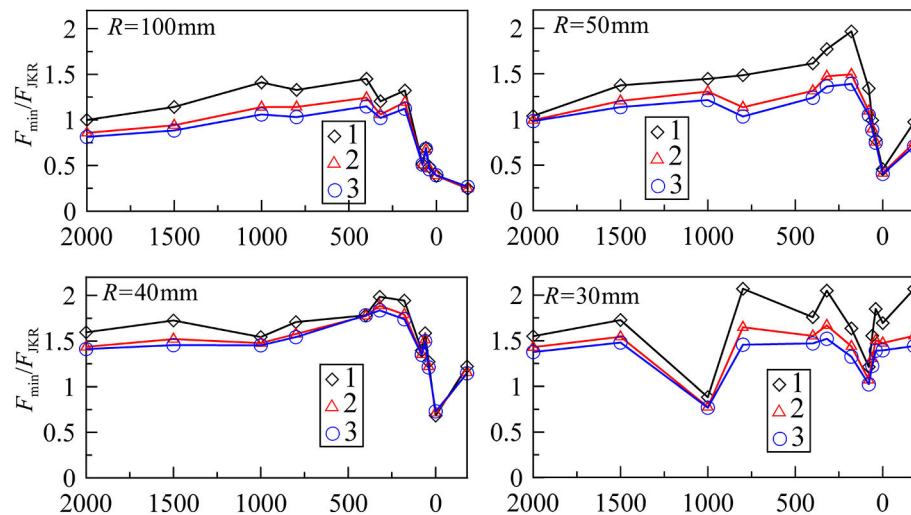


FIGURE 6 | Dependencies of adhesive strength measured in experiments vs. grit number of sandpaper for indenters with radius $R = 100, 50, 40$, and 30 mm. The symbols with numbers 1, 2, and 3 in the figure are first (diamonds), second (triangles), and third (circles) cycles of indentation.

number. Please note that the far-right case at -180 corresponds to our “P0 + 180” sample.

In **Figure 6**, we show the normalized values of normal forces F_{\min}/F_{JKR} . In the case of the half-space, JKR theory gives the

value of minimal adhesive force with the simple formula $F_{\text{JKR}} = -3\pi R\gamma_{12}/2$, and for the cases $R = 100, 50, 40$, and 30 mm, we obtained the critical forces $F_{\text{JKR}} = -0.424, -0.212, -0.170$, and -0.127 N, respectively. To find these values, we used surface

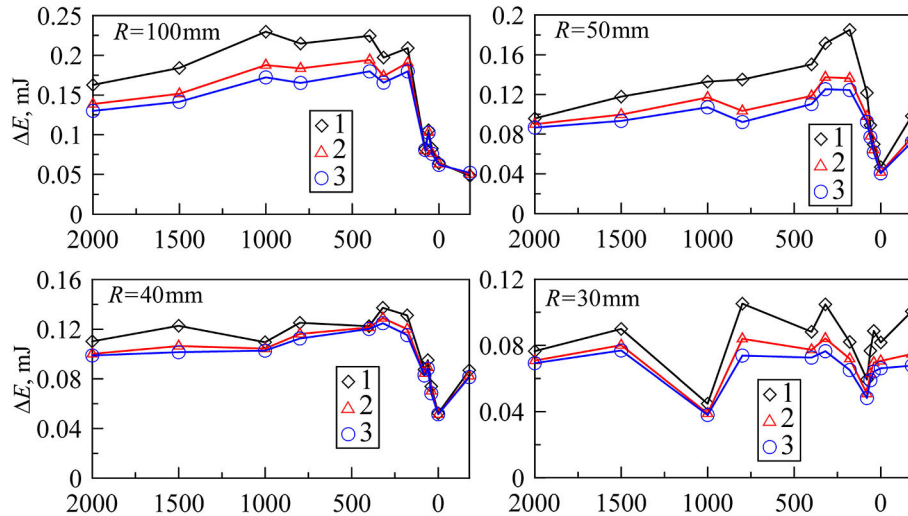


FIGURE 7 | Dependencies of energy dissipation ΔE (mJ) due to adhesion in the full cycle of indentation vs. grit number of the sandpaper for indenters with radius $R = 100, 50, 40$, and 30 mm. The symbols with numbers 1, 2, and 3 in the figure are first (diamonds), second (triangles), and third (circles) cycles of indentation.

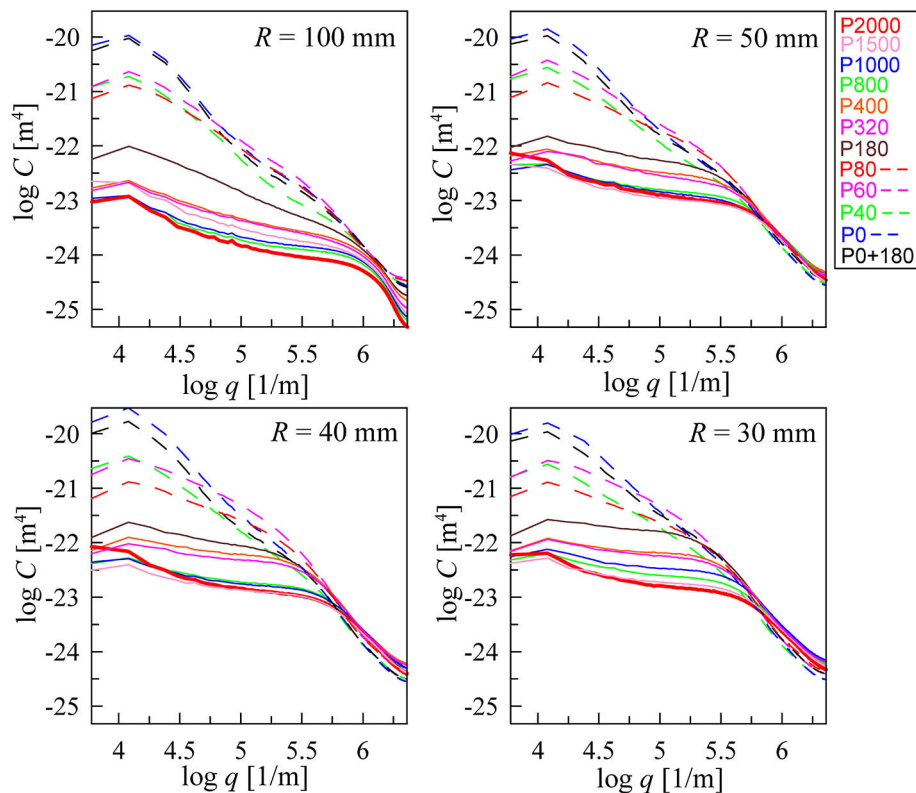


FIGURE 8 | Dependencies of calculated $C(q)$ for all indenters and roughness. The curves with different colors correspond to the curves in **Figure 5** with the same colors and line styles (solid and dashed lines). The solid red lines in all figures correspond to the situation with smallest roughness P2000.

energy $\gamma_{12} = 0.9 \text{ J/m}^2$ (see **Figure 4** and the explanation of this figure). However, we did all experiments with the rubber layer of thickness $h = 5 \text{ mm}$, so we need to obtain proper values of critical forces directly from BEM modeling (for the case $R = 100 \text{ mm}$, the results of modeling are shown in **Figure 4**). In modeling, we obtained other critical values of the forces -0.522 , -0.226 , -0.178 , and -0.131 , in comparison with half-space results. In all situations under consideration, the adhesive strength was bigger in the case of the layer. As can be seen, with a decrease of the indenter radius R , the difference between results for half-space and layers decreases also because when the thickness of the layer is much bigger than the indenter radius, we are in the limit of half-space approximation. The values of F_{\min} in **Figure 6** were found directly from the experiment (see **Figure 5**). With the exception of $R = 30 \text{ mm}$, it can be seen from **Figures 5, 6** that an increase in the roughness first leads to increased adhesive strength. At big roughness values, the adhesion strength significantly decreases. Consequently, there must be an optimal value of roughness, at which the adhesive strength takes maximal values. In our case, the greatest adhesive strength corresponds to processing of the surface with sandpaper around P320. In **Figure 6**, for the cases $R = 100 \text{ mm}$ and $R = 50 \text{ mm}$, the normalized strength for P2000 corresponds to the value $F_{\min}/F_{\text{JKR}} \approx 1$, but for another 40 and 30 mm, it is $F_{\min}/F_{\text{JKR}} \approx 1.5$ for the cases of flat surfaces with roughness of P2000. It happens because surface energy γ_{12} can significantly vary in value in different experiments. Surface energy is a function of the contact duration, temperature, surface chemical composition, and so on. However, we used the same value of surface energy to find the F_{JKR} for all experiments. Because of this, we observe such deviations.

Comparing the different repetitions of the experiment, we found that the strength of the contact has maximal values at first detachment and is slightly reduced at the next contact cycles. We assume that this is due to the freshly cleaned surface of the indenter. In the subsequent indentations, the indenter surface may already contain traces of rubber, and surface energy is potentially decreased. Also note that our indenter with radius $R = 30 \text{ mm}$ did not show a very clear dependence of adhesive

strength vs. roughness (see **Figure 6**). We assume that this is due to the very small contact area at this radius. Here individual stochastic particularities in the center of the indenter determine the adhesive strength.

Figure 7 shows the calculated values of energy dissipation ΔE during the indentation cycle for all situations as a function of sandpaper number. These dependencies are similar to **Figure 6**: with an increase of adhesive strength, dissipation also increases. This is to be expected since an increase of the adhesive strength leads to the growth of the hysteresis loop width. Note that the indenter with radius $R = 100 \text{ mm}$ had an indentation depth of $d = 0.3 \text{ mm}$, while it was 0.4 mm in all other cases.

Surface Topography Measurements

After subtraction of the polynomial approximation of the macroscopic shape, we calculated the power spectra density $C(q)$ of the surface topography in a standard way (Persson, 2014) and averaged this dependency for all sub-areas of the indenter.

Figure 8 shows curves $C(q)$ for all indenters.

Based on $C(q)$, both the averaged root mean square roughness h_{rms} and the averaged slope κ using standard definitions (Persson, 2014) were calculated:

$$h_{\text{rms}}^2 = 2\pi \int_{q_{\min}}^{q_{\max}} q C(q) dq,$$

$$\kappa^2 = 2\pi \int_{q_{\min}}^{q_{\max}} q^3 C(q) dq.$$

Figure 9 shows these values for all indenters and cases that we studied. In Liashenko and Lyashenko (2020), we described the numerical algorithm for calculating $C(q)$ in detail.

Evolution of the Contact Area

Furthermore, we analyzed the area of contact vs. indentation depth for a different roughness of indenter with radius $R = 40 \text{ mm}$ because, for this radius, the obtained pictures had the best contrast. We chose to normalize the contact area A with its largest value A_{\max} , which was different for each experiment. The contact

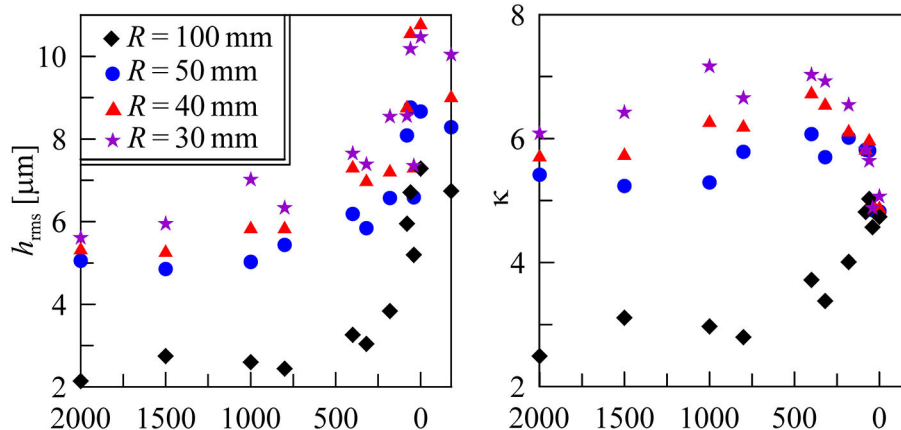
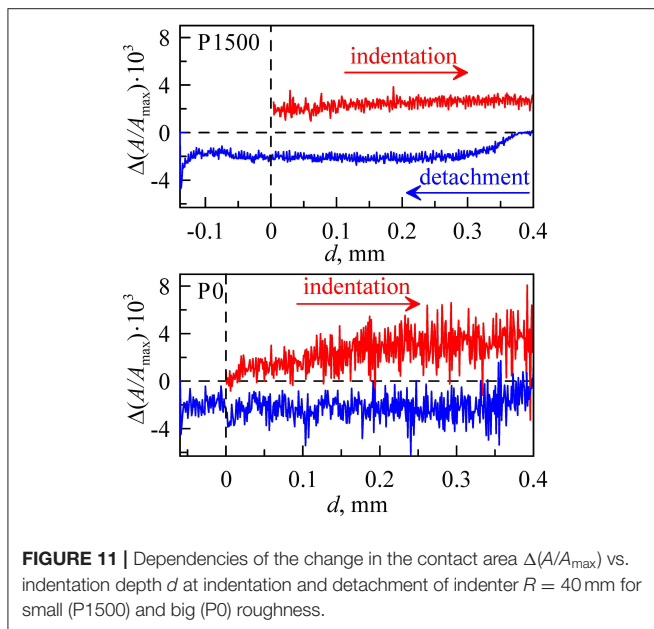
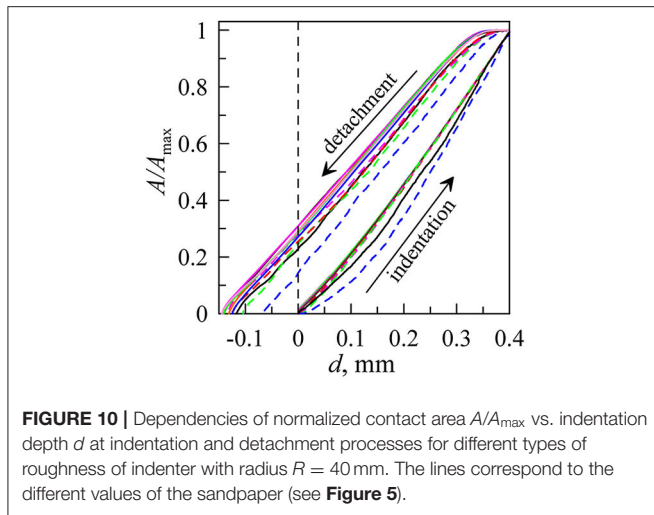
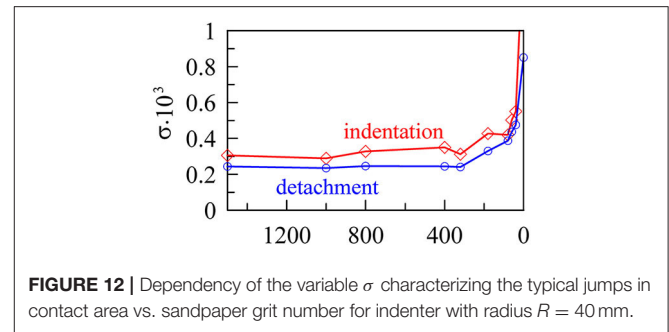


FIGURE 9 | Measured values of h_{rms} and κ for all the cases that we considered, obtained by using $C(q)$. We showed the last value “P0+180” as a negative value “-180” because “P0 + 180” is not a number.



areas were monitored for both indentation and detachment phases. The biggest area corresponds to the maximal indentation depth $d = 0.4$ mm (0.3 mm for $R = 100$ mm). In **Figure 10**, the dependencies of the normalized contact area A/A_{\max} vs. indentation depth d are shown. For all curves, a very pronounced hysteresis of the contact area for the two different directions is observed.

Figure 10 depicts the dependencies for all indenters except P2000 because the mirror-like properties of these surfaces (in the case $R = 40$ mm) prevented us from using the analysis tool. It can be seen that, during the indentation phase, only the cases P0 (dashed blue lines) and P0 + 180 (black solid lines) deviate from universal behavior. During the detachment phase, P0 has a significantly smaller A/A_{\max} , while the difference is less pronounced for P0 + 180. Note that the case of very large roughness has already shown deviations from universal behavior of the normal force as a function of indentation depth (see **Figure 5**).



As we noted before, the indentation process is described well by the Hertz theory (Hertz, 1881), where contact radius is $a = (Rd)^{1/2}$ and contact area $A = \pi a^2$, therefore $A = \pi R d$. We should thus expect to see a clear linear behavior in the $A/A_{\max}(d)$ dependency. However, the two lower curves in **Figure 10** at indentation significantly decline from this behavior. This is also expected since the extreme roughness dominates over the spherical shape in small d (Pohrt and Popov, 2013).

Local Stability of Indentation and Detachment

Let us now have a closer look at the way new contact spots appear during indentation and how they are lost during detachment. We investigated the change in the contact area $\Delta \tilde{A} = \Delta(A/A_{\max})$ for each time step $\Delta t = 1$ s, corresponding to $\Delta d = 1$ μ m. Such dependencies are shown in **Figure 11** for small (P1500) and big (P0) roughness. It can be seen that, in these two cases, the contact area increases with the appearance of new areas at the boundary of the contact. For the rougher surface, the new contact areas tend to have significantly bigger sizes, in comparison with the smoother indenter. During detachment, the contact area of P0 decreases much more homogeneously. This can be explained by the local instabilities of asperities (Li et al., 2019; Popov, 2019).

To quantify this, we calculated standard deviation $\sigma = \langle (\Delta \tilde{A} - \langle \Delta \tilde{A} \rangle)^2 \rangle^{1/2}$ in each case. We chose to only look at $0.2 < d < 0.4$ for the indentation phase and at $0 < d < 0.2$ for detachment. In these regions, $\Delta \tilde{A}$ only fluctuates around a certain value but has no global trend. For the dependency of σ on the grit number, see **Figure 12**.

Here we can clearly see that detachment is more homogeneous than indentation for all cases under consideration. Interestingly, the value σ does not change significantly with the grit number of sandpaper, except when this number exceeds a critical value. Starting from approximately P320, σ increases strongly when the grit number is decreased. Please note that this value coincides with the maximum adhesion strength, as shown in **Figure 6**.

CONCLUSIONS

We have tested the adhesive strength of rough steel indenters in contact with soft rubber using different radii and different sandpapers to produce roughness. We found that, in contrast to JKR theory, adhesive hysteresis was always observable. The maximum value of the attractive force (adhesive strength) was

found to depend on surface roughness. The increase over a smooth indenter reaches 25% for the case $R = 40$ mm, 32% for $R = 100$ mm, and 90% for $R = 50$ mm, respectively. Observations of the contact area evolution showed that, for rough surfaces, the attachment and the detachment of surface patches happen more abruptly, especially during the indentation phase. When the sandpaper grit number used for surface preparation is below 320, this effect is very pronounced.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

RP and IL conceived the study and built the experimental setup. IL carried out the experiments. Both authors drafted and reviewed the manuscript.

FUNDING

We acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG PO 810-55-1). We also acknowledge the support by the German Research Foundation and the Open Access Publication Fund of TU Berlin.

REFERENCES

- Benz, M., Rosenberg, K. J., Kramer, E. J., and Israelachvili, J. N. (2006). The deformation and adhesion of randomly rough and patterned surfaces. *J. Phys. Chem. B* 110, 11884–11893. doi: 10.1021/jp0602880
- Ciavarella, M. (2015). Adhesive rough contacts near complete contact. *Int. J. Mech. Sci.* 104, 104–111. doi: 10.1016/j.ijmecsci.2015.10.005
- Ciavarella, M., Joe, J., Papangelo, A., and Barber, J. R. (2019). The role of adhesion in contact mechanics. *J. R. Soc. Interface* 16, 1–22. doi: 10.1098/rsif.2018.0738
- Deng, W., and Kesari, H. (2019). Depth-dependent hysteresis in adhesive elastic contacts at large surface roughness. *Sci. Rep.* 9:1639. doi: 10.1038/s41598-018-38212-z
- Guduru, P. R. (2007). Detachment of a rigid solid from an elastic wavy surface: theory. *J. Mech. Phys. Solids* 55, 445–472. doi: 10.1016/j.jmps.2006.09.004
- Hertz, H. (1881). Über die Berührung fester elastischer Körper. *J. Die Reine Angew. Mathem.* 92, 156–171. doi: 10.1515/crll.1882.92.156
- Jin, F., Guo, X., and Zhang, W. (2013). A unified treatment of axisymmetric adhesive contact on a power-law graded elastic halfspace. *J. Appl. Mech.* 80, 1–9. doi: 10.1115/1.4023980
- Johnson, K. L. (1995). The adhesion of two elastic bodies with slightly wavy surfaces. *Int. J. Solids Struct.* 32, 423–430. doi: 10.1016/0020-7683(94)00111-9
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. Royal Soc. Lond. A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Li, Q., Pohrt, R., Lyashenko, I. A., and Popov, V. L. (2020). Boundary element method for nonadhesive and adhesive contacts of a coated elastic half-space. *Proc. Inst. Mech. Eng. Part J.-J. Eng. Tribol.* 234, 73–83. doi: 10.1177/1350650119854250
- Li, Q., Pohrt, R., and Popov, V. L. (2019). Adhesive strength of contacts of rough spheres. *Front. Mech. Eng.* 5:7. doi: 10.3389/fmech.2019.00007
- Liashenko, Z. M., and Lyashenko, I. A. (2020). Influence of spatial inhomogeneity on the formation of chaotic modes at the self-organization process. *Ukr. J. Phys.* 65, 130–142. doi: 10.15407/ujpe65.2.130
- Liu, Z., Lu, H., Zheng, Y., Tao, D., Meng, Y., and Tian, Y. (2018). Transient adhesion in a non-fully detached contact. *Sci. Rep.* 8:6147. doi: 10.1038/s41598-018-24587-6
- Lyashenko, I. A., and Popov, V. L. (2019). Mechanics of adhesive contacts: experiment and theory. *AIP Conf. Proc.* 2167, 1–4. doi: 10.1063/1.5132068
- Papangelo, A., and Ciavarella, M. (2017). A Maugis-Dugdale cohesive solution for adhesion of a surface with a dimple. *J. R. Soc. Interface* 14, 1–9. doi: 10.1098/rsif.2016.0996
- Papangelo, A., and Ciavarella, M. (2018). Adhesion of surfaces with wavy roughness and a shallow depression. *Mech. Mater.* 118, 11–16. doi: 10.1016/j.mechmat.2017.12.005
- Papangelo, A., Scheibert, J., Sahli, R., Pallares, G., and Ciavarella, M. (2019). Shear-induced contact area anisotropy explained by a fracture mechanics model. *Phys. Rev. E* 99:053005. doi: 10.1103/PhysRevE.99.053005
- Pastewka, L., and Robbins, M. O. (2014). Contact between rough surfaces and a criterion for macroscopic adhesion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3298–3303. doi: 10.1073/pnas.1320846111
- Persson, B. N. J. (2006). On the fractal dimension of rough surfaces. *Surf. Sci. Rep.* 61, 201–227. doi: 10.1016/j.surfrep.2006.04.001
- Persson, B. N. J. (2014). On the fractal dimension of rough surfaces. *Tribol. Lett.* 54, 99–106. doi: 10.1007/s11249-014-0313-4
- Persson, B. N. J., and Scaraggi, M. (2014). Theory of adhesion: role of surface roughness. *J. Chem. Phys.* 141, 1–14. doi: 10.1063/1.4895789
- Pohrt, R., and Popov, V. L. (2013). Contact mechanics of rough spheres: crossover from fractal to Hertzian behavior. *Adv. Tribol.* 2013:974178. doi: 10.1155/2013/974178
- Popov, V. L. (2019). Adhesive contribution to friction. *AIP Conf. Proc.* 2167:020286. doi: 10.1063/1.5132153
- Popov, V. L. (2020). Adhesion hysteresis due to chemical heterogeneity. *Preprints*. doi: 10.20944/preprints202003.0131.v1
- Popov, V. L., Pohrt, R., and Li, Q. (2017). Strength of adhesive contacts: influence of contact geometry and material gradients. *Friction* 5, 308–325. doi: 10.1007/s40544-017-0177-3
- Popova, E., and Popov, V. L. (2018). Note on the history of contact mechanics and friction: interplay of electrostatics, theory of gravitation and elasticity from Coulomb to Johnson–Kendall–Roberts theory of adhesion. *Phys. Mesomech.* 21, 1–5. doi: 10.1134/S1029959918010010

ACKNOWLEDGMENTS

The authors thank Valentin L. Popov for the suggestion of basic ideas, which have been experimentally proven in the paper, as well as for many fruitful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmech.2020.00049/full#supplementary-material>

Supplementary Video 1 | Movies of two experiments for indentation of steel parabolic indenter with radius $R = 40$ mm in the elastic layer of transparent rubber TARNAC CRG N3005. The velocity of indenter motion for both cases is $v = 1$ $\mu\text{m/s}$. Panels from left to right show real pictures of the contact area, dependencies of normal force F_N vs. indentation depth d , and dependencies of normalized contact area A/A_{max} vs. indentation depth d . The three upper panels correspond to indenter with surface processed by sandpaper P1500. The lower three panels correspond to indenter with highest roughness “P0”. In the pictures of contact area, the time of experiment duration in seconds is shown. Three consistent cycles of indentation are shown. For the first cycle, in the case of P1500, the adhesion strength (absolute value of minimal normal force at phase of detachment) $F_N = 0.307$ N is 2.52 times bigger value than in the case P0, where the corresponding value is $F_N = 0.122$ N. For bigger roughness as in P0, the contact disappears earlier than for the case P1500. These two movies correspond to **Figure 11** in the article.

- Rey, V., Anciaux, G., and Molinari, J. -F. (2017). Normal adhesive contact on rough surfaces: efficient algorithm for FFT-based BEM resolution. *Comput. Mech.* 60, 69–81. doi: 10.1007/s00466-017-1392-5
- Sahli, R., Pallares, G., Papangelo, A., Ciavarella, M., Ducottet, C., Ponthus, N., et al. (2019). Shear-induced anisotropy in rough elastomer contact. *Phys. Rev. Lett.* 122, 1–6. doi: 10.1103/PhysRevLett.122.214301
- Tiwari, A., Dorogin, L., Bennett, A. I., Schulze, K. D., Sawyer, W. G., Tahir, M., et al. (2017). The effect of surface roughness and viscoelasticity on rubber adhesion. *Soft. Matter* 13, 3602–3621. doi: 10.1039/C7SM00177K
- Waters, J. F., and Guduru, P. R. (2010). Mode-mixity-dependent adhesive contact of a sphere on a plane surface.

Proc. R. Soc. A 466, 1303–1325. doi: 10.1098/rspa.2009.0461

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Lyashenko and Pohrt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Viscoelastic and Adhesion Properties of New Poly(Ether-Urethane) Pressure-Sensitive Adhesives

Mónica Fuensanta and José Miguel Martín-Martínez*

Adhesion and Adhesives Laboratory, Department Inorganic Chemistry, University of Alicante, Alicante, Spain

OPEN ACCESS

Edited by:

Andrey I. Dmitriev,
Institute of Strength Physics and
Materials Science (ISPMS SB
RAS), Russia

Reviewed by:

Valentin L. Popov,
Technical University of
Berlin, Germany
Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine

*Correspondence:

José Miguel Martín-Martínez
jm.martin@ua.es

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 30 March 2020

Accepted: 04 May 2020

Published: 24 June 2020

Citation:

Fuensanta M and Martín-Martínez JM
(2020) Viscoelastic and Adhesion
Properties of New
Poly(Ether-Urethane)
Pressure-Sensitive Adhesives.
Front. Mech. Eng. 6:34.
doi: 10.3389/fmech.2020.00034

New thermoplastic poly(ether-urethane) (TPUs) pressure-sensitive adhesives (PSAs) have been synthesized with the prepolymer method by reacting methylene diisocyanate, different blends of polyether diols—poly(propyleneglycol) (PPG) and poly(tetramethylene ether glycol) (PTMEG)—and 1,4-butanediol chain extender; different NCO/OH ratios have been used. The properties of the TPU PSAs depended on both the NCO/OH ratio and the PTMEG content in the blend. The addition of PTMEG polyol produced semicrystalline regions in the soft segments of the TPUs and inhibited the mobility of the polymeric chains; this led to improved cohesion of the TPU PSAs; however, similar degrees of phase separation were obtained in all TPUs synthesized with different PTMEG contents. The increase of the PTMEG content in the polyols blend improved both the cohesion and the adhesion but decreased the tack of the TPU PSAs. The optimal balance between the adhesion and cohesion properties was found in the TPU PSA synthesized with 50 wt% PPG + 50 wt% PTMEG and an NCO/OH ratio of 1.20.

Keywords: polyurethane, pressure-sensitive adhesive, viscoelastic properties, probe tack, peel, shear under creep

INTRODUCTION

Pressure-sensitive adhesives (PSAs) are able to bond different substrates by applying light pressure for a short time, and they must be removed without leaving any residue. Therefore, the viscoelastic properties of the PSAs determined their performance; that is, the viscous component must be balanced for allowing adequate wettability and bonding, and the elastic component must be sufficient for allowing clean debonding from the substrate surface (Feldstein and Siegel, 2012). Pressure-sensitive adhesives are used as labels, tapes, protective films, and so on, and they are common in medical products (patches, surgical bandages, biomedical electrodes, medical plasters, etc.), in which good biocompatibility with human skin, excellent reversible adhesion, and permeability to water and air are important properties (Satas and Satas, 1989).

The main properties of the PSAs related to their viscoelastic properties are the tack, shear, and peel adhesion. The tack is the initial immediate bond of a PSA with a substrate upon applying light pressure. The shear adhesion is the resistance of the PSA to flow under creep and is essential to ensure an adequate easy debonding from the substrate without leaving adhesive residues; that is, it is related to the cohesion of the PSA. The force required to separate a PSA from a substrate surface is the peel adhesion. In general, the PSA for medical applications requires a careful reversible application to the clean skin surface, so medium–low tack is necessary.

The PSAs for medical applications are made of natural and synthetic rubbers, polyisobutylene (PIB), acrylates, or silicones (Czech et al., 2011). Because of allergic issues, the acrylic and silicone PSAs were replaced by PIBs, but PIB PSAs oxidized easily and showed limited permeability to water and air (Higgins et al., 1989). Polyurethane-based PSAs are scarcely used for medical devices but polyurethane PSAs intended for transdermal drug delivery have been developed (Chen et al., 2009), and more recently, novel solvent-free poly(ether-urethane)-silicone cross-linked PSAs have been synthesized for drug reservoir (Tombs et al., 2018).

Thermoplastic polyurethanes (TPUs) have a great potential for designing PSAs because of their segmented structure constituted by hard and soft segments. The hard segments are formed by reacting a diisocyanate and a short diol (chain extender), and they are polar and short, whereas the soft segments are made of the polyol chains, and they are less polar and have higher length than the hard segments. Therefore, the hard and the soft segments are incompatible and tend to segregate in immiscible domains, and the hard segments determine the cohesion properties of the TPUs, whereas the soft segments are related to their tack and adhesion properties (Yilgör and Yilgör, 2015). Thus, the structure of the TPUs can be designed by changing the composition and the extent of the interactions between the hard and the soft segments; this may allow finding the right balance between their adhesion and cohesion properties; however, typically the TPUs do not show tack, an important property for PSAs.

Recently, different TPU PSAs synthesized with blends of polypropylene glycols (PPGs) of different molecular weights and 4,4'-diphenylmethane diisocyanate (MDI) showed good tack at 10–37°C but low peel strength and cohesion (Fuensanta and Martín-Martínez, 2018). In order to improve the cohesion of the TPU PSAs, the hard segments content of the TPUs was changed by using different blends of PPGs of different molecular weights (Fuensanta and Martín-Martínez, 2019), and it was shown that the TPU PSAs synthesized with lower hard segments content had poor shear strength but high tack and peel strength, whereas the increase of the hard segments increased the shear strength and the peel strength values, but tack was low.

In this study, the segmented structure and the degree of phase separation of the TPUs have been modified by changing the extent of the interactions between the soft segments (blends of polyether polyols with and without alkyl pendant groups; **Scheme 1**) and the hard segments content (NCO/OH ratio = 1.05–1.35). The absence of the methyl pendant group in the poly(tetramethylene ether glycol) (PTMEG) will allow more net van der Waals interactions between the soft segments in the TPU reducing the mobility of the polymer chains, but increasing its cohesion and mechanical properties. The mixing of PTMEG with PPG, which has no pendant group, during the synthesis of the polyurethanes will modify the extent of interactions between the soft segments. The structure of the TPUs synthesized with different blends of PPG and PTMEG polyols of similar molecular weights was determined by infrared spectroscopy, differential scanning calorimetry (DSC), and thermal gravimetric analysis (TGA), and their viscoelastic

properties were obtained by plate–plate rheology. Thermoplastic polyurethane PSAs were prepared by placing TPU coatings on PET (polyethylene terephthalate) film (Fuensanta and Martín-Martínez, 2018), and the tack, shear under creep, and peel adhesion properties were measured.

MATERIALS AND EXPERIMENTAL TECHNIQUES

Materials

The TPUs were synthesized by reacting MDI [Desmodur® 44MC flakes supplied by Covestro (Leverkusen, Germany)], different blends of PPG of molecular weight 2,000 g/mol [Alcupol® D2021 supplied by Repsol (Madrid, Spain)], and PTMEG of molecular weight 2,000 g/mol (PTMEG) supplied by Sigma-Aldrich (St. Louis, MO, USA). The polyols were melted and dried at 80°C under reduced pressure (300 mbar) for 2 h before used. Dibutyl tin dilaurate (DBTDL) was used as catalyst, and 1,4-butanediol (BD) was used as chain extender; both were supplied by Sigma-Aldrich. Methyl ethyl ketone (MEK) supplied by Jaber S.A. (Almansa, Spain) was used to dissolve the TPU for allowing adequate coating on PET films (TPU PSAs).

Synthesis of the Thermoplastic Poly(Ether-Urethane)s

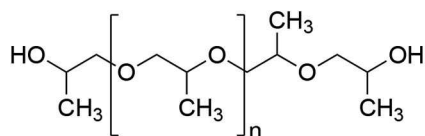
The thermoplastic thermoplastic poly(ether-urethane)s were synthesized by using the prepolymer method in 500 mL glass reactor under inert atmosphere (dried nitrogen), and an anchor shaped stirrer coupled to Heidolph overhead stirrer RZR-2000 (Kelheim, Germany) was used (Fuensanta and Martín-Martínez, 2018). MDI was melted at 80°C in the reactor, and PPG, PTMEG, or PPG + PTMEG blends were added under stirring at 250 revolutions/min (rpm) for 30 min. Afterward, 0.04 mmol catalyst (DBTDL) was added, and the stirring was decreased to 80 rpm. The reaction lasted for 2 h, and the amount of free NCO content was determined by dibutylamine titration. Then, the chain extender (BD) was added under stirring at 80°C and 80 rpm for 5 min. The scheme of the synthesis of the TPUs is shown in **Figure 1**.

Poly(ether-urethane) PSAs (TPU PSAs) were prepared according the procedure described in Fuensanta and Martín-Martínez (2019). One hundred milliliters of MEK solution containing 40 g TPU was spread on PET film (50 µm thick), and the thickness of the TPU coating was adjusted with a metering rod of 400 µm. The solvent evaporated at room temperature for 72 h for obtaining the TPU PSAs. The thicknesses of the TPUs on the PET films were 40 to 50 µm.

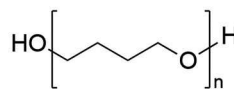
Experimental Techniques

Attenuated Total Reflection Infrared Spectroscopy

Attenuated total reflection infrared (ATR-IR) spectra of the TPUs were obtained in absorbance mode in a Tensor 27 FT-IR spectrometer (Bruker Optik GmbH, Erlanger, Germany) by using Golden Gate single reflection diamond, recording 64 scans with a resolution of 4 cm⁻¹ in the range of wavenumbers from 4,000 to 400 cm⁻¹ (Fuensanta and Martín-Martínez, 2018).



Poly(propyleneglycol) - PPG



Poly(tetramethylene ether glycol) - PTMEG

SCHEME 1 | Chemical structures of the polyether polyols.

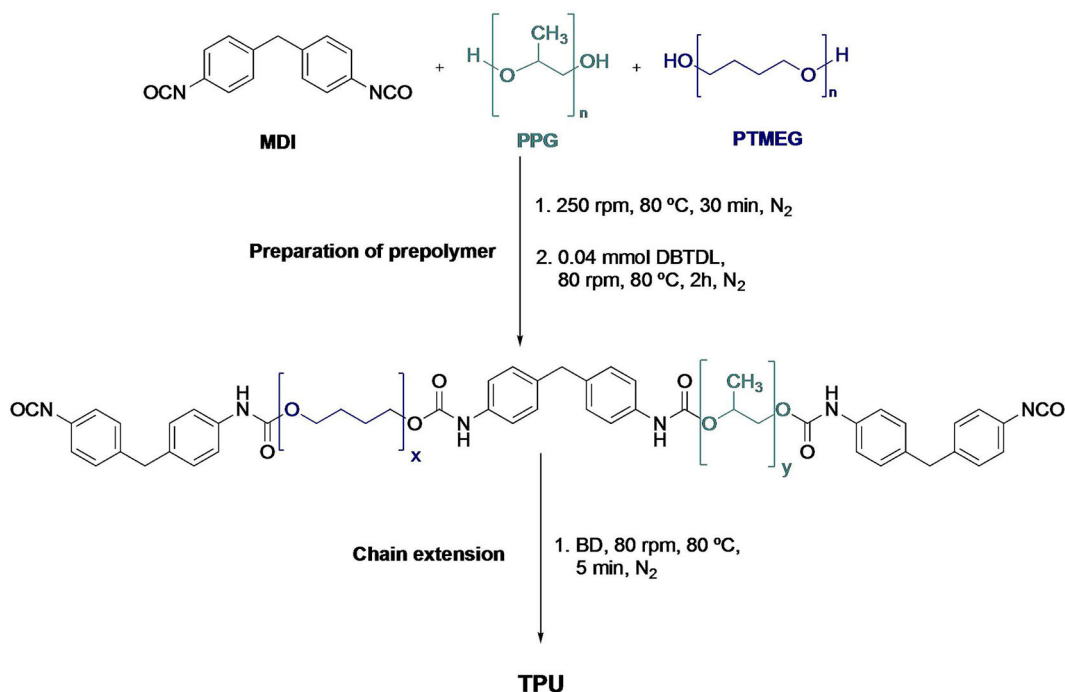


FIGURE 1 | Scheme of the synthesis of the thermoplastic poly(ether-urethane)s.

Differential Scanning Calorimetry

Differential scanning calorimetry was used to obtain the structural properties of the TPUs, a DSC Q100 calorimeter (TA Instrument, New Castle, DE, USA) under nitrogen atmosphere (flow rate = 50 mL/min) was used. Seven- to 8-mg sample was heated in hermetic sealed aluminum pan from -80 to 150°C , cooled at -80°C , and heated again from -80 to 200°C by using heating and cooling rates of $10^{\circ}\text{C}/\text{min}$ (Fuensanta and Martín-Martínez, 2018). The glass transition temperatures and the thermal events of the TPUs were obtained from the second DSC heating run.

Thermal Gravimetric Analysis

Thermal gravimetric analysis was also used to obtain the structural properties of the TPUs, a TGA Q500 equipment (TA Instruments) under nitrogen atmosphere (flow rate = 50 mL/min) was used. A 10-mg sample was placed in platinum crucible, and it was heated from 35 to 800°C by using a heating rate of $10^{\circ}\text{C}/\text{min}$ (Fuensanta and Martín-Martínez, 2018).

Plate-Plate Rheology

The viscoelastic properties of the TPUs were assessed in a DHR-2 rheometer (TA Instruments) using parallel plate-plate geometry; a stainless-steel plate of 20-mm diameter and a gap of 0.40 mm were used. Temperature sweep experiments were carried out from -5 to 120°C , by using a heating rate of $5^{\circ}\text{C}/\text{min}$ and a frequency of 1 Hz. Furthermore, frequency sweep experiments were carried out at 25°C by using 2.5% strain amplitude in the angular frequency range from 0.01 to 100 rad/s (Fuensanta and Martín-Martínez, 2018).

Adhesion Properties

The adhesion properties of the TPU PSAs were assessed at 25°C by probe tack, 180° peel strength, and creep test under shear (Figure 2).

The probe tack of the TPU PSAs was measured at 25°C by using a flat-end cylindrical stainless-steel probe of 3-mm diameter in a TA.XT2i Texture Analyzer (Stable Micro Systems, Surrey, UK) (Figure 2) (Fuensanta and Martín-Martínez, 2018).

The probe was approached slowly to the TPU PSA surface applying a load of 5 N for 1 s, and it was pulled out at 10 mm/s. The maximum of the stress–strain curve was taken as the tack of the TPU PSA. At least five replicates were carried out and averaged.

A 180° peel strength of aluminum 5754/TPU PSA joints was carried out in an Instron 4411 universal testing machine (Instron Ltd., Buckinghamshire, UK); the pulling rate was 152 mm/min (Fuensanta and Martín-Martínez, 2018). The TPU PSA strips have dimensions of 30 × 300 × 0.5 mm, and they were joined to clean aluminum 5754 pieces of dimensions 30 × 150 × 1.5 mm; the joints were made by passing 30 times a 2-kg rubber-coated roller (Figure 2). Five replicates were tested and averaged for each joint.

The creep tests under shear of the TPU PSAs were carried out in a Shear-10 equipment (ChemInstruments, Fairfield, OH, USA). Thermoplastic polyurethane PSA strips of 2.4 × 20 cm were attached to the center of a clean polished stainless-steel 304 piece. Thermoplastic polyurethane PSA area of 2.5 × 2.5 mm was joined to the stainless steel, and a 2-kg rubber-coated roller was passed over the joint. The, the TPU PSA stainless-steel joint was placed on the holder of the equipment hanging a weight of 1 kg at the bottom (Fuensanta and Martín-Martínez, 2019). The creep resistance at 25°C of the TPU PSAs is related to their cohesion and was obtained as the “holding time,” that is, the time needed

for the TPU PSA strip to fall down. Three replicates were tested for each TPU PSA, and the results obtained were averaged.

RESULTS AND DISCUSSION

The innovative strategy of this study is the synthesis of TPUs by using blends of two polyether polyols of different structure for being used as PSAs. According to Scheme 1, PPG is a polyether with methyl pendant groups that disturb the interactions between the chains, whereas PTMEG has no pendant groups that allow more net interaction between the chains. In the Supplementary

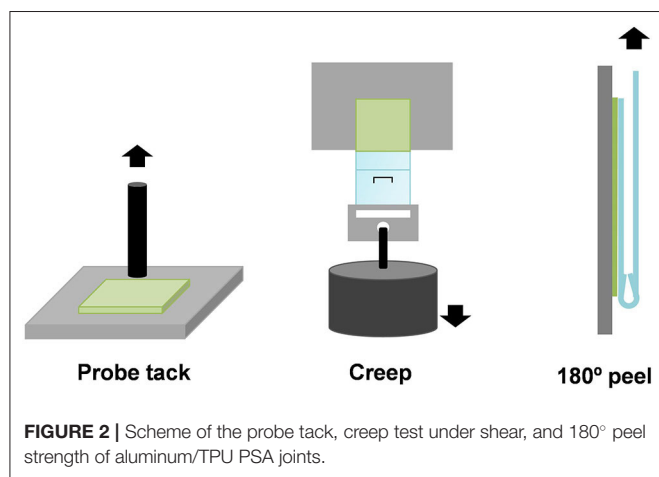


TABLE 1 | Nomenclature and composition of the poly(ether-urethane)s synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

Nomenclature	PPG (wt%)	PTMEG (wt%)	NCO/OH ratio	HS (%) ^a
1.05-50PPG50PTMEG	50	50	1.05	11.7
1.10-50PPG50PTMEG	50	50	1.10	12.5
1.20-50PPG50PTMEG	50	50	1.20	13.8
1.35-50PPG50PTMEG	50	50	1.35	15.8

^aHard segments content. HS = 100 × [weight (MDI) + weight (BD)]/[total weight].

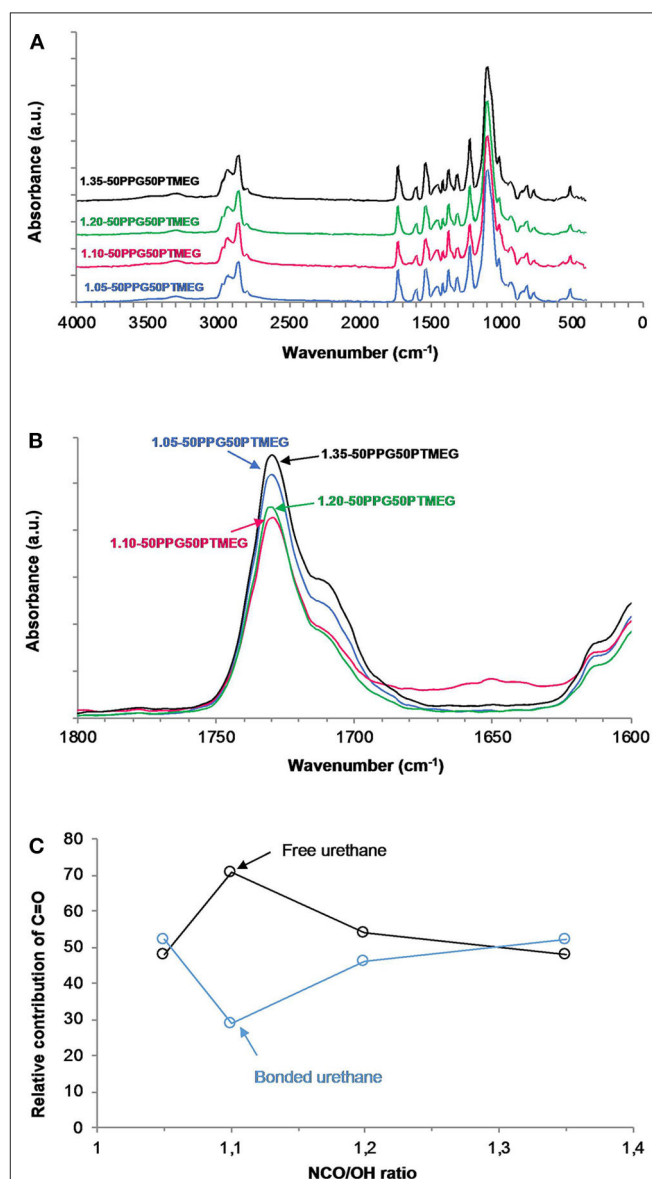


FIGURE 3 | (A) ATR-IR spectra, (B) carbonyl region (1,800–1,600 cm⁻¹) of the ATR-IR spectra, and (C) variation of the free and hydrogen-bonded urethane groups as a function of the NCO/OH ratio for TPUs synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

Materials (Figures S1–S3, Tables S1, S2), the characterization of the two polyether polyols is shown. Poly(propyleneglycol) shows one glass transition temperature at -69°C and one thermal decomposition at 350°C , whereas PTMEG shows more intense C-O-C band at $1,101\text{ cm}^{-1}$ in the ATR-IR spectrum and one melting process at 26°C with noticeable enthalpy due to the interaction between the chains; furthermore, PTMEG shows two thermal decompositions at 280 and 417°C (the main); the higher temperature of the main decomposition than in PPG can be associated to the interactions between the chains. The use of blends of PPG and PTMEG should

change the interactions between the soft segments, which will modify the degree of phase separation and the properties of the TPU PSAs.

Modification of the Hard Segments Content by Changing the NCO/OH Ratio of the Poly(Ether-Urethane)s

For increasing the cohesion of the TPUs, their hard segments content was increased from 11.7 to 15.8% by synthesizing different TPUs with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios (1.05–1.35) (Table 1).

The chemical structure of the TPUs was analyzed by ATR-IR spectroscopy (Figure 3A). All ATR-IR spectra were normalized with respect to the most intense absorption band of the C-O-C group at $1,099\text{ cm}^{-1}$. The absorption bands at 3,506 to $3,498\text{ cm}^{-1}$ and $3,296\text{ cm}^{-1}$ are attributed to the N-H stretching of the urethane group; the N-H bending band appears at $1,534\text{ cm}^{-1}$, and the C-N stretching is found at $1,222$ to $1,221\text{ cm}^{-1}$. The main characteristic bands of the polyols (PPG and PTMEG) appear at 2,976 to $2,856\text{ cm}^{-1}$ (C-H stretching of the CH_3 and CH_2 groups), $1,449$ to $1,448\text{ cm}^{-1}$ (C-H bending), $1,372\text{ cm}^{-1}$ (C-H rocking of the methyl group), and $1,099$ to $1,098\text{ cm}^{-1}$ —the most intense—due to C-O-C stretching. The C=O stretching band of the urethane group is located at $1,730\text{ cm}^{-1}$, and the absorptions bands at $1,599$ and $1,412\text{ cm}^{-1}$ correspond to C-C and C-H stretching of the aromatic ring of MDI, respectively.

The ATR-IRs of all TPUs are quite similar, but they differ in the carbonyl region (Figure 3B) in which the percentages of the associated by hydrogen bonding urethane (bonded urethane) groups are different depending on the NCO/OH ratio. The free urethane and the ether groups bonded to N-H hydrogen appear at $1,730$ to $1,729\text{ cm}^{-1}$, whereas the hydrogen-bonded urethane group appears at $1,711$ to $1,710\text{ cm}^{-1}$ (Strikovskiy and Zharkov, 1993; Mattia and Painter, 2007). In order to quantify the percentages of the free and bonded urethane groups in the TPUs, the carbonyl region was curve fitted, assuming a Gaussian function (Fuensanta and Martín-Martínez, 2019). According to Figure 3C, the percentages of the free and bonded urethane groups are similar in all TPUs, except in the one synthesized with NCO/OH ratio of 1.10 in which the free urethane groups are dominant. Therefore, 1.10-50PPG50PTMEG shows the highest

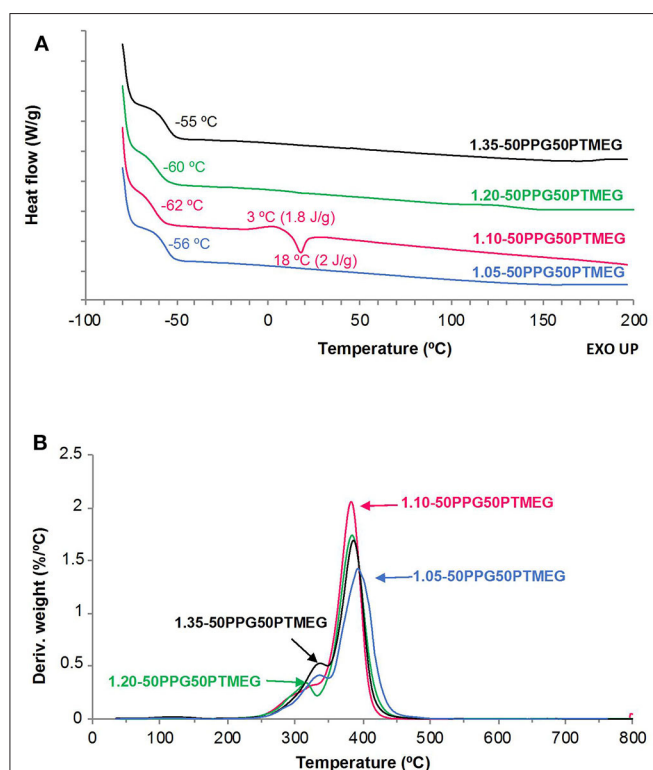


FIGURE 4 | (A) DSC traces (second heating run) and **(B)** variation of the derivative of the weight as a function of the temperature for TPUs synthesized with 50 wt% PPG + 50 wt% PTMEG and different NCO/OH ratios.

TABLE 2 | Temperatures and weight losses of the thermal decompositions of the TPUs synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

TPU PSA	1st degradation		2nd degradation		3rd degradation		Residue (wt%)
	T ₁ (°C)	Weight loss ₁ (%)	T ₂ (°C)	Weight loss ₂ (%)	T ₃ (°C)	Weight loss ₃ (%)	
1.05-50PPG50PTMEG	283	4	337	18	393	76	2
1.10-50PPG50PTMEG	297	9	325	8	383	83	0
1.20-50PPG50PTMEG	—	—	315	16	385	83	1
1.35-50PPG50PTMEG	—	—	333	28	387	70	2

DTGA experiments.

degree of phase separation between the hard and soft segments, whereas similar degree of phase separation is found in the other TPUs.

The structure and thermal properties of the TPUs made with different NCO/OH ratios have been assessed by DSC and TGA. The DSC traces (**Figure 4A**) show the glass transition temperature of the soft segments (T_g) between -55 and -62°C ; the lowest and similar T_g corresponds to 1.05-50PPG50PTMEG and 1.35-50PPG50PTMEG indicating lower degree of phase separation than in the other TPUs. Thus, one of the requirements in the adhesive in PSAs is obeyed; that is, the T_g must be substantially lower than the temperature of use. Furthermore, 1.10-50PPG50PTMEG shows a cold crystallization of the soft segments at 3°C followed by a melting peak of the soft segments at 18°C , which derives from the interactions between the PTMEG chains. This confirms the different structure of 1.10-50PPG50PTMEG with respect to the other TPUs, in agreement with the results of ATR-IR spectroscopy. On the other hand, the DTGA plots of the TPUs (**Figure 4B**) show three thermal degradations due to the decomposition of the free urethane hard domains (283 – 297°C), hydrogen-bonded urethane domains (315 – 337°C), and soft domains (383 – 393°C) (Ferguson and Petrovic, 1976). Because the amounts of PPG and PTMEG are similar in all TPUs, the differences in their structure and thermal properties can be ascribed to different interactions due to the hard segments. The existence of free urethane domains can only be distinguished in the TPUs made with NCO/OH ratios of 1.05 and 1.10, and in general, the percentage of by hydrogen-bonded urethane domains increases by increasing the NCO/OH ratio; 1.10-50PPG50PTMEG is an exception (**Table 2**). The percentages of the soft domains are more noticeable in the TPUs made with NCO/OH ratios of 1.10 and 1.20. Therefore, 1.35-50PPG50PTMEG shows the highest percentage of hydrogen-bonded urethane domains and the lowest percentage of soft domains, indicating less mobility of the polymeric chains and higher cohesion than in the other TPUs. On the contrary, 1.10-50PPG50PTMEG and 1.20-50PPG50PTMEG have the highest percentages of soft domains, indicating higher mobility of the polymeric chains, particularly for 1.10-50PPG50PTMEG in which dominant free urethane domains are distinguished.

The extent of mobility of the polymeric chains in the TPUs can be estimated from their rheological and viscoelastic properties. **Figure 5A** shows a decrease of the storage modulus (G') by increasing the temperature in all TPUs; the decrease is more noticeable in 1.35-50PPG50PTMEG, and the rheological curves are quite similar in all TPUs except in 1.35-50PPG50PTMEG, likely due to the thermal rupture of higher amounts of hydrogen bonds between the hard domains. On the other hand, all TPUs show a crossover of the storage and loss moduli (**Figure S4**), and the values of temperature ($T_{\text{crossover}}$) and modulus ($G_{\text{crossover}}$) at the crossover are given in **Table 3**. Similar values of $G_{\text{crossover}}$ ($1.2 \cdot 10^5$ – $1.6 \cdot 10^5$ Pa) are obtained in all TPUs indicating that they fulfill the Dahlquist criterion at room temperature (Dahlquist, 1969), one of the requirements in PSAs. However, the values of $T_{\text{crossover}}$ of the TPUs decrease by increasing the NCO/OH

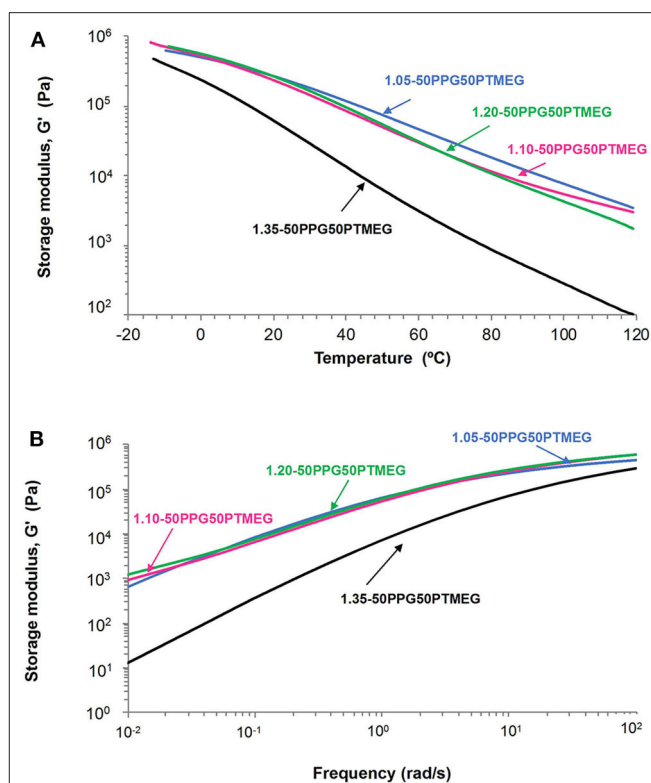


FIGURE 5 | (A) Variation of the storage modulus (G') as function of the temperature and **(B)** variation of the storage modulus (G') as a function of the frequency for TPUs synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

TABLE 3 | Values of temperature and modulus at the crossover of G' and G'' of the TPUs synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

TPU	$T_{\text{crossover}}$ ($^\circ\text{C}$)	$G_{\text{crossover}}$ (Pa)
1.05-50PPG50PTMEG	40	$1.2 \cdot 10^5$
1.10-50PPG50PTMEG	27	$1.6 \cdot 10^5$
1.20-50PPG50PTMEG	33	$1.4 \cdot 10^5$
1.35-50PPG50PTMEG	10	$1.3 \cdot 10^5$

Plate-plate rheology experiments (temperature sweep).

ratio; 1.10-50PPG50PTMEG is an exception because of its particular structure. Furthermore, 1.35-50PPG50PTMEG shows a temperature at the crossover lower than 25°C , indicating that at room temperature the elastic properties are dominant, and lower mobility of the polymeric chains than in the other TPUs can be expected.

The most common typical applications of the PSAs are at room temperature, and therefore, the viscoelastic properties of the TPUs were also determined at 25°C by oscillatory frequency sweep rheological experiments. It has been shown elsewhere (Derail and Marin, 2009) that the tack and the shear strength of the PSAs correlated well with their storage moduli (G') values at lower frequency, whereas the peel strength was related

TABLE 4 | Probe tack at 25°C, holding time at 25°C, and 180° peel strength at 25°C of aluminum 5754/TPU PSA joints.

TPU PSA	Probe tack (kPa)	Holding time (min)	180° peel strength (N/cm) ^a
1.05-50PPG50PTMEG	278 ± 17	27 ± 5	0.63 ± 0.03
1.10-50PPG50PTMEG	874 ± 18	188 ± 27	9.25 ± 0.35
1.20-50PPG50PTMEG	576 ± 17	1,518 ± 167	9.91 ± 0.56
1.35-50PPG50PTMEG	185 ± 8	331 ± 71	1.75 ± 0.03

Thermoplastic polyurethane PSAs synthesized with 50 wt% PPG + 50 wt% PTMEG blend and different NCO/OH ratios.

^aCohesive failure of the adhesive was always obtained.

TABLE 5 | Nomenclature and composition of the poly(ether-urethane)s synthesized with different PPG + PTMEG blends.

Nomenclature	PPG (wt%)	PTMEG (wt%)	NCO/OH ratio	HS (%) ^a
1.20-100PPG	100	0	1.20	13.9
1.20-75PPG25PTMEG	75	25	1.20	13.8
1.20-50PPG50PTMEG	50	50	1.20	13.8
1.20-25PPG75PTMEG	25	75	1.20	13.8
1.20-100PTMEG	0	100	1.20	15.8

NCO/OH ratio = 1.20.

^aHard segments content. HS = $100 \times [\text{weight (MDI)} + \text{weight (BD)}] / [\text{total weight}]$.

to their G' values at high frequency. **Figure 5B** shows that 1.35-50PPG50PTMEG should exhibit excellent tack but poor cohesion due to its low G' value, and the rest of TPUs have similar rheological plots and higher G' values at lower frequency, anticipating lower tack but higher cohesion. On the other hand, all TPUs should show good peel strength because of the high values of G' at higher frequency.

The adhesion properties of the TPU PSAs made with different NCO/OH ratios were evaluated by probe tack, 180° peel strength, and creep test under shear (**Table 4**). In general, the TPU PSA made with NCO/OH ratio of 1.05 has low tack, low cohesion, and low peel strength due to its lower degree of phase separation, whereas 1.35-50PPG50PTMEG has good cohesion and peel strength, but low tack, likely due to the high content of the hydrogen-bonded urethane domains and low G' value at lower frequencies in the rheological curves. Excellent tack and peel strength values are obtained for 1.10-50PPG50PTMEG, but the cohesion is moderate because of its particular structure, and 1.20-50PPG50PTMEG shows good tack, high cohesion, and high peel strength.

Modification of the Interactions Between the Soft Segments by Changing the PTMEG Content in the Poly(Ether-Urethane)s

In this study, the interactions between the soft segments in the TPUs have been changed by mixing different amounts of PPG

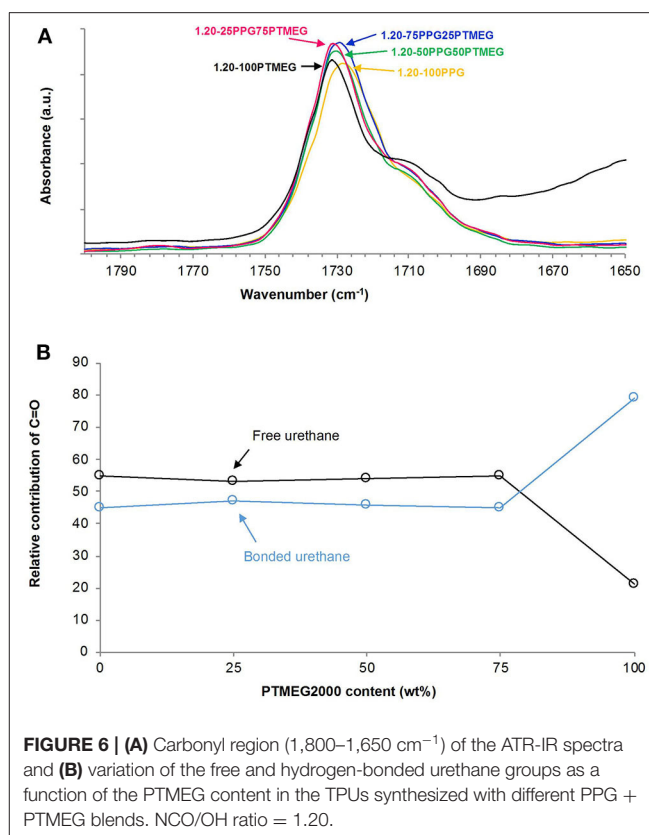


FIGURE 6 | (A) Carbonyl region (1,800–1,650 cm^{-1}) of the ATR-IR spectra and **(B)** variation of the free and hydrogen-bonded urethane groups as a function of the PTMEG content in the TPUs synthesized with different PPG + PTMEG blends. NCO/OH ratio = 1.20.

and PTMEG polyols during their syntheses; an NCO/OH ratio of 1.20 was used; this NCO/OH ratio was selected because of the adequate properties of the TPU PSAs, shown in **Table 4**. **Table 5** summarizes the composition of the TPUs PSAs in which the hard segments content is similar (~14%); 1.20-100PTMEG is an exception.

The structural characterization of the TPUs was assessed by ATR-IR spectroscopy, DSC, and TGA. All ATR-IR spectra of the TPUs show similar absorption bands, but the intensities of the C-H bands due to the polyols—2,970 to 2,940 and 2,868 to 2,854 cm^{-1} (C-H stretching), 1,454 to 1,447 cm^{-1} (C-H bending), 1,372 to 1,370 cm^{-1} (C-H rocking of methyl group)—and 1,098 to 1,099 cm^{-1} (C-O-C stretching) (**Figure S5**) change. On the other hand, the carbonyl region of the ATR-IR spectra (**Figure 6A**) is similar in all TPUs synthesized with PPG + PTMEG blends, although the wavenumber in the maximum shifts slightly from 1,728 to 1,731 cm^{-1} by increasing the PTMEG content. The percentages of the free and hydrogen-bonded urethane groups obtained by curve fitting of the carbonyl bands (**Figure S6**) are shown in **Figure 6B**, the percentages of free urethane are slightly higher than the ones of the hydrogen-bonded urethane groups in all TPUs, irrespective of the PPG + PTMEG blend, except in 1.20-100PPG in which the hydrogen-bonded urethane groups are dominant.

The DSC traces of the TPUs (**Figure 7A**) show the T_g of the soft segments between -66 and -50°C, and the increase

of the PTMEG content decreases the T_g value of the TPU. Furthermore, the DSC traces of the TPUs synthesized with PTMEG and 25 wt% PPG + 75 wt% PTMEG blend show the melting of the soft segments indicating the existence of semicrystalline domains due to the interactions between the PTMEG soft segments. On the other hand, the TGA thermograms of the TPUs (**Figure S7**) and the temperatures at which 5 and 50 wt% are lost increase by increasing the PTMEG content (**Table 6**). Furthermore, the amounts of the soft domains in the TPUs made with PPG + PTMEG blends are similar, but the temperature of decomposition

of the soft domains increases by increasing their PTMEG content, indicating higher interactions in the TPUs. The thermal decomposition of the hydrogen-bonded urethane hard domains can only be distinguished in the TPUs synthesized with more than 50 wt% PTMEG, and the percentages of free urethane hard domains in the TPUs decrease by increasing their PTMEG content, indicating stronger interactions between the hard domains. Additionally, the temperature at which the free urethane hard domains of the TPUs decomposes decreases by increasing the PTMEG content because of the stronger interactions between the hydrogen-bonded urethane hard domains.

The viscoelastic properties of the TPUs synthesized with PPG + PTMEG blends were studied by temperature sweep plate-plate rheology experiments (**Figure S8**), and the storage modulus (G') increases by increasing the PTMEG content, although the viscoelastic curves of the TPUs made with PPG or PTMEG only show different trend. All TPUs exhibit a crossover between the G' and G'' moduli (**Figure S9**), and **Table 7** shows that the temperatures at the crossover of the TPUs increase by increasing their PTMEG content, but the moduli at the crossover are similar in all TPUs.

Frequency sweep plate-plate rheology experiments were carried out to assess the viscoelastic properties at 25°C of the TPUs (**Figure 8A**). The rheological experiment of 1.20-100PTMEG could not be done because of the sliding of the plates during the experiment. The storage moduli (G') of

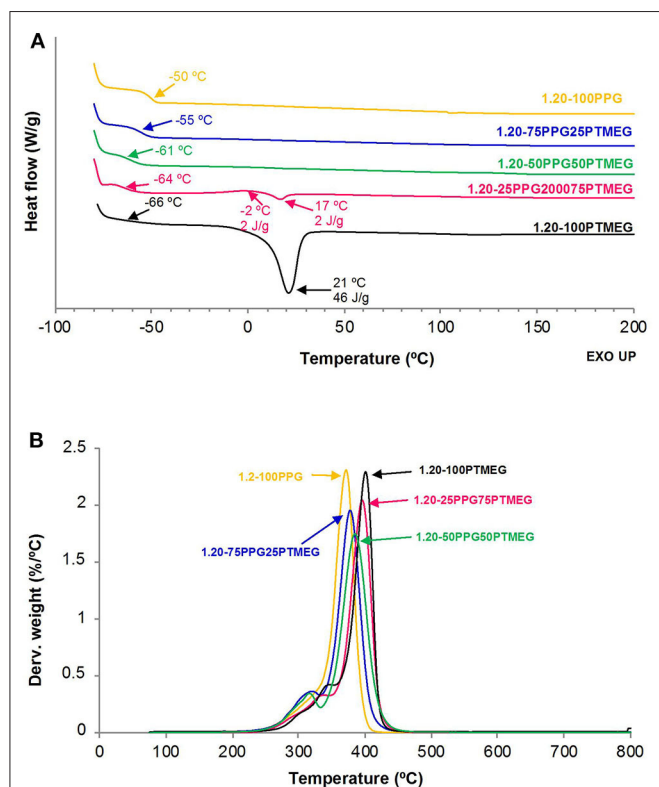


FIGURE 7 | (A) DSC traces (second heating run) and **(B)** variation of the derivative of the weight as a function of the temperature for TPUs synthesized with different PPG + PTMEG blends. NCO/OH ratio = 1.20.

TABLE 7 | Values of temperature and modulus at the crossover of G' and G'' of the TPUs synthesized with different PPG + PTMEG blends.

TPU	$T_{\text{crossover}} (^{\circ}\text{C})$	$G_{\text{crossover}} (\text{Pa})$
1.20-100PPG	1	$1.2 \cdot 10^5$
1.20-75PPG25PTMEG	25	$1.2 \cdot 10^5$
1.20-50PPG50PTMEG	33	$1.4 \cdot 10^5$
1.20-25PPG75PTMEG	49	$1.5 \cdot 10^5$
1.20-100PTMEG	25	$1.6 \cdot 10^5$

NCO/OH ratio = 1.20.

Plate-plate rheology experiments (temperature sweep).

TABLE 6 | Temperatures at which 5 ($T_{5\%}$) and 50 wt% ($T_{50\%}$) are lost, and temperatures and weight losses of the thermal decompositions of the TPUs synthesized with different PPG + PTMEG blends.

TPU PSA	$T_{5\%} (^{\circ}\text{C})$	$T_{50\%} (^{\circ}\text{C})$	1st degradation		2nd degradation		3rd degradation	
			$T_1 (^{\circ}\text{C})$	Weight loss ₁ (%)	$T_2 (^{\circ}\text{C})$	Weight loss ₂ (%)	$T_3 (^{\circ}\text{C})$	Weight loss ₃ (%)
1.20-100PPG	287	365	—	—	—	—	371	99
1.20-75PPG25PTMEG	289	371	317	18	—	—	378	80
1.20-50PPG50PTMEG	291	379	314	16	—	—	385	82
1.20-25PPG75PTMEG	298	388	287	5	332	13	396	80
1.20-100PTMEG	311	391	299	5	342	15	401	79

NCO/OH ratio = 1.20.

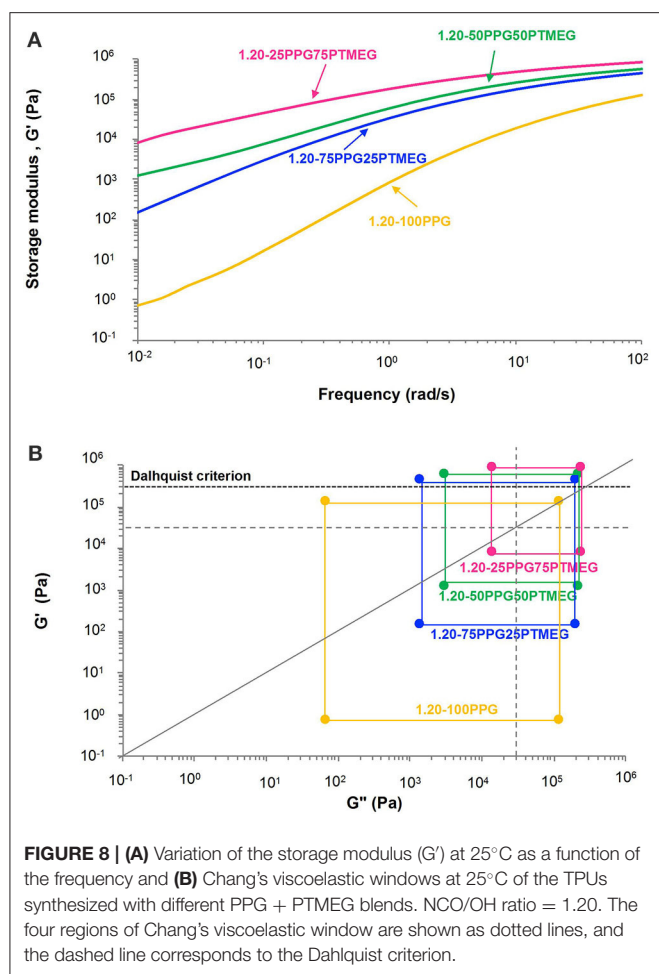


TABLE 8 | Probe tack at 25°C, holding time at 25°C, and 180° peel strength at 25°C of aluminum 5754/TPU PSA joints.

TPU PSA	Probe tack (kPa)	Holding time (min)	180° peel strength (N/cm) ^a
1.20-100PPG	438 ± 20	5 ± 0	0.21 ± 0.03 (CA)
1.20-75PPG25PTMEG	661 ± 39	240 ± 0	4.62 ± 0.60 (CA)
1.20-50PPG50PTMEG	576 ± 17	1,518 ± 167	9.91 ± 0.56 (CA)
1.20-25PPG75PTMEG	148 ± 13	> 3 days	12.03 ± 2.45 (AF)

Thermoplastic polyurethanes synthesized with different PPG + PTMEG blends. NCO/OH ratio = 1.20.

^aCA, cohesive failure of the adhesive; AF, adhesion failure.

the TPUs increase by increasing their PTMEG content, as a consequence of the existence of semicrystalline domains of PTMEG soft segments, which decrease the molecular flexibility and the mobility of the polymeric chains (Inescu, 2005). The kind of PSA can be assessed by means of Chang's (1991) viscoelastic windows, which are built from the values of the storage (G') and loss (G'') moduli at low (10^{-2}

rad/s) and high (10^2 rad/s) frequencies. Chang proposed the existence of four quadrants to classify the different types of PSAs, depending on their location of the viscoelastic window. Chang's viscoelastic windows of the TPU PSAs are shown in **Figure 8B**, in which all them are below the Dahlquist criterion line indicating a good contact efficiency. By increasing the PTMEG content of the TPU PSAs, Chang's viscoelastic windows shift from the lower left quadrant characterized by low storage modulus and low dissipation (removable PSA) (1.20-100PPG), to the central region characterized by medium storage modulus and medium dissipation (general purpose PSA) (1.20-25PPG75PTMEG).

The adhesion properties of the TPU PSAs are summarized in **Table 8**. The increase of the PTMEG content increases the cohesion (or holding time) and the peel strength of the TPU PSAs, but decreases the tack. This trend can be ascribed to the interactions between the PTMEG soft segments, which produce semicrystalline domains and reduce the mobility of the polyurethane chains. However, 1.20-50PPG50PTMEG PSA shows an excellent compromise between tack, holding time, and 180° peel strength, and it can be used as removable PSA in labels and medical patches.

CONCLUSIONS

Different TPUs intended for PSAs have been prepared by using blends of polyols with and without methyl pendant groups. The NCO/OH ratio was varied for increasing the hard segments contents, and the structure of the TPUs was similar, except in the one made with NCO/OH ratios of 1.10 and 1.35. The increase of the percentage of PTMEG polyol produced an increase of the thermal stability and the storage moduli of the TPUs due to the creation of semicrystalline regions due to the interactions between the soft segments. Excellent tack, shear strength, and peel strength were obtained in 1.20-50PPG50PTMEG PSA.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MF carried out the experiments and write the first draft of the manuscript. JM-M designed the experiments, contributed to the discussion of the results, and wrote the final manuscript. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmech.2020.00034/full#supplementary-material>

REFERENCES

- Chang, E. P. (1991). Viscoelastic windows of pressure-sensitive adhesives. *J. Adhes.* 34, 189–200. doi: 10.1080/00218469108026513
- Chen, X., Liu, W., Zhao, Y., Jiang, L., Xu, H., and Yang, X. (2009). Preparation and characterization of PEG-modified polyurethane pressure-sensitive adhesives for transdermal drug delivery. *Drug Dev. Ind. Pharm.* 35, 704–711. doi: 10.1080/03639040802512235
- Czech, Z., Kowalczyk, A., and Swiderska, J. (2011). “Pressure sensitive adhesives for medical applications,” in *Wide Spectra of Quality Control*, ed I. Akyar (Rijeka: InTech), 309–332.
- Dahlquist, C. A. (1969). “Pressure-sensitive adhesives,” in *Treatise on Adhesion and Adhesives*. ed R. L. Patrick (New York, NY: Marcel Dekker), 219–260.
- Deraill, C., and Marin, G. (2009). “Role of viscoelastic behavior of pressure-sensitive adhesives in the course of bonding and debonding processes,” in *Handbook of Pressure-Sensitive Adhesives and Products. Fundamentals of Pressure Sensitivity*. Eds I. Benedek and M. M. Feldstein (Boca Raton: CRC Press Chapter 4), 1–26.
- Feldstein, M. M., and Siegel, R. A. (2012). Molecular and nanoscale factors governing pressure-sensitive adhesion strength of viscoelastic polymers. *J. Polym. Sci. Part B: Polym. Phys.* 50, 739–772. doi: 10.1002/polb.23065
- Ferguson, J., and Petrovic, Z. (1976). Thermal stability of segmented polyurethanes. *Eur. Polym. J.* 12, 177–181. doi: 10.1016/0014-3057(76)90050-1
- Fuensanta, M., and Martín-Martínez, J. M. (2018). Thermoplastic polyurethane coatings made with mixtures of polyethers of different molecular weights with pressure sensitive adhesion property. *Prog. Org. Coat.* 118, 148–156. doi: 10.1016/j.porgcoat.2017.11.021
- Fuensanta, M., and Martín-Martínez, J. M. (2019). Thermoplastic polyurethane pressure sensitive adhesives made with mixtures of polypropylene glycols of different molecular weights. *Int. J. Adhes. Adhes.* 88, 81–90. doi: 10.1016/j.ijadhadh.2018.11.002
- Higgins, J. J., Jagisch, F. C., and Stucker, N. E. (1989). “Butyl rubber and polyisobutylene,” in *Handbook of Pressure Sensitive Adhesive Technology*. ed D. Satas (New York, NY: Van Nostrand Reinhold), 374–395.
- Inescu, M. (2005). “Relationships between the oligo-polyol structure and polyurethane properties,” in *Chemistry and Technology of Polyols for Polyurethanes*. ed M. Ionescu (Shawbury: Rapra Technology Limited), 535–550.
- Mattia, J., and Painter, P. (2007). A comparison of hydrogen bonding and order in a polyurethane and poly(urethane-urea) and their blends with poly(ethylene glycol). *Macromolecules*. 40, 1546–54. doi: 10.1021/ma0626362
- Satas, D., and Satas, A. M. (1989). “Hospital and first aid products,” in *Handbook of Pressure Sensitive Adhesive Technology*. ed D. Satas (New York, NY: Van Nostrand Reinhold), 627–642.
- Strikovskiy, A. G., and Zharkov, V. V. (1993). Infra-red spectroscopy study of equilibrium association of urethane groups in poly(ether urethane)s. *Polymer*. 34, 3397–3401. doi: 10.1016/0032-3861(93)90467-O
- Tombs, E. L., Nikolaou, V., Nurumbetov, G., and Haddleton, D. M. (2018). Transdermal delivery of ibuprofen utilizing a novel solvent-free pressure-sensitive adhesive (PSA): TEPI[®] technology. *J. Pharm. Innov.* 13, 48–57. doi: 10.1007/s12247-017-9305-x
- Yilgör, I., Yilgör, E., and Wilker, G. L. (2015). Critical parameters in designing segmented polyurethanes and their effect on morphology and properties: A comprehensive review. *Polymer*. 58, A1–A36. doi: 10.1016/j.polymer.2014.12.014

Conflict of Interest: The authors declare that the PPG polyol sample was donated by Repsol (Madrid, Spain). The company was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Copyright © 2020 Fuensanta and Martín-Martínez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effect of Surface Roughness on Adhesive Instabilities for the Elastic Layer

Junki Joe^{1*}, M. D. Thouless^{1,2} and J. R. Barber¹

¹ Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, United States, ² Department of Materials Science and Engineering, University of Michigan, Ann Arbor, MI, United States

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine
Elena Torskaya,
Institute for Problems in Mechanics
(RAS), Russia

*Correspondence:

Junki Joe
jkjoe@umich.edu

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 19 March 2020

Accepted: 27 April 2020

Published: 22 May 2020

Citation:

Joe J, Thouless MD and Barber JR
(2020) Effect of Surface Roughness
on Adhesive Instabilities for the Elastic
Layer. *Front. Mech. Eng.* 6:31.
doi: 10.3389/fmech.2020.00031

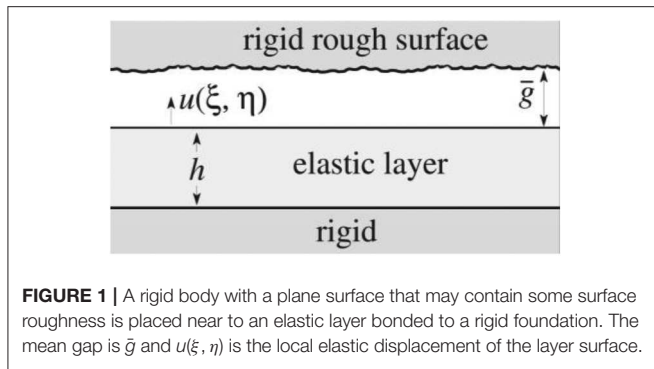
If an incompressible elastic layer bonded to a rigid plane is placed close to a second rigid plane, adhesive interactions between the surfaces can cause elastic instabilities. These lead to spatially non-uniform traction and gap distributions which exhibit a regular pattern with a characteristic wavenumber. However, real surfaces are never completely plane. In this paper, we consider the influence of surface roughness on the instability, with particular reference to the force-displacement relation. With random roughness profiles, the traction distributions are always spatially irregular, so the onset of instability is more difficult to define. One approach is to monitor the amplitude of the power spectrum of the distribution near the characteristic wavenumber. Since surface roughness generally reduces the mean adhesive traction, we might expect it to exert a stabilizing effect. Numerical results confirm this for moderate to large RMS amplitudes, but show that low RMS roughness can actually trigger the instability in ranges where the uniform layer would be stable. The resulting traction-displacement relation is then found to be approximately linear with a slope close to that at the point where the uniform solution loses stability.

Keywords: contact mechanics, surface roughness, adhesion, elastic layers, patterning

INTRODUCTION

If two bodies with plane surfaces are placed close together, they may experience attractive [e.g., van der Waals'] forces, or forces involving both attractive and repulsive ranges (Jones, 1924; Maugis, 2013). Since the attractive forces must eventually decay with increasing separation, they have the character of a “negative spring,” which can trigger an elastic instability. If the bodies are incompressible [Poisson’s ratio $\nu = 0.5$], or if a body comprising a thin elastic layer bonded to a rigid plane surface is attracted to another rigid plane surface, the instability may result in a non-uniform [typically periodic] pattern of alternating regions of contact and separation. Patterns of this kind have been observed experimentally (Mönch and Herminghaus, 2001; Gonuguntla et al., 2006a), and predicted theoretically, based on energetic arguments (Shenoy and Sharma, 2001; Sarkar et al., 2004). In particular, the characteristic length scale of the pattern correlates with the unstable wavelength in a linear perturbation of the uniform state. The patterning instability also modifies the mean traction-separation characteristic for the layered system, generally leading to different behavior during approach and separation and consequent hysteresis losses (Ciavarella et al., 2019).

The instability permits self-assembly processes such as elastic contact lithography [ECL], where the pattern in a polymer film is fixed by UV curing or by lowering the temperature (Sarkar and Sharma, 2010; Ghosh et al., 2016, 2017). In ECL, the periodicity and size of the pattern are critical parameters and various methods have been proposed to control them, including the use of a curved



substrate (Ghosh et al., 2018), an imprinted stamp (Gonuguntla et al., 2006b; Mukherjee et al., 2007; Bhandaru et al., 2017), or a pre-strained substrate (Davis-Purcell et al., 2018). Also, electric fields can be used to extend the range of attraction relative to van der Waals' forces and to provide greater control of the process, whilst retaining a similar morphology (Arun et al., 2006; Sarkar et al., 2008; Sahoo et al., 2019).

Real surfaces are of course never perfectly smooth, and surface roughness generally reduces both the maximum pull-off traction, and the maximum negative slope of the effective traction-separation law (Joe et al., 2018). This should reduce the tendency for patterning in the contact of layered bodies. However, sufficiently small amplitude roughness might also serve as an initial perturbation to trigger an instability. In this paper, we shall therefore use a numerical solution to examine the effect of roughness on both the generation of patterns and the mean traction-separation relation. In particular, we shall examine the extent to which the effect of roughness can be captured by using a modified adhesive traction law developed for the contact of rough elastic half spaces.

DEFORMATION OF A THIN LAYER

We consider an elastic layer of thickness h bonded to a rigid plane, and define Cartesian coordinates (x, y) in the plane of the layer surface and corresponding dimensionless coordinates $\xi = x/h, \eta = y/h$. If a second rigid plane is placed a distance \bar{g} away from the undeformed surface of the layer as shown in **Figure 1**, the local gap between the surfaces will then be

$$g(\xi, \eta) = \bar{g} - u(\xi, \eta) \quad (1)$$

where $u(\xi, \eta)$ is the local outward normal elastic displacement of the layer surface.

Interface Energy

We assume that the adhesive tractions between the surfaces can be described by a traction law $\sigma(g)$, where g is the local value of the gap. We can then also define the mean interface energy per unit area as

$$\Gamma = \langle \gamma(g(\xi, \eta)) \rangle \text{ where } \gamma(g) = - \int_g^\infty \sigma(s) ds \quad (2)$$

and a stable final configuration will be one that minimizes the total potential energy $\Pi = U + \Gamma$, where U is the mean elastic strain energy per unit area.

Elastic Strain Energy

The normal traction $\sigma(\xi, \eta)$ at the free surface of the layer needed to produce a sinusoidal normal elastic displacement $u(\xi, \eta) = u_\zeta \cos(\zeta \xi)$ is

$$\sigma(\xi, \eta) = \frac{Ef(\zeta)u_\zeta}{h} \cos(\zeta \xi) \quad (3)$$

where

$$f(\zeta) = \frac{\zeta [(3 - 4\nu) \cosh(2\zeta) + 2\zeta^2 + 5 - 12\nu + 8\nu^2]}{2(1 - \nu^2) [(3 - 4\nu) \sinh(2\zeta) - 2\zeta]} \quad (4)$$

(Hannah, 1951), and we recall that $\xi = x/h$, so ζ is a dimensionless wavenumber.

In this paper, we shall restrict attention to incompressible layers [$\nu = 0.5$], for which the corresponding dimensionless compliance $1/f(\zeta)$ is shown as a function of dimensionless wavenumber ζ in **Figure 2**. The curve exhibits a maximum of ~ 0.482 at a wavenumber $\zeta \approx 2.1$, and zero compliance for uniform loading [$f(\zeta) \rightarrow \infty$ as $\zeta \rightarrow 0$]. One consequence of this is that with general loading $\sigma(\xi, \eta)$, the mean value of u is zero and hence \bar{g} is determined by the rigid-body approach which is a controlled parameter.

The mean elastic strain energy per unit area associated with the deformation (3) is

$$U(\zeta) = \frac{1}{2} \langle \sigma(\xi, \eta) u_\zeta(\xi, \eta) \rangle = \frac{Ef(\zeta)u_\zeta^2}{4h} \quad (5)$$

The elastic strain energy for more general displacement distributions can be obtained by writing $u(\xi, \eta)$ as a double Fourier series or Fourier transform and convoluting the resulting transform with (5).

Stability Criterion

If both surfaces are plane [i.e., smooth], the state $u(\xi, \eta) = 0, g(\xi, \eta) = \bar{g}, \sigma(\xi, \eta) = \sigma(\bar{g})$ is clearly an equilibrium state, but it will be unstable to small sinusoidal perturbations of dimensionless wavenumber ζ if there exists any ζ such that

$$- \left(\frac{\partial \sigma}{\partial g} \right)_{g=\bar{g}} > \frac{Ef(\zeta)}{h} \quad (6)$$

The critical wavenumber is defined by the maximum of the curve in **Figure 2**, from which we deduce that the uniform solution will be unstable if and only if $-\sigma'(\bar{g}) > E/0.482h$.

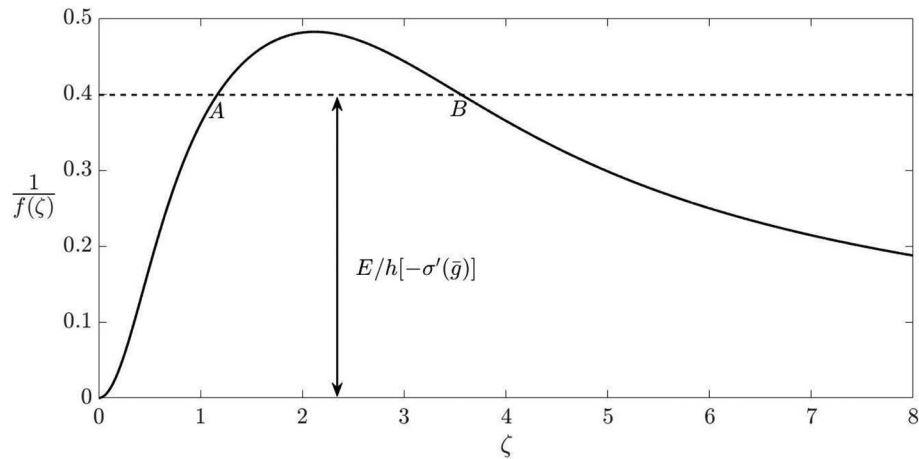


FIGURE 2 | Dimensionless layer compliance as a function of wavenumber for an incompressible elastic layer bonded to rigid foundation. For a given value of $E/h[-\sigma'(\bar{g})]$ satisfying condition (6), wavenumbers in the range $\zeta_A < \zeta < \zeta_B$ are unstable.

Solution Method

More general displacement distributions for a square domain $0 < x < L, 0 < y < L$ can be written in the form

$$u(x, y) = \Re \sum_{m=-N}^N \sum_{n=-N}^N A_{mn} \exp \left[\frac{2\pi i(mx + ny)}{L} \right], \quad (7)$$

and equation (5) can then be used to obtain the mean strain energy per unit area U as

$$U = \sum_{m=-N}^N \sum_{n=-N}^N \frac{E|A_{mn}|^2 f(\zeta_{mn})}{4h} \quad (8)$$

where

$$\zeta_{mn} = \frac{2\pi h \sqrt{m^2 + n^2}}{L} \quad (9)$$

Adding the corresponding interface energy from equation (2), we obtain the total potential energy Π , which must be a minimum at a stable equilibrium state. We used the gradient descent method to identify the values of the Fourier coefficients A_{mn} for a local energy minimum. Notice that the upper limit N must be chosen so as to provide an adequate number of integration points in the evaluation of Γ .

RESULTS FOR SMOOTH SURFACES

If the two surfaces are smooth, the uniform state is always an equilibrium solution at which energy gradients are zero, and since the material is incompressible [$\nu = 0.5$], this corresponds to $u(\xi, \eta) = 0$. Even when the criterion (6) is satisfied, the numerical solution may remain at the uniform state unless some small perturbation is introduced.

We used the traction law

$$\sigma(g) = \frac{8\Delta\gamma}{3\varepsilon} \left[\left(\frac{\varepsilon}{g} \right)^3 - \left(\frac{\varepsilon}{g} \right)^9 \right] \quad (10)$$

(Maugis, 2013) derived from Lennard-Jones molecular force law (Jones, 1924), where ε is an interatomic length scale and $\Delta\gamma = \gamma(\varepsilon)$ is the interface energy per unit volume at the equilibrium spacing $g = \varepsilon$. The maximum tensile traction occurs at $g = 3^{1/6}\varepsilon$ and is of magnitude $\sigma_0 \approx 1.06\Delta\gamma\varepsilon$.

If the maximum negative slope $[-\sigma'(g)]_{\max}$ satisfies the condition

$$\frac{h}{E} [-\sigma'(g)]_{\max} > \frac{1}{0.482} \quad (11)$$

there will exist a bounded range $g_1 < \bar{g} < g_2$ in which the uniform solution is unstable. Also, for a value of \bar{g} strictly within this range, there will exist a range of unstable wavenumbers $\zeta_A < \zeta < \zeta_B$, including but not limited to $\zeta \approx 2.1$. This range can be identified by drawing a horizontal line at the height $E/h[-\sigma'(\bar{g})]$ in **Figure 2** as shown, and finding its intersection with the curve.

We start the solution procedure with a value of \bar{g} outside the unstable range and then change \bar{g} by small increments, using the solution at the previous step as an initial guess for the gradient descent solution. This is expected to mimic the behavior of the physical system under controlled displacement conditions. Numerical noise might also be expected to emulate the effect of noise [e.g., vibration] in an experimental system.

We characterize the inverse thickness of the layer by the dimensionless parameter

$$\beta = \frac{E\varepsilon^2}{h\Delta\gamma} \quad (12)$$

Figure 3 compares a typical relation between the mean traction $\bar{\sigma} = \langle \sigma(x, y) \rangle$ and the mean gap [approach] \bar{g} with the local traction law (1) for the case where $\beta = 0.25$. The domain

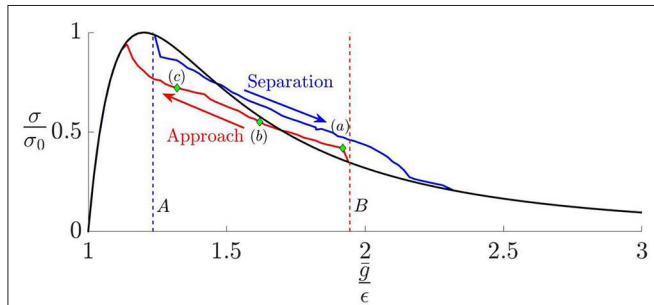


FIGURE 3 | Mean traction $\bar{\sigma}$ as a function of mean gap \bar{g} for a smooth layer for $\beta = 0.25$ and domain size $L = 8\pi h$, showing instabilities during approach and separation. The Lennard-Jones law (1) is shown in black. The dashed lines **A**, **B** define the region in which the uniform traction solution is unstable.

size was chosen so as to define a fundamental wavenumber [e.g., ζ_{01}] equal to 0.25, so that $L = 8\pi h$. Results are shown for both separation $d\bar{g}/dt > 0$ and approach $d\bar{g}/dt < 0$. In each case, the uniform state is preserved up to the appropriate stability boundary [denoted by **A** and **B**, respectively], but the non-uniform solution then persists significantly beyond the point at which the uniform solution reverts to stability. We deduce that even in the stable range there exist local energy minima corresponding to non-uniform states, and that these states are separated from the lower energy uniform state by energy barriers.

Notice that both the approach and separation curves in **Figure 3** are approximately straight lines with slope close to the critical slope defined by (6).

The non-uniform deformation states are characterized by the development of regular patterns. **Figures 4A–C** shows contours of local gap $g(x, y)$ corresponding to the points on the approach curve labeled (a), (b) and (c) in **Figure 3**. A “labyrinth” [i.e., a connected system of passageways (high g) separated by walls of “contact” (low g)] develops at the onset of instability (a) and

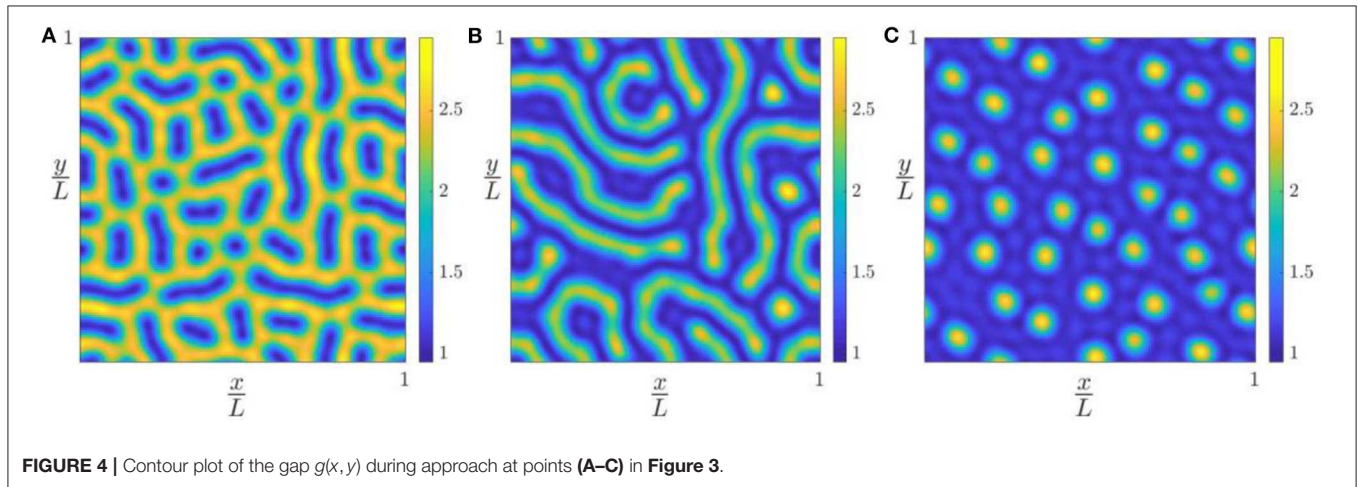


FIGURE 4 | Contour plot of the gap $g(x, y)$ during approach at points **(A–C)** in **Figure 3**.

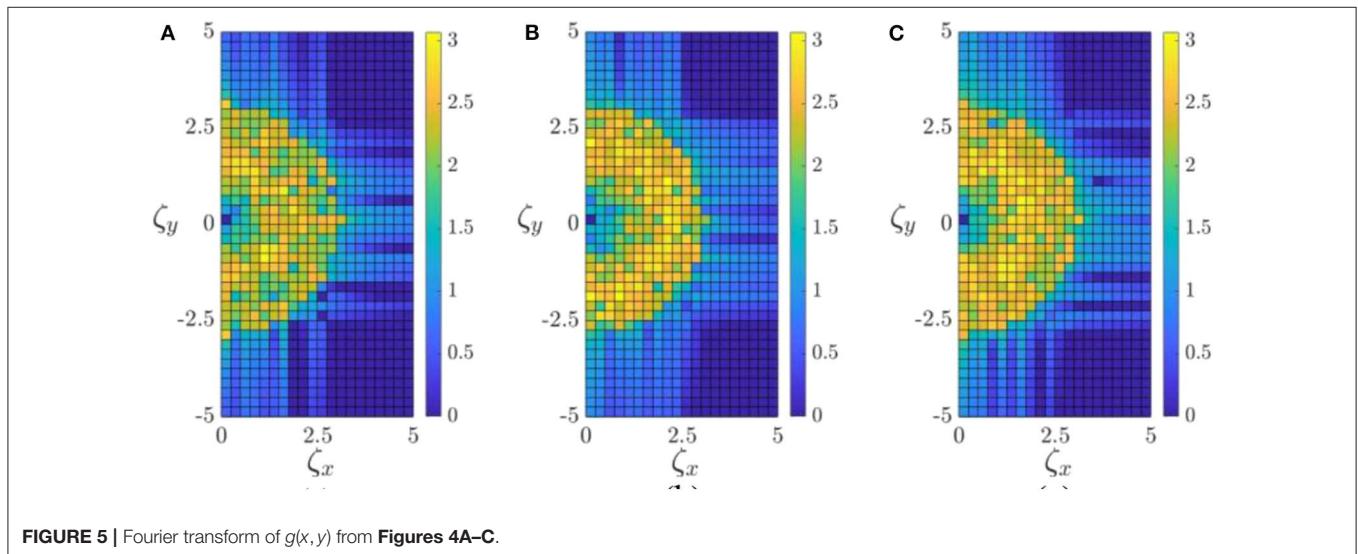


FIGURE 5 | Fourier transform of $g(x, y)$ from **Figures 4A–C**.

then morphs into an inverse labyrinth at (b) and into an array of isolated regions of separation [“holes”] at (c).

Figures 5A–C shows the same results in Fourier transform space. In all cases we see high values clustered near the most unstable wavenumber $\zeta = 2.1$ and the distribution is axisymmetric within the limits of statistical variance, indicating that the pattern is statistically isotropic.

EFFECT OF SURFACE ROUGHNESS

We assume that the surface is rough with a power spectral density [PSD] of the form

$$P(\zeta) = B\zeta^{-2-2H}; \quad \zeta_1 < \zeta < \zeta_2, \quad (13)$$

where B is a constant, H is the Hurst exponent, here taken as $H = 0.2$, and ζ_1, ζ_2 define the range of dimensionless wavenumbers in

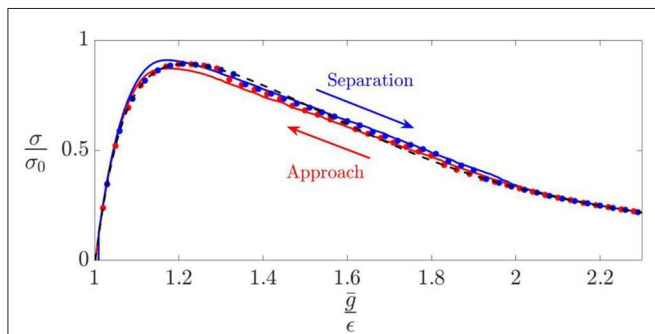


FIGURE 6 | Mean traction $\bar{\sigma}$ as a function of mean gap \bar{g} for a layer with fine-scale roughness defined by equation (14) with $\beta = 0.35$, $\zeta_1 = 6.5$, $\zeta_2 = 8.0$ and $m_0 = 10^{-1.8}\epsilon^2$. The solid lines correspond to a direct numerical solution for the rough surface using the Lennard-Jones traction law, whilst the circles were obtained by approximating the effect of roughness through a modified traction law $\sigma_M(g)$ (Joe et al., 2018). [shown here as a dashed line].

the spectrum, outside which the spectral content is zero. For the finite domain $L \times L$, a realization of this PSD can be written as

$$u_0(x, y) = \Re \sum_m \sum_n B_{mn} \exp \left[\frac{2\pi i(mx + ny)}{L} \right] \quad (14)$$

where $u_0(x, y)$ describes the deviation of the surface from the mean plane in the undeformed state, and

$$\frac{\zeta_1 L}{2\pi h} < \sqrt{m^2 + n^2} < \frac{\zeta_2 L}{2\pi h} \quad (15)$$

The magnitudes of the coefficients $|B_{mn}|$ were chosen so as to ensure that the resulting surface PSD was of the form (13) and the corresponding arguments [phases] were chosen randomly.

With this definition, the gap can be written

$$g(x, y) = \bar{g} + u_0(x, y) + u(x, y), \quad (16)$$

where $u(x, y)$ is given by (7). Notice that we arbitrarily assign the value $B_{00} = 0$, so that the roughness makes no contribution to the mean gap. The interface energy is then determined from (2) and the constants A_{mn} are chosen so as to minimize Π as in the smooth surface case.

A Two-Scale Approximation

We anticipate that the wavelengths in the roughness spectrum will generally be significantly smaller than the layer thickness, and this suggests the possibility of a scale-separation approach. Compared with the roughness scale, the thickness of the layer is large, so local effects can be approximated by those in a corresponding half space. The effect of the roughness, as compared with a corresponding smooth surface, can therefore be described in terms of a modified traction law.

An inductive method for estimating this law is described in (Joe et al., 2018). If the modified traction law $\sigma_M(g, \zeta_0)$ is known

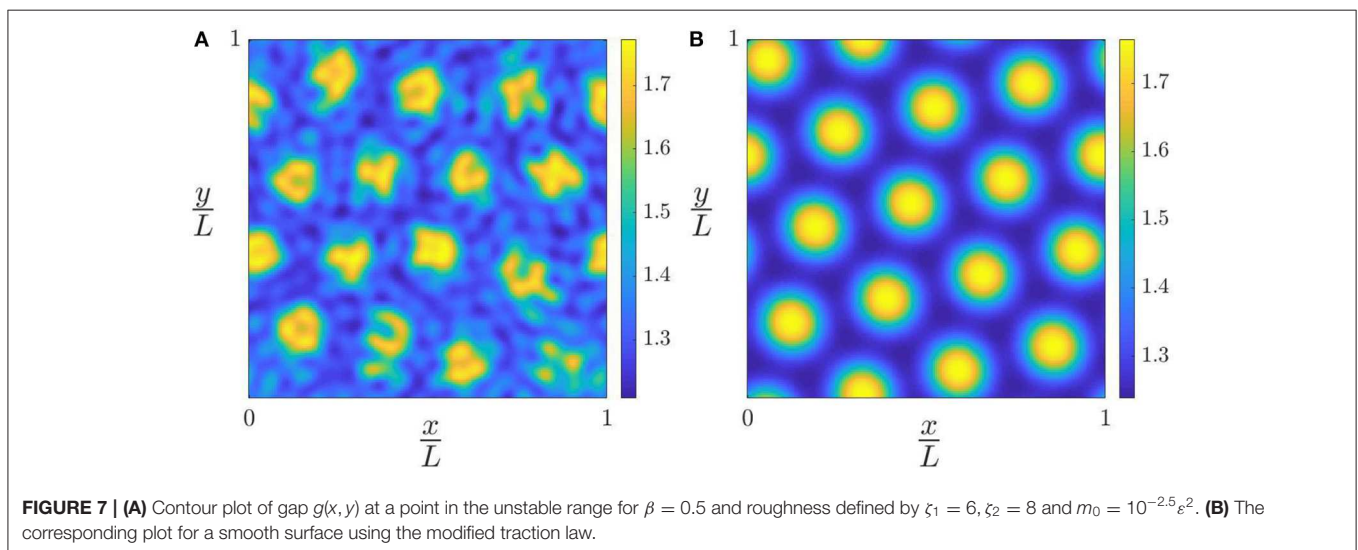


FIGURE 7 | (A) Contour plot of gap $g(x, y)$ at a point in the unstable range for $\beta = 0.5$ and roughness defined by $\zeta_1 = 6$, $\zeta_2 = 8$ and $m_0 = 10^{-2.5}\epsilon^2$. **(B)** The corresponding plot for a smooth surface using the modified traction law.

for a surface with spectral content only in $\zeta_0 < \zeta < \zeta_2$, where $\zeta_0 > \zeta_1$, the corresponding law for $\sigma_M(g, \zeta_0 - \delta\zeta)$ can be obtained by convoluting $\sigma_M(g, \zeta_0)$ with the additional roughness tranche $\zeta_0 - \delta\zeta < \zeta < \zeta_0$. Successive applications of this technique allow us to determine the modified law for the entire spectrum. In Joe et al. (2018), this procedure was implemented using discrete tranches of the spectrum, but the same approach can be used to develop a partial differential equation for $\sigma_M(g, \zeta)$, following the methodology of Persson (2001). The modified traction law for the complete roughness spectrum is then defined by $\sigma_M(g) = \sigma_M(g, \zeta_1)$.

On the scale of the layer thickness, the effect of the surface roughness can then be approximated by using $\sigma_M(g)$ in place of equation (1), and treating the layer surface as smooth.

Results

The two-scale approximation is likely to be most accurate when the most unstable wavenumber $\zeta \approx 2.1$ is much smaller than the smallest wavenumber ζ_1 in the PSD [recall that wavenumbers are normalized by the layer thickness h]. However, this degree of scale separation is difficult to achieve in a direct numerical simulation. **Figure 6** shows the relation between mean traction and mean gap for an incompressible elastic half space with roughness of the form (13) with $\zeta_1 = 6.5, \zeta_2 = 8.0$ and height variance

$$m_0 = 2\pi \int_{\zeta_1}^{\zeta_2} \zeta P(\zeta) d\zeta = 10^{-1.8} \varepsilon^2 \quad (17)$$

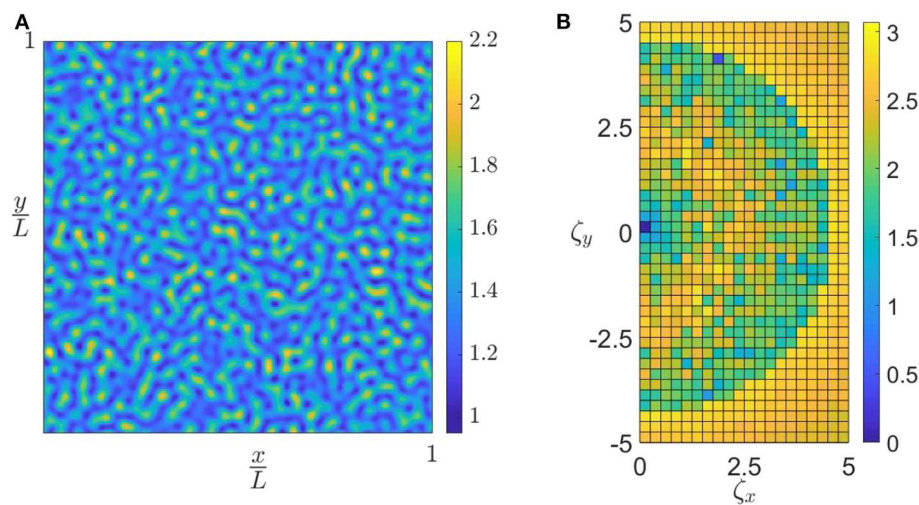


FIGURE 8 | (A) Contour plots of gap $g(x, y)$ in the unstable range for $\beta = 0.5$ and roughness defined by $\zeta_1 = 4.5, \zeta_2 = 8$ and $m_0 = 10^{-2} \varepsilon^2$. **(B)** Fourier domain plot for the gap distribution $g(x, y)$ from **(A)**.

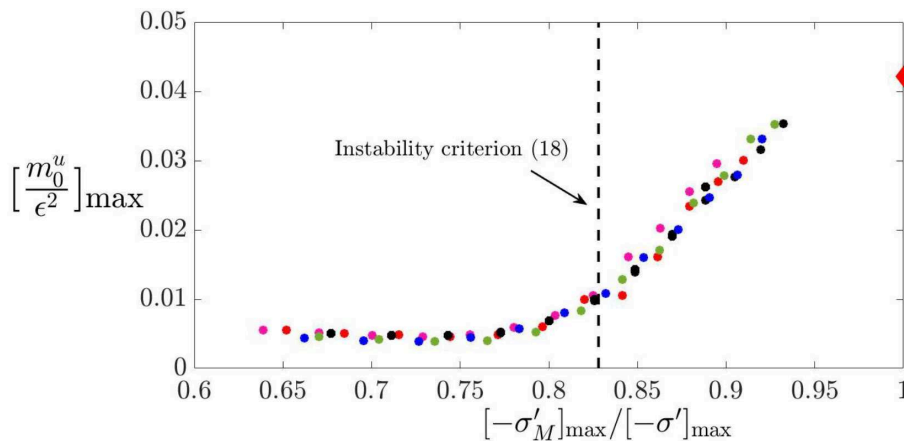


FIGURE 9 | Normalized spectral content $[m_0^u / \varepsilon^2]_{\max}$ for the gap $g(x, y)$ in the unstable wavenumber range as a function of the maximum negative slope of the modified traction law $[-\sigma'_M(g)]_{\max}$. Consistency between several roughness spectra shows that $[-\sigma'_M(g)]_{\max}$ is a good indicator of the effect of roughness on instability. The point at the top right defines $[m_0^u / \varepsilon^2]_{\max}$ for a smooth layer.

Barber (2018). In this figure, we compare two methods of solution: (i) a direct numerical solution of the problem [solid line] by minimizing the total energy for a gap defined by (16) using the Lennard-Jones traction law and (ii) a solution in which the layer is smooth [$u_0(x, y) = 0$], but the traction law is modified to describe the effect of roughness [circles]. The agreement is clearly extremely good. The advantage of using the modified traction law is that the deformation of the layer can then be adequately described on a much coarser grid, or equivalently, by a more severely truncated Fourier series, and this is particularly useful if the spectral range $[\zeta_1, \zeta_2]$ is broad.

Identifying Instability Effects

The argument of the previous section suggests that pattern instabilities should occur if and only if the maximum slope of the modified traction law $[\sigma_M(\bar{g})]$ satisfies the condition (11) — i.e.,

$$\frac{h}{E} [-\sigma'_M(\bar{g})]_{\max} > \frac{1}{0.482} \quad (18)$$

The modified traction law $\sigma_M(\bar{g})$ is independent of h , so instabilities are more likely for thicker layers, always assuming the finite linear dimensions of the interface are sufficient to accommodate a wavelength in the unstable range.

In some cases, pattern instabilities can be detected by examining the corresponding contour plots for $g(x, y)$. For example, **Figure 7A** shows the contour plot for the actual rough surface during separation at a mean gap $\bar{g} = 1.4$, and **Figure 7B** shows the corresponding pattern predicted for a smooth surface with the modified traction law. There is some blurring of the observed patterns, but the overall morphology is clearly similar.

However, if the roughness amplitude is larger, patterns become more blurred, a typical example being shown in **Figure 8A**. In this case, detection of instability from gap contours is more difficult, but the corresponding Fourier plot of **Figure 8B** clearly shows substantial spectral content in the unstable range near $\zeta = 2.1$. This suggests that we might quantify the extent of pattern formation in the numerical solution for the rough surface using the dimensionless parameter m_0^u/ε^2 , where m_0^u is the variance of that part of the gap PSD that lies in the unstable range in the “smooth” solution, identified in **Figure 2**. We assume here that the roughness PSD has no content in this wavenumber range, since otherwise it would be difficult to distinguish the separate effects of instability and roughness.

In **Figure 9** we plot $[m_0^u/\varepsilon^2]_{\max}$, obtained from the numerical solution, as a function of the maximum negative slope of the modified traction law $[-\sigma'_M(\bar{g})]_{\max}$, normalized by the corresponding expression for the Lennard-Jones law. Each set of points corresponds to a different value of the lower wavenumber of the roughness $\zeta_1 = 4.5:0.5:6.5$ and a range of roughness variances $-2.7 < \log_{10}(m_0/\varepsilon^2) < -1.6$. Results were

obtained under both approach and separation conditions, but no significant difference was observed. The vertical dashed line in **Figure 9** corresponds to the criterion (18), below which the two-scale approximation would predict $m_0^u = 0$, though the direct numerical results exhibit a level of noise as one might expect. However, the results exhibit a remarkable level of consistency, showing that $[-\sigma'_M(\bar{g})]_{\max}$ is a very good indicator of the effect of roughness on the instability, and more generally that the two-scale approach to the layer problem defines a good approximation to important features of the system behavior.

CONCLUSIONS

If a smooth elastic layer is placed close to a plane surface, elastic instabilities due to adhesive tractions lead to the development of patterns, and to a modification of the traction-separation law. However, roughness with RMS amplitude comparable with the range of the adhesive force law can have a significant effect on this process. Here we have described a model for analyzing the contact of both rough and smooth surfaces using a double Fourier series.

We also developed a two-scale approximation to the rough contact behavior, by (i) estimating the effect of roughness on the mean traction between two half spaces, using a previously published method (Joe et al., 2018), and then (ii) using this modified traction law in the analysis of a smooth elastic layer. Results show that this gives a very good approximation to the traction-separation law obtained by direct numerical simulation. In particular, the development of patterns is predicted if the maximum slope of the modified traction law satisfies the inequality (18) and the corresponding results correlate extremely well with a criterion based on the spectral content in the unstable range from the numerical solution.

Local layer deformations decay spatially at a rate linked to the layer thickness, so this method is expected to give good predictions for bodies of finite size sufficient to support wavelengths in the unstable range.

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/supplementary files.

AUTHOR CONTRIBUTIONS

JJ wrote the first draft of the manuscript, performed the statistical analysis and organized the database. JB wrote sections of the manuscript. All authors contributed to conception and design of the study, manuscript revision, read, and approved the submitted version.

REFERENCES

- Arun, N., Sharma, A., Shenoy, V. B., Narayan K. (2006). Electric-field-controlled surface instabilities in soft elastic films. *Adv. Mater.* 1, 660–663. doi: 10.1002/adma.200502199
- Barber, J. R. (2018). *Contact Mechanics*. Cham: Springer. doi: 10.1007/978-3-319-70939-0
- Bhandaru, N., Sharma, A., and Mukherjee, R. (2017). Programmable nanopatterns by controlled debonding of soft elastic films. *ACS Appl. Mater. Interfaces* 9, 19409–19416. doi: 10.1021/acsami.6b09127

- Ciavarella, M., Joe, J., Papangelom A., Barber, J. R. (2019). The role of adhesion in contact mechanics. *J. R. Soc. Interface* 16:20180738. doi: 10.1098/rsif.2018.0738
- Davis-Purcell, B., Soulard, P., Salez, T., Raphael, E., and Dalnoki-Veress, k., (2018). Adhesion-induced fingering instability in thin elastic films under strain. *Eur. Phys. J. E* 41:36. doi: 10.1140/epje/i2018-11643-6
- Ghosh, A., Bandyopadhyay, D., Sarkar, J., and Sharma, A., (2017). Hierarchical micro- and nanofabrication by pattern-directed contact instabilities of thin viscoelastic films. *Phys. Rev. Fluids* 2:124004. doi: 10.1103/PhysRevFluids.2.124004
- Ghosh, A., Bandyopadhyay, D., and Sharma, A. (2016). Influence of the mutable kinetic parameters on the adhesion and debonding of thin viscoelastic films. *J. Colloid Interface Sci.* 477, 109–122. doi: 10.1016/j.jcis.2016.05.036
- Ghosh, A., Bandyopadhyay, D., and Sharma, A. (2018). Micro-patterning of coatings on a fiber surface exploiting the contact instabilities of thin viscoelastic films. *Phys. Fluids*. 30:114101. doi: 10.1063/1.5053797
- Gonuguntla, M., Sharma, A., Mukherjee, R., and Subramanian, S. A. (2006a). Control of self-organized contact instability and patterning in soft elastic films. *Langmuir* 22, 7066–7071. doi: 10.1021/la0600696
- Gonuguntla, M., Sharma, A., and Subramanian, S. A. (2006b). Elastic contact induced self-organized patterning of hydrogel films. *Macromolecules* 39, 3365–3368. doi: 10.1021/ma0600411
- Hannah, M. (1951). Contact stress and deformation in a thin elastic layer. *Quart. J. Mech. Appl. Mathematics* 4, 94–105. doi: 10.1093/qjmam/4.1.94
- Joe, J., Thouless, M. D., and Barber, J. R. (2018). Effect of roughness on the adhesive tractions between contacting bodies. *J. Mech. Phys. Solids* 118, 365–373. doi: 10.1016/j.jmps.2018.06.005
- Jones, J. E. (1924). On the determination of molecular fields.—II. from the equation of state of a gas. *Proc. R. Soc. London Series A* 106, 463–477. doi: 10.1098/rspa.1924.0082
- Maugis, D. (2013). *Contact, Adhesion and Rupture of Elastic Solids*. New York, NY: Springer. doi: 10.1007/978-3-662-04125-3
- Mönch, W., and Herminghaus, S. (2001). Elastic instability of rubber films between solid bodies. *EPL* 53:525. doi: 10.1209/epl/i2001-00184-7
- Mukherjee, R., Pangule, R. C., Sharma, A., and Banerjee, I. (2007). Contact instability of thin elastic films on patterned substrates. *J. Chem. Phys.* 127:064703. doi: 10.1063/1.2752499
- Persson, B. N. J. (2001). Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115, 3840–3861. doi: 10.1063/1.1388626
- Sahoo, S., Bhandaru, N., and Mukherjee, R. (2019). Reversible morphological switching and deformation hysteresis in electric field mediated instability of thin elastic films. *Soft Matter* 15, 3828–3834. doi: 10.1039/C8SM02622J
- Sarkar, J., and Sharma, A. (2010). A unified theory of instabilities in viscoelastic thin films: from wetting to confined films, from viscous to elastic films, and from short to long waves. *Langmuir* 26, 8464–8473. doi: 10.1021/la9049007
- Sarkar, J., Sharma, A., and Shenoy, V. B. (2008). Electric-field induced instabilities and morphological phase transitions in soft elastic films. *Phys. Rev. E* 77:031604. doi: 10.1103/PhysRevE.77.031604
- Sarkar, J., Shenoy, V., and Sharma, A. (2004). Patterns, forces, and metastable pathways in debonding of elastic films. *Phys. Rev. Lett.* 93:018302. doi: 10.1103/PhysRevLett.93.018302
- Shenoy, V., and Sharma, A. (2001). Pattern formation in a thin solid film with interactions. *Phys. Rev. Lett.* 86:119. doi: 10.1103/PhysRevLett.86.119

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Joe, Thouless and Barber. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Onset of Sliding of Elastomer Multicontacts: Failure of a Model of Independent Asperities to Match Experiments

Julien Scheibert^{1*}, Riad Sahli¹ and Michel Peyrard²

¹ Univ Lyon, Ecole Centrale de Lyon, ENISE, ENTPE, CNRS, Laboratoire de Tribologie et Dynamique des Systèmes LTDS, UMR5513, Écully, France, ² Université de Lyon, Ecole Normale Supérieure de Lyon, Laboratoire de Physique, CNRS, UMR 5672, Lyon Cedex 7, France

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of Berlin,
Germany

Reviewed by:

Lucia Nicola,
University of Padova, Italy
Carminé Putignano,
Politecnico di Bari, Italy

*Correspondence:

Julien Scheibert
julien.scheibert@ec-lyon.fr

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 13 November 2019

Accepted: 27 March 2020

Published: 17 April 2020

Citation:

Scheibert J, Sahli R and Peyrard M
(2020) Onset of Sliding of Elastomer
Multicontacts: Failure of a Model of
Independent Asperities to Match
Experiments. *Front. Mech. Eng.* 6:18.
doi: 10.3389/fmech.2020.00018

Modeling of rough frictional interfaces is often based on asperity models, in which the behavior of individual microjunctions is assumed. In the absence of local measurements at the microjunction scale, quantitative comparison of such models with experiments is usually based only on macroscopic quantities, like the total tangential load resisted by the interface. Recently however, a new experimental dataset was presented on the onset of sliding of rough elastomeric interfaces, which includes local measurements of the contact area of the individual microjunctions. Here, we use this more comprehensive dataset to test the possibility of quantitatively matching the measurements with a model of independent asperities, enriched with experimental information about the area of microjunctions and its evolution under shear. We show that, despite using parameter values and behavior laws constrained and inspired by experiments, our model does not quantitatively match the macroscopic measurements. We discuss the possible origins of this failure.

Keywords: rough contact, elastomer friction, onset of sliding, asperity model, shear-induced area reduction, stick-slip, elastic interactions

1. INTRODUCTION

The mechanical behavior of contact interfaces between rough solids is crucial to understand their tribological properties. The rough contact mechanics community has been developing models in two main directions (see Vakis et al., 2018 for a recent review). First, *asperity models* in which the contact interface is divided into well-defined microjunctions actually carrying the normal and tangential loads applied to the contacting solids (Braun and Röder, 2002; Ciavarella et al., 2006; Violano and Afferrante, 2019). Each microjunction is ascribed a set of individual properties (e.g., its height, radius of curvature or friction coefficient) necessary to apply some assumed behavior laws [e.g., any contact (Johnson, 1987) or friction law (Le Bot et al., 2019)] when submitted to an external stimulus. The macroscopic behavior of the interface is then the emerging, collective response of the population of microjunctions (Trømborg et al., 2014; Braun and Peyrard, 2018; Costagliola et al., 2018). Second, *continuum models* in which the input quantity is the full topography of the rough surfaces, and an exact solution of the unilateral contact and friction problem is sought (Pastewka and Robbins, 2014; Yastrebov et al., 2017; Ponthus et al., 2019), again under some assumptions on the interfacial behavior, concerning, e.g., elasticity, friction, and adhesion.

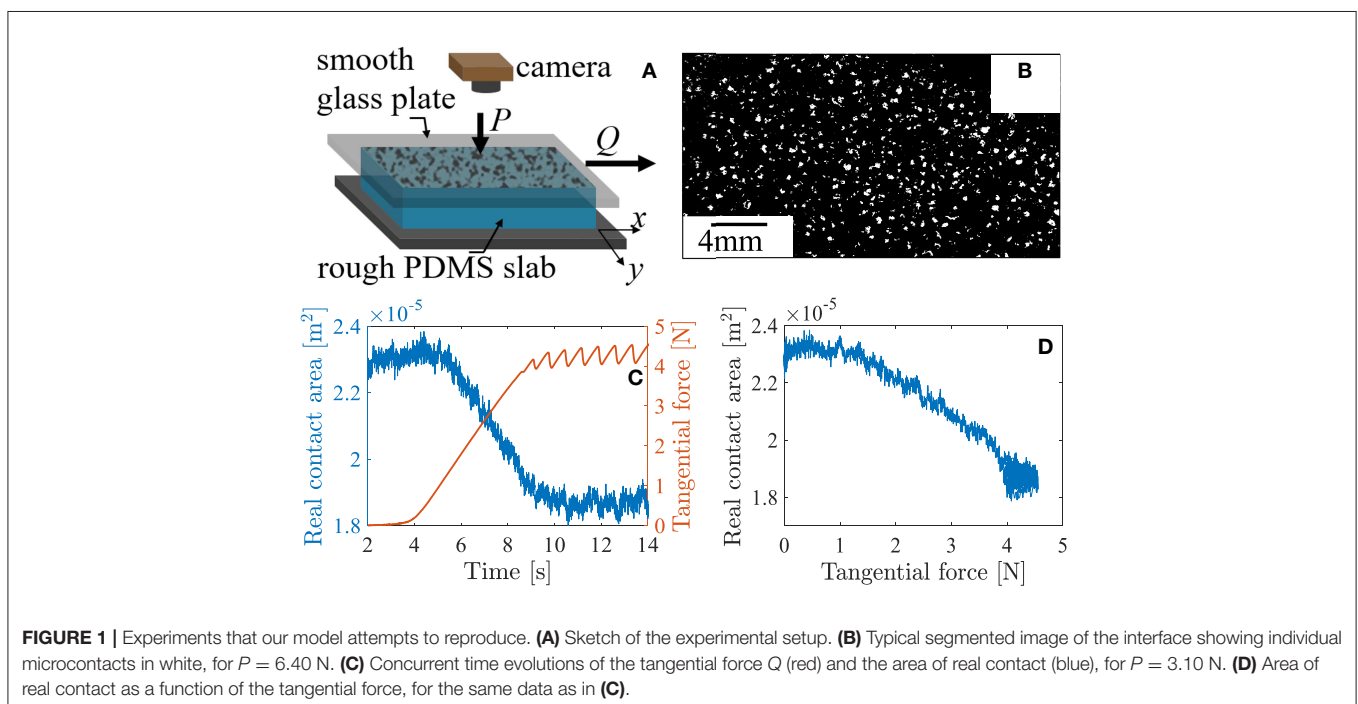
Each approach can be used to produce two types of results, either deterministic or statistical. Deterministic results are obtained for a given topography (for continuum models) or a given set of model parameters (for asperity models, including the properties of each microjunction) and are thus specific to those input data. They are relevant for quantitative comparison with a particular experiment. In contrast, statistical results are the expected results of a large number of deterministic calculations performed on statistically similar random surfaces. In asperity models, statistical results are obtained when using probability density functions (pdfs) of the microjunction properties (Greenwood and Williamson, 1966; Braun and Peyrard, 2008; Thøgersen et al., 2014). In continuum models, they are usually obtained using the power spectrum density (psd) of the topographies under study (Persson, 2001). In the following, we aim at finding a quantitative match with a specific set of measurements, so we will consider deterministic results.

Both asperity and continuum models have been widely explored in the context of rough contacts under purely normal load, with a recent study explicitly comparing the relative merits of the two approaches (Müser et al., 2017). Several studies aimed at a quantitative comparison between deterministic model results and local, microjunction level measurements (see e.g., McGhee et al., 2017; Acito et al., 2019). In contrast, to our best knowledge, such comparisons have not been reported in the case of sheared multicontacts. Here we will attempt to build an asperity model able to quantitatively match recent measurements performed on the incipient tangential loading and onset of sliding of a rough elastomer slab in contact with a smooth glass plate (Sahli et al., 2018, 2019) (**Figure 1A**). Those measurements (see a typical example in **Figures 1C,D**) are particularly interesting and constraining for models because,

in addition to the macroscopic loads on the interface, they include the evolution under shear of the individual contact areas and shapes of the many microjunctions forming the interface (**Figure 1B**).

The philosophy of this work is to start with a model of independent asperities like the earthquake model of Braun and Peyrard (2008), enrich it with the recently identified phenomenology of shear-induced area reduction, and genuinely ask the question whether such a model is sufficient to quantitatively match a particular experimental dataset. In other words, we do not aim at a definitive model of the incipient tangential loading and onset of sliding of rough elastomer contacts. Rather, we make one single step ahead compared to the models in the literature, and try to conclude whether this step (including shear-induced area reduction) is sufficient or not. Such an approach can only be fruitful if the values of the model parameters are sufficiently constrained by the experimental dataset, so that one avoids fortuitous agreement. This can be achieved (i) by limiting to the strict minimum the number of parameters that cannot be directly measured experimentally, and (ii) by performing a thorough exploration of the parameter space for those remaining, unconstrained parameters. In this work, we did our best to apply this strategy, which in our case leads to an unsatisfactory agreement. This result is nevertheless a progress in the sense that it clarifies the range of assumptions that remain to be questioned and improved in future studies.

In section 2.1, we describe the asperity model and provide the experimental constraints on the model parameters in section 2.2. Quantitative comparisons between the model and measurements are given in section 3, while in section 4 we discuss the possible reasons for the absence of good matching between the two.



2. MATERIALS AND METHODS

2.1. Model Description

We consider the frictional interface between a slider of mass M and a track. The tangential displacement of the slider, $X(t)$, is assumed to obey the following equation of motion:

$$M\ddot{X} + M\eta\dot{X} = k_L(vt - X) - F, \quad (1)$$

where M is the slider's mass, k_L is the stiffness of the loading spring through which the slider is pulled at constant velocity v , η is an effective viscous parameter accounting for dissipation, e.g., in the air or in the loading spring, the dot indicates the time derivative, and F is the resistive force due to interfacial/adhesive friction.

We assume that the interface is a multicontact made of N independent individual microjunctions, each resisting a force f_i , so that $F = \sum_{i=1}^N f_i$. Each microjunction can be in either of two states. First, a *pinned* state during which the junction acts like a (time-evolving) elastic spring of stiffness k_i , so that $f_i = k_i(X(t) - x_i)$, with $x_i = 0$ as long as junction i has never been slipping). When a threshold force f_{si} is reached, the junction enters a *slipping* state, during which $f_i = \epsilon f_{si}$. Note that $\epsilon < 1$, so that f_{si} and ϵf_{si} are the analogs, at the junction level, of a static and a dynamic friction force, respectively.

The mechanical behavior of individual junctions is inspired by experimental observations made on the same setup and materials in contact as in **Figure 1A**, but when the rough slab is replaced by a single smooth sphere (Sahli et al., 2018, 2019; Mergel et al., 2019). The resulting sphere/plane contact is assumed to be representative of an individual microjunction within a multicontact like that of **Figure 1B**. Those experiments, carried out both for large normal loads (Sahli et al., 2019) and for small (even negative) normal loads (Mergel et al., 2019), have shown that, under increasing shear, the initially circular contact shrinks anisotropically and becomes increasingly ellipse-like. As shown in Sahli et al. (2019) the shrinking minor axis of the ellipse is parallel to the shear loading direction, while the variations of the major axis (in the direction orthogonal to shear) can be neglected.

Defining $\ell_{\parallel i}$ and $\ell_{\perp i}$ the sizes of an elliptic microjunction along and orthogonal to shear, respectively, we can define its area as $A_i = \frac{\pi}{4} \ell_{\parallel i} \ell_{\perp i}$. Following Mindlin (1949), the stiffness of such an elliptic contact along the shear direction is, assuming no-slip contact conditions:

$$k_i = \frac{\frac{\pi}{2} \ell_{\perp i} E}{(1 + \nu) \left[\mathbf{K}(e) - \frac{\nu}{e^2} (\mathbf{K}(e) - \mathbf{E}(e)) \right]}, \quad (2)$$

with E and ν being the Young's modulus and Poisson's ratio of the material that constitutes the microjunctions, $e = \sqrt{1 - \frac{\ell_{\parallel i}^2}{\ell_{\perp i}^2}}$ is the excentricity of the junction, \mathbf{K} and \mathbf{E} are the elliptic integrals of the first and second type, respectively. Note that assuming that microjunctions are elliptic is the simplest increment of realism compared to a circular assumption, in order to account for the complex shapes observed for microjunctions in the experiments.

Assuming that each microjunction is initially circular, we can define the common initial value, ℓ_{0i} , of $\ell_{\perp i}$ and $\ell_{\parallel i}$ from its initial individual area A_{0i} as $\ell_{0i} = \sqrt{\frac{4A_{0i}}{\pi}}$. As already mentioned, $\ell_{\perp i}$ varies negligibly under shear, so we will consider that $\ell_{\perp i} = \ell_{0i}$ at all times. The evolution of each $\ell_{\parallel i}$ is then deduced from the shear-induced area reduction reported in Sahli et al. (2018):

$$A_i = A_{0i} - \alpha_b \frac{1}{A_{0i}^p} f_i^2, \quad (3)$$

with α_b and the exponent p two constant parameters of the model. The size of junction i along the shear direction is thus simply $\ell_{\parallel i} = \frac{4A_i}{\pi \ell_{0i}}$. Note that there is currently no rigorous contact mechanics theory for the evolution of the shear stiffness of a sheared sphere/plane contact that would incorporate anisotropic contact area reduction. Here, such a behavior is approximated at all times by the combination of Equation (2), which is valid under no-slip assumption, and of Equation (3), which was empirically found at macroscale. Doing so, we assume that Equation (3) also applies at microscale, as suggested by the existence of common values of α_b and p for both the macro- and micro-scales (Sahli et al., 2018).

For each microjunction, Equation (3) is used from the beginning of the experiment, when f_i assumed to be 0, up to when the junction first starts to slip (when $f_i = f_{si}$). At that instant, A_i takes the value $A_{si} = A_{0i} - \alpha_b \frac{1}{A_{0i}^p} f_{si}^2$. For later times, based on the observation of the typical behavior of A_i during the experiments of Sahli et al. (2018) (see **Figure 2**), we assume that A_i always remains equal to A_{si} .

In contrast, the force resisted by a microjunction can vary with time after the first onset of slipping. When the slider's velocity, $\dot{X}(t)$, gets smaller than a minimum value $\dot{X}_{\min} = c_{\min} \times v$, with c_{\min} a scalar parameter, we assume that all the slipping contacts

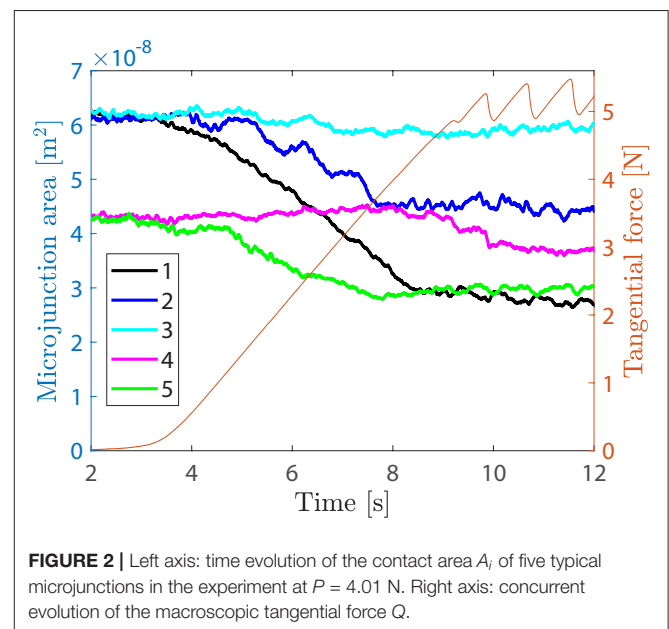


FIGURE 2 | Left axis: time evolution of the contact area A_i of five typical microjunctions in the experiment at $P = 4.01$ N. Right axis: concurrent evolution of the macroscopic tangential force Q .

will repin, with a position $x_i = X - \epsilon f_{si}/k_i$. Doing so, at repinning, there is no force discontinuity, as the repinning force $k_i(X - x_i)$ is equal to the slipping force ϵf_{si} .

Following Sahli et al. (2018), the threshold force f_{si} at which the microjunction starts to slip is assumed to be proportional to its area at the same instant, i.e., $f_{si} = \sigma A_{si}$, with σ the frictional shear strength of the contact.

The algorithm used to solve the model numerically is provided in **Appendix 1**.

2.2. Experimental Constraints

In order to quantitatively compare the model results with the multicontact experiments reported in Sahli et al. (2018, 2019), we need to feed the model with parameter values based on the measurements. In the following list, we first provide all constant parameter values that are directly accessible experimentally, with the error bars when relevant.

- $M = M_0 + M_1$, where $M_0 = 100$ g is the mass of the slider and $M_1 = 0, 55, 111, 215, 308, 552$ g an additional mass, for the six experiments performed. All masses are given at ± 1 g.
- $k_L = 9,200 \pm 200$ N/m (Sahli et al., 2018).
- $v = 0.1$ mm/s.
- $E = 1.6 \pm 0.1$ MPa (Sahli et al., 2018). This value and the associated error bar are the mean value and standard deviation over 32 estimates using 5 different spherical PDMS samples prepared in the same conditions.
- ν is assumed to be equal to 0.5, as is classically done for elastomers.
- the individual values of the initial areas of all microjunctions, A_{0i} , are extracted from the initial image (for $Q = 0$), segmented as described in Sahli et al. (2018). The fact that all microjunctions have different areas is the result of the random nature of the elastomer surface topography (see a typical Power Spectrum Density in Supplemental Material of reference Sahli et al., 2019) and of its elastic contact interaction with the rigid glass plate.
- N is also extracted from the same segmented image.
- $\sigma_{exp} = 0.23 \pm 0.02$ MPa (Sahli et al., 2018), is the experimental value of the frictional shear strength of the interface, determined from a linear fit of (A_s, Q_s) for the 6 experiments. Q_s is the macroscopic static friction (peak) force and A_s is the total area of real contact at the same instant. We will discuss below how the value of σ in the model is related to σ_{exp} .

There are three model parameters which cannot be directly measured in experiments: η , ϵ , and c_{min} .

η is introduced to enable energy dissipation in the system, thus avoiding spurious oscillations of the slider. However, the value of η should not be too large, because it would prevent the possibility of stick-slip in the model. We found that stick-slip exists up to η between 180 and 200, but for those large values, the initial stick-slip cycles are significantly different from the experiments. In practice, we found that

$$\eta = 100 \quad (4)$$

is a good compromise between oscillation reduction and a reasonable reproduction of the stick-slip sequence. The results are rather insensitive to the precise value of η , since $\eta = 50$ gives virtually identical results.

ϵ has a leading order control on the amplitude and period of the tangential force fluctuations during stick-slip. Systematic tests of the model for various values of ϵ led us to choose

$$\epsilon = 0.90. \quad (5)$$

In particular, this value is sufficiently small to enable stick-slip for all six normal loads (as observed experimentally), while reproducing reasonably well the amplitude and period of the stick-slip sequences in all cases.

Note that in the model, if there was no stick-slip, the steady-state sliding friction force would be $\sum_{i=1}^N \epsilon \sigma A_{si}$ (all microjunctions are in their slipping state). In order for this value to match the macroscopically measured value $Q_s = \sigma_{exp} A_s$, one has to impose that

$$\sigma = \frac{\sigma_{exp}}{\epsilon}, \quad (6)$$

and this is what we do in the following.

Our tests showed that the value of c_{min} has no impact on the results as far as it is sufficiently small. For instance, simulations with $c_{min} = 10^{-5}$ are essentially undistinguishable from those using 0.01. The reason is that, when $|\dot{X}(t)|$ crosses the value $c_{min} \times v$, the velocity drop is so fast that the time at which the crossing occurs is almost independent on the value of c_{min} . In our calculations, we will use

$$c_{min} = 0.01. \quad (7)$$

Extracting values for p and α_b in Equation (3) requires fitting the power law relationship between the individual area reduction parameters, α_i , and the initial areas, A_{0i} , presented as purple squares in Figure 3 of Sahli et al. (2018). Such a fitting is actually difficult due to the large dispersion of the data, as can be inferred from the large difference in total area decay of microjunctions 1–3 in **Figure 2**, although they start with almost identical areas. A fit letting both α_b and p as fitting parameters gives 95% confidence error bars as large as 600% for the optimum value for α_b , which is not a viable option. We then tried to fix the value of p and fit the data with α_b being the only fitting parameter. We found that the quality of the fit (quantified by its R^2 value) was essentially independent of p (as long as it is not too different from the value 3/2 proposed in Sahli et al., 2018), preventing any objective choice of p .

Based on those observations regarding the determination of p and α_b from experimental data, in our model studies we decided to fix p and, for each value of p , we determined the value of α_b that gives the best agreement between the area decay predicted by the model and that measured in the experiments. To do that, we fitted both the experimental and model version of the curve $A(Q)$ by a quadratic function of the form $A = A_0 - \alpha Q^2$. A_0 being the same in the model as in the experiment (because $A_0 = \sum_{i=1}^N A_{0i}$), the fitting procedure enables identification of

an α_b which provides an exact match between the two quadratic decays. Importantly, we found that, for all tested values of p close to $3/2$ (the value suggested in Sahli et al., 2018), the model results (when using the corresponding fitted α_b) were almost undistinguishable. So, in practice, we chose $p = 3/2$, for which the model studies give an optimal $\alpha_b = 0.45 \cdot 10^{-15} \text{ m}^5/\text{N}^2$ for the experiment with the smallest normal load, and $\alpha_b = 1.00 \cdot 10^{-15} \text{ m}^5/\text{N}^2$ for the experiment with the largest normal load. We then adopted the average value between both, $\alpha_b = 0.725 \cdot 10^{-15} \text{ m}^5/\text{N}^2$, as a constant to be used for all experiments.

3. RESULTS: QUANTITATIVE COMPARISON

We run the model of section 2.1 with the parameter values described in section 2.2, for the six different PDMS/glass multicontact experiments reported in Sahli et al. (2018). **Figure 3** compares, for two different normal loads, the measured time evolution of the area of real contact and tangential force to their corresponding model predictions. Note that the initial real contact area is essentially proportional to the normal load, as widely discussed in the contact mechanics modeling literature [see e.g., the reviews (Persson et al., 2005; Vakis et al., 2018)] and confirmed in the experiments discussed here (see Figure S2 of Sahli et al., 2018). To facilitate comparison between model prediction and measurement, the time origin of the experimental data has been offset by the amount necessary to superimpose the measured and predicted force curves in the central portion of their initial increase. Note that the initial non-linear increase of the measured force is due to the non-vanishing bending stiffness of the steel wire used to pull the slider, when it first bends around a pulley before a significant tension arises along the wire. The apparent difference between the measured and predicted values of the initial area of real contact is due to the above mentioned time offset: the initial predicted value exactly corresponds to the measured value from the first image, but the latter image now corresponds to a negative time and is thus not shown in the figure. The observed difference is of the order of the area measurement noise, presumably due to temporal fluctuations in the illumination and noise in the camera's sensor.

Figure 4 then shows, for all normal loads, the evolution of the area of real contact as a function of the tangential force, for both the measured and predicted data. This figure is similar to Figure 2A in Sahli et al. (2018), but shows all measurements points rather than just 1 of 130. Note that stick-slip is responsible for the accumulation of nearly horizontal cycles close to the minimum area/maximum force point of each curve. Also note that the model forces can transiently exceed the value $\sigma_{exp}A$, but always remain smaller than $\sigma A = \frac{\sigma_{exp}}{\epsilon}A$, as expected.

4. DISCUSSION

Although other combinations of model ingredients may have been proposed, we believe that our model incorporates all of the currently available knowledge on the system that we tried to reproduce. As such, it can be seen as the most comprehensive independent asperities model of shear multicontacts so far, to be

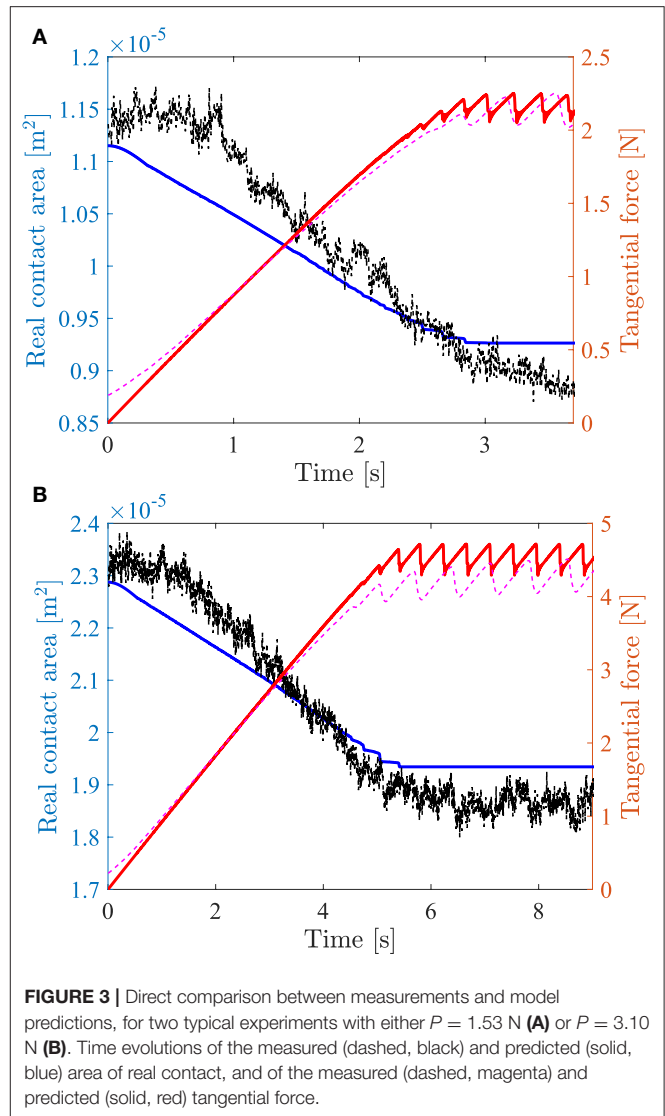
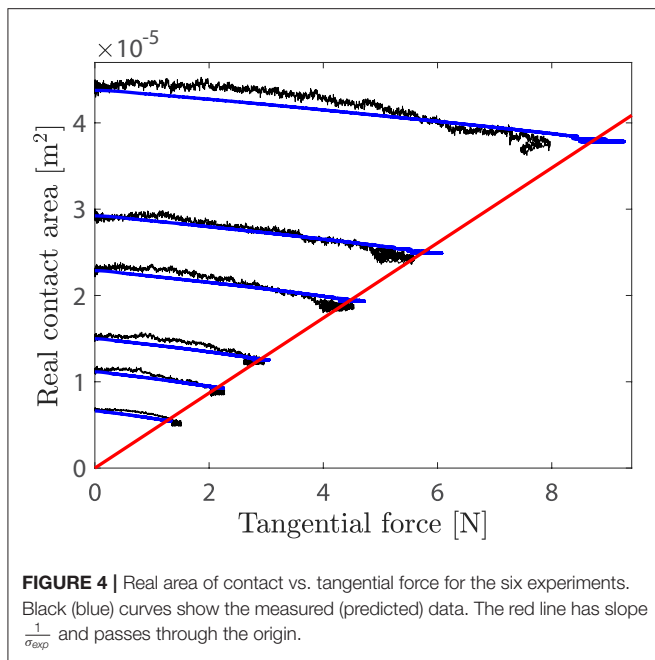


FIGURE 3 | Direct comparison between measurements and model predictions, for two typical experiments with either $P = 1.53 \text{ N}$ (A) or $P = 3.10 \text{ N}$ (B). Time evolutions of the measured (dashed, black) and predicted (solid, blue) area of real contact, and of the measured (dashed, magenta) and predicted (solid, red) tangential force.

used for deterministic comparison with the experiments of Sahli et al. (2018, 2019).

Most of the model parameters (M , k_L , v , E , v , A_{0j} , N , σ_{exp}) take their value directly from the measurements. Three adjustable parameters have been systematically varied to choose the most relevant value: c_{min} has no effect on the results, while η and ϵ have been adjusted to reproduce at best the stick-slip regime. Ideally, p and α_b should not be adjustable, but the dispersion in the experimental estimates of α_i is such that their values were not sufficiently constrained. In practice, the value of p was chosen equal to the one suggested from experiments incorporating not only microjunctions within multicontacts, but also millimetric smooth sphere/plane individual contacts (Sahli et al., 2018). The value of α_b was then adjusted to best match the overall decay of real contact area during the incipient loading of the interface.

With those values, the time evolution of the tangential load Q is quite well-reproduced (see **Figure 3**). In particular, the slope of



the incipient loading is correct, which suggests that the stiffness used for the individual microjunctions is also correct. In contrast, the time evolution of the real contact area is not satisfactory. Of course, the total *amplitude* of the real area decay, from the initial contact to macroscopic sliding, is correct, because we start from the measured initial value ($\sum_{i=1}^N A_{i0}$), and we adjusted α_b to get the correct final value. So we argue that the quality of the comparison between the model and experimental results can only be assessed through the *shape* of the real area decay. And as can be seen from **Figure 4**, while the shape of the experimental curves $A(Q)$ is essentially quadratic, that of the model curves is much more linear [except from the very beginning, when all microjunctions are pinned and thus decay quadratically according to Equation (3)]. We emphasize that this quasi-linear shape is a very robust feature of our model, because we found that the predictions are essentially unaffected by changes in the model assumptions (elliptic vs. circular microjunctions, Equation (3) applied at all times or only before the first depinning event) and in the parameter values (for values of η and ϵ enabling stick-slip or not).

The shape of the curve $A(Q)$ results from a sum of a large number (N) of complex individual behaviors (non-linear area decay while pinned, constant area while slipping) with distributed parameters (initial area, stiffness, threshold), and is therefore unlikely amenable to a simple explanation. We can however mention an instructive particular case where all microjunctions would have the same initial area. In those (unrealistic) conditions, all microjunctions would behave identically when submitted to a common displacement X and thus depin at the same instant. The total area decay would be the sum of N identical quadratic decays, and thus be itself quadratic with the total shear load, until macroscopic sliding. With those specific (but wrong) initial conditions, we would

recover a macroscopic area curve with the correct quadratic shape and a simple adjustment of the value of α_b would allow us to provide a good matching with the measurements. This example illustrates the major influence of the distribution of initial areas on the final shape of $A(Q)$. It also clearly shows that the fact that we did not succeed in reproducing a quadratic area decay is not a generic problem of our model, but partly relates to the initial conditions (through the A_{0i}) imposed by the experimental dataset.

Could there be other reasons for the failure of our model to reproduce the evolution of the real contact area? The main model ingredient responsible for this evolution is Equation (3). The first possibility is that, despite the evidence brought in Sahli et al. (2018, 2019), the anisotropic area reduction under shear would not follow a single behavior law at all junction scales, from millimeter- to micrometer-sized junctions. This possibility is indeed suggested by a recent adhesion-based model of sheared sphere/plane junctions (Papangelo et al., 2019), where the authors find that the exponent p varies systematically, for a given sphere, with the normal load applied, and hence the initial area. Here we did not try to apply the model of Papangelo et al. (2019), because it would require the knowledge of the characteristic radius of curvature of, and normal load on, each individual microjunction. In contrast, experimental measurements only provide a combination of both quantities, through the area of the microjunction.

We now argue that the solution to the failure of our model will presumably be much more complex than a mere improvement of the form of Equation (3). The problem may very well be that the predicted individual force f_i is significantly different from the one that really applies on the microjunction. This is substantiated by **Figure 2** which shows the time evolution of the contact area of various microjunctions. Two of them (4 and 5) were selected to show that the time window over which the area decay occurs can be very different: microjunction 4 has not started to shrink yet when the decay of microjunction 5 is already complete. This observation suggests that the individual tangential forces f_4 and f_5 evolve very differently during the experiment, although they have very similar initial areas and are submitted to the same tangential displacement by the glass substrate. We speculate that such a difference may be the result of elastic interactions between microjunctions, with junctions in a crowded neighborhood evolving differently from those far from neighboring junctions¹.

Those interactions are ignored in our model of independent microjunctions. We thus believe that, in order for an asperity model to have a chance to quantitatively match experiments like those considered here, tangential elastic interactions must be accounted for to describe the shear behavior of individual microjunctions. Such improved models may incorporate those tangential interactions in ways similar to models already developed for the normal interactions during normal loading of rough surfaces (see e.g., Ciavarella et al., 2006; Afferrante et al.,

¹The slight initial increase in area of junction 4 in **Figure 2** may be due not only to elastic interactions but also to a slight aging due to viscous creep.

2012) or for the friction of multicontacts (see e.g., Braun and Scheibert, 2014; Trømborg et al., 2014; Braun and Peyrard, 2018).

5. CONCLUSION

We developed an independent asperity model for the incipient shear loading and onset of sliding of dry multicontact interfaces between a rough elastic solid and a smooth rigid surface. We used it to attempt the first deterministic comparison with experiments which, in addition to the macroscopic loads and displacements, also considers the individual areas of the many microjunctions forming the interface.

The main outcome is that, although we did our best to incorporate experimentally-based behavior laws, parameter values and initial conditions into the model, it fails to quantitatively reproduce the measurements of Sahli et al. (2018, 2019). Based on observations at the microjunction scale, we suggest that an interesting starting point for future attempts to improve the quantitative deterministic comparison between asperity models and experiments, may be to incorporate a description of the tangential elastic interactions between microjunctions.

Nevertheless, we anticipate that asperity-based friction models, although accounting for tangential elastic interactions, may suffer from the same limitations as their counterparts for purely normal contact (in particular the difficulty to define asperities when a continuum of length scales is involved in the topography, see e.g., Müser et al., 2017; Vakis et al., 2018), and may still be unsuccessful to quantitatively match friction experiments. We thus urge for the concurrent development of

continuum models suitable to reproduce friction experiments like those of Sahli et al. (2018, 2019).

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

JS and MP designed and analyzed the results of the model. MP developed the numerical implementation of the model, and produced the model data shown in **Figures 3, 4**. RS performed the experiments and produced the experimental data shown in **Figures 1–4**. JS wrote the manuscript. All authors commented on the manuscript.

FUNDING

This work was supported by LABEX MANUTECH-SISE (ANR-10-LABX-0075) of Université de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). It has been supported by CNRS-Ukraine PICS Grant No. 7422.

ACKNOWLEDGMENTS

We thank Oleg M. Braun, who initiated the work, but passed away before its completion.

REFERENCES

- Acito, V., Ciavarella, M., Prevost, A. M., and Chateauminois, A. (2019). Adhesive contact of model randomly rough rubber surfaces. *Tribol. Lett.* 67:54. doi: 10.1007/s11249-019-1164-9
- Afferrante, L., Carbone, G., and Demelio, G. (2012). Interacting and coalescing Hertzian asperities: a new multisasperity contact model. *Wear* 278–279, 28–33. doi: 10.1016/j.wear.2011.12.013
- Braun, O. M., and Peyrard, M. (2008). Modeling friction on a mesoscale: master equation for the earthquakelike model. *Phys. Rev. Lett.* 100:125501. doi: 10.1103/PhysRevLett.100.125501
- Braun, O. M., and Peyrard, M. (2018). Seismic quiescence in a frictional earthquake model. *Geophys. J. Int.* 213, 676–683. doi: 10.1093/gji/ggy008
- Braun, O. M., and Röder, J. (2002). Transition from stick-slip to smooth sliding: an earthquakelike model. *Phys. Rev. Lett.* 88:096102. doi: 10.1103/PhysRevLett.88.096102
- Braun, O. M., and Scheibert, J. (2014). Propagation length of self-healing slip pulses at the onset of sliding: a toy model. *Tribol. Lett.* 56, 553–562. doi: 10.1007/s11249-014-0432-y
- Carnahan, B., Luther, H. A., and Wilkes, J. O. (1969). *Applied Numerical Methods*. New-York, NY: Wiley.
- Ciavarella, M., Delfine, V., and Demelio, G. (2006). A “re-vitalized” Greenwood and Williamson model of elastic contact between fractal surfaces. *J. Mech. Phys. Solids* 54, 2569–2591. doi: 10.1016/j.jmps.2006.05.006
- Costagliola, G., Bosia, F., and Pugno, N. M. (2018). A 2-D model for friction of complex anisotropic surfaces. *J. Mech. Phys. Solids* 112, 50–65. doi: 10.1016/j.jmps.2017.11.015
- Greenwood, J. A., and Williamson, J. B. P. (1966). Contact of nominally flat surfaces. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* 295, 300–319.
- Johnson, K. L. (1987). *Contact Mechanics*. Cambridge: Cambridge University Press.
- Le Bot, A., Scheibert, J., Vasko, A. A., and Braun, O. M. (2019). Relaxation tribometry: a generic method to identify the nature of contact forces. *Tribol. Lett.* 67:53. doi: 10.1007/s11249-019-1168-5
- McGhee, A. J., Pitenis, A. A., Bennett, A. I., Harris, K. L., Schulze, K. D., Urueña, J. M., et al. (2017). Contact and deformation of randomly rough surfaces with varying root-mean-square gradient. *Tribol. Lett.* 65:157. doi: 10.1007/s11249-017-0942-5
- Mergel, J. C., Sahli, R., Scheibert, J., and Sauer, R. A. (2019). Continuum contact models for coupled adhesion and friction. *J. Adhes.* 95, 1101–1133. doi: 10.1080/00218464.2018.1479258
- Mindlin, R. D. (1949). Compliance of elastic bodies in contact. *J. Appl. Mech. ASME* 16, 259–268.
- Müser, M. H., Dapp, W. B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T. A., et al. (2017). Meeting the contact-mechanics challenge. *Tribol. Lett.* 65:118. doi: 10.1007/s11249-017-0900-2
- Papangelo, A., Scheibert, J., Sahli, R., Pallares, G., and Ciavarella, M. (2019). Shear-induced contact area anisotropy explained by a fracture mechanics model. *Phys. Rev. E* 99:053005. doi: 10.1103/PhysRevE.99.053005
- Pastewka, L., and Robbins, M. O. (2014). Contact between rough surfaces and a criterion for macroscopic adhesion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 3298–3303. doi: 10.1073/pnas.1320846111
- Persson, B. N. J. (2001). Theory of rubber friction and contact mechanics. *J. Chem. Phys.* 115, 3840–3861. doi: 10.1063/1.1388626
- Persson, B. N. J., Albohr, O., Tartaglino, U., Volokitin, A. I., and Tosatti, E. (2005). On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *J. Phys. Condens. Matter* 17, R1–R62. doi: 10.1088/0953-8984/17/1/R01

- Ponthus, N., Scheibert, J., Thøgersen, K., Malthe-Sørenssen, A., and Perret-Liaudet, J. (2019). Statistics of the separation between sliding rigid rough surfaces: simulations and extreme value theory approach. *Phys. Rev. E* 99:023004. doi: 10.1103/PhysRevE.99.023004
- Sahli, R., Pallares, G., Ducottet, C., Ben Ali, I. E., Al Akhrass, S., Guibert, M., et al. (2018). Evolution of real contact area under shear and the value of static friction of soft materials. *Proc. Natl. Acad. Sci. U.S.A.* 115, 471–476. doi: 10.1073/pnas.1706434115
- Sahli, R., Pallares, G., Papangelo, A., Ciavarella, M., Ducottet, C., Ponthus, N., et al. (2019). Shear-induced anisotropy in rough elastomer contact. *Phys. Rev. Lett.* 122:214301. doi: 10.1103/PhysRevLett.122.214301
- Thøgersen, K., Trømborg, J. K., Sveinsson, H. A., Malthe-Sørenssen, A., and Scheibert, J. (2014). History-dependent friction and slow slip from time-dependent microscopic junction laws studied in a statistical framework. *Phys. Rev. E* 89:052401. doi: 10.1103/PhysRevE.89.052401
- Trømborg, J. K., Sveinsson, H. A., Scheibert, J., Thøgersen, K., Amundsen, D. S., and Malthe-Sørenssen, A. (2014). Slow slip and the transition from fast to slow fronts in the rupture of frictional interfaces. *Proc. Natl. Acad. Sci. U.S.A.* 111, 8764–8769. doi: 10.1073/pnas.1321752111
- Vakis, A. I., Yastrebov, V. A., Scheibert, J., Nicola, L., Dini, D., Minfray, C., et al. (2018). Modeling and simulation in tribology across scales: an overview. *Tribol. Int.* 125, 169–199. doi: 10.1016/j.triboint.2018.02.005
- Violano, G., and Afferrante, L. (2019). Contact of rough surfaces: modeling adhesion in advanced multiasperity models. *Proc. Inst. Mech. Eng. J. Eng. Tribol.* 233, 1585–1593. doi: 10.1177/1350650119838669
- Yastrebov, V. A., Anciaux, G., and Molinari, J.-F. (2017). The role of the roughness spectral breadth in elastic contact of rough surfaces. *J. Mech. Phys. Solids* 107, 469–493. doi: 10.1016/j.jmps.2017.07.016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Scheibert, Sahli and Peyrard. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX 1: INTEGRATION OF THE EQUATION OF MOTION OF THE SLIDER

Integration Scheme for the Differential Equation Giving $X(t)$

Equation (1) is integrated by a second order Leapfrog algorithm.

$$\gamma(t) = \ddot{X}(t) = \frac{1}{M} \left[k_L(vt - X) - \sum_{i=1}^N f_i \right] - \eta \dot{X} \quad (\text{A1})$$

is the acceleration of the slider at the beginning of an integration step. $\gamma(t = 0)$ has to be calculated at the start of the simulation. In our case, we start with $X(t = 0) = 0$ and $\dot{X}(t = 0) = 0$ so that $\gamma(t = 0) = 0$.

Let dt be the time step. The algorithm first updates $X(t)$ according to:

$$X(t + dt) = X(t) + dt \dot{X}(t) + \frac{1}{2} dt^2 \gamma(t). \quad (\text{A2})$$

With this updated value of X we calculate the new acceleration $\gamma(t + dt)$. In the second stage of the algorithm \dot{X} is updated according to

$$\dot{X}(t + dt) = \dot{X}(t) + \frac{1}{2} dt (\gamma(t) + \gamma(t + dt)), \quad (\text{A3})$$

and then everything is ready for the next step starting at $t + dt$.

To properly select dt we calculate the elasticity constant k_i for each contact at the start of the simulation. The angular frequency of oscillation of the slider of mass M under the total force of the contacts is $\Omega_0 = \sqrt{(\sum_{i=1}^N k_i)/M}$ and the corresponding period is $T_0 = 2\pi/\Omega_0$. The calculation uses $dt = T_0/C_t$ with $C_t = 10^4$. The exact value of dt depends on the experiment, but a typical value is $dt = 1 \mu\text{s}$. This value is sufficiently small with respect to all the frequencies entering in the dynamics of the slider so that even a second order algorithm gives a very good numerical accuracy. We have however run some calculations with a 4th order Runge-Kutta method (Carnahan et al., 1969), which is significantly slower, but has errors that decay as dt^4 , to check the accuracy of our calculations.

Algorithm of the Subroutine Which Calculates γ

To compute $\gamma(t + dt)$, $X(t + dt)$ and x_i are known. The main point is to compute all the forces f_i on the junctions. The state of each junction is recorded with two flags: θ_i records its instantaneous state, $\theta_i = 1$ for a pinned junction, $\theta_i = 0$ for a slipping junction, and h_i keeps track of its history, $h_i = 1$ for a junction which has never been slipping switches to $h_i = 0$ the first time the junction starts to slip, when $f_i \geq f_{si}$.

The program scans all the junctions and performs the following steps:

- Compute f_i for each junction
 - ★ if $h_i = 1$ the area of the junction depends on f_i according to Equation (3), which determines $\ell_{\parallel i}$ and then $k_i[A_i(f_i)]$ according to Equation (2). Thus

$$f_i = k_i[A_i(f_i)] (X(t + dt) - x_i) \quad (\text{A4})$$

gives an equation for f_i . It is too complex to be solved analytically. We solve it by an iterative process, using a dichotomy method starting from the value of f_i from the previous step. Once f_i is known we update $A_i(f_i)$, $k_i(A_i)$ and $f_{si} = \sigma A_i$ for further steps.

- ★ if $h_i = 0$
 - if $\theta_i = 1$ the junction is pinned but A_i is fixed, as well as $k_i(A_i)$, and they are known from previous iterations so that $f_i = k_i(X(t) - x_i)$.
 - if $\theta_i = 0$ the junction is slipping, $f_i = \epsilon f_{si}$.
- Check for transitions in the junction state
 - ★ if $\theta_i = 1$ (pinned junction) then if $f_i \geq f_{si}$ the junction starts to slip so that θ_i switches to 0, $f_i = \epsilon f_{si}$, h_i switches to 0 if it was still equal to 1.
 - ★ if $\theta_i = 0$ (slipping junction), if $|\dot{X}(t)| < c_{min} \times v$ the junction repins, θ_i switches to 1 and we set $x_i = X - \epsilon f_{si}/k_i$ so that the junction starts in the pinned state with $f_i = \epsilon f_{si}$.

Once all junctions have been scanned and all f_i are determined, we can compute γ from Equation (A1).



Hallmarks of Life in Single Cell Contact Mechanics: Outstanding Challenges and Perspectives

Susana Moreno-Flores*

Independent Researcher, Vienna, Austria

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Alexander Filippov,
Donetsk Institute for Physics and
Engineering, Ukraine
Ken Nakano,
Yokohama National University, Japan

*Correspondence:

Susana Moreno-Flores
smf8097@gmail.com

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 14 May 2020

Accepted: 23 June 2020

Published: 31 July 2020

Citation:

Moreno-Flores S (2020) Hallmarks of
Life in Single Cell Contact Mechanics:
Outstanding Challenges and
Perspectives. *Front. Mech. Eng.* 6:58.
doi: 10.3389/fmech.2020.00058

Cell mechanics is a fascinating subject that has witnessed a boost of scientific interest in the most diverse disciplines. Evidence of its biological relevance is continuously mounting, unveiling a real-life scenario ever more intricate and complex, and ever less universal. This review revisits the topic at the same time that appraises routine approaches and the synchronous combination of experimental methodologies to tackle cell mechanics in all its dimensionality and complexity. The focus is placed on cultured mammalian cells and on highly sensitive experimental techniques that rely on contact mechanics, namely scanning-probe and traction force microscopies. The importance of time as key variable in theory and experiment is particularly highlighted, together with the need to unambiguously identify the active and physiological contributions to cell behavior, and provide suitable mechanisms that dynamically interconnect relevant events, cellular structures and organelles to thoroughly understand the mechanobiology of cells. A special consideration is given to the role of friction and the importance of cell-cell interactions in scaling up the mechanical behavior from single cells to tissues. The topic is in constant demand of crossdisciplinarity, and in that sense, this review also serves the purpose of bringing the subject nearer to the mechanical physics and engineering community.

Keywords: cell mechanics, scanning probe microscopy, traction force microscopy, mechanotransduction, mechanosensing, mechanobiology, cell viscoelasticity, combined micro-mechanical techniques

INTRODUCTION

Cells are the smallest living units/machines engineered by Nature. In a volume that may vary from 30 to 4 $10^6 \mu\text{m}^3$ they comprise the most refined machineries for self-division, molecular synthesis, assembly and metabolism, that on the whole regulate their development and secure their survival (Alberts et al., 2008). But all those activities would not be possible if the cells were not capable of interacting with their surroundings. Many of these interactions, like cellular adhesion, are mediated by mechanical cues (Schwarz and Safran, 2013), the mechanisms of which are so important and complex, and yet mostly unknown. It is for that reason that Cell Mechanobiology has become a major focus of continuous research and review, with (Kamm et al., 2017; Dufrene and Persat, 2020; Hallou and Brunet, 2020; Miller and Hu, 2020) being just a few of the most recent and multiple examples.

Indeed, cells can sense and generate forces and other mechanical stimuli as a result of the interaction with their surroundings, which usually trigger a functional response or are triggered by a cellular activity (Huang et al., 2004; Vogel and Sheetz, 2006). The response or activity can

be physiological, transformative or motional (**Figure 1**) and involve the interplay of intracellular and extracellular components at different spatio-temporal scales. The mechanisms that convert mechanical cues into biochemical responses or viceversa, define the capacity of the cell to transduce mechanical signals and, in doing so, regulate its physiological activity and maintain the chemical balance. Cell mechanotransduction is hence key in sustaining life.

Mechanics is intertwined and evolves in parallel with cellular activity in these highly dynamic systems. This fact, together with the relatively high variability in cell behavior, adds a level of complexity that is not present in inanimate materials. Notwithstanding, the mechanical tests traditionally employed in the study of material properties have also been the experimental approach of choice for cells. Those methods became routine in their mechanical characterization, even though the results produced by different groups and with different techniques may differ from one another (Nawaz et al., 2012). The reason for this lack of consensus has been subject of thorough discussion (Hoffman and Crocker, 2009), but mainly attributed to the specifics of the technique used and the experimental conditions, which on the whole prompt the manifestation of a particular behavioral aspect above others. Therefore, it has been difficult to identify a universal behavior, even for cells of the same type. A particularly illustrative example of the latter is the work of Wu et al. (2018). The result of a large-scale, multi-institutional cooperation, a total of seven different techniques were put in place to measure the mechanical properties of MCF-7 cells, only to demonstrate the discrepancy between the experimental data. In stating the variability in cell behavior, even for those of the same type, the work actually evinces the fundamental issue of characterizing cells by a unique set of mechanical parameters, on the one hand. On the other hand, it demonstrates the limited capability of single techniques and routine experimental protocols for the detection and qualitative assessment of cell mechanical behavior as a complex whole. Does this/the above mean that we might require an arsenal of techniques for a complete assessment of the cell mechanics? In a way, it does. If the purpose is to find a thorough description of general validity.

Despite the continuous effort of the last few decades to understand the mechanical behavior of cells, there are still many open questions and the scientific challenge remains high. In the context of theory, we might be facing a similar situation, if not more severe: “... cell mechanics is but the tip of the iceberg, while a more profound approach would have to consider inelastic or plastic rather than elastic deformations and kinematic hardening rather than differential stiffening as the most salient mechanical characteristics of live cells.” (Kroy, 2008). Similarly, from the experimental perspective, a study concluded that: “...the mechanical properties of a cell are not static but dynamic and responsive to environmental conditions. Therefore, mechanical models of a cell must include a dependence on time and physiological conditions” (Pelling et al., 2007). These statements are now more than 10 years ago, and yet has as much validity as if they were made yesterday.

This review does not aim to describe as to build up on the state of the art in single cell mechanics, scrutinizing common

scientific approaches and pinpointing current issues, outstanding challenges and directions worth further exploring.

STATE OF THE ART

Experimental Techniques and the Importance of Combined and Synchronous Detection

In comparison with inanimate matter, cells are a very special kind of materials; they possess a unique architecture of interconnected membrane compartments and molecular networks that are in constant evolution and physiological activity, ready to respond to intra and extracellular cues of the most diverse kind, via multiple processes that may occur sequentially or simultaneously, and at different locations in a three-dimensional space.

Methods exist that allow to exert the smallest mechanical perturbations with great precision and sensitivity, as well as quantitatively detect and/or visualize the mechanical response at different times and precise locations. Most of them make use of micro-sized or nano-sized probes, either mechanical, optical, electric or magnetic, to apply and/or quantitatively detect, even map, forces or deformations, and some kind of optical microscopy to precisely locate the application of stimuli and visualize the mechanical response. But none of the experimental, state-of-the-art techniques can, separately, apply and detect *any* mechanical perturbation, *anytime* and *anywhere*, in ways and magnitudes that can be *relevant* for cells (**Figure 2**). In this context, relevant to define cellular behavior as completely and thoroughly as possible.

It is for that reason that the behavioral complexity of cells demands a proportionate and suitable combination of measuring setups for the *in-situ* detection of relevant mechanical phenomena, in two and three dimensions, and at the *relevant* mechanical, spatial, and temporal scales.

In the context of single cells, probe-based microscopies deserve special mention (**Figure 3A**). Collectively, this family of techniques can precisely apply and register mechanical forces in the pico- and nanonewton range, deformations of the order of nano and micrometers, as well as optically track phenomena at the nano and micrometer scales, ranging from the millisecond to the hour. The atomic force microscope (AFM) can nominally apply compressive and tensile loads and detect the normal and horizontal components of forces and vertical deformations at the upper surface of the cells. These should be preferably attached to a substrate, meaning that the technique is sensitive to phenomena occurring at or at the vicinity of the apical membrane and, in any case, close enough to the cantilever-probe. The in-plane spatial resolution is hence determined, and limited, by the size of the probe as it applies tensions or compressions across the cell surface. But the major limitation is that the technique alone cannot provide a direct measure of the contact area, let alone the geometry of the contact region between cell and probe. To that end, the integration of an optical-interference technique may be required, such as total reflection interference contrast microscopy (TRICM). The combined setup has been earlier applied to study ligand-receptor interactions

CELL MECHANOBIOLOGY - communication between cells and bioassemblies via mechanical cues

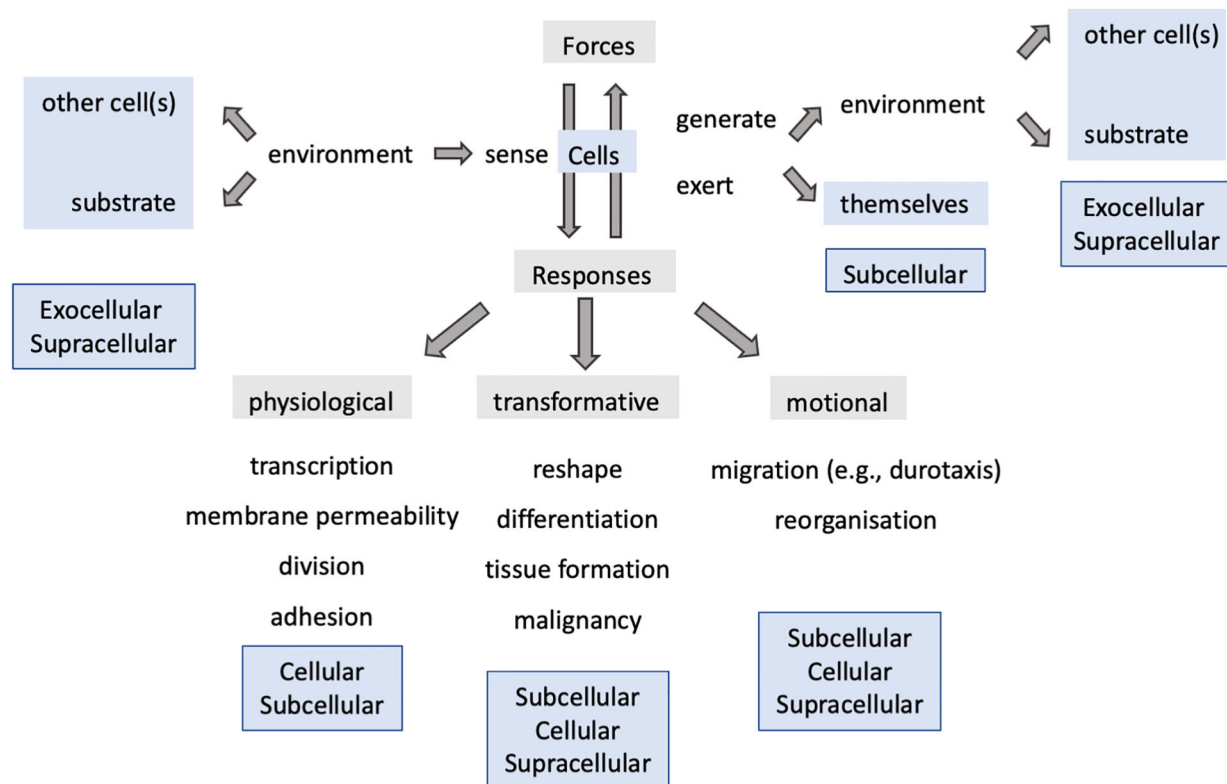
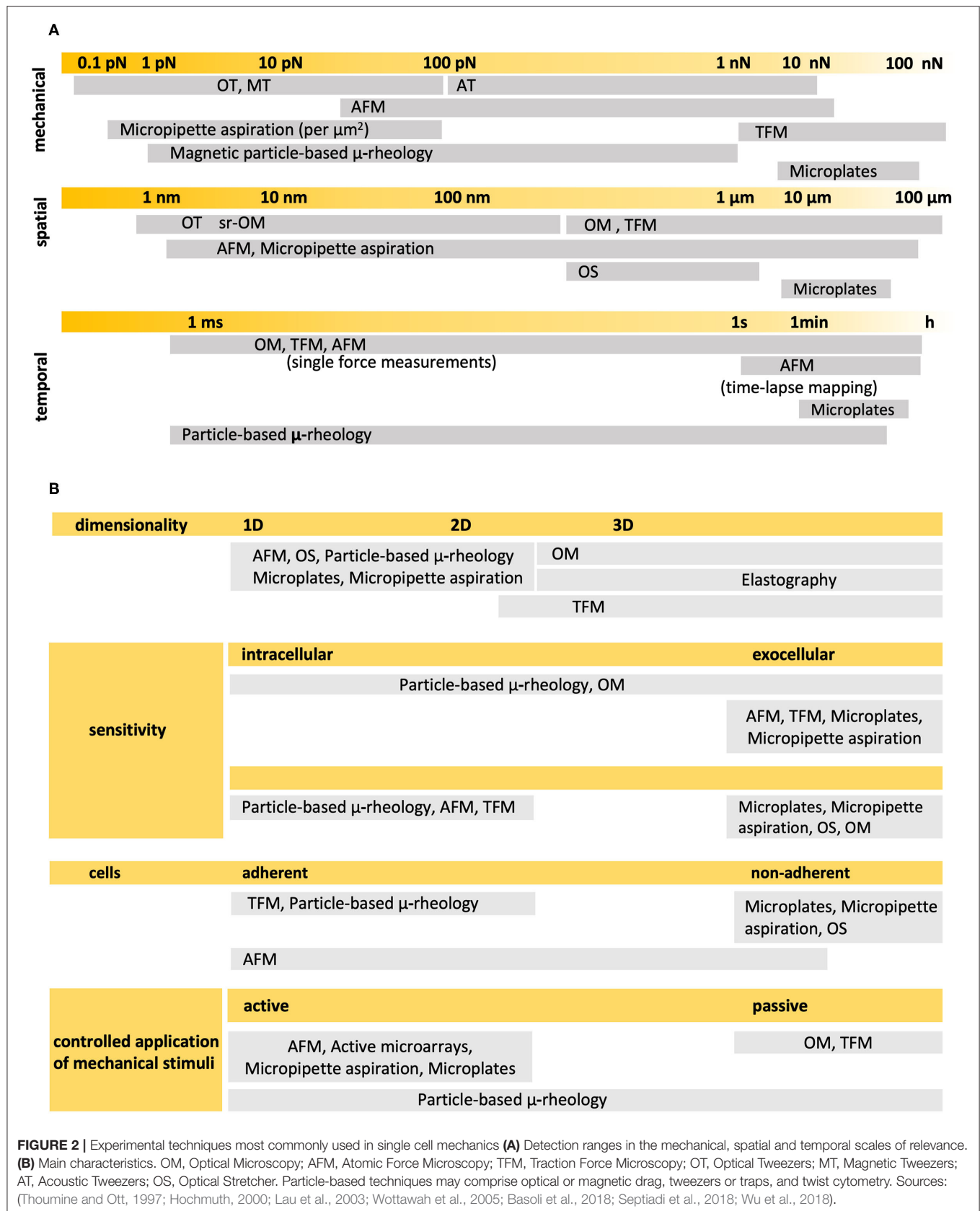


FIGURE 1 | Landscape of cellular mechanobiology. The discipline revolves about single cell behavior and encompasses phenomena involving supracellular, intracellular and exocellular bioassemblies.

at the molecular level (Stuart and Hlady, 1999), leveraging the capacity of AFM and RICM to detect forces and separation distances, respectively, with sub-nanoscale precision. Previously, RICM had been proven effective in measuring gap distances between cell basal membranes and the underlying substrate, in the calculation of contact areas (Curtis, 1964; Verschuere, 1985), and of their changes in cells under stress, either in a stand-alone configuration (Hategan et al., 2003), or in combination with the method of micropipette aspiration (Heinrich et al., 2001). But despite its high measuring potential, the tandem AFM-RICM has not been so much applied in the study of cell mechanics as in the investigation of inert materials (Dubreuil et al., 2003). Instead, contact areas are calculated based on geometrical assumptions, which restrict the probes to a limited range of sizes and geometries of high symmetry, and in turn limits the measuring potential of the technique. The cantilever itself, poses a limitation in the rates at which forces and deformations can be applied: loading rates should be fast enough to minimize the instrumental drift, and slow enough to avoid the dominance of the hydrodynamic drag, and to quantise its effect in the measurement of forces (Alcaraz et al., 2003; Mahaffy

et al., 2004). Also, the amplitudes of oscillatory stimuli are kept relatively low so that it is possible to reasonably assume that cells behave as linearly viscoelastic materials, which in turn allows for the determination of meaningful dynamic elastic moduli. As a result, the maximal forces and deformations applied are of the order of hundreds of pN or a few nanometres (2–50 nm), respectively, with frequency windows between 0.1 and 10^2 Hz, which limits the overall detection range of the technique (Alcaraz et al., 2003; Mahaffy et al., 2004; Hiratsuka et al., 2009). On the other hand, Traction Force Microscopy (TFM) (Figure 3B) heavily relies on optical microscopy to detect cell deformations, primarily, through the lateral displacement of fluorescent probes embedded in the substrate onto which the cells are adhered (Polacheck and Chen, 2016). The optical device is usually an inverted microscope that provides bottom-side views of cells at a particular (focal) plane or set of planes (should a confocal or a structured illumination microscope be used). Hence, the technique is particularly appropriate to detect and track lateral deformations and obtain in-plane components of cell-generated forces at the vicinity of the basal membrane of adherent cells. In comparison with AFM, TFM is faster in optically detecting



concurrent mechanical events at different cell locations, and the hydrodynamic drag nor the obtention of contact areas poses critical issues. However, the spatial resolution lies at its best in the (sub)micrometer range, and the data processing, including the computation of traction forces from observed in-plane displacements, is highly laborious, challenging and time-consuming. Henceforth, the technique is particularly responsive in detecting displacements at the (sub)cellular level, but not in tracking events that occur at the molecular scale and cannot generate mechanical maps on the fly. Also, the technique is not most appropriate to track out-of-plane events, i.e., normal components of forces or concurrent out-of-plane deformations. The limitations of TFM are common to other similar techniques that trace external deformations (i.e., by means of micropillars), or particle-based techniques that trace internal displacements of cell structures and organelles.

Other techniques, such as optical tweezers or the recently developed molecular tension fluorescence microscopy (MTFM) (Liu et al., 2017), are sensitive to mechanical events at the molecular scale, and in that sense have the highest spatial and temporal resolution. But also in these cases, the information they provide is restricted to one or, at the most, two spatial dimensions, and lack comparable significance at the (sub)cellular level.

As shown above, there is indeed a panoply of high-end technologies from which to define suitable combinations that allow a thorough investigation of the mechanical behavior of cells in the widest detection range. These combinations should assist in finding missing links between mechanics, structure and physiology, going beyond purely observational correlations into the causal relations and underpinning mechanisms.

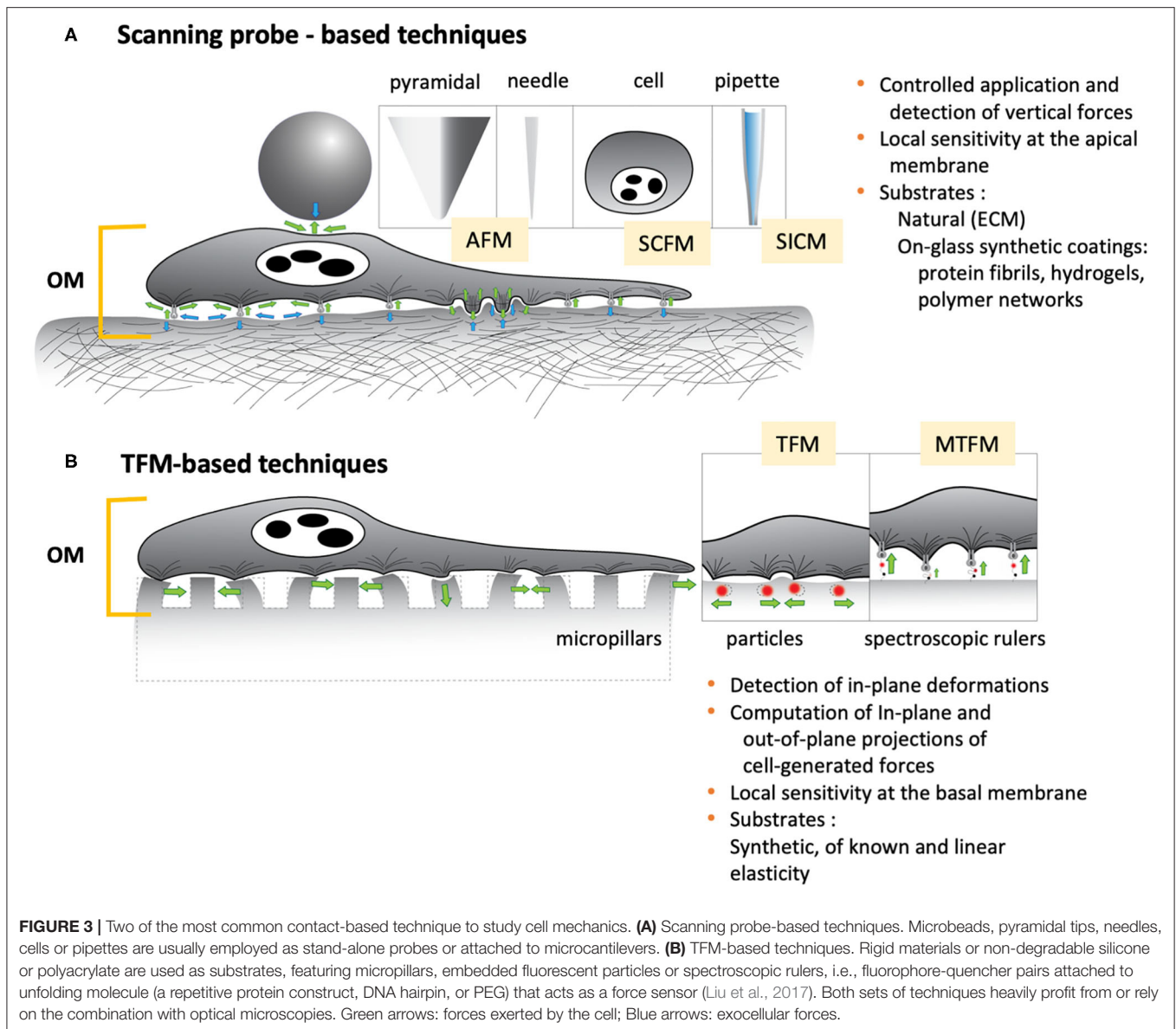
To date, some promising efforts to combine techniques have been recently reported and to some extent, provided significant evidence in that sense. For example, Schierbaum et al. (2019) recently employed a combined AFM-TFM setup to find statistical correlations between viscoelastic parameters and contractile stresses in endothelial cells (**Figure 4**). Although the observations clearly indicate that, on average, stiffer cells exert higher traction forces, at a subcellular scale this does not seem to be the case. Indeed, for a given cell, the highest traction forces were located mainly around the more compliant and fluid regions. This apparent discrepancy of results across length scales questions the significance of such correlations, especially if the specifications of each measuring technique, their synchronized use, and the underlying mechanism(s) of force transmission inside cells are not sufficient considered. In this particular case, stiffness and fluidity were obtained from creep maps at the apical region of the cells, whereas traction moments were computed from local displacements of fluorescent beads in proximity with the basal membrane. The set-up combination was not so much used synchronously as sequentially, at different unspecified times, and given the disparate spatial resolution of the techniques employed, most presumably on different subcellular regions of dissimilar size.

A significant step in achieving synchronicity of mechanical and optical data, has been made by Beicker et al. (2018). In their combined set-up, they were able to laterally visualize

the morphology of a single cell as it was mechanically tested with an AFM (**Figure 5**). The integration of vertical light sheet illumination and sideways microscopy in an AFM literally introduces a new perspective and makes it possible to directly observe the influence of the mechanical stimuli and the nature of the response in relation to the cell morphology, structure and behavior alongside the lateral region and as a function of time. Through this combination, the authors show evidence of structural and mechanical coupling between the cell membrane and the nucleus, an unprecedented finding of significant relevance that potentially shows the interplay of cell organelles in the transmission of forces within the cell.

More recently, the work of Skamrahl et al. (2019) has provided further evidence of dynamical mechano-adaptation of cells by a suitable combination of AFM and optical fluorescence microscopy (**Figure 6**). Indeed, the report shows alterations in the turnover rate and length of basal F-actin stress fibers as cells are subjected to apical loads between 0.1 and 1 nN. The findings constitute a clear evidence of apical-basal force transmission or “action at a distance,” whereby the application of local mechanical stress at the top of a cell results in the modulation of cytoskeletal activity and remodeling, i.e., an active response, at the bottom. In this respect, the work adds to the pioneering findings (Mathur et al., 2000), and follows the experimental approach of Charras and Horton (2002a). Marthur et al. had reported on a similar manifestation of apical-basal force transmission, whereby apical loads applied on single HUVEC cells induce the reorganization of focal adhesions at the immediacy of their basal membranes within minutes. Charras and Horton on the other hand, combined AFM and confocal microscopy to unveil for the first time the effect that mechanical strains have in modulating Ca-mediated mechanotransduction pathways in osteoblasts. Through this approach, they were able to identify threshold cellular strains triggering the chemical response, to precisely define the experimental timeframe for the whole process within seconds, to assess the contribution of the cytoskeletal structures, and to propose a sensible model for the mechanism by which bone cells sense mechanical strains. Although Skamrahl et al. provide characteristic time scales of stress fiber dynamics, they do not propose a molecular model or mechanism in line with their results and those reported by Mathur et al. (2000), nor do they provide a measure of the timeframe for the force transmission and transduction events. The question remains open, as to when and how, in the interval of 2 min between the application of the mechanical stress and the time when the observation begins, force propagates and triggers the physiological response.

Although a complete mechanistic picture remains elusive at the cellular scale, it may not be the case at the subcellular level. A remarkable example of the latter is the work of Bouissou et al. (2017), which provides a detailed mechanism of the way podosomes work in motile cells. Built upon the podosome architecture, and consistent with the experimental results, the authors manage to relate mechanical function to key molecules and structures. The work is exemplary in that it manages to obtain a mechanism of how a subcellular structure mechanically works, via an original experimental approach that

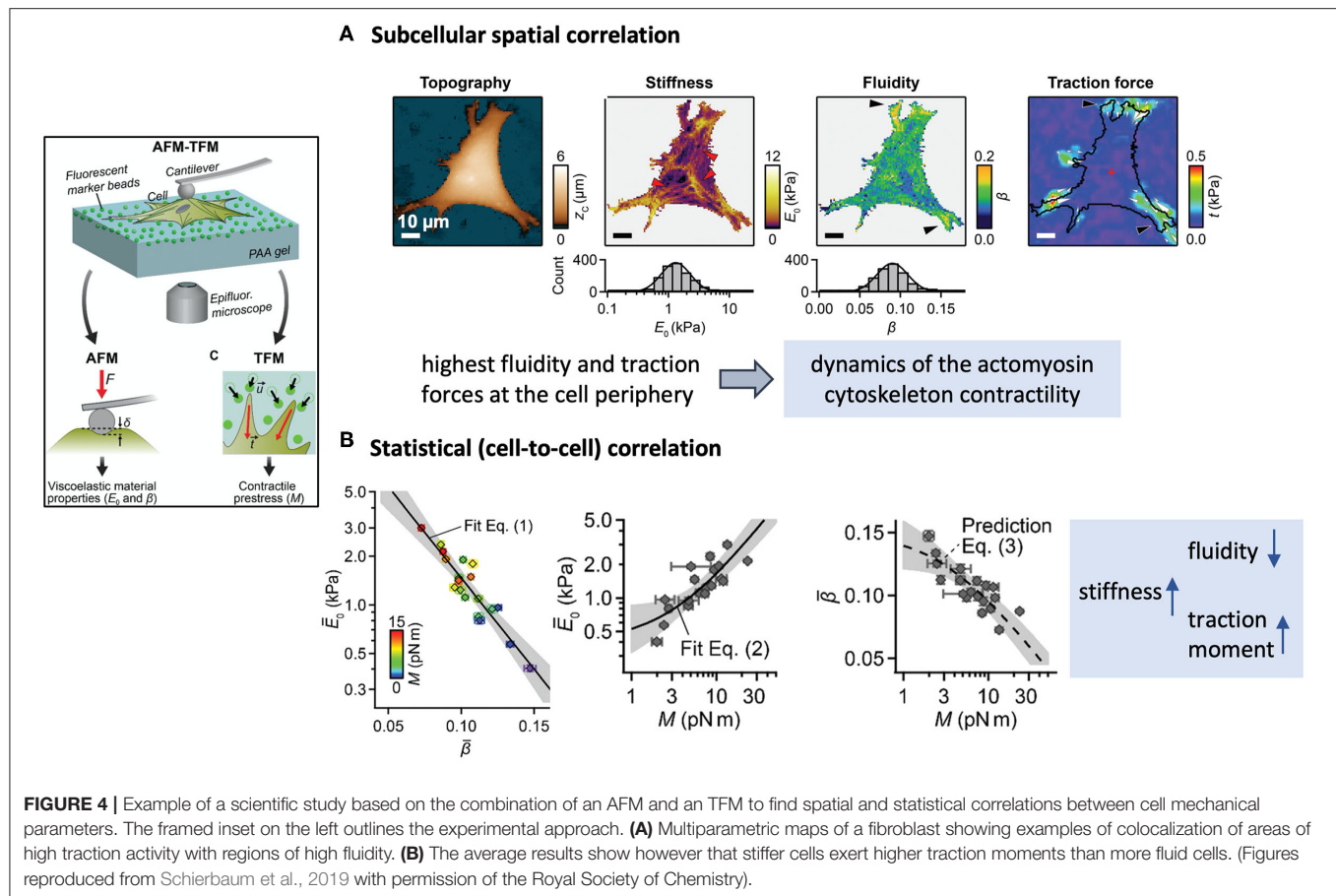


comprises the combination of AFM, super-resolution optical microscopy with numerical simulations and protein expression knock-down techniques (Figure 7, see also section Experimental Approaches and Protocols). Though the work eludes discussing the podosome as a dynamic structure that emerges and disappears with time, as the time-lapse AFM experiments show, studies of such nature and caliber very well deserve to be considered as gold standards in the study of cell mechanics.

Outside the AFM context, a recent study combining TFM and arrays of micropillar actuators particularly stands out (Shi et al., 2019). Unlike the classical micro-rheological techniques, the experimental set-up allows to obtain rheological maps of cell basal regions, along with spatial distributions of cellular fluctuations (Figure 8). Such a configuration was applied to thoroughly describe the cytoskeletal dynamics of single fibroblasts as a function of time and space. Via a sound and

comprehensive statistical analysis of the experimental data, the authors were able to detect and locate distinctive behavioral traits of the cell cortex and stress fibers, and establish meaningful spatial correlations in the observed displacements, which showed on the one hand, the highly cooperative dynamics of the actomyosin network, and on the other hand, the avalanche-like dynamics of the actin cortex. The results were not discussed in comparable depth, and although they seemed to be in agreement with the theoretical predictions of Ferrero et al. (2014) and Hwang et al. (2016), they are based on different dynamic mechanisms, neither of which could be ruled out.

The existing reports reveal the great measuring and detection potential of combining state-of-the-art techniques, and the incipient effort in that direction. But also, that their application in cell mechanical studies is far from being fully exploited. Other relevant aspects to consider are the simultaneous determination

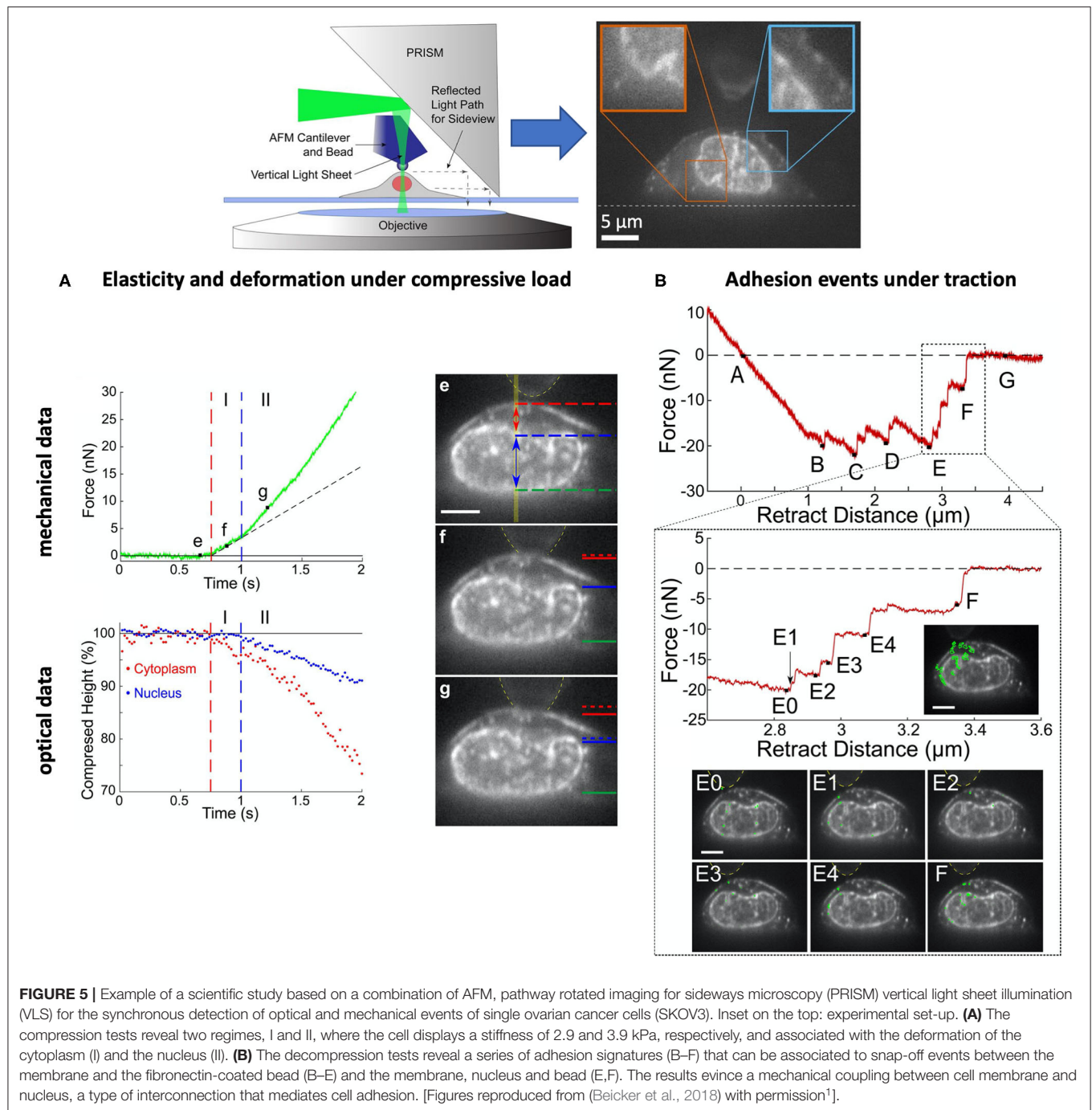


of contact areas and the influence of lateral forces in cell behavior. AFM, being one of the most versatile techniques in its diverse operational modes (Figure 9), can also detect the lateral component of forces, but these are usually ignored in non-rheological studies, most of the time under the assumption, not fully justified, that the forces are applied normally and hence the horizontal component is negligible or irrelevant. This assumption may appear valid when the experiment is performed at the apex of the cells, but not necessarily, and much less elsewhere. The reason is simple: the cell surface is not horizontally flat, its curvature profile is complex, adaptive and site-dependent. Experimentally, this magnitude is not directly measurable, and can only be assumed a-priori, or, at its best, indirectly inferred from optical cross sections of limited spatial resolution and asynchronously obtained. As a consequence, the experimental study of friction at the cellular scale remains as challenging as unexplored, with a potential biological relevance still in process of being discovered.

Methodologies and Protocols

As in the case of inert materials, mechanical assays alone do not suffice to understand the relation between behavior and structure in cells. To ascertain the relative influence of individual variables, the usual approach has consisted in manipulating the material, either chemically or physically, in a controlled manner

and in such a way that either the nature of its constituents, relative composition, or a particular environmental condition during its synthesis or treatment is selectively altered. The effect of such manipulation can be visualized optically, via the use of molecular dyes (i.e., fluorophores) that selectively label specific cellular structures or organelles. In the case of cells, mechanical assays are often accompanied by chemical or genetic interventions, whereby the cells are subjected to the action of certain drugs of known effect or to manipulation treatments of their genetic material or gene expression machinery. In both cases, the interventions are mostly focused on causing alterations of diverse nature in their cytoskeletal structure, and most recently, dynamics. The interventions are mainly used to test molecular hypotheses, usually with a substantial biological background, whereby the molecule or structure believed to play a decisive role in the process is selectively altered. The mechanical essay is then used to either confirm or refute the working hypothesis and draw conclusions in this regard. How generally valid this conclusion might be, depends on the level of statistical significance of the observation (reproducibility or repeatability of a particular behavior) across cells of the same type and across cell types. Unlike inanimate materials, making generalizations on cellular behavior or on their mechanical properties from experimental results is audacious: cell-to-cell variability is an intrinsic feature that is far more superior and depends on far



more numerous factors as batch-to-batch variability in inanimate matter. Hence, reproducible behavior must be interpreted in statistical terms, and based on a proportionate number of experiments on comparably numerous cell strains and cell types.

The vast majority of studies have been mainly focussed on unveiling the connection between mechanics and certain components of the cellular architecture and/or the cellular

membrane; in other words, the main structural elements. In this regard, and in the context of chemical interventions, the use of disruptive or stabilizing agents of actin filaments and microtubules in the experimental protocols has become commonplace ever since the early works of Rotsch and Radmacher (2000) and Charras and Horton (2002a). Drugs such as cytochalasin B, D or latrunculin A are often employed to inhibit the formation of actin fibers, whereas nocodazole, colchicine and colcemide exert a similar effect on microtubules. Paclitaxel, on the other hand, has been the

¹Granted by the terms and conditions of the Creative Commons CC BY License (<https://creativecommons.org/licenses/>).

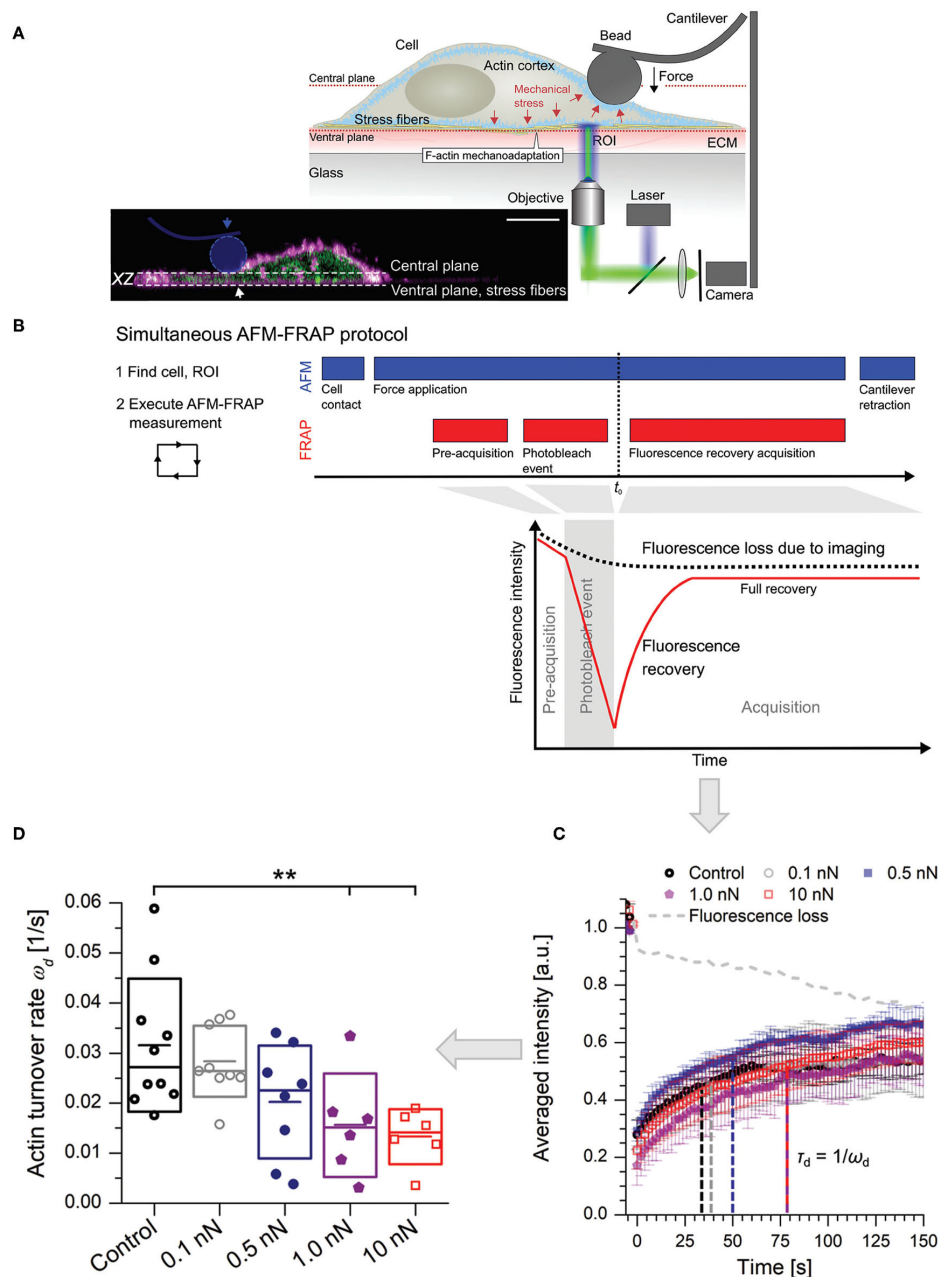


FIGURE 6 | Combined AFM-FRAP for the detection of a physiological response in the immediacy of the basal membrane of cells subjected to apical stress. **(A)** The experimental set-up. The inset features a real-life cross-section of an adherent cell. (horizontal scale bar = 10 μm , distance between planes = 4 μm). **(B)** Experimental protocol. Force is kept constant and applied at least two minutes before acquisition of fluorescence recovery data. **(C)** Typical fluorescence recovery curves as a function of applied force. The turnover rate is indicated in relation with the characteristic recovery time. **(D)** Statistical results of turnover rates (**, level of significance $p < 0.01$). (Figures reproduced from Skamrahl et al., 2019 with permission²).

drug of choice to stabilize microtubules. Less common, but important if the purpose is to disrupt cytoskeletal activity or mechanotransduction pathways, is the employment of enzyme or pathway inhibitors. Blebbistatin and ML-7 disrupt the activity of enzymes associated to the actomyosin network, whereas

gadolinium ions, verapamil or thapsigargin suppress the calcium entry pathway, each in a different way (Charras and Horton, 2002a). Interestingly, these chemical agents do not disrupt cellular structures so much as physiological processes in a selective manner. In the first case, it is the mechanism of cytoskeletal contractility as example of active cellular response or hallmark of dynamic cellular behavior. In the second, the trafficking of Ca ions as part of a mechano-transducing pathway.

²According to the terms of Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0/>.

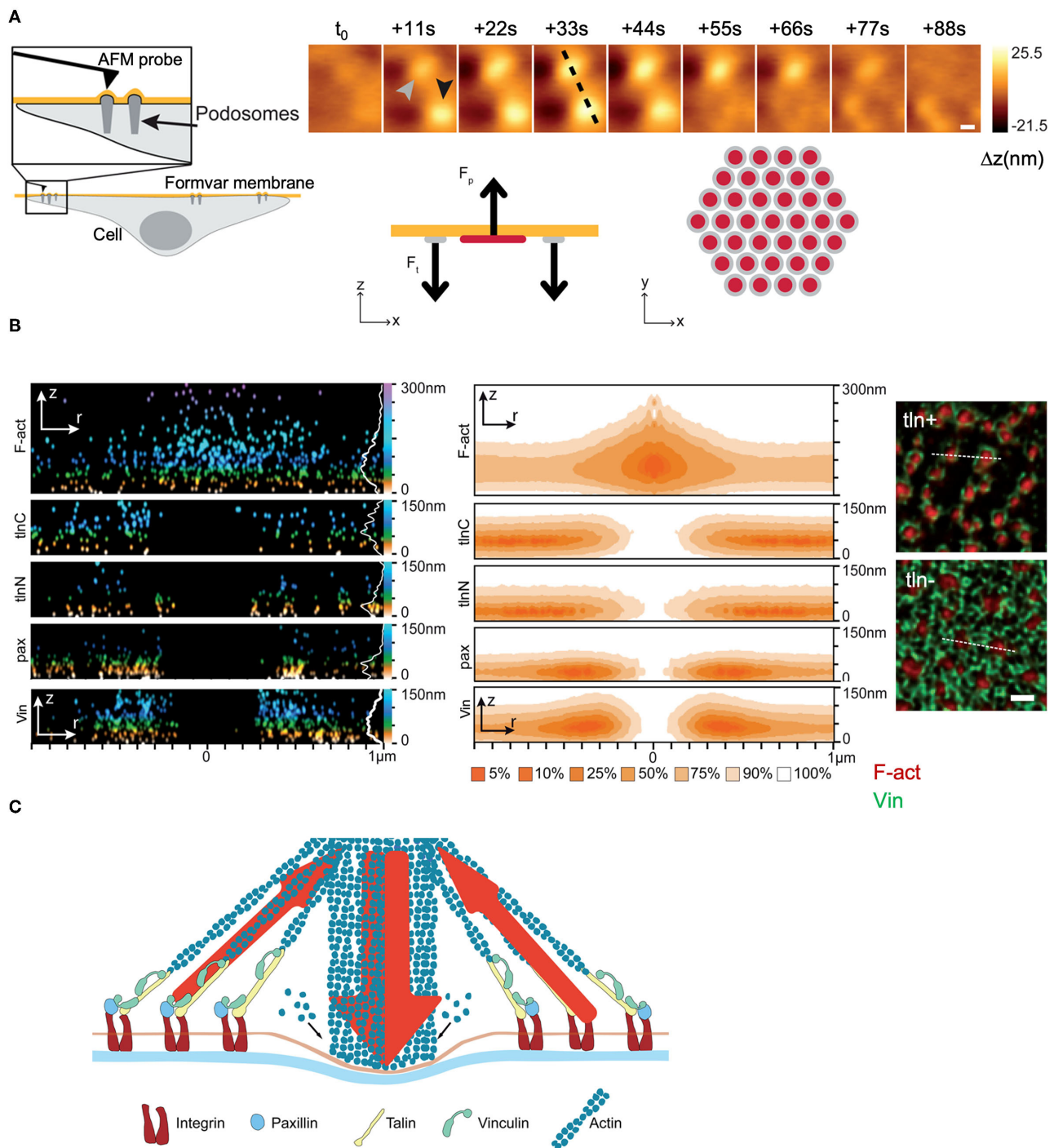
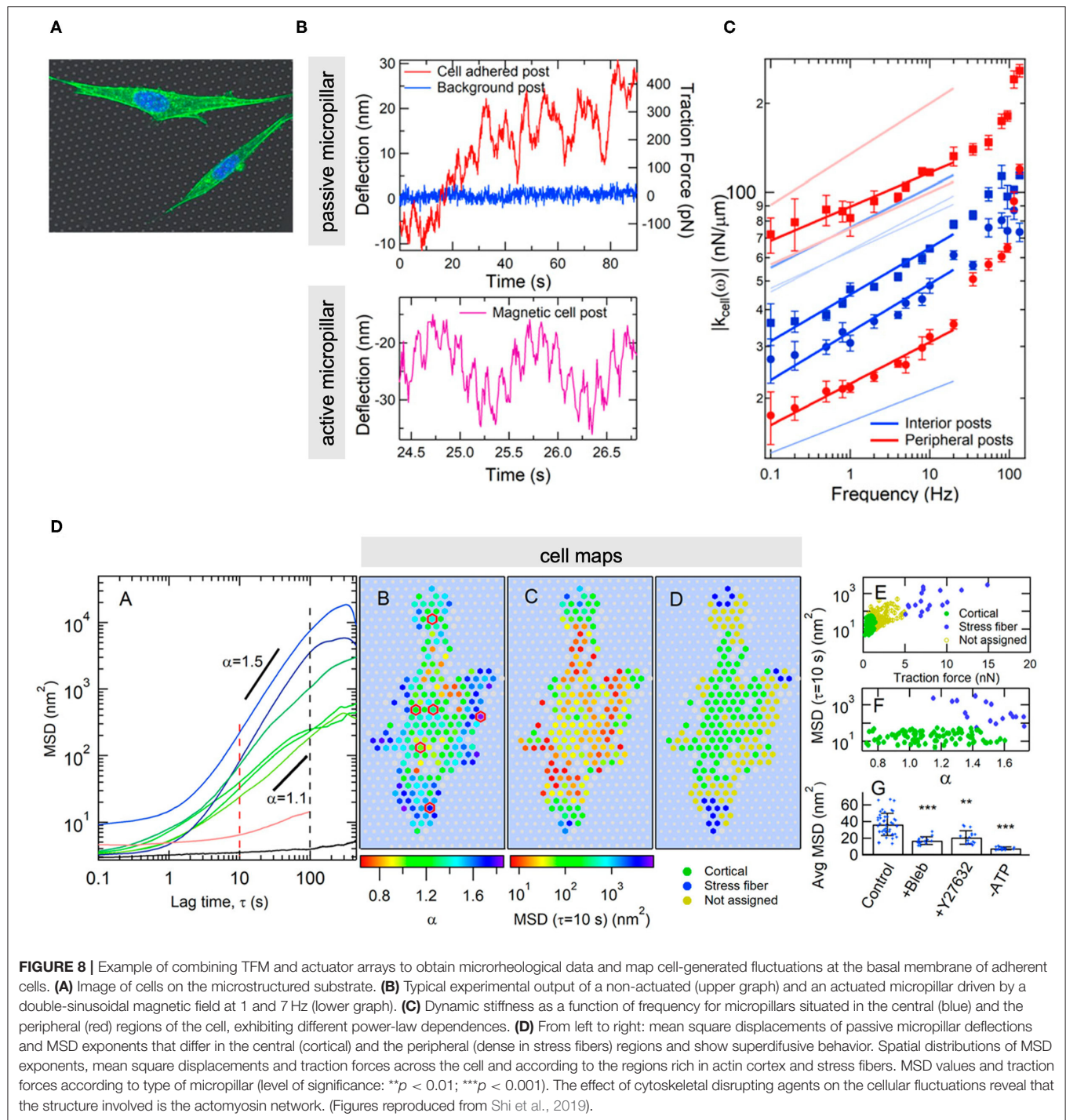


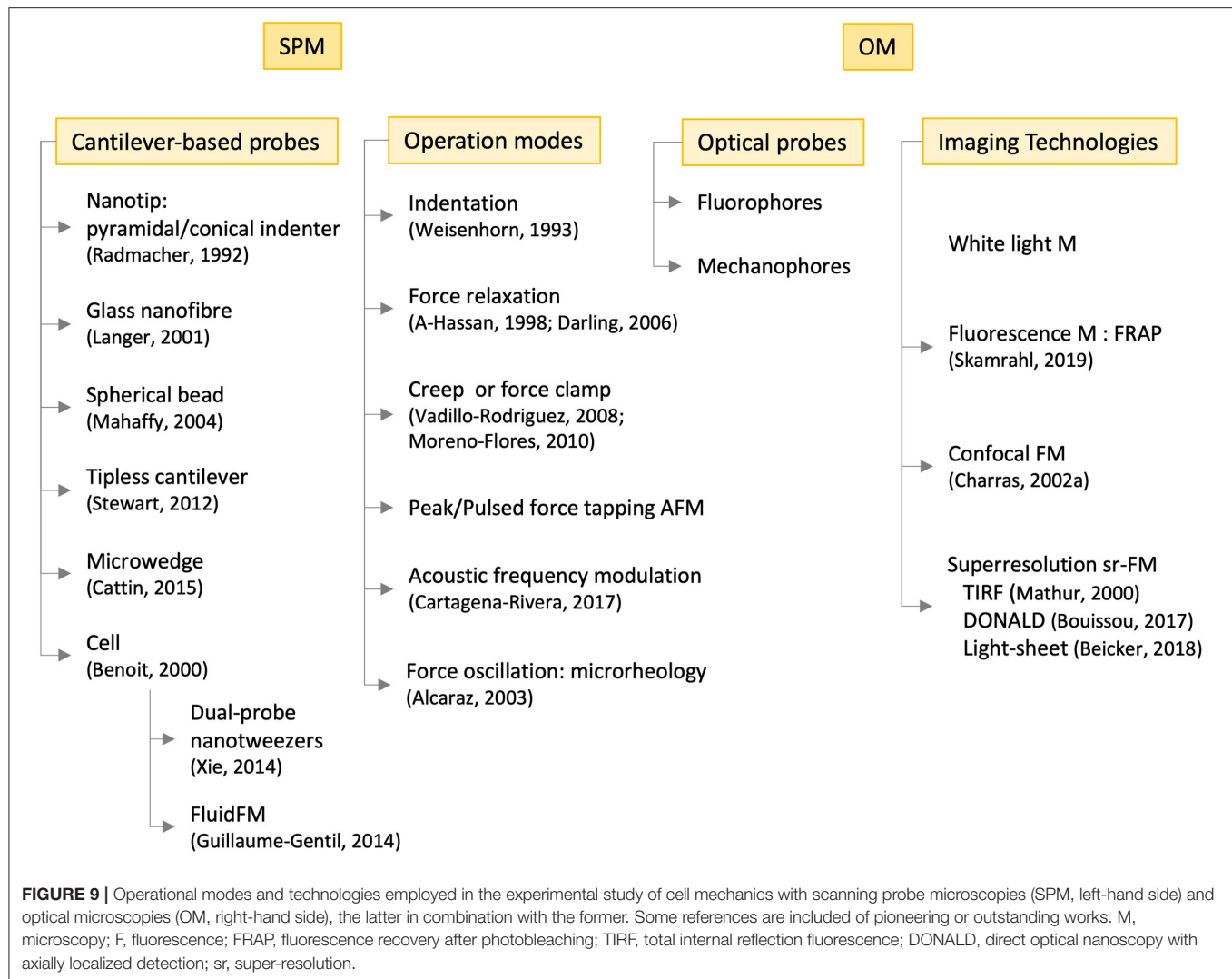
FIGURE 7 | Experimental study of the podosome structure and working mechanism. **(A)** AFM measurements: Topology of a pair of podosomes from fingerprint tracks left on a Formvar membrane. Simulation results: Protrusion-traction model of a single podosome in agreement with the experimental results. The stability of the podosome at a certain time is the result of a mechanical balance between protrusion forces (F_p) at the center of the podosome (red) and peripheral traction forces (F_i) (gray). **(B)** DONALD measurements. Spatial distribution of key structures and molecules (F-actin, talin, paxillin and vinculin) in a single podosome (left), and their corresponding probability density distributions (center). The immunofluorescence images (right) show the distribution of F-actin and vinculin in single podosomes and evince the contribution of talin in co-maintaining the annular structure. **(C)** Proposed general model of a podosome, featuring the molecular structure and the mechanical configuration of exerted forces. (Figures reprinted and adapted with permission from Bouissou et al., 2017). Copyright 2017 American Chemical Society.



In this sense, those substances and similar others with tuneable and controllable effects and well-known mechanisms of action at the single cell level, hold great scientific potential and hence are much in need. They enable to ascertain the relevance and influence of physiologically active processes, and hence to understand the mechanical behavior of cells as living entities.

Chemical intervention can be easily incorporated in the mechanical assays; drugs can be added or removed *in-situ*

without major practical complications, allowing the study of their effect on the same cells and in one single experiment. A major objection to the use of drugs is that these may cause perturbations on cells that may not be controllable, measurable, reproducible or selective enough at the cellular scale. Besides, a substantial knowledge of their effects and mechanism of action, also in the appropriate dose(s), is required prior to their utilization. Factors such as the extent of drug intake or



resistance at the single cell level are critical but likely to vary among cells, resulting in an added variability of behavior that cannot be experimentally controlled. In this regard, the physical investigation heavily relies on (and lags behind) the biological and pharmaceutical investigations pertaining drug development and associated effects *in-vitro*. As positive as this interaction may be in enhancing cross-disciplinarity, it also limits the gain in physical knowledge in that it restricts the nature and number of hypotheses to test.

Genetic manipulation may be more precise and a method of choice when it comes to modulate the presence of *key* molecules inside cells. These molecules are proteins, and usually constituents of the cellular skeleton or of molecular complexes putatively relevant in determining the mechanical characteristics of cells, such as cell-substrate (e.g., focal adhesions) and cell-cell anchors (e.g., cell-cell junctions). The manipulations, which usually occur at the DNA level, aim at the up-regulation, down-regulation or suppression (*knock-out*) of the expression of the corresponding target protein. Correspondingly, the process usually results in the increase, decrease or absence of intracellular

protein, with the corresponding alterations in the processes or structures the protein is part of. In this sense, the genetic approach introduces a higher level of control and versatility in the type and number of processes it can modify. However, the manipulations are complicated and laborious, which rules out a simple, *in-situ* implementation, require numerous materials and dedicated equipment, not to mention the expertise and know-how of geneticists and biologists. Also in this case, the biological and genetic knowledge is key in defining the appropriate course of action (i.e., choice of relevant proteins and genes to manipulate and how) for the intended outcome (alteration of cellular activity or structure).

Data Analysis and Interpretation

Despite the advancement in the measuring technologies and the level of sophistication of the experimental methodologies, the analysis and interpretation of (opto-)mechanical data have not witnessed a comparable development. Indeed, the current techniques can generate big sets of data with relative ease, but only a relatively small portion can

be easily explained or is susceptible of routine analysis and interpretation.

Also in the context of living matter, a typical mechanical experiment allows to obtain the relations between forces and deformations, also as a function of time or of frequency. Reports abound in which the behavior of many types of cells have been investigated via “indentation” tests³, which have become a routine procedures, or through stress relaxation, creep or rheological experiments at the microscale. In all these cases, the results have been analyzed in analogous ways as for inanimate materials, borrowing the traditional models of contact mechanics, in particular of the theory of linear viscoelasticity (**Figure 10**). It is hence not uncommon to find elastic or complex moduli (usually tensile or shear), as characterizing parameters for various cell types or subcellular regions. The models range from those of purely elastic and homogeneous bodies, tension-based models whereby cells are considered as pre-stressed entities characterized by a homogenous membrane tension, to those of linear viscoelastic bodies, either homogeneous or not. When the cells are treated as viscoelastic bodies, the models employed vary according to the type of mechanical assay performed; in the case of transient-type experiments (stress relaxation, creep), the spring-dashpot-based representation and its variants is usually the model of choice, with Kelvin-type or Voigt-type elements representing the contribution of cell components to the observed behavior. Whichever cell components these may be, remain a matter of data interpretation, usually in the form of hypotheses that may or not be confirmed by chemically intervention experiments of the type described above. In dynamic microrheological tests, the power-law model is predominantly applied, whereby the elastic component of the shear modulus (G') is expected to depend linearly on the logarithm of the frequency. The slope, denoted as β , is taken as a measure of the *fluidity* of the material, a value between 0 (linear elastic solid) and 1 (linear fluid). Typical β values for cells are 0.1–0.3 at low frequencies, increasing with frequency toward the value of 0.75 (Hoffman and Crocker, 2009).

The significance of giving mechanical constants as intensive magnitudes is relative in the sense that they allow to easily compare and classify cell types based on their mechanical characteristics, as for any other kind of material. But, to which extent these values are a faithful and complete representation of the cellular mechanical behavior is a matter that, surprisingly, has not been settled or discussed extensively enough. And it should, for various reasons:

Calculation of moduli and compliance. The problem of the contact area. The fundamental relations are based on intrinsic magnitudes, stress and strains that, in order to be properly and accurately calculated, require the knowledge of the contact area between the cell and the surface, or surfaces, where forces and deformations are applied. A parameter that cannot be obtained experimentally via the techniques usually employed

and previously described. Indeed, the geometry of the contact region is either unknown or poorly defined, and prone to vary among locations on the cell surface, if local probes are used in the mechanical tests. For the sake of simplification, the type of mechanical probes for the experiments are intently chosen to be highly symmetrical or geometrically simple, mainly flat planes or spheres. But still assumptions need to be made that concern the contact mechanical behavior of cells, including values for parameters such as the Poisson's ratio, with typical values ranging between 0.35 and 0.5 (Trickey et al., 2006; Liu et al., 2019). This is nothing other than paradoxical, as that is precisely what one wants to find out through the experiment. The procedure appears at times inconsistent, especially when contact areas or Young's moduli are calculated from indentation-type experiments under the assumption that cells behave elastically, while, at the same time, transient data, from force relaxation or creep compliance tests, are being analyzed with viscoelastic models (Gullekson et al., 2017; Liu et al., 2019). Another challenge related to the contact area is, particularly in the case of cells and in long experiments, its *dependence with time*. This aspect is most frequently overlooked, but its relevance can compromise the validity of the assumptions and the analysis approach as a whole.

The intentional search for simplicity. The use and abuse of Hertzian models. Many experimental studies have been done under conditions, usually small deformations (i.e., few nm) and/or forces (i.e., < 1 nN) and during short periods of time (i.e., < 1 –2 s), whereby the *instantaneous* response of cells is registered and found to be similar to that of linearly elastic, even rigid and semi continuous bodies. As much as this approach simplifies the data treatment and interpretation, a question remains regarding the actual relevance and significance of the results and the calculated parameters. The treatment has become routine to an extent that it is now possible to obtain stiffness maps of whole cells automatically (**Figure 11**), without so much as a simple calculation of slopes, or of Young's moduli, mostly according to the Hertz model (Hertz, 1882). The Hertzian approach for a spherical indenter, and to a less extent the later modifications introduced by Sneddon and Bilodeau for axisymmetric (Sneddon, 1965) and pyramidal indenters (Bilodeau, 1992), have gained widespread acceptance and use in cellular studies, although their validity is far from evident in this experimental context. Indeed, the issue of applying Hertz-based models in cell mechanics has been discussed as early as 2002 (McElfresh et al., 2002): these models work on the assumption that the material is relatively hard, linearly elastic, and represented as a continuous, isotropic and homogeneous half-space. Besides, the material should have a flat and non-adhesive surface for test and analysis. But the vast experimental evidence on cells amassed in the last 30 years clearly indicates that, in rigor, none of these assumptions can faithfully approximate the nature and behavior of cells. More realistic contact mechanical models have been reported that consider the effect of surface (Ding et al., 2018), i.e., membrane tension (references in **Figure 10**) and glycocalyx (Sokolov et al., 2013; Guz et al., 2014) and that have been applied to cells. Despite their good performance, they have not gained the widespread popularity of the hertzian models. In the same line of thought

³ Although the term *indentation* could be in this case misleading, as this may not be what it actually occur when cells are subjected to normal loads, and especially if these forces are non-local.

	Data	Model	Result
passive behaviour	Force, F – deformation, d $F = f(d)$	linear elastic model $f(\text{probe geometry})$ (sph: Radmacher, 1996; cone: Domke, 2000 Cone+sph: Mathur, 2001)	global/local tensile modulus, E
		brush model $f(\text{probe geometry})$ (sph+cone: Guz, 2014)	
		tension-based model $f(\text{probe geometry})$ (prestress $\neq 0$) (sph: Lomakina, 2004 ; cone: Sen, 2005; py: Rosenbluth, 2006)	(apical) cortical tension area compressibility modulus fluidity
	F – tether elongation, l $F = f(l)$	tension-based model (prestress $\neq 0$) (Pietuch, 2013b)	(apical) membrane tension
	F – time, t d – time, t $F, d = f(t)$	linear viscoelastic model (spring-dashpot based) (Bausch, 1998; Moreno-Flores, 2010)	E, G, μ (single component) E_i, G_i, μ_i (multicomponent)
	d – time, t	liquid-droplet model (prestress $\neq 0$) (Evans and Yeung, 1989)	cortical tension area compressibility modulus
	F – frequency, ω d – frequency, ω $F, d = f(\omega)$	linear viscoelastic model (power law) (Stamenović, 2008; Hoffman, 2009)	E', E'' G', G''

FIGURE 10 | Summary of the phenomenological descriptions most commonly used to analyse experimental data of living cells. The methods are traditional of Contact Mechanics and used to characterize at least the passive behavior of materials. Representative references are included in brackets. sph, cone, py stand for spherical, conical and pyramidal indenter, respectively.

and to the author's knowledge, classical models that consider the effect of adhesion, such as the Johnson-Kendall-Roberts (JKR) model (Johnson et al., 1971), the Derjaguin-Muller-Toporow (DMT) model (Derjaguin et al., 1975) or the generalization of Maugis and Pollock (1984), their solutions for different contact geometries and approaches for viscoelastic materials (Popov et al., 2019), as well as derived models (e.g., Hui et al., 2015; Long et al., 2016) have so far not been applied to cells.

Passive vs. active responses. Influence of structure and physiology. Cells are active and living materials, and in this sense, they constitute a special kind of smart matter.

Irrespective of the categorization in elastic or viscoelastic materials, the mechanism of the response may be active or passive, a mixture of both and non-linear. In this sense, the experiments have not been conclusive enough, and the methods for data analysis routinely employed do not allow making distinctions of such kind. Although chemical and genetic interventions can throw some light on the influence of certain structural elements and molecular roles, the issue lacks sufficient experimental evidence. Instead, it remains a subject of interpretation and hence, debatable.

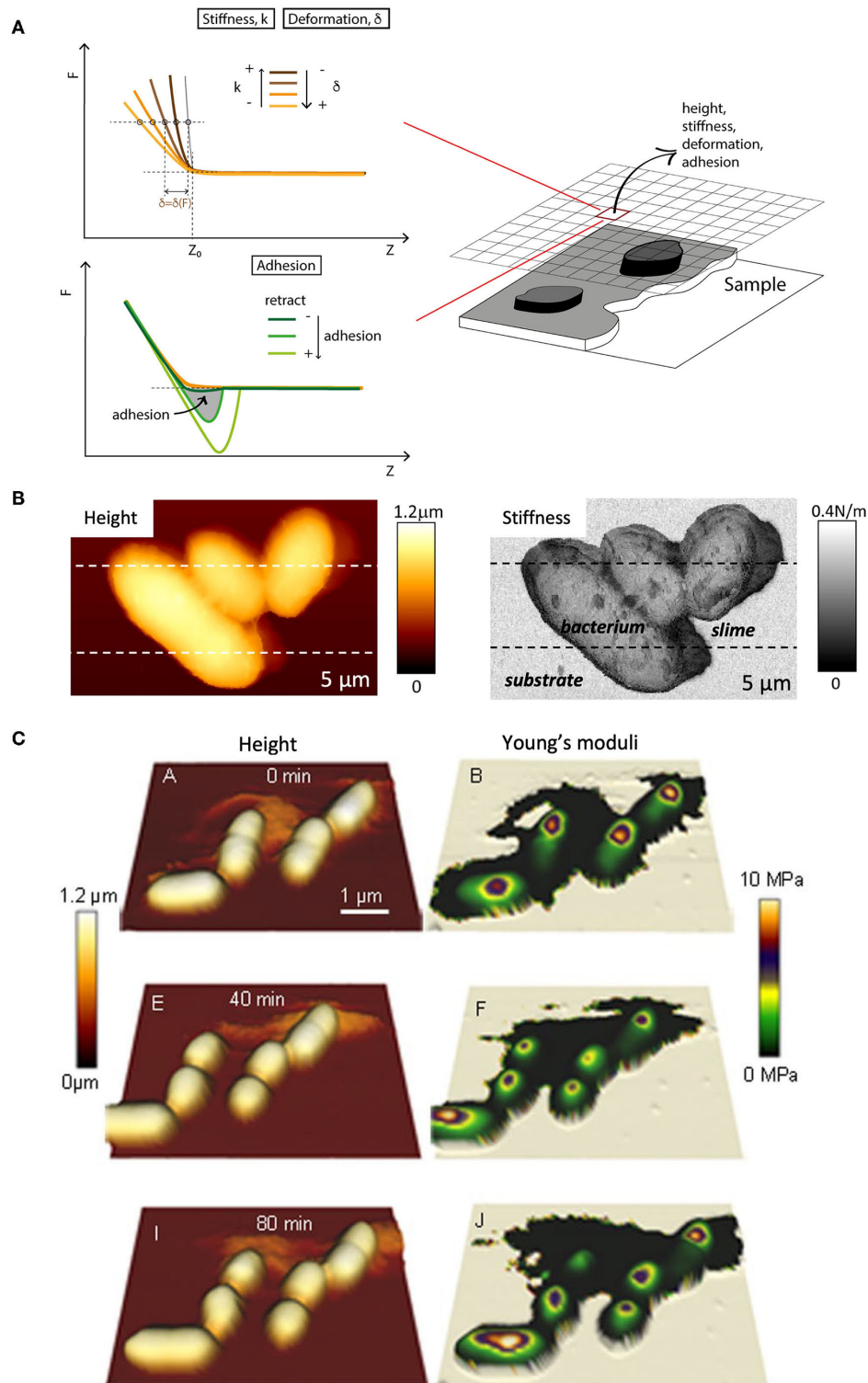


FIGURE 11 | Multiparametric imaging on the fly. **(A)** Generation of multiplexed data from AFM force-displacement curves. Stiffness are computed as slopes from linear fits of parts of the loading curve contact region (traces in shades of orange). The fitting limits are either preset or readjusted a posteriori (graphics adapted from Moreno-Flores and Toca-Herrera, 2013, fig. 7.9). **(B)** Height and stiffness maps of motile bacteria leaving a trace of, presumably, slime of distinctive softness (figure reproduced from Dhahri et al., 2013 with permission²). **(C)** Time-lapse topological and mechanical maps of single bacteria (*E. Coli*) undergoing cell division. The Young's moduli were calculated according to the Hertz model via a data analysis software of the AFM manufacturer (figure reproduced from Bhat et al., 2018 with permission²).

Cell heterogeneity. Influence of the immediate environment and time. As mentioned before, a contact-based technique such as AFM can map the viscoelastic properties of whole cells by systematically and *sequentially* probing their local response to an applied force or deformation at specific locations across the cell body. Cells are not homogeneous materials nor do they have a defined shape, and in this respect, the maps reveal certain spatial variability of the mechanical behavior. This variability may be attributed to multiple factors, such as local variations in the morphology and structure, variations in cell thickness (i.e., cell height) and cell adherence. But neither are cells inert, and their dynamics and interactions with their immediate surroundings can play a significant role in the observed behavioral diversity. In this sense, the underlying substrate, its nature and mechanics, as well as the presence of neighboring cells (cell confluency) are relevant aspects that should be considered. In particular, whether these factors are intrinsic or extrinsic to the cellular behavior are issues worth debating, as only in the latter case, and depending on the technique employed, may constitute an environmental factor that should be at least considered in the analysis (Managuli and Roy, 2017; Ding et al., 2018), rather than corrected. In the case of tissue-forming adherent cells, such as epithelial cells, for which the contact with substrates and other cells is essential in order to properly develop function, both cells and substrate are constitutive of the system and their role an integral part of the cellular response. Contrarily, suspension cells, such as blood cells, reproductive cells or certain types of bacteria, the existence of which does not depend on contact with substrates or with other cells, those factors may not be part of the system but can influence the response in one way or another. It remains to consider if this response has any biological significance, and in this sense, the cellular response should be correctly identified and interpreted.

Cells as in vitro samples. Reproducibility of cell behavior. Cells are no ordinary materials, and this also extends to the ways in which the cells samples are produced, which should also be considered. Animal cells for *in vitro* studies can be purchased as batches of cell lines, or obtained from primary cultures. The former are “modified versions” of naturally-occurring cells, which can be maintained at the laboratory, in principle indefinitely (Freshney, 2005). This is possible due to the capability of these cells to continuously grow and proliferate, if cultured with the appropriate growing medium and in suitable amounts. The cell lines, called immortal for that reason, share characteristics with and originate from tumors. Cell lines are maintained *in-vitro* via a procedure named cell passaging, in which the cells, when reaching a certain limiting confluence, are detached from the old substrate, dispersed, diluted and transferred to a new substrate, in a fresh medium. Cell passaging is a routine practice, but relative aggressive to cells, which in practice limits the number of times it can be performed without irreversibly and uncontrollably altering the integrity of cultured cells. The frequency at which cells should be passaged, as well as maximum number of cell passages vary with each cell line, and when applicable, are recommendations determined by biological practice and integrated in the experimental protocols of cell culture. Though expected to bring changes, the variability

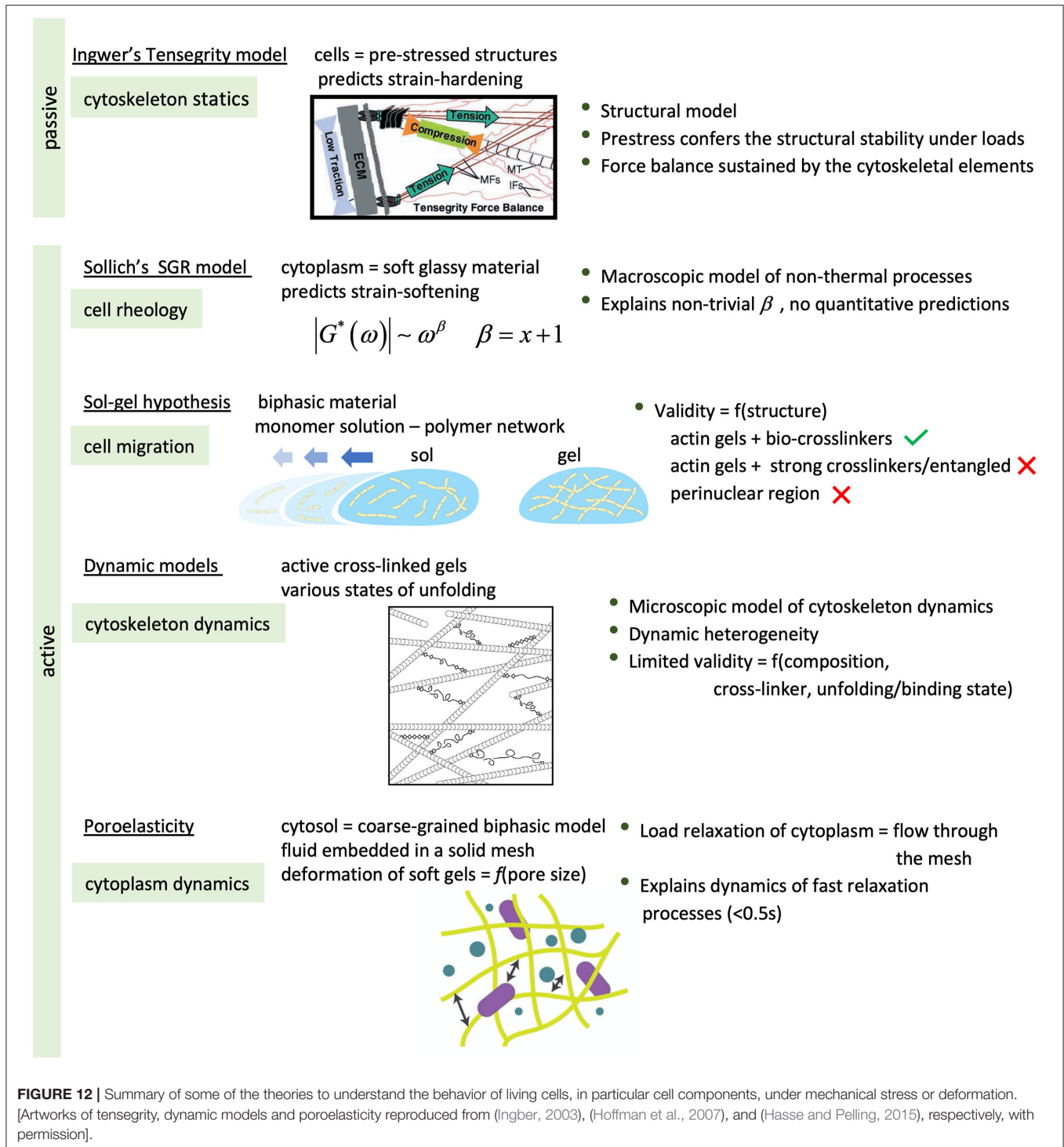
in biological behavior and cell function, also phenotype and genotype, between cells of the same type but from different passages is not well determined. This includes the physical behavior. Contrarily, primary cells are directly extracted from living tissue, tumor or not, and have limited capacity to survive. They are much more sensitive to passaging, and hence do not stand *in-vitro* maintenance. Also in this case, primary cells may display different behavior depending on the tissue, the subject they have been extracted from, as well as the procedure of preparation and maintenance of the integrity of the samples prior to experiment. Taking the above into account, it is reasonable to ask the extent to which the parameters experimentally obtained are actually influenced by the preparation method and in-culture pre-treatment.

Different experimental methodologies and models of analysis. As suggested before, different technologies have distinctive experimental sensitivities and hence likely to provide divergent perspectives of the cellular mechanical behavior. Therefore, it is reasonable to expect that the respective results may not be necessarily coincident or comparable in quantitative terms. Besides, the mechanical parameters are likely to differ if disparate methods of data analysis and interpretation are chosen for the same kind of experimental data.

Theory and Simulation

Theoretical models. Just as with notions of the theory of linear (visco)elasticity to obtain mechanical properties of cells, single cell biophysics has also borrowed pre-existent theories and models of material science to explain and understand the mechanical behavior (see the summary of **Figure 12**). Among the latter, rigid scaffolds and various types of soft matter such as polymer-based gels, soft glassy (foams and emulsions) or biphasic (sol-gel) materials have been considered as references with which to compare the behavioral features of certain cell components, especially the cytoskeleton. Indeed, the theories/models employed so far have been able to explain particular aspects of the cellular behavior, but none of them has stood out as a “theory for everything” in cell mechanics; a model that can account for the behavior of cells in all its experimental manifestations and complexity.

Ingber’s theory of Tensegrity (Ingber, 1993) explains the static mechanics of structural scaffolds. In this sense the model accounts for the passive behavior of cells, as long as these are viewed as pre-stressed structures with capacity to maintain their mechanical and structural stability under loads. In the model, the cytoskeletal network is the resilient structure that sustains the mechanical forces and preserves the mechanical equilibrium, with the actin microfilaments and microtubules the elements supporting tensions and compressions, respectively. The disruption of either of these elements disturbs the force balance between the cells and their surroundings, increasing tractions on the substrate (MT) or in the cell (MF). The theory provides a macroscopic view of the passive behavior of biopolymer fibers and cytoskeletal networks and in this sense, it can explain strain-induced stiffening exhibited by some type of cells. But it fails to explain their dynamics and hence the active



behavior to compensate the mechanical unbalance, the strain-softening behavior of other cells, also from a microscopic point of view (Hoffman and Crocker, 2009; Hasse and Pelling, 2015).

The mechano-dynamic aspects of cellular behavior have been usually identified with those of the cytoplasm. From a macroscopic point of view, the interior of cells can be considered as made of a (homogeneous) material capable of dynamically

responding to transient stimuli. In this case, cells have been found to behave as soft glassy materials when subjected to periodic shear⁴, but also relax under steady loads as though they were poroelastic (Moeendarbary et al., 2013). In each case,

⁴ An enlightening summary of experimental and theoretical works supporting the statement can be found in Hwang et al. (2016) and Shi et al. (2019).

the mechanical response could be described by a power-law behavior in frequency and time, in qualitative agreement with the predictions of Sollich, or of the poroelasticity theory (Biot, 1941), respectively. Indeed, the SGR theory of Solich (1998) predicts a power-law-like behavior, whereby the availability of non-thermal⁵ free energy, rather than temperature, is the determining factor that elicits the dynamic response. This is defined by the parameter x , with values lower than 1 characteristic of a glassy state, of frozen dynamics, and values between 1 and 2 defining the dynamic range. x is related to the power-law exponent, β , by a simple relation ($\beta = x + 1$). In this sense, the SGR theory can be useful in explaining the mechanical behavior of cells driven by ATP-dependent processes. Although able to explain some relevant experimental observations such as strain-induced softening or non-trivial rheological behavior, the quantitative predictions do not seem to square with the observations (Hoffman and Crocker, 2009). More successful in its quantitative predictions has been the poroelastic model (Moeendarbary et al., 2013). In it, the cytoplasm is coarsely-grained modeled as a bicomponent material; an elastic, porous mesh embedded in an aqueous fluid. The dynamics is determined by the movement of liquid through the mesh pores in response to an applied load, resulting in characteristic times defined by the pore size, the elasticity of the mesh, and the properties of the fluid.

Neither of these models take into consideration the microscopic dimension, and cannot provide a mechanistic view that integrates, at least, the most relevant molecular components. Consequently, they cannot fully explain the dynamic heterogeneity observed in single cells, as well as among cells and cell types. The microscopic models such as the sol-gel hypothesis, the glassy worm like chain model (GWLC) or active gels, consider the interior of cells as dynamic networks of protein fibers, interconnected or not, in a fluid medium. In each respective model, the dynamic response is viewed as a consequence of reversible molecular processes of fibrillar assembly and disassembly, of hindered fibrillar motion and, in the case of active gels, either of cross-link formation/disruption, or of the activity of motor proteins/binding proteins (Lau et al., 2003; Mizuno et al., 2007). However, these models can only explain a relatively narrow range of non-universal cellular behavior, strain-softening (Kroy, 2008) or strain-stiffening (Mizuno et al., 2007), provide a partial view of the structural dynamics and heterogeneity of the cytoskeletal network, and they are structure-specific⁶, which limits their validity. Besides, the sol-gel hypothesis cannot provide a rational explanation of non-thermally driven dynamics and fails to describe cell crawling (Hoffman and Crocker, 2009; Hasse and Pelling, 2015). Also, the GWLC model and the model of an active biopolymer gel differ notably in their fundamental assumptions despite their structural similarity, with one attributing the dynamic molecular response to equilibrium fluctuations and the other to non-equilibrium fluctuations. And in all cases, the influence

of non-cytoskeletal and non-cytoplasmic constituents in the mechanical response is either overlooked or not considered. This might result in an oversimplification, taking into account that organelles such as the nucleus accounts for 10–20% of the total cell volume (e.g., 16% in NIH/3T3 fibroblasts), exhibit distinctive mechanical characteristics of singular physiological relevance and responsiveness (Guilak et al., 2000; Lherbette et al., 2017; Stephens et al., 2017), are actively involved in mechanotransduction processes (Burrige et al., 2019) and mechanically coupled with other structural elements, as suggested by the work of Beicker et al. (2018).

Simulations. Computational approaches such as the Finite Element Method (FEM) have been regularly employed to reproduce the response (i.e., force and deformations) of materials to mechanical stimuli. Basically, these materials are modeled as/discretized as bi- or tridimensional meshed objects with a specific geometry and material properties, and a well-defined set of boundary conditions or assumptions. In the context of cell mechanics, this type of simulations has provided a “proof of model,” confirming or refuting the working hypotheses of the wet-bench (experimental) studies. Cells have been represented as axisymmetric constructs (Unnikrishnan et al., 2007), spheres or ellipsoids (Liu et al., 2019), with or without structural components (Bursa and Fuis, 2009). The former are the structural models, whereby some of the cell constituents, namely the membrane cortex, cytoplasm and nucleus, are modeled as homogeneous continua or discrete elements. The non-structural models were relatively simple and the first to be developed; these represented the cell as a homogeneous, isotropic continuum (Charras and Horton, 2002b). Both the structural and non-structural models have been combined in the so-called bendotensegrity model (Bansod et al., 2018). The model combines continuous and discrete elements to simulate the nucleus and cytoplasm on the one hand, and the cytoskeletal fibrillar network (actin filaments, microtubules and intermediate filaments) on the other hand. In all the previous cases, calculations have been made under the assumption that the mechanical characteristics of the object or of its *critical* components are known, and they focus on simulating the structural response based on an assumed architecture. Something that, in the case of cells and in view of the above and the observed behavioral variability, is still a matter of discussion and speculation, as is the identification and contribution of those critical components. Other assumptions include the nature of the contact region, the interaction between simulated components, and the Poisson’s ratio. In this sense, simulations rely heavily on experimental background data, and mechanical parameters obtained through experiment. Consequently, the simulation results may be as reliable or questionable as the experimental data, and hence subjected to a similar kind of discussion.

DISCUSSION

Single cell mechanics has undergone a considerable and significant development, that continues today. The non-stopping effort and intensive research have led to numerous and significant findings as regards the importance of mechanics in shaping cell type and behavior. Although with certain lack of consensus in

⁵Meaning chemical and ATP-dependent.

⁶The sol-gel hypothesis relies on the dynamics of the fibre assembly/disassembly to explain a particular type of active response of cells (cell migration, cell deformation), whereas the behaviour of cross-linked gels is determined by the polymer composition, the nature of the cross-linker or the active molecule.

the quantitative details and important questions still open, it is generally agreed that, from the materials point of view, cells are heterogeneously complex, viscoelastic materials, that exhibit linear and non-linear behavior depending on the environmental conditions. Indeed, the mechanical behavior of cells highly depends, among other factors:

- on the time during which the mechanical stimulus is applied, and on the speed with which this is applied
- on the integrity and dynamics of interconnected structural elements
- on cell type, topology and density (confluency).

Likewise, the cellular elasticity and adherence have proved mechanical signatures of cell type, tumor cell malignancy, and cell-cell communication, that are mainly determined by the cytoskeletal network, in particular the microfilaments, or F-actin. Also, the substrate stiffness and the mechanical state of the cell can affect its shape, pre-stress, adhesion on substrates, differentiation, and motility.

The body of work supporting the statements above is considerable. If anything, it evinces the high complexity and variability of cellular behavior, including the interaction with its surroundings, and the interconnection between external and internal, structural and non-structural factors. It has enabled to establish *statistical* correlations between cellular structure, behavior and function, but without hardly providing a mechanistic view or proof of mechanisms that can account for all the observations as yet. How could one tackle the outstanding challenges?

Time matters in cell mechanics. In principle, many of the experimental methodologies available today are capable of capturing events occurring in different locations inside and outside the cell, as well as in the course of time (**Figure 2**). Given that cells are highly dynamic systems, this kind of experiments can be paramount in finding *causal* correlations out of the time sequence of physical, chemical and physiological events that take place inside the cell, and in identifying the agents (molecules, organelles and structures) actually involved. This type of information may prove valuable in discovering the mechanisms by which cells sense, generate, transduce and transmit mechanical cues. So far, the efforts have been mainly focussed in finding statistical correlations between cell mechanical parameters and the functionality of the cell cytoskeleton, particularly the actin-actomyosin network. In most of these works, time is not considered as experimental variable, and hence the degree of interconnection between the different behaviors is missing in those studies.

Time may also be behind the question of how important the cytoskeletal microtubules are in shaping the mechanical response of cells. The pivotal work of Rotsch and Radmacher (2000) set an influential precedent in experimentally showing that microtubules are not particularly relevant in defining cell elasticity of fibroblasts, as much as the microfilament network is, and quite a number of later reports have supported this idea. The conclusions are mainly based on loading tests under conditions—low compressive loads (of the order of 10^2 pN), short times (barely tens of seconds to seconds)—where cells

may mostly reveal an elastic response of structural nature, mainly determined by the actin cortex. A discussion in terms of dynamics and the interconnectivity of the microfilament and microtubular networks are in most cases, if not in all, missing. And yet, there is sufficient evidence of the importance that the architecture and dynamics of the microtubular network has in cell division⁷, polarization and locomotion (Alberts et al., 2008), in sensing sound-induced vibrations (Schwander et al., 2010), as well as in sustaining high compressive loads (Brangwynne et al., 2006), in determining the response to mechanical strains (Charras and Horton, 2002a), or in maintaining local stiffness (Pelling et al., 2007). Likewise, there is an established importance that intermediate filaments have in shaping mechanics of cells and tissues (Broussard et al., 2020). It is thus reasonable to suppose that, as the works described in the previous section, the mechanical role and dynamics of microtubules as well as intermediate filaments manifest at larger mechanical and spatio-temporal scales that have been either scarcely or not thoroughly explored on the experimental ground, nor have they been studied in mutual interconnection, or in relation to other molecular complexes and cell organelles.

Is Cell Plasticity Actually Plasticity? The Passive, Active or a Physiological Contributions

Cells are living entities, and unlike inanimate matter, have an added level of behavioral complexity. The structural, dynamical and physiological nature of the cell mechanical response is ever-present and intertwined in yet unknown ways. Although experiments can be devised in such a way that cell responsiveness is predominantly inert and hence determined by structure, any conclusions that may be drawn from these studies may have very limited, if not questionable, validity.

It is therefore essential to include the dynamical and physiological aspects of cell behavior in the study of cell mechanics, if the purpose is to gain a solid and comprehensive knowledge. In this respect, some fundamental questions arise: to what extent is the mechanical behavior of cells a consequence of dynamics or physiology? How all these contributions interconnect to one another and deployed? Is it possible to identify them experimentally? All this boils down to prove the following hypothesis: that the physiological processes have their own mechanical fingerprint, and this is dependent on the timing and duration of each process. As far as cytoskeleton remodeling is concerned, it has been evinced that stem cells are particularly sensitive to shear oscillations of low frequency (i.e., 0.1–0.5 Hz) at the basal membrane, modulating their adherence and triggering their differentiation accordingly (Kang et al.,

⁷Though not explicitly shown (Matzke et al., 2001) hints at the relevance that microtubules may well have in shaping cell shape and stiffness prior cell division by the following brief remark: “Adding of 10 μ M nocodazole blocked the increases in both height and stiffness that occur before furrowing and cell division (data not shown)”. The comment appears even though the discussion uniquely revolves about actin microfilaments being key in the process. Nocodazole does not disrupt microfilaments so much as microtubules (see section “Experimental Approaches and Protocols”), and yet has such a distinctive effect that nonetheless is left out from the discussion.

2017); oscillations of slightly higher frequency, i.e., 2 Hz, have not such effect, which suggests a possible frequency window for the manifestation of such physiological response. For the particular case of cell mitotic division, the proof of concept of the hypothesis mentioned above, was early reported by Matzke et al. (2001), and only much later was it further explored by Stewart et al. (2011) and Cattin et al. (2015). At the level of molecules, the outstanding work of Guo et al. (2014) and particularly their original experimental approach, deserves special mention. The authors smartly leveraged the measuring potential of optical traps and tweezers and combined active and passive microrheology, in order to obtain force spectra out of spatial fluctuations of intracellular particle tracers. In this way, they were capable of identifying the active forces caused by motor proteins in the cytoskeleton and detect differences in the cytoplasm activity between normal and cancer cells of the same type. From this perspective, the work sets a remarkable precedent and evinces once more the great benefit of combining experimental capabilities and analytical resources in gaining fundamental knowledge of the active mechanics of cells. In consequence, it is to be expected that the integration of multiple techniques for synchronous detection, measurement and data analysis holds great potential in providing qualitative and quantitative evidence of interrelated processes, as well as in identifying mechanical signatures and characteristic time scales.

Cell confluency: the role of cell-cell interactions in the mechanical response. Cells that form tissues need to establish connections with other cells in order to develop properly. Therefore, it is rational to expect that the mechanical behavior of these cells be different if studied in isolation or in the presence of other neighboring cells. In this respect, the cell density can be an important parameter to consider in the study of cell mechanics (Brückner and Janshoff, 2018; Broussard et al., 2020). In the context of two-dimensional *in-vitro* cell culture models, cell confluency, or the percentage of substrate area covered by a monolayer of cells, has proven to be an influential factor in determining the mechanics of healthy epithelial cells, as compared to invasive tumor cells of the same type (Schierbaum et al., 2017), in enhancing membrane tension (Pietuch et al., 2013a,b) or in modifying the cell elasticity, although with quite different results in the latter case (Efremov et al., 2013; Schierbaum et al., 2017). In the context of 3D *in vitro* cell culture models, recent attempts have been reported to characterize the elasticity and cell-cell interfacial tensions of multicellular spheroids by cavitation rheology (Blumlein et al., 2017), as well as the overall viscoelasticity of epithelial cysts (Shen et al., 2017) and elasticity of multicellular spheres of mesenchymal stromal cells by AFM (Tietze et al., 2019). Both latter works provide evidence of characteristic relaxation dynamics and indentation mechanics that differ substantially to those of single cells. It is thus reasonable to expect “transitions” in the mechanical behavior as cell proliferate and develop into mature spheroids (Moreno-Flores and Küpcü, 2015).

These findings are in line with the underlying idea that the mechanical signatures of cells evolve as they network and eventually develop into tissues and organelles, or in case of cancer cells, as they develop tumors with a certain

invasiveness. Establishing the connection between mechanics and cell development at the fundamental level (Hallou and Brunet, 2020), can thus be critical in gaining a deeper understanding of morphogenesis (Keller, 2012), as well as tumor development and metastasis (Kumar and Weaver, 2009; Wirtz et al., 2011). In this regard, the clinical application of the cell *mechanotype* and of micromechanical techniques in cancer detection does no longer seem to be a far-fetched possibility (Nautiyal et al., 2018; Stylianou et al., 2018).

The 3D view. Cell anisotropy is a property that varies according to cell type and may evolve in the course of time, i.e., cell life cycle and physiological activity. The basal-apical polarity of epithelial cells, and the dendrite-to-axon transmission of electrical signals in neurons are two classical examples. Likewise, cells *in vivo* are subjected to, and exert, a characteristic set of mechanical cues that determine their development, function and shape in a three-dimensional space. The *in-vivo* scenario heavily contrasts with most of the experimental studies in cell mechanics, whereby the application of stimuli and the observation of the response are limited to a single direction or plane. Hence, technologies enabling 3D visualization and mapping of micromechanical interactions between cells and their environment in a dynamic manner would be much needed. These in combination with 3D scaffolds of controlled structure—pore size, fiber length and diameter—dynamic behavior and degradability, capable of active sensing and responsiveness, fluidization and remodeling (Kennedy et al., 2017; Lemma et al., 2019), may prove promising in defining the new state of the art in the methodology for cell mechanobiology.

Single cell tribology. Friction can cause or aggravate tissue damage and inflammation, induce bone, tooth wear and blister formation, and lead to commonplace pathologies such as osteoarthritis. Therefore, a considerable effort has been put in place to understand tissue resilience and degradation, as well as to investigate restorative treatments or, alternatively and whenever applicable, replacement materials for implants (Gebeshuber et al., 2008; Correa and Lietman, 2017; Pina et al., 2018). A recurrent model material in the characterization and engineering studies has been the articular cartilage, for which a wealth of reported research exists (Correa and Lietman, 2017; Pina et al., 2018). However, much is still unknown at the level of single cells, in particular about the relation between cell rheology and cell tribology and the impact that friction may have on the mechanical and physiological behavior. A significant contribution in that direction suggests a relation between cell deformability and surface friction, and that both characteristics may determine the metastatic potential of cancer cells (Byun et al., 2013). Just recently, friction-induced mechanisms for cellular inflammation and cellular death have been proposed for corneal epithelial cells and chondrocytes from multicellular experiments (Bonnievie et al., 2018; Pitenis et al., 2018), and, in view of the relevance of the findings, this type of investigations are expected to gain momentum. In this respect, active microrheological techniques able to apply and detect shear stress and strains inside and outside single cells, in combination with microfluidic approaches can make a significant difference.

CONCLUSIONS AND OUTLOOK

Despite the impressive record of research work, single cell mechanics has still a long way to go in reaching consensus and thorough understanding. There is great need in putting the reported findings in relation to one another, and from this perspective, revise the vast knowledge amassed so far. On the other hand, tackling the questions still open in the field requires integrative approaches and methodologies on a par with the still unfathomed cell complexity. The search for a complete and universal mechanistic view of cell mechanics remains to be a formidable task and an outstanding challenge that absolutely depends on extensive crossdisciplinarity to be fruitful. A roadmap toward such aim should consider the variable time as key experimental and theoretical parameter, the development of rigorous gold standards for the routine

mechanical characterization of cells, and the conception of dynamic models that integrate the biological, physical and chemical knowledge, as well as resources in computation and engineering. Synchronous or combined multi-instrumental and multidimensional methods of cell manipulation and detection emerge as key experimental approaches to ascertain and identify the plausible active and physiological contributions to the overall cell behavior, the anisotropic nature, the relevant structural and functional interconnections, the role of friction, as well as the implications to cell communication and development.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- A-Hassan, E., Heinz, W. F., Antonik, M. D., D'Costa, N. P., Nageswaran, S., Schoenenberger, C.-A., et al. (1998). Relative microelasticity mapping of living cells by atomic force microscopy. *Biophys. J.* 74, 1564–1578. doi: 10.1016/S0006-3495(98)77868-3
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). *Molecular Biology of the Cell, 5th Edn.* New York, NY: Garland Science.
- Alcaraz, J., Buscemi, L., Grabulosa, M., Trepas, X., Fabry, B., Farré, R., et al. (2003). Microrheology of human lung epithelial cells measured by atomic force microscopy. *Biophys. J.* 84, 2071–2079. doi: 10.1016/S0006-3495(03)75014-0
- Bansod, Y. D., Matsumoto, T., Nagayama, K., and Bursa, J. (2018). A finite element bendo-tensegrity model of eukaryotic cell. *J. Biomech. Eng.* 140:101001. doi: 10.1115/1.4040246
- Basoli, F., Giannitelli, S. M., Gori, M., Mozetic, P., Bonfanti, A., Trombetta, M., et al. (2018). Biomechanical characterization at the cell scale: present and prospects. *Front. Physiol.* 9:1449. doi: 10.3389/fphys.2018.01449
- Bausch, A. R., Ziemann, F., Boulbitch, A. A., Jacobson, K., and Sackmann, E. (1998). Local measurements of viscoelastic parameters of adherent cell surfaces by magnetic bead microrheometry. *Biophys. J.* 75, 2038–2049. doi: 10.1016/S0006-3495(98)77646-5
- Beicker, K., O'Brien, E. T. III, Falvo, M. R., and Superfine, R. (2018). Vertical light sheet enhanced side-view imaging for AFM cell mechanics studies. *Sci. Rep.* 8:1504. doi: 10.1038/s41598-018-19791-3
- Benoit, M., Gabriel, D., Gerisch, G., and Gaub, H. E. (2000). Discrete interactions in cell adhesion measured by single-molecule force spectroscopy. *Nat. Cell Biol.* 2, 313–317. doi: 10.1038/35014000
- Bhat, S. V., Sultana, T., Körnig, A., McGrath, S., Shahina, Z., and Dahms, T. E. S. (2018). Correlative atomic force microscopy quantitative imaging-laser scanning confocal microscopy quantifies the impact of stressors on live cells in real-time. *Sci. Reports* 8:8305. doi: 10.1038/s41598-018-26433-1
- Bilodeau, G. G. (1992). Regular pyramid punch problem. *J. Appl. Mech.* 59, 519–523. doi: 10.1115/1.2893754
- Biot, M. A. (1941). General theory of three-dimensional consolidation. *J. Appl. Phys.* 12, 155–164. doi: 10.1063/1.1712886
- Blumlein, A., Williams, N., and McManus, J. J. (2017). The mechanical properties of individual cell spheroids. *Sci. Rep.* 7:7346. doi: 10.1038/s41598-017-07813-5
- Bonnevie, E. D., Delco, M. L., Bartell, L. R., Jasty, N., Cohen, I., Fortier, L. A., et al. (2018). Microscale frictional strains determine chondrocyte fate in loaded cartilage. *J. Biomech.* 74, 72–78. doi: 10.1016/j.jbiomech.2018.04.020
- Bouissou, A., Proag, A., Bourg, N., Pingris, K., Cabriel, C., Balor, S., et al. (2017). Podosome force generation machinery: a local balance between protrusion at the core and traction at the ring. *ACS Nano* 11, 4028–4040. doi: 10.1021/acsnano.7b00622
- Brangwynne, C. P., MacKintosh, F. C., Kumar, S., Geisse, N. A., Talbot, J., Mahadevan, L., et al. (2006). Microtubules can bear enhanced compressive loads in living cells because of lateral reinforcement. *J. Cell Biol.* 173, 733–741. doi: 10.1083/jcb.200601060
- Broussard, J. A., Jaiganesh, A., Zarhoob, H., Conway, D. E., Dunn, A. R., Espinosa, H. D., et al. (2020). Scaling up single-cell mechanics to multicellular tissues – the role of the intermediate filament-desmosome network. *J. Cell Sci.* 133:jcs228031. doi: 10.1242/jcs.228031
- Brückner, B. R., and Janshoff, A. (2018). Importance of integrity of cell-cell junctions for the mechanics of confluent MDCK II cells. *Sci. Rep.* 8:14117. doi: 10.1038/s41598-018-32421-2
- Burridge, K., Monaghan-Benson, E., and Graham, D. M. (2019). Mechanotransduction: from the cell surface to the nucleus via RhoA. *Phil. Trans. R. Soc. B* 374:0229. doi: 10.1098/rstb.2018.0229
- Bursa, J., and Fuis, V. (2009). "Finite element simulation of mechanical tests of individual cells," in *IFMBE Proceedings WC, Vol 25*, eds O. Dössel and W. C. Schlegel (Springer Verlag), 16–19.
- Byun, S., Son, S., Amodei, D., Cermak, N., Shaw, J., Kang, J. H., et al. (2013). Characterizing deformability and surface friction of cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, 7580–7585. doi: 10.1073/pnas.1218806110
- Cartagena-Rivera, A. X., Van Itallie, C. M., Anderson, J. M., and Chadwick, R. S. (2017). Apical surface supracellular mechanical properties in polarized epithelium using noninvasive acoustic force spectroscopy. *Nature Comm.* 8:1030. doi: 10.1038/s41467-017-01145-8
- Cattin, C. J., Düggelin, M., Martinez-Martin, D., Gerber, C., Müller, D. J., and Stewart, M. P. (2015). Mechanical control of mitotic progression in single animal cells. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11258–11263. doi: 10.1073/pnas.1502029112
- Charras, G. T., and Horton, M. A. (2002a). Single cell mechanotransduction and its modulation analyzed by atomic force microscope indentation. *Biophys. J.* 82, 2970–2981. doi: 10.1016/S0006-3495(02)75638-5
- Charras, G. T., and Horton, M. A. (2002b). Determination of cellular strains by combined atomic force microscopy and finite element modeling. *Biophys. J.* 83, 858–879. doi: 10.1016/s0006-3495(02)75214-4
- Correa, D., and Lietman, S. A. (2017). Articular cartilage repair: current needs, methods and research directions. *Sem. Cell Develop. Biol.* 62, 67–77. doi: 10.1016/j.semcdb.2016.07.013
- Curtis, A. S. G. (1964). The mechanism of adhesion of cells to glass: a study by interference reflection microscopy. *J. Cell Biol.* 20, 199–215. doi: 10.1083/jcb.20.2.199
- Darling, E. M., Zauscher, S., and Guilak, F. (2006). Viscoelastic properties of zonal articular chondrocytes measured by atomic force microscopy. *Osteoarthritis Cartilage* 14, 571–579. doi: 10.1016/j.joca.2005.12.003
- Derjaguin, B. V., Muller, V. M., and Toporov, V. P. (1975). Effect of contact deformations on the adhesion of particles. *J. Colloid Interface Sci.* 53, 314–325. doi: 10.1016/0021-9797(75)90018-1
- Dhahri, S., Ramonda, M., and Marlière, C. (2013). *In-situ* determination of the mechanical properties of gliding or non-motile bacteria by atomic force

- microscopy under physiological conditions without immobilization. *PLoS ONE* 8:e61663. doi: 10.1371/journal.pone.0061663
- Ding, Y., Wang, J., Xu, G.-K., and Wang, G.-F. (2018). Are elastic moduli of biological cells depth dependent or not? Another explanation using a contact mechanics model with surface tension. *Soft Matter* 14, 7534–7541. doi: 10.1039/C8SM01216D
- Domke, J., Dannöhl, S., Parak, W. J., Müller, O., Aicher, W. K., and Radmacher, M. (2000). Substrate dependent differences in morphology and elasticity of living osteoblasts investigated by atomic force microscopy. *Coll. Int. B Bioint.* 19, 367–379. doi: 10.1016/S0927-7765(00)00145-4
- Dubreuil, F., Elsner, N., and Fery, A. (2003). Elastic properties of polyelectrolyte capsules studied by atomic-force microscopy and RCM. *Eur. Phys. J. E* 12, 215–221. doi: 10.1140/epje/i2003-10056-0
- Dufrene, Y. D., and Persat, A. (2020). Mechanomicrobiology: how bacteria sense and respond to forces. *Nat. Rev. Microbiol.* 18, 227–240. doi: 10.1038/s41579-019-0314-2
- Efremov, Y. M., Dokrunova, A. A., Bagrov, D. V., Kudryashova, K. S., Sokolova, O. S., and Shaitan, K. V. (2013). The effects of confluency on cell mechanical properties. *J. Biomech.* 46, 1081–1087. doi: 10.1016/j.jbiomech.2013.01.022
- Evans, E., and Yeung, A. (1989). Apparent viscosity and cortical tension of blood granulocytes determined by micropipet aspiration. *Biophys. J.* 56, 151–160. doi: 10.1016/S0006-3495(89)82660-8
- Ferrero, E. E., Martens, K., and Barrat, J. L. (2014). Relaxation in yield stress systems through elastically interacting activated events. *Phys. Rev. Lett.* 113:248301. doi: 10.1103/PhysRevLett.113.248301
- Freshney, R. I. (2005). *Culture of Animal Cells: A Manual of Basic Techniques*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9780471747598
- Gebeshuber, I. C., Drack, M., and Scherge, M. (2008). Tribology in biology. *Tribology* 2, 200–212. doi: 10.1179/175158308X383206
- Guilak, F., Tedrow, J. R., and Burgkart, R. (2000). Viscoelastic properties of the cell nucleus. *Biochem. Biophys. Res. Comm.* 269, 781–786. doi: 10.1006/bbrc.2000.2360
- Guillaume-Gentil, O., Potthoff, E., Ossola, D., Franz, C. M., Zambelli, T., and Vorholt, J. A. (2014). Force-controlled manipulation of single cells: from AFM to FluidFM. *Trends Biotech.* 32, 381–388. doi: 10.1016/j.tibtech.2014.04.008
- Gullekson, C., Cojoc, G., Schürmann, M., Guck, J., and Pelling, A. (2017). Mechanical mismatch between Ras transformed and untransformed epithelial cells. *Soft Matter* 13, 8483–8491. doi: 10.1039/C7SM01396E
- Guo, M., Ehrlicher, A. J., Jensen, M. H., Renz, M., Moore, G. R. D. et al. (2014). Probing the stochastic, motor-driven properties of the cytoplasm using force spectrum microscopy. *Cell* 158, 822–832. doi: 10.1016/j.cell.2014.06.051
- Guz, N., Dokukin, M., Kalaparthy, V., and Sokolov, I. (2014). If cell mechanics can be described by elastic modulus: study of different models and probes used in indentation experiments. *Biophys. J.* 107, 564–575. doi: 10.1016/j.bpj.2014.06.033
- Hallou, A., and Brunet, T. (2020). On growth and force: mechanical forces in development. *Development* 147:dev187302. doi: 10.1242/dev.187302
- Hasse, K., and Pelling, A. E. (2015). Investigating cell mechanics with atomic force microscopy. *J. R. Soc. Interface* 12:20140970. doi: 10.1098/rsif.2014.0970
- Hategan, A., Law, R., Kahn, S., and Discher, D. E. (2003). Adhesively-tensed cell membranes: lysis kinetics and atomic force microscopy probing. *Biophys. J.* 85, 2746–2759. doi: 10.1016/S0006-3495(03)74697-9
- Heinrich, V., Ritchie, K., Mohandas, N., and Evans, E. (2001). Elastic thickness compressibility of the red cell membrane. *Biophys. J.* 81, 1452–1463. doi: 10.1016/S0006-3495(01)75800-6
- Hertz, H. (1882). Über die Berührung fester elastischer Körper. *J. Reine Angew. Math.* 92, 156–171. doi: 10.1515/crll.1882.92.156
- Hiratsuka, S., Mizutani, Y., Tsuchiya, M., Kawahara, K., Tokumoto, H., and Okajima, T. (2009). The number distribution of complex shear modulus of single cells measured by atomic force microscopy. *Ultramicroscopy* 109, 937–941. doi: 10.1016/j.ultramicro.2009.03.008
- Hochmuth, R. M. (2000). Micropipette aspiration of living cells. *J. Biomech.* 33, 15–22. doi: 10.1016/S0021-9290(99)00175-X
- Hoffman, B. D., and Crocker, J. C. (2009). Cell mechanics: dissecting the physical responses of cells to force. *Annu. Rev. Biomed. Eng.* 11, 259–288. doi: 10.1146/annurev.bioeng.10.061807.160511
- Hoffman, B. D., Massiera, G., and Crocker, J. C. (2007). Fragility and mechanosensing in a thermalized cytoskeleton model with forced protein unfolding. *Phys. Rev. E* 76:051906. doi: 10.1103/PhysRevE.76.051906
- Huang, H., Kamm, R. D., and Lee, R. T. (2004). Cell mechanics and mechanotransduction: pathways, probes, and physiology. *Am. J. Physiol. Cell Physiol.* 287, 1–11. doi: 10.1152/ajpcell.00559.2003
- Hui, C.-Y., Liu, T., Salez, T., Raphael, E., and Jagota, A. (2015). Indentation of a rigid sphere into an elastic substrate with surface tension and adhesion. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 471:727. doi: 10.1098/rspa.2014.0727
- Hwang, H. J., Riggleman, R. A., and Crocker, J. C. (2016). Understanding soft glassy materials using an energy landscape approach. *Nat. Mater.* 15, 1031–1036. doi: 10.1038/nmat4663
- Ingber, D. E. (1993). Cellular tensegrity: defining new rules of biological design that govern the cytoskeleton. *J. Cell Sci.* 104, 613–627.
- Ingber, D. E. (2003). Tensegrity I. Cell structure and hierarchical systems. *J. Cell Sci.* 116, 1157–1173. doi: 10.1242/jcs.00359
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. R. Soc. Lond. A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Kamm, R. D., Lammerding, J., and Mofrad, M. R. K. (2017). “Cellular Nanomechanics,” in *Springer Handbook of Nanotechnology*, ed B. Bhushan (Berlin: Springer Handbooks; Heidelberg: Springer), 1069–1100.
- Kang, H., Wong, D. S. H., Yan, X., Jung, H. J., Kim, S., Lin, S., et al. (2017). Remote control of multimodal nanoscale ligand oscillations regulates stem cell adhesion and differentiation. *ACS Nano* 11, 9636–9649. doi: 10.1021/acsnano.7b02857
- Keller, R. (2012). Physical biology returns to morphogenesis. *Science* 338, 201–203. doi: 10.1126/science.1230718
- Kennedy, K. M., Bhaw-Lusimon, A., and Jhurry, D. (2017). Cell-matrix mechanical interaction in electrospun polymeric scaffolds for tissue engineering: Implications for scaffold design and performance. *Acta Biomaterialia* 50, 41–55. doi: 10.1016/j.actbio.2016.12.034
- Kroy, K. (2008). Dynamics of wormlike and glassy wormlike chains. *Soft Matter* 4, 2323–2330. doi: 10.1039/b807018k
- Kumar, S., and Weaver, V. M. (2009). Mechanics, malignancy and metastasis: the force journey of a tumor cell. *Cancer Metastasis Rev.* 28, 113–127. doi: 10.1007/s10555-008-9173-4
- Langer, M. G., Fink, S., Koitschev, A., Rexhausen, U., Heinrich Hörber, J. K., and Ruppertsberg, J. P. (2001). Lateral mechanical coupling of stereocilia in cochlear hair bundles. *Biophys. J.* 80, 2608–2621. doi: 10.1016/S0006-3495(01)76231-5
- Lau, A. W. C., Hoffman, B. D., Davies, A., Crocker, J. C., and Lubensky, T. C. (2003). Microrheology, stress fluctuations and active behavior of living cells. *Phys. Rev. Lett.* 91:198101. doi: 10.1103/PhysRevLett.91.198101
- Lemma, E. D., Spagnolo, B., De Vittorio, M., and Pisanello, F. (2019). Studying cell mechanobiology in 3D: the two-photon lithography approach. *Trends in Biotech.* 37, 358–372. doi: 10.1016/j.tibtech.2018.09.008
- Lherbette, M., dos Santos, A., Hari-Gupta, Y., Fili, N., Toseland, C. P., and Schaap, I. A. T. (2017). Atomic Force Microscopy microrheology reveals large structural inhomogeneities in single cell-nuclei. *Sci. Rep.* 7:8116. doi: 10.1038/s41598-017-08517-6
- Liu, Y., Galior, K., Ma, V. P.-Y., and Salaita, K. (2017). Molecular tension probes for imaging forces at the cell surface. *Acc. Chem. Res.* 50, 2915–2924. doi: 10.1021/acs.accounts.7b00305
- Liu, Y., Mollaeian, K., and Ren, J. (2019). Finite element modeling of living cells for AFM indentation-based biomechanical characterization. *Micron* 116, 108–115. doi: 10.1016/j.micron.2018.10.004
- Lomakina, E. B., Spillmann, C. M., King, M. R., and Waugh, R. E. (2004). Rheological analysis and measurement of neutrophil indentation. *Biophys. J.* 87:4246–4258. doi: 10.1529/biophysj.103.031765
- Long, J., Wang, G., Feng, X.-Q., and Yu, S. (2016). Effects of surface tension on the adhesive contact between a hard sphere and a soft substrate. *Int. J. Solids Struct.* 84, 133–138. doi: 10.1016/j.ijsolstr.2016.01.021
- Mahaffy, R. E., Park, S., Gerde, E., Käs, J., and Shih, C. K. (2004). Quantitative analysis of the viscoelastic properties of thin regions of fibroblasts using atomic force microscopy. *Biophys. J.* 86, 1777–1793. doi: 10.1016/S0006-3495(04)74245-9
- Managuli, V., and Roy, S. (2017). Simultaneous analysis of elastic and nonspecific adhesive properties of thin sample and biological cell considering bottom substrate effect. *J. Biomech. Eng.* 139:9. doi: 10.1115/1.4037289
- Mathur, A. B., Collinsworth, A. M., Reichert, W. M., Kraus, W. E., and Truskey, G. A. (2001). Endothelial, cardiac muscle and skeletal muscle exhibit different viscous and elastic properties as determined by atomic force microscopy. *J. Biomech.* 34, 1545–1553. doi: 10.1016/S0021-9290(01)00149-X

- Mathur, A. B., Truskey, G. A., and Reichert, W. M. (2000). Atomic force and total internal reflection fluorescence microscopy for the study of force transmission in endothelial cells. *Biophys. J.* 78, 1725–1735. doi: 10.1016/S0006-3495(00)76724-5
- Matzke, R., Jacobson, K., and Radmacher, M. (2001). Direct, high-resolution measurement of furrow stiffening during division of adherent cells. *Nature Cell Biol.* 3, 607–610. doi: 10.1038/35078583
- Maugis, D., and Pollock, H. M. (1984). Surface forces, deformation and adherence at metal microcontacts. *Acta Metal.* 32, 1323–1334. doi: 10.1016/0001-6160(84)90078-6
- McElfresh, M., Baesu, E., Balhorn, R., Belak, J., Allen, M. J., and Rudd, R. E. (2002). Combining constitutive materials modeling with atomic force microscopy to understand the mechanical properties of living cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6493–6497. doi: 10.1073/pnas.082520599
- Miller, A. E., Hu, Pi, and Barker, T. H. (2020). Feeling things out: bidirectional signaling of the cell-ECM interface, implications in the mechanobiology of cell spreading, migration, proliferation, and differentiation. *Adv. Healthcare Mat.* 9:1901445. doi: 10.1002/adhm.201901445
- Mizuno, D., Tardin, C., Schmidt, C. F., and MackKintosh, F. C. (2007). Nonequilibrium mechanics of active cytoskeletal networks. *Science* 315, 370–373. doi: 10.1126/science.1134404
- Moeendarbary, E., Valon, L., Frizsche, M., Harris, A. R., Moulding, D. A., Thrasher, A. J., et al. (2013). The cytoplasm of living cells behaves as a poroelastic material. *Nat. Mater.* 12, 253–261. doi: 10.1038/nmat3517
- Moreno-Flores, S., Benitez, R., Vivanco, M., and Toca-Herrera, J. L. (2010). Stress relaxation and creep on living cells with the atomic force microscope: a means to calculate elastic moduli and viscosities of cell components. *Nanotechnology* 21:445101. doi: 10.1088/0957-4484/21/44/445101
- Moreno-Flores, S., and Küpcü, S. (2015). 2D protein arrays induce 3D *in vivo*-like assemblies of cells. *Soft Matter* 11, 1259–1264. doi: 10.1039/C4SM02278E
- Moreno-Flores, S., and Toca-Herrera, J. L. (2013). *Hybridizing Surface Probe Microscopies. Toward a Full Description of the Meso- and Nanoworlds*. Boca Raton, FL: CRC Press.
- Nautiyal, P., Alam, F., Balani, K., and Agarwal, A. (2018). The role of nanomechanics in healthcare. *Adv. Healthcare Mat.* 7:1700793. doi: 10.1002/adhm.201700793
- Nawaz, S., Sánchez, P. L., Bodensiek, K., Li, S., Simons, M., and Shaap, I. A. (2012). Cell viscoelasticity measured with AFM and optical trapping at sub-micrometer deformations. *PLoS ONE* 7:e45297. doi: 10.1371/journal.pone.0045297
- Pelling, A. E., Dawson, D. W., Carreon, D. M., Chirstiansen, J. J., Shen, R. R., Teittel, M. A., et al. (2007). Distinct contributions of microtubule subtypes to cell membrane shape and stability. *Nanomed. Nanotech. Biol. Med.* 3, 43–52. doi: 10.1016/j.nano.2006.11.006
- Pietuch, A., Brückner, B. R., and Janshoff, A. (2013b). Membrane tension homeostasis of epithelial cells through surface area regulation in response to osmotic stress. *Biochim. Biophys. Acta: Mol. Cell Res.* 1833, 712–722. doi: 10.1016/j.bbamcr.2012.11.006
- Pietuch, A., Rouven Brückner, B., Fine, T., Mey, I., and Janshoff, A. (2013a). Elastic properties of cells in the context of confluent cell monolayers: impact of tension and surface area regulation. *Soft Matter* 9:11490. doi: 10.1039/c3sm51610e
- Pina, S., Rebelo, R., Correlo, V. M., Oliveira, J. M., and Reis, R. L. (2018). Bioceramics for Osteochondral Tissue Engineering and Regeneration. *Adv. Exp. Med. Biol.* 1058:53–75. doi: 10.1007/978-3-319-76711-6_3
- Pitenis, A. A., Urueña, J. M., Hart, S. M., O'Bryan, C. S., Marshall, S. L., Levings, P. P., et al. (2018). Friction-Induced Inflammation. *Tribol. Lett.* 66:81. doi: 10.1007/s11249-018-1029-7
- Polacheck, W., and Chen, C. S. (2016). Measuring cell-generated forces: a guide to the available tools. *Nat. Methods* 13, 415–423. doi: 10.1038/nmeth.3834
- Popov, V. L., He, B. M., and Willert, E. (2019). *Handbook of Contact Mechanics – Exact Solutions of Axisymmetric Contact Problems*. Berlin: Springer. doi: 10.1007/978-3-662-58709-6
- Radmacher, M., Fritz, M., Kacher, C. M., Cleveland, J. P., and Hansa, P. K. (1996). Measuring the viscoelastic properties of human platelets with the atomic force microscopy. *Biophys. J.* 70, 556–567. doi: 10.1016/S0006-3495(96)79602-9
- Radmacher, M., Tillmann, R. W., Fritz, M., and Gaub, H. E. (1992). From molecules to cells: imaging soft samples with the atomic force microscope. *Science* 257, 1900–1905. doi: 10.1126/science.1411505
- Rosenbluth, M. J., Lam, W. A., and Fletcher, D. A. (2006). Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. *Biophys. J.* 90, 2994–3003. doi: 10.1529/biophysj.105.067496
- Rotsch, C., and Radmacher, M. (2000). Drug-induced changes of cytoskeletal structure and mechanics in fibroblasts: an atomic microscopy study. *Biophys. J.* 78, 520–535. doi: 10.1016/S0006-3495(00)76614-8
- Schierbaum, N., Rheinlaender, J., and Schäffer, T. E. (2017). Viscoelastic properties of normal and cancerous human breast cells are affected differently by contact to adjacent cells. *Acta Biomaterialia* 55, 239–248. doi: 10.1016/j.actbio.2017.04.006
- Schierbaum, N., Rheinlaender, J., and Schäffer, T. E. (2019). Combined atomic force microscopy (AFM) and traction force microscopy (TFM) reveals a correlation between viscoelastic material properties and contractile prestress of living cells. *Soft Matter* 15, 1721–1729. doi: 10.1039/C8SM01585F
- Schwander, M., Kachar, B., and Müller, U. (2010). The cell biology of hearing. *J. Cell Biol.* 190, 9–20. doi: 10.1083/jcb.201001138
- Schwarz, U. S., and Safran, S. A. (2013). Physics of adherent cells. *Rev. Mod. Phys.* 85, 1327–1381. doi: 10.1103/RevModPhys.85.1327
- Sen, S., Subramanian, S., and Discher, D. E. (2005). Indentation and adhesive probing of a cell membrane with afm: theoretical model and experiments. *Biophys. J.* 89, 3203–3213. doi: 10.1529/biophysj.105.063826
- Septiadi, D., Crippa, F., Moore, T. L., Rothen-Rutishauser, B., and Petri-Fink, A. (2018). Nanoparticle-cell interaction: a cell mechanics perspective. *Adv. Mat.* 30:1704463. doi: 10.1002/adma.201704463
- Shen, Y., Guan, D., Serien, D., Takeuchi, S., Tong, P., Yobas, L., et al. (2017). Mechanical characterization of microengineered epithelial cysts by using atomic force microscopy. *Biophys. J.* 112, 398–409. doi: 10.1016/j.bpj.2016.12.026
- Shi, Y., Porter, C. L., Crocker, J. C., and Reich, D. H. (2019). Dissecting fat-tailed fluctuations in the cytoskeleton with active micropost arrays. *Proc. Natl. Acad. Sci. U.S.A.* 116, 13839–13846. doi: 10.1073/pnas.1900963116
- Skamrah, M., Colin-York, H., Barbieri, L., and Fritzsche, M. (2019). Simultaneous quantification of the interplay between molecular turnover and cell mechanics by AFM-FRAP. *Small* 15:1902202. doi: 10.1002/smll.201902202
- Sneddon, I. N. (1965). The relation between load and penetration in the axisymmetric Boussinesq problem for a punch of arbitrary profile. *Int. J. Eng. Sci.* 3, 47–57. doi: 10.1016/0020-7225(65)90019-4
- Sokolov, I., Dokukin, M. E., and Guz, N. V. (2013). Method for quantitative measurements of the elastic modulus of biological cells in AFM indentation experiments. *Methods* 60, 202–213. doi: 10.1016/j.ymeth.2013.03.037
- Solich, P. (1998). Rheological constitutive equation for a model of soft glassy materials. *Physical Rev E* 58, 738–759. doi: 10.1103/PhysRevE.58.738
- Stamenović, D. (2008). Rheological behavior of mammalian cells. *Cell. Mol. Life Sci.* 65, 3592–3605. doi: 10.1007/s00018-008-8292-y
- Stephens, A. D., Banigan, E. J., Adam, S. A., Goldman, R. D., and Marko, J. F. (2017). Chromatin and lamin A determine two different mechanical response regimes of the cell nucleus. *Mol. Biol. Cell* 28, 1984–1996. doi: 10.1091/mbc.e16-09-0653
- Stewart, M. P., Helenius, J., Toyoda, Y., Ramanathna, S. P., Müller, D. J., and Hyman, A. A. (2011). Hydrostatic pressure and the actomyosin cortex drive mitotic cell rounding. *Nature* 469, 226–230. doi: 10.1038/nature09642
- Stewart, M. P., Toyoda, Y., Hyman, A. A., and Müller, D. J. (2012). Tracking mechanics and volume of globular cells with atomic force microscopy using a constant-height clamp. *Nature Protocols* 7, 143–154. doi: 10.1038/nprot.2011.434
- Stuart, J. K., and Hlady, V. (1999). Reflection interference contrast microscopy combined with scanning force microscopy verifies the nature of protein-ligand interaction force measurements. *Biophys. J.* 76, 500–508. doi: 10.1016/S0006-3495(99)77218-8
- Stylianou, A., Lekka, M., and Stylianopoulos, T. (2018). AFM assessing of nanomechanical fingerprints for cancer early diagnosis and classification: from single cell to tissue level. *Nanoscale* 10:20930. doi: 10.1039/C8NR06146G
- Thoumine, O., and Ott, A. (1997). Time scale dependent viscoelastic and contractile regimes in fibroblasts probed by microplate manipulation. *J. Cell Sci.* 110, 2109–2116.
- Tietze, S., Kräter, M., Jacobi, A., Taubenberger, A., Herbig, M., Wehner, R., et al. (2019). Spheroid culture of mesenchymal stromal cells results in

- morphorheological properties appropriate for improved microcirculation. *Adv. Sci.* 6:1802104. doi: 10.1002/advs.201802104
- Trickey, W. R., Baaijens, F. P., Laursen, T. A., Alexopoulos, L. G., and Guilak, F. (2006). Determination of the Poisson's ratio of the cell: recovery properties of chondrocytes after release from complete micropipette aspiration. *J. Biomech.* 39, 78–87. doi: 10.1016/j.jbiomech.2004.11.006
- Unnikrishnan, G. U., Unnikrishnan, V. U., and Reddy, J. N. (2007). Constitutive material modeling of cell: a micromechanics approach. *J. Biomech. Eng.* 129, 315–323. doi: 10.1115/1.2720908
- Vadillo-Rodriguez, V., Beveridge, T. J., and Dutcher, J. R. (2008). Surface viscoelasticity of individual gram-negative bacterial cells measured using atomic force microscopy. *J. Bacteriol.* 190, 4225–4232. doi: 10.1128/JB.00132-08
- Verschueren, H. (1985). Interference reflection microscopy in cell biology: methodology and applications. *J. Cell Sci.* 75, 279–301.
- Vogel, V., and Sheetz, M. (2006). Local force and geometry sensing regulate cell functions. *Nat. Rev. Mol. Cell Biol.* 7, 265–275. doi: 10.1038/nrm1890
- Weisenhorn, A. L., Khorsandi, M., Kasas, S., Gotzos, V., and Butt, H.-J. (1993). Deformation and height anomaly of soft surfaces studied with an AFM. *Nanotechnology* 4, 106–113. doi: 10.1088/0957-4484/4/2/006
- Wirtz, D., Konstantopoulos, K., and Searson, P. C. (2011). The physics of cancer: the role of physical interactions and mechanical forces in metastasis. *Nat. Rev. Cancer.* 11, 512–522. doi: 10.1038/nrc3080
- Wottawah, F., Schinkinger, S., Lincoln, B., Ananthakrishnan, R., Romeyke, M., Guck, J., et al. (2005). Optical rheology of biological cells. *Phys. Rev. Lett.* 94:098103. doi: 10.1103/PhysRevLett.94.098103
- Wu, P.-H., Aroush, D. R.-B., Asnacios, A., Chen, W.-C., Dokukin, M. E., Doss, B. L., et al. (2018). A comparison of methods to assess cell mechanical properties. *Nat. Methods* 15, 491–498. doi: 10.1038/s41592-018-0015-1
- Xie, H., Yin, M., Rong, W., and Sun, L. (2014). *In situ* quantification of living cell adhesion forces: single cell force spectroscopy with a nanotweezer. *Langmuir* 30, 2952–2959. doi: 10.1021/la500045q

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Moreno-Flores. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Macroscopic Modeling of Fingerpad Friction Under Electroadhesion: Possibilities and Limitations

Markus Heß* and Fabian Forsbach

Institute of Mechanics, Technische Universität Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Irina Goryacheva,
Institute for Problems in Mechanics
(RAS), Russia

Reviewed by:

Cagatay Basdogan,
Koç University, Turkey
Emile Van Der Heide,
University of Twente, Netherlands

*Correspondence:

Markus Heß
markus.hess@tu-berlin.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 29 May 2020

Accepted: 20 August 2020

Published: 21 October 2020

Citation:

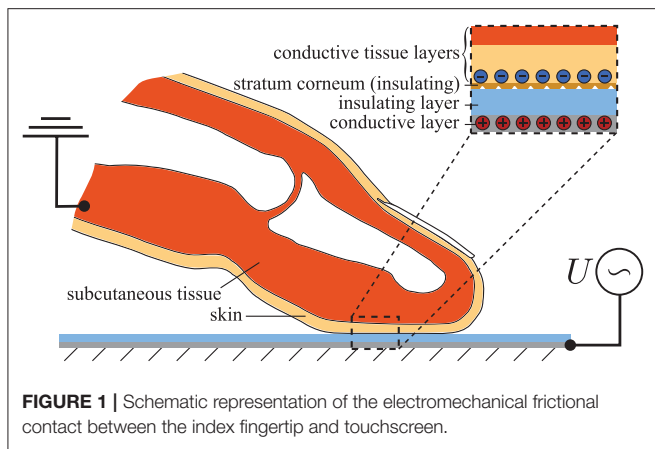
Heß M and Forsbach F (2020)
Macroscopic Modeling of Fingerpad
Friction Under Electroadhesion:
Possibilities and Limitations.
Front. Mech. Eng. 6:567386.
doi: 10.3389/fmech.2020.567386

Electrovibration is one of the key technologies in surface haptics. By inducing controlled electrostatic forces, the friction within a sliding contact between the human finger and a capacitive screen is modulated, which in turn gives effective tactile feedback to the user. Such powerful haptic displays can be built into mobile phones, tablets, navigation devices, games consoles and many other devices of consumer electronics. However, due to the layered structure and complex material of human skin, the underlying contact mechanical processes have not yet been fully understood. This work provides new continuum-based approaches to macroscopic modeling of the electro-adhesive frictional contact. A solution of pure normal contact between a human finger and a rigid, smooth plane under electroadhesion is derived by applying Shull's compliance method in the extended regime of large deformations. Based on these results and assuming pressure-controlled friction, a model for the sliding electro-adhesive contact is developed, which adequately predicts the friction force and coefficient of friction over the whole range of relevant voltages and applied normal forces. The experimentally observed area reduction caused by the tangential force is incorporated in a more empirical than profound contact mechanical way. This effect is studied with the help of a two-dimensional finite element model of the fingertip, assuming non-linear elastic material for the skin tissue. Although the simulations are restricted to non-adhesive tangential contacts, they show a significant reduction of the contact area, which is caused by large deformations of the non-linear elastic material around the distal phalanx. This result indicates that adhesion is only of secondary importance for the area reduction.

Keywords: friction, adhesion, electrovibration, surface haptics, finite element method—FEM, compliance method, hyperelastic material

INTRODUCTION

Understanding contact mechanics and friction of human skin is a great challenge for the tribological community. Human skin is characterized by a complex layered structure of non-linear viscoelastic material and a specific surface topography. In addition, its hydration level as well as moisture at its surface can strongly influence grip and touch properties. Especially with regard to tactile perception skin tribology is not yet fully understood (Derler and Gerhardt, 2012; van Kuilenburg et al., 2015). In this respect, improved knowledge is urgently needed as it plays a major role in the rapidly growing field of robotic and haptic applications. One key technology in surface haptics is electrovibration, which is based on the polarization of a fingerpad pressed in contact



with an AC voltage supplied surface coated by an insulating layer. When the fingerpad is moved over the substrate, the user perceives a characteristic feeling which can be altered by controlling the shape, amplitude and frequency of the voltage (Vardar et al., 2017). In this way, the user can get effective tactile feedback. A schematic representation of such an electromechanical frictional contact between the fingerpad and touchscreen under electrovibration is depicted in **Figure 1**.

Despite numerous experimental studies in this area, several effects are not yet sufficiently understood (Sirin et al., 2019b). Above all, there is a lack of well-founded models that correctly reflect the interaction between contact mechanics and electrodynamics. These models should not only provide good results for a single measured quantity like the friction force, but rather all other contact mechanical quantities, particularly the contact area, must also be correctly mapped on the way there. Excellent modeling from an electrodynamic point of view can be found in the works by Shultz et al. (2015) and Shultz et al. (2018) as well as Nakamura and Yamamoto (2017). From the point of view of contact mechanics, however, the simplest approaches are chosen. Promising multiscale approaches that cover both electrodynamics and contact mechanics in a suitable manner include the works by Persson (2018) and Sirin et al. (2019a).

Following the current Research Topic “contact mechanics perspective of tribology,” we focus exclusively on macroscopic modeling. We define a model as “macroscopic” if it is based on the apparent or ridge contact area. Smaller scales are not taken into account! On this macro scale some effort has been made to map voltage-induced friction as well. In this context, reference is made to Vodlak et al. (2016), Heß and Popov (2019) as well as Argatov and Borodich (2020). The work by Vodlak et al. focus on the assessment of two analytical models of electrovibration based on the parallel-plate capacitor by comparisons with experimental results published in literature. The approach proposed by Heß and Popov exploits the close analogy of electroadhesive contacts to classical adhesion theories based on van der Waals forces. However, this model provides insufficient results with respect to the contact area as a function of the normal force, since the original theory by Johnson et al.

(1971) is applied. The interesting extension in the work by Argatov and Borodich is that it also takes non-linear elastic material behavior into account. However, a simple Winkler-Fuss model is mainly used, which is why it should generally be checked whether the three-dimensional contact mechanical behavior of the adhesive fingerpad contact can be mapped correctly. During the preparation phase of this manuscript another work on the same topic was published by Basdogan et al. (2020). Their model is based on the original theory of Johnson, Kendall and Roberts applicable for parabolic contacts of linear elastic materials. Therefore, the inhomogeneous, non-linear elastic finger material is replaced by a (fictive) homogeneous linear elastic one. A further characteristic approach of the model is the assumed proportionality between the real and apparent contact area, which is chosen in accordance with the results of recent multi-scale calculations (Ayyildiz et al., 2018; Sirin et al., 2019a).

As stated above, the aim of the present work is to develop a “macroscale” model which correctly reproduces all contact mechanical quantities and effects arising from the electro-adhesive frictional contact between the fingerpad and touchscreen. Since the tangential contact model is based on the solution of the pure normal contact, it is necessary to derive a robust model for the normal contact under electroadhesion.

Therefore, the present manuscript is structured as follows: First a novel model for the pure normal contact under electroadhesion is developed by application of the compliance method in the extended regime of large deformations and non-linear elastic materials. The integration of electroadhesion is realized by an idea of Popov and Heß (2018). Based on the resulting function of the ridge contact area in terms of applied voltage and normal force, an extended model for the sliding electro-adhesive contact is developed in Chapter Tangential Contact with Electroadhesion. This chapter begins with a study of the origins of the experimentally observed area reduction in frictional contact by means of adhesion theory and a non-adhesive two-dimensional finite element model of the fingertip accounting for the large deformations and non-linear elastic material behavior. In agreement with the FE results and recent studies, the area reduction is then incorporated in a model for pressure-controlled sliding friction in an empirical way. Finally, the developed model is compared to recent experimental results. Some conclusive remarks and a short discussion close the manuscript.

NORMAL CONTACT WITH ELECTROADHESION

Although the main objective of this study is to develop a model for sliding friction of a fingerpad over a smooth surface under electroadhesion, the preliminary investigation of pure normal contact is mandatory. The solution of the normal contact problem must be reproduced correctly in the limit of a vanishing tangential force, for both cases, with and without electroadhesion. In particular, the model for calculating the reduction of the contact area in the state of full slip requires precise knowledge of the contact area under pure normal loading

with switched-on electroadhesion. While section Theoretical Background is devoted to the repetition and discussion of the theoretical principles to be applied, in section Application to Fingerpad in Contact with Capacitive Screen they are used to solve the normal contact between finger and capacitive screen under electroadhesion.

Theoretical Background

The model for normal contact under electroadhesion is essentially based on two fundamental principles, which are briefly repeated here. From a contact mechanical point of view, the focus lies on Shull's compliance method, which was originally designed for linear elastic material behavior. Its extended application to non-linear elastic material is not very well-known. According to Heß and Popov (2019), the incorporation of electrostatic attraction into the model is done by calculating the work of electroadhesion as well as the electrostatic force per unit area, which also includes the concept of the equivalent air gap.

Shull's Compliance Method—Generalized Version of the JKR-Theory

In analogy to the energy-based derivation of the original theory by Johnson et al. (1971), Shull and coworkers (Shull et al., 1998; Shull, 2002) developed a method which enables obtaining the solution of a more arbitrary adhesive normal contact problem from the known solution of the corresponding non-adhesive one. This method is called “compliance method” and represents a generalized version of the JKR-theory. Its applicability is neither restricted to homogeneous materials nor to circular contact areas, but linear material behavior is required. The main results of the compliance method are the following expressions for the elastic energy release rate G :

$$G = \frac{(F_1 - F_N)^2}{2S^2} \frac{dS}{dA} \quad \text{and} \quad G = \frac{(\delta_1 - \delta)^2}{2} \frac{dS}{dA}, \quad (1)$$

where F_N and δ are the normal force and indentation depth of the adhesive contact. F_1 and δ_1 refer to the values of the corresponding non-adhesive contact and S denotes the contact stiffness defined by:

$$S := \frac{dF_1}{d\delta_1}. \quad (2)$$

After equating the energy release rate G with the thermodynamic work of adhesion w , Equation (1) leads to:

$$F_N(A) = F_1(A) - S(A) \sqrt{2w \frac{dS}{dA}}, \quad (3)$$

$$\delta(A) = \delta_1(A) - \sqrt{2w \frac{dS}{dA}}. \quad (4)$$

Recently Equations (3) and (4) have been rediscovered and more precise restrictions concerning their applicability have been added (Ciavarella, 2018; Popov, 2018). The main assumption is that the sequence of contact configurations in an adhesive contact should be the same as that of contact configurations in a non-adhesive one. For this reason, the method cannot be generally applied to rough contacts. However, in some cases it

seems to provide a good approximation of the adhesive solution. Furthermore, it should be stressed that the application to non-linear elastic material behavior, which characterizes human skin tissue, is only permitted under certain conditions, which are addressed at the beginning of section Application to Fingerpad in Contact with Capacitive Screen.

Application to power-law relationships between non-adhesive quantities

Typically, experimental results of the non-adhesive fingerpad in normal contact with a smooth rigid plane predict the following power-law relationships between the contact area and normal force as well as indentation depth:

$$A(F_1) = \alpha F_1^m, \quad (5)$$

$$A(\delta_1) = \beta \delta_1^n. \quad (6)$$

The corresponding adhesive solution can be obtained from Equations (3) and (4). From Equations (5) and (6) we first determine the stiffness according to Equation (2) by using the chain rule and its derivation with respect to the contact area:

$$S(A) = \frac{n}{m} \beta^{1/n} \alpha^{-1/m} A^{1/m-1/n}, \quad (7)$$

$$\frac{dS}{dA} = \frac{n-m}{nm} \frac{S(A)}{A}. \quad (8)$$

After inserting Equations (7), (8) in (3), (4) and taking into account the non-adhesive relationships, the solution of the adhesive normal contact is found:

$$F_N(A) = \alpha^{-1/m} A^{1/m} - \sqrt{\frac{2wn^2}{n-m} \beta^{1/n} \alpha^{-1/m} A^{1+1/m-1/n}}, \quad (9)$$

$$\delta(A) = \beta^{-1/n} A^{1/n} - \sqrt{\frac{2wm^2}{n-m} \beta^{-1/n} \alpha^{1/m} A^{1-1/m+1/n}}. \quad (10)$$

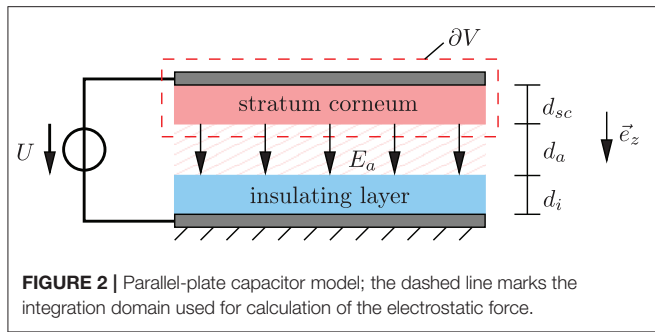
In particular, Equation (9) is used in section Application to Fingerpad in Contact with Capacitive Screen to calculate both the apparent contact area and the ridge contact area when electroadhesion is switched on. It should be noted that the exponents m and n are generally not independent of one another but are related due to the geometric and material properties of the contact. For instance, for axisymmetric normal contact problems of linear elastic homogeneous half-spaces (with a compact contact area) the exponents are connected by:

$$m = \frac{2n}{n+2} \quad (11)$$

and Equation (7) yields the well-known relationship $S \sim \sqrt{A}$.

The classical JKR-theory

As an example, let us rederive the Equations of the classical JKR-theory. For this purpose, we take the solution of the non-adhesive contact between two parabolically shaped elastic bodies with elastic moduli E_1 and E_2 , Poisson's ratios ν_1 and ν_2 as well as radii of curvature R_1 and R_2 from Hertz theory. The contact radius is



denoted by a . By a comparison with the predefined relationships (5) and (6) the following parameters can be identified:

$$m = 2/3, \quad n = 1, \quad \beta = \pi R^*, \quad \alpha = \pi \left(\frac{3R^*}{4E^*} \right)^{2/3}, \quad (12)$$

where $\frac{1}{R^*} = \frac{1}{R_1} + \frac{1}{R_2}$ and $\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}$. Inserting these parameters into Equations (9) and (10) results in:

$$F_N(a) = \frac{4}{3} \frac{E^* a^3}{R^*} - \sqrt{8\pi E^* w a^3}, \quad (13)$$

$$\delta(a) = \frac{a^2}{R^*} - \sqrt{\frac{2\pi w a}{E^*}}, \quad (14)$$

which indeed represent the classical JKR solution.

Electrostatic Force and Work of Electroadhesion

The most common approach to modeling the electrostatic contact between the fingerpad and the touchscreen is based on the parallel-plate capacitor shown in **Figure 2**. The conductive tissue of the skin as well as the conductive layer of the screen form the electrodes of the capacitor. Its space is filled by the stratum corneum, an air layer and the insulating layer of the screen. Here, the stratum corneum is assumed to be a perfect non-conducting layer, although it generally has a finite resistivity. However, if we focus on the AC case and the frequency of the applied alternating voltage is high enough, the assumption is justified. For extended approaches from the electrodynamic point of view, which study the frequency-dependence of the frictional force, the reader is referred to the works of Meyer et al. (2013), Vezzoli et al. (2014), and Shultz et al. (2015).

The electrostatic force onto the upper part of the capacitive system consisting of the upper plate and the stratum corneum (see **Figure 2**) can be calculated from the general definition, that is, by integration of the Maxwell stress tensor \mathbf{T} over the surface of the enclosed volume:

$$\vec{F}_{el} = \oint_{\partial V} \mathbf{T} \cdot d\vec{A}, \quad (15)$$

$$\text{with } \mathbf{T} = \varepsilon \left[\vec{E} \vec{E} - \frac{1}{2} (\vec{E} \cdot \vec{E}) \mathbf{I} \right], \quad (16)$$

where \vec{E} denotes the electric field, \mathbf{I} is the unit tensor of second order and ε the absolute permittivity. In our simple capacitive

system, the only contribution to the electrostatic force on the upper part comes from the electric field of the air gap \vec{E}_a , which points in the z -direction and thus perpendicular to the relevant surface. Therefore, only the zz -component of the Maxwell stress tensor is required to calculate the electrostatic force according to Equation (15), which leads to

$$\vec{F}_{el} = \int_A T_{a,zz} dA_z \vec{e}_z = \int_A \frac{1}{2} \varepsilon_a E_a^2 dA_z \vec{e}_z = \frac{1}{2} \varepsilon_a A E_a^2 \vec{e}_z. \quad (17)$$

From the continuity of the normal component of the electrical displacement at the interfaces between stratum corneum and air as well as air and insulating layer of the screen, the following relationships hold:

$$\varepsilon_{sc} E_{sc} = \varepsilon_a E_a = \varepsilon_i E_i, \quad (18)$$

where the abbreviations “sc” and “i” stand for “stratum corneum” and “insulating layer,” respectively. In addition, the voltage between the plates can be determined from the line integration of the electrical field, which leads to the following result:

$$U = E_{sc} d_{sc} + E_a d_a + E_i d_i. \quad (19)$$

Herein, d_{sc} , d_a , and d_i represent the thicknesses of the stratum corneum, air gap and insulating layer of the screen. From Equations (18) and (19) the electric field in the air gap can be determined and after substituting the result into Equation (17) the magnitude of the electrostatic force yields:

$$\begin{aligned} F_{el} &= |\vec{F}_{el}| = \frac{U^2 A}{2\varepsilon_a} \left(\frac{d_{sc}}{\varepsilon_{sc}} + \frac{d_a}{\varepsilon_a} + \frac{d_i}{\varepsilon_i} \right)^{-2} \\ &= \frac{\varepsilon_0 U^2 A}{2\varepsilon_{r,a}} \left(\frac{d_{sc}}{\varepsilon_{r,sc}} + \frac{d_a}{\varepsilon_{r,a}} + \frac{d_i}{\varepsilon_{r,i}} \right)^{-2}, \end{aligned} \quad (20)$$

where we have introduced the relative permittivities $\varepsilon_{r,sc}$, $\varepsilon_{r,a}$, and $\varepsilon_{r,i}$ as well as the permittivity of free space ε_0 on the right side. After dividing Equation (20) by the plate area, the electrostatic force per unit area is found:

$$\sigma_{el} = \frac{F_{el}}{A} = \frac{\varepsilon_0 U^2}{2\varepsilon_{r,a}} \left(\frac{d_{sc}}{\varepsilon_{r,sc}} + \frac{d_a}{\varepsilon_{r,a}} + \frac{d_i}{\varepsilon_{r,i}} \right)^{-2}. \quad (21)$$

Analogous to the work of adhesion that comes from the van der Waals forces, the work of electro-adhesion was introduced by Popov and Heß (2018). It represents the required work per unit area to separate the plates and can be calculated by:

$$w = \int_{d_a}^{\infty} \sigma_{el}(\tilde{d}_a) d\tilde{d}_a = \frac{\varepsilon_0 U^2}{2} \left(\frac{d_{sc}}{\varepsilon_{r,sc}} + \frac{d_a}{\varepsilon_{r,a}} + \frac{d_i}{\varepsilon_{r,i}} \right)^{-1}. \quad (22)$$

Let us briefly take a closer look at the limiting case of “direct contact,” i.e., a disappearing air gap. The corresponding electrostatic force results from Equation (20) taking into account $d_a = 0$:

$$F_{el}(d_a = 0) = \frac{\varepsilon_0 U^2 A}{2\varepsilon_{r,a}} \left(\frac{d_{sc}}{\varepsilon_{r,sc}} + \frac{d_i}{\varepsilon_{r,i}} \right)^{-2}. \quad (23)$$

If we make use of the usual estimation $\varepsilon_{r,a} \approx 1$, then Equation (23) exactly agrees with the formula proposed by Strong and Troxel (1970) in their pioneering work:

$$F_{el}^{ST} = \frac{\varepsilon_0 U^2 A}{2} \left(\frac{d_{sc}}{\varepsilon_{r,sc}} + \frac{d_i}{\varepsilon_{r,i}} \right)^{-2} \quad (24)$$

It should be noted that alternative approaches to Equations (24) and (20) were proposed by Kaczmarek et al. (2006) and Vodlak et al. (2016), which are supported by several authors (see e.g., Radivojevic et al., 2012; Giraud et al., 2013; Vezzoli et al., 2014; Liu G. et al., 2018). It goes without saying that these approaches can alternatively be integrated into the new model. Nevertheless, this work makes use of Equations (20–22). It is assumed that there is always an air gap even in the in-contact state, which includes not only the interstitial spaces between the ridges but also non-contacting areas on smaller length scales resulting from roughness. It is well-known that even the ridges themselves are far away from being smooth. They are punctuated by many concave shaped sweat pores openings. According to measurements by Liu et al. (2013), the number of sweat ducts considerably varies between subjects and lies between 300 and 1,000 per cm^2 . Therefore, some authors have recently introduced the junction area as a measure of the real contact area (Dzidek et al., 2016). Current purely theoretical investigations using mean-field models based on multiscale contact mechanics take even smaller scales into account which results in a further, significant decrease of the predicted real contact area. A ratio $A/A_0 < 10^{-3}$ is given by Ayyildiz et al. (2018) for instance.

Concept of an equivalent air gap

As mentioned above, the real contact area is made up of various micro-asperity contacts and is generally much smaller than the apparent or ridge contact area. This results in a non-uniform interfacial air gap between the surfaces of the stratum corneum and insulating screen layer. As part of a macroscopic model, we want to capture the whole influence of the non-uniform air gap on the electrostatic force by introducing an equivalent air gap of constant thickness. It is worth noting that the real non-uniform interfacial air gap and thus also the thickness of the equivalent air gap strongly depends on both the normal force and the applied voltage. If the real non-uniform air gap corresponding to a given normal force and voltage were known, the equivalent air gap would be obtained from:

$$\sigma_{el}(d_{a,eq}) \stackrel{!}{=} \frac{1}{A_0} \iint_{A_0} \sigma_{el}(d_a(x,y)) dx dy, \quad (25)$$

where $d_{a,eq}$ denotes the equivalent air gap. Unfortunately, it is impossible to determine the non-uniform air gap. However, the equivalent air gap concept can be used in another way. Some scientists calculate the equivalent air gap from accessible experimental data on the frictional force in electroadhesive contacts to incorporate it into a suitable substitute model. For example, Guo et al. (2019) measured the friction force of the finger sliding on a 3M touchscreen at different normal forces but under a *fixed apparent contact area*. From their experimental results they estimated the electrostatic force, which increased

significantly with increasing normal force. The authors explain this effect through an existing (equivalent) air gap between the fingertip and screen, whose thickness decreases with increasing normal force by a power function. At 150 V peak-to-peak voltage, the thickness decreased from 2.5 to 1.5 micrometers when the normal force has been increased from 0.5 N to 4.5 N. Nakamura and Yamamoto (2016) have proposed a multi-user visuo-haptic system, which integrates an additional rubber-like pad between the fingertip and touchscreen surface that has a surface-insulated electrode on its bottom. By means of an electrically activated and insulated screen electrode an electrostatic force acts on the pad and is then transferred to the fingertip placed on the pad. The electrostatic component to the friction force obtained from measurements showed different behavior at small and large voltages. Thus, they included an equivalent air gap in their parallel-plate capacitor model that varies between 0 and 6 micrometers depending on the applied voltage. We would also like to highlight the work of Shultz et al. (2015), who succeeded in unifying the DC based Johnson Rahbek and AC based electrovibration force models. They clearly show that Coulombic attraction force across the very small interfacial air gap is the origin of both. An alternative possibility is to calculate the thickness of the air gap from knowledge of the measured capacitance, as implemented by Nakamura and Yamamoto (2017) or Shultz et al. (2018). In summary, all the above-mentioned works predict a thickness of the equivalent gap in the order of a few micrometers. Keeping this order of magnitude, the equivalent air gap as a function of the normal force is used in section Sliding Friction to fit the friction force resulting from the tangential contact model onto experimental data.

Application to Fingerpad in Contact With Capacitive Screen

In the following, the developed theoretical principles are applied to the normal contact between the fingerpad and the screen under electrovibration. Therefore, the solution of the corresponding non-adhesive contact is required, whereby “non-adhesive” means that the voltage is turned off. Adhesion due to van der Waals forces are excluded. One critical point must be discussed in advance. The compliance method is based on the principle of superposition, hence its application is restricted to linear elasticity. Human skin, however, shows non-linear material behavior and the contact between finger and screen is associated with large deformations. Although these non-linearities indicate a breakdown of superposition, Lin and Chen (2006) have demonstrated that under comparatively weak adhesion the compliance method can still be applied in the large-deformation regime assuming non-linear elastic materials. In this case, the adhesive part of the solution to be superposed must be understood as a linear perturbation of the non-linear non-adhesive one. Hence, the applicability of this so-called large-deformation JKR (LDJKR) theory is linked to the validity of the assumption:

$$\frac{F_N - F_1(A)}{\delta - \delta_1(A)} \approx \left. \frac{dF_1}{d\delta_1} \right|_A. \quad (26)$$

By using the finite element method, Lin and Chen (2006) studied two adhesive contact problems involving hyperelastic material: the contact between a hyperelastic hemisphere and a smooth rigid substrate as well as the contact between a smooth rigid spherical indenter and a hyperelastic half-space. In both cases they checked the accuracy of Equation (26) which actually coincided with the simulation results. Further applications of the LDJKR to Neo-hookean layers can be found in the work by Lin et al. (2008). Without further proof, but based on the above explanations, it is assumed that the compliance method can be applied to the contact between the fingerpad and the screen, too.

The solution of the non-adhesive normal contact between the index finger and screen is required as input for the compliance method. For this purpose, experimental results from accessible literature are used. We focus on the range of small normal forces between 0 and 2 N, which are relevant regarding electrovibration. Although some reported experimental data in literature vary significantly, there is general consensus that both the dependence of the apparent contact area A_0 and the ridge contact area A_R on the normal force can be fitted to power functions according to Equation (5). Given exponents corresponding to the apparent contact area in the low force regime range between 0.36 and 0.42 and those corresponding to the ridge contact area range between 0.42 and 0.58 (Warman and Ennos, 2009; van Kuilenburg et al., 2013a; Lin et al., 2015; Dzidek et al., 2017; Liu X. et al., 2018). Taking into account that the relationship between area and load depends on many things like the fingerpad inclination angle, measurement methods (ink printing or optical method), environmental conditions (temperature and humidity) and individual properties of the finger (influenced by age and gender of subjects) the differences in exponents are still relatively small. Soneda and Nakano (2010) determined slightly higher exponents from measurements with an optical method, 0.52 for the apparent and 0.68 for the ridge contact area, but they consider a different range of forces lying between 0.1 and 5 N. For further investigations, values from the work by Dzidek et al. (2017) are taken, since it also provides the required relationship between the contact area and indentation depth according to Equation (6).

Approach Based on Apparent Contact Area

The measurement results of Dzidek et al. (2017) stem from the left index finger of a 27-year-old female subject. With regard to the power functions according to Equations (5) and (6), at a finger inclination angle of 30 degrees relative to the smooth countersurface, the following parameters were determined:

$$m_0 = 0.37, \alpha_0 = 91.9 \text{ mm}^2 \text{N}^{-m_0}, n_0 = 1, \beta_0 = 64.4 \text{ mm}^{2-n_0}, \quad (27)$$

where the subscript “0” signifies parameters corresponding to the apparent contact area. After inserting these parameters into Equations (9) and (10), the corresponding solution of the electroadhesive contact is found. In **Figure 3**, the apparent contact area as a function of the normal force is plotted for

different values of the work of electroadhesion. Curves over the entire loading range are shown on the left, whereas the plot on the right gives an enlarged view of the pull-off region associated with negative loads. The solid black lines represent the original power-law fits to the experimental measurements by Dzidek et al. (2017). The other two solid lines illustrate the solutions under electroadhesion corresponding to different values of the work of adhesion, $w = 0.132 \text{ J/m}^2$ and $w = 0.532 \text{ J/m}^2$. Both are calculated from Equation (22) considering a voltage of 200 V but in one case a realistic equivalent air gap thickness of $1 \mu\text{m}$ is assumed whereas the other one takes into account unlikely complete contact, characterized by an disappearing equivalent air gap. At first glance, especially if one assumes a moderate air gap thickness, the change in the contact area when switching on the voltage appears so small that one tends to neglect it. However, it can be clearly seen from **Figure 3** on the right that this is not permitted for the range of very small normal forces (combined with higher voltages), which is definitely still of interest for electrovibration. The curve associated with $w = 0.132 \text{ J/m}^2$ predicts a pull-off force of 6.74 mN. In addition, the pure voltage-induced contact (no external normal force is applied) creates a contact area of 24.14 mm^2 . Unfortunately, to the best of the authors' knowledge, no experimental data are available in this interesting range.

The percentage change in the apparent contact area decreases with increasing normal force. At a normal force of 0.5 N, the change is still around 9 %. Nevertheless, in recent measurements by Sirin et al. (2019b) for characteristic normal forces of 0.5 N, 1 N, and 2 N, no significant difference was observed between the (initial) apparent contact areas with and without electrovibration. Thus, it can be assumed that switching on the voltage only results in an enlargement of the ridge contact area. This is the main reason why we decided to develop a model based on the ridge contact area instead of the apparent one.

Approach Based on Ridge Contact Area

Under the assumption that the LDJKR theory remains valid for applications to contact problems involving non-linear elastic human skin material, we were able to derive a solution for the electro-adhesive contact between the finger and the screen. It was tacitly assumed that the apparent contact area is compact and approximately circular. The applicability of the LDJKR theory to the ridge contact area requires an additional, very strict assumption. As previously mentioned, each contact area configuration under electroadhesion must be the same as that of the corresponding non-adhesive contact loaded by an appropriately chosen higher normal force (Ciavarella, 2018; Popov, 2018). This condition is definitely not fulfilled here! However, since more suitable simple methods do not exist, the compliance method is used once again to obtain a (rough) approximate solution.

In the following, we proceed in exactly the same way as in the last section. First, the parameters for the power functions are taken from the work by Dzidek et al. (2017).

$$m_R = 0.52, \alpha_R = 54.9 \text{ mm}^2 \text{N}^{-m_R}, n_R = 1.41, \beta_R = 33.3 \text{ mm}^{2-n_R}, \quad (28)$$

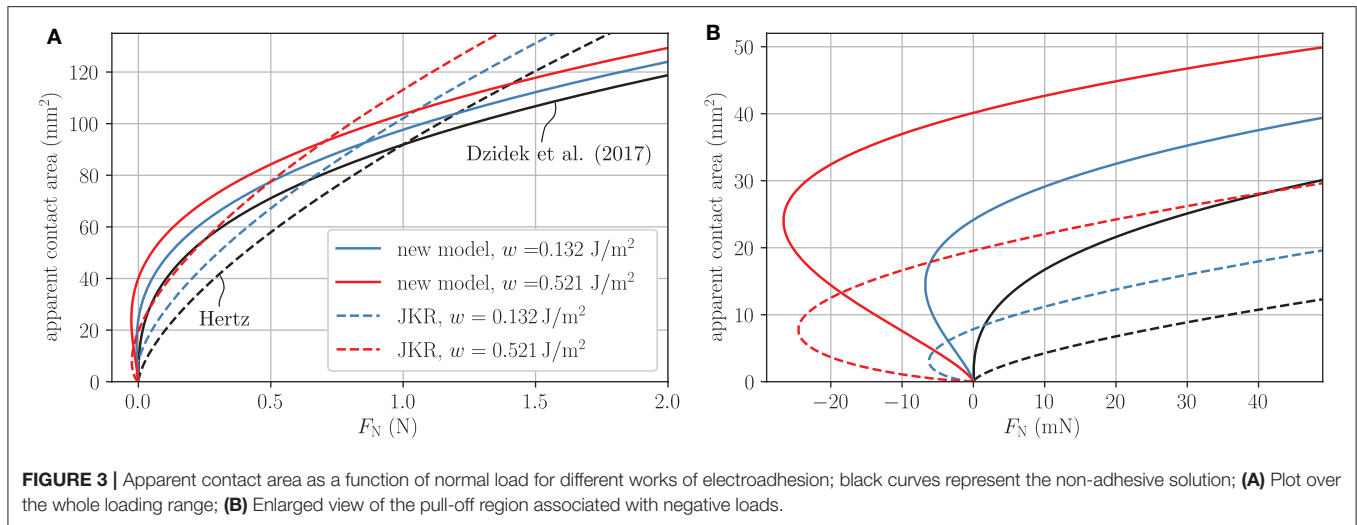


FIGURE 3 | Apparent contact area as a function of normal load for different works of electroadhesion; black curves represent the non-adhesive solution; **(A)** Plot over the whole loading range; **(B)** Enlarged view of the pull-off region associated with negative loads.

where the subscript “R” indicates parameters corresponding to the ridge contact area. Subsequently, these parameters are inserted into Equations (9) and (10) which yields the solution under electroadhesion. The ridge contact area as a function of the normal force is plotted in **Figure 4**, again distinguishing between the three characteristic cases: non-adhesive contact, contact under electroadhesion taking into account an equivalent airgap of 1 μm as well as complete contact associated with a vanishing equivalent air gap.

In the latter two electroadhesive cases, pull-off forces of 2.3 and 11.7 mN are obtained, as illustrated by **Figure 4B**. The change in the ridge contact area at a normal force of 0.5 N reads 11.2% for an air gap of one micron and 23.8% for complete contact.

As already stated, there appears to be no experimental data in the literature to evaluate the quality of our model and adjust the equivalent air gap thickness from it. In this context, however, we would like to point out that models based on Hertz or the original JKR theory assuming a parabolic profile and a constant equivalent elastic modulus cannot map the apparent or ridge contact area as a function of the normal force correctly. To show this, the apparent contact area for the cases of Hertz and original JKR [see Equation (13)] are included in **Figure 3**. The effective radius was estimated with $R^* \approx 1$ cm and an effective elastic modulus of $E^* \approx 47.4$ kPa is chosen such that the apparent contact areas for the non-adhesive normal contact agree at $F_N = 1$ N. The values for the work of (electro-)adhesion are chosen as for the new model. The Hertzian prediction differs significantly from the experimentally verified non-adhesive curve by Dzidek et al. (2017). Of course, these large differences are inherited in the JKR cases. At most, Hertz based models that take into account an equivalent effective elastic modulus that varies with the contact area (or indentation depth) would be qualified for modeling (van Kuilenburg et al., 2013a,b). Nevertheless, such models would still have to be suitably expanded to include the effect of electroadhesion.

TANGENTIAL CONTACT WITH ELECTROADHESION

A suitable, macroscopic model for mapping the electroadhesive frictional contact between the fingerpad and screen should fulfill three main characteristics:

- I. It should be based on a model for normal contact under electroadhesion.
- II. It must be able to reproduce the experimentally observed contact area reduction caused by the frictional force in the state of full slip.
- III. It requires a macroscopic approach for the frictional force.

Regardless of these specifications, the quality of a model can only be ensured by experimental verification of all relevant correlations. To meet the first point, the normal contact model based on the ridge contact area presented in section Application to Fingerpad in Contact with Capacitive Screen is applied. It is essentially described by Equation (9) with exponents $m_R = 0.52$ and $n_R = 1.41$ originating from the experiments by Dzidek et al. (2017). The occurring work of electroadhesion w is defined by Equation (22).

To satisfy requirement II, most current studies exploit adhesion theories based on fracture mechanics concepts. For this reason, section Tangential Contact under Full Stick assumption—Peeling starts with a discussion about peeling. However, by means of simulations with a two-dimensional FE model of the fingertip in section Transition from Stick to Slip—the Evolution of Contact Area in the Non-adhesive Case, it is shown, that the area reduction is mainly caused by large deformations of the non-linear elastic material around the distal phalanx. Therefore, in consistence with the FE results and available experimental investigations by Sahli et al. (2018) the area reduction is incorporated empirically [see Equation (43)].

On the defined macroscopic scale, a pressure-controlled friction law is assumed in section Sliding Friction. In connection with the specific parameter optimization, this leads to a good

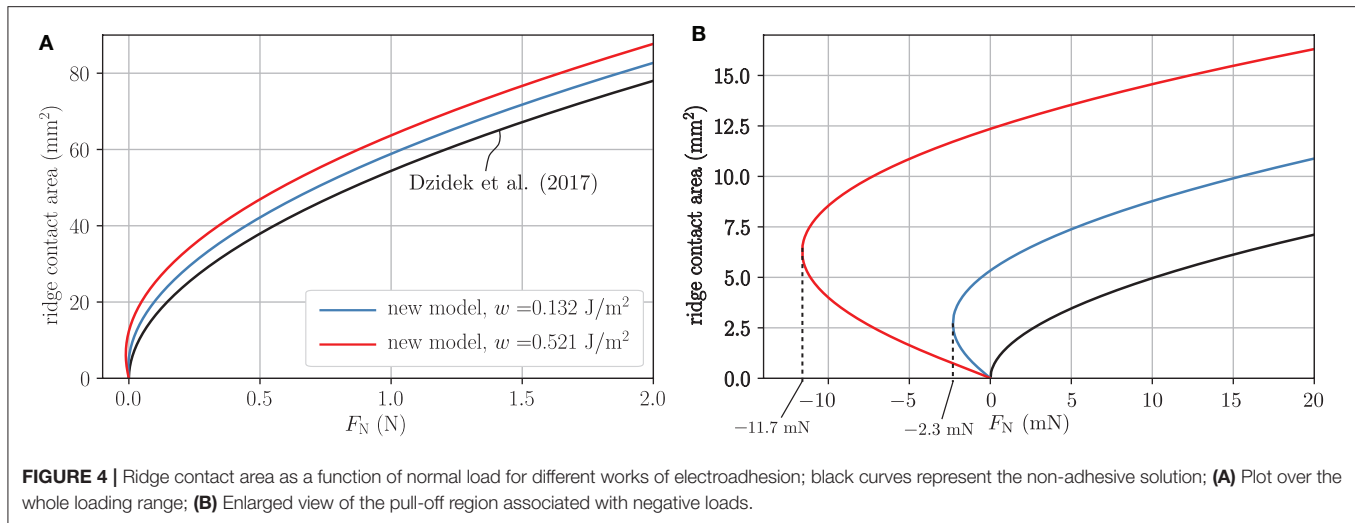


FIGURE 4 | Ridge contact area as a function of normal load for different works of electroadhesion; black curves represent the non-adhesive solution; **(A)** Plot over the whole loading range; **(B)** Enlarged view of the pull-off region associated with negative loads.

agreement of the friction force and the friction coefficient with experimental data. However, the incorporation of extended approaches consisting of both a pressure-based and an adhesion-based term is in principle possible.

Tangential Contact Under Full Stick Assumption—Peeling

Most current theoretical studies on tangential contacts of soft materials try to explain the experimentally observed contact area reduction by fracture mechanics concepts. They focus on the adhesive contact between elastically similar materials loaded by a small tangential force that does not cause any slip in the contact area. Under this full-stick condition all material points within the contact area undergo the same tangential displacement δ_T with respect to points remote from the contact area. In this case, the compliance method can easily be extended by including corresponding energetic terms to the elastic strain energy as well as the mechanical potential energy of the external load. This results in an energy release rate given by:

$$G = \frac{(F_1 - F_N)^2}{2S^2} \frac{dS}{dA} + \frac{F_T^2}{2S_T^2} \frac{dS_T}{dA} \quad \text{and} \quad (29)$$

$$G = \frac{(\delta_1 - \delta)^2}{2} \frac{dS}{dA} + \frac{\delta_T^2}{2} \frac{dS_T}{dA}$$

where $S_T := dF_T/d\delta_T$ denotes the tangential contact stiffness. After setting the energy release rate G equal to the work of (electro)adhesion w the general solution of the (no-slip) tangential contact under electroadhesion is obtained. Note, that the applicability of the extended compliance method is neither restricted to homogeneous materials nor to the half-space approximation or circular contact areas, but linear material behavior is required. In principle, the same requirements apply as for the theory of pure normal contact with adhesion. In this context, Equation (29) represents a novelty to the best of the authors' knowledge. Unfortunately, its applicability to non-linear elastic materials is constrained by very small tangential forces.

Since the tangential stiffness of the contact does not emerge from any literature, a direct extension of the promising results for the normal contact with adhesion from section Application to Fingerpad in Contact With Capacitive Screen is not readily possible. In order to discuss the influence of the tangential force on the apparent contact area between the finger and the screen, the (homogenized) Hertzian-based contact model is used as a rough approximation instead. In this case the normal and tangential contact stiffness read (Popov et al., 2019).

$$S(A_0) = 2E^* \sqrt{\frac{A_0}{\pi}}, \quad S_T(A_0) = 2G^* \sqrt{\frac{A_0}{\pi}} \quad \text{with} \quad (30)$$

$$\frac{1}{G^*} := \frac{2 - \nu_1}{4G_1} + \frac{2 - \nu_2}{4G_2}.$$

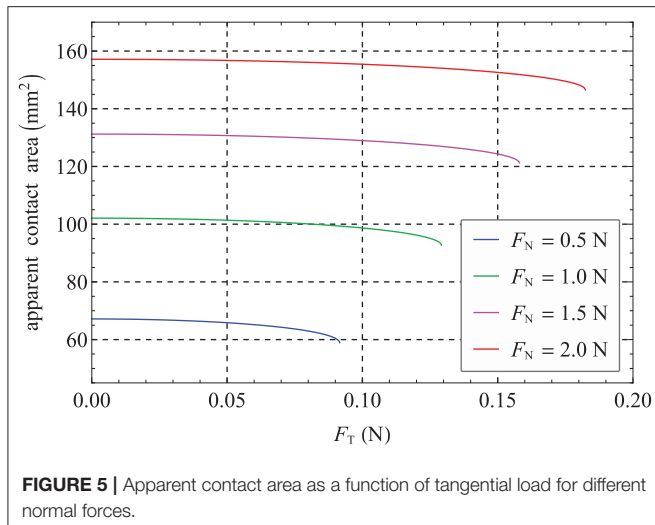
By substituting the results of Equation (30) into Equation (29) and taking into account $G = w$, the apparent contact area as a function of the applied external forces is determined after some rearranging

$$A_0(F_T, F_N) = \pi \left[\frac{3R^*}{4E^*} \left(F_N + 3\pi R^* w + \sqrt{6\pi R^* w F_N + 9\pi^2 R^{*2} w^2 - \frac{E^*}{G^*} F_T^2} \right) \right]^{2/3} \quad (31)$$

According to Equation (31), the contact area decreases with increasing tangential force. Since the tangential force causes a mode I separation, this effect is termed “peeling.” Stable peeling proceeds until the critical tangential force $F_{T,c}$ is reached:

$$F_{T,c} = \sqrt{\frac{G^*}{E^*} (6\pi R^* w F_N + 9\pi^2 R^{*2} w^2)}. \quad (32)$$

Note, that the expressions given in Equations (31) and (32) are equivalent to results of the pioneering work by Savkoor and Briggs (1977). As a rough estimate, the effective values for the radius, $R^* \approx 1$ cm, and the elastic modulus, $E^* \approx 47.4$ kPa, that are already used in section Approach Based on Ridge Contact Area are again used and the work of electroadhesion is taken as w



$= 0.132 \text{ J/m}^2$ which corresponds to an equivalent air gap of one micron (see section Approach Based on Apparent Contact Area). Since the insulating layer of the screen is much stiffer than the soft finger material the effective elastic modulus is determined by the latter. In addition, incompressible skin material is assumed to meet the requirement of elastic similarity which prescribes $G^*/E^* = 2/3$. The apparent contact area as a function of the applied tangential force for different normal forces is illustrated in **Figure 5**. The values of the tangential forces at the end points of the curves correspond to the critical tangential forces according to Equation (32).

The reduction is clearly too small compared to experimental results by Sirin et al. (2019b). The theory also significantly underpredicts the maximum tangential load, i.e., the reduction of the contact area cannot be explained satisfactorily by means of this simplified theory. It could be remedied by appropriate modeling of the transition regime from peeling to complete sliding. Peeling itself represents only the initial stage of static friction. However, the transition regime itself has not been sufficiently understood until today and is beyond of the scope of this work. Instead, we would like to briefly address another option. In the model by Savkoor and Briggs (1977), the work of adhesion is assumed to be constant throughout the loading. However, from experimental results it is known, that the work of adhesion can significantly increase under combined normal and tangential loading. In this case, the constant work of adhesion for pure mode I loading w must be replaced by a mode-mixity-dependent \tilde{w} :

$$w \rightarrow \tilde{w} = w \cdot f(\psi) \quad \text{with} \\ \psi := \arctan\left(\frac{K_{II}}{K_I}\right) = \arctan\left(\frac{F_T}{F_I - F_N}\right), \quad (33)$$

where K_I and K_{II} denote the mode I and mode II stress intensity factors. It is common to use one of the three phenomenological functions for interface cracks proposed by Hutchinson and Suo (1991) as normalized interfacial toughness function $f(\psi)$. They

include only one empirical parameter which can be used to fit onto measurement results (Johnson, 1996; Waters and Guduru, 2010; Papangelo and Ciavarella, 2019). The same procedure could now be followed to improve the adaptation of the model to the measurement data concerning the apparent area reduction in the electroadhesive tangential contact between the finger and the screen. However, this should be deliberately avoided because the main cause for the reduction of the contact area when applying a tangential force is *not adhesion*! In the next section, this claim is supported by a simplified FEM calculation for a non-adhesive, tangential contact between a finger and smooth rigid plate, taking into account non-linear elastic material behavior. Note, that very recently a similar finding concerning a smooth contact between a soft cylindrical cap and a rigid plate was made by Mergel et al. (2020). They have shown that the contact length decreased under tangential shear even in the absence of adhesion.

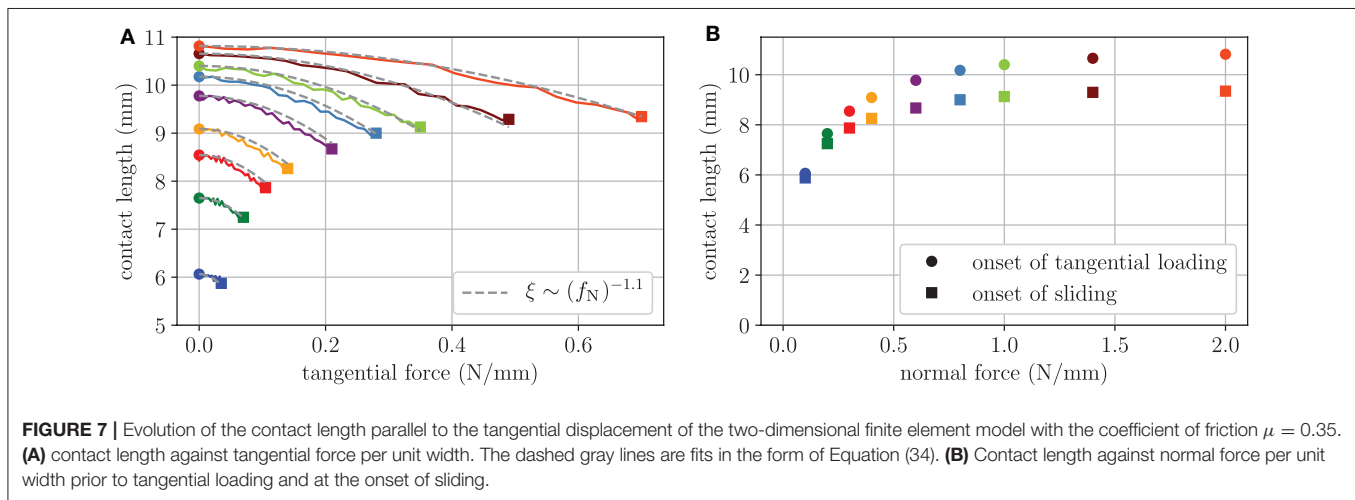
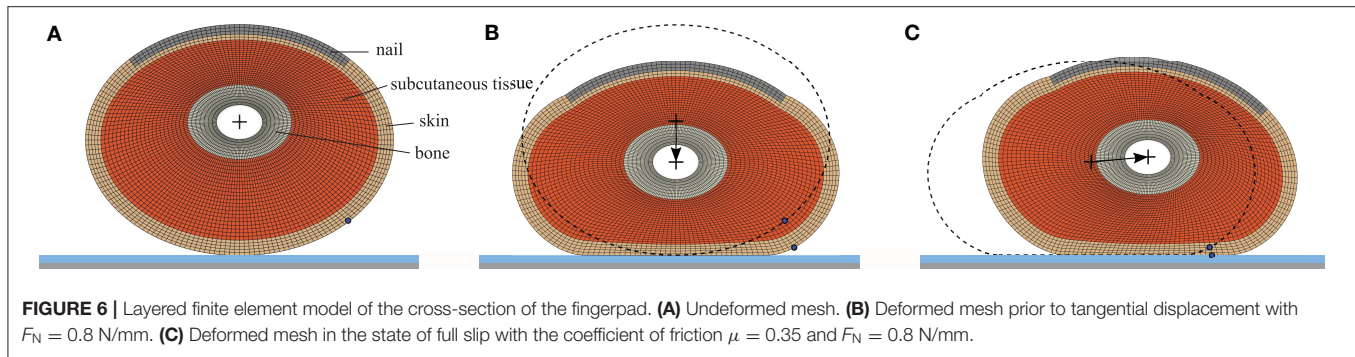
Transition From Stick to Slip—the Evolution of Contact Area in the Non-adhesive Case

Several experimental studies report a significant decrease of apparent contact area when a tangential force is applied to the fingerpad under constant normal loading (Delhay et al., 2014; Sirin et al., 2019b). This area reduction is already significant without electroadhesion, but is further increased under the influence of electroadhesion (Sirin et al., 2019b). Contrary to the peeling mechanism described in the previous section, this much larger reduction is accompanied by the development of local slip in the contact zone. It is further shown that the contact area does not shrink uniformly, but that primarily the length parallel to the tangential loading is reduced. Sahli et al. (2018) studied the apparent contact area as well as the real contact area and found that both reduce with the same reduction mechanisms. The investigated real contact area is of the order of magnitude of the ridge contact area discussed in section Approach Based on Ridge Contact Area.

As already stressed in the previous section, we believe that adhesion is not the decisive factor for the area reduction. It is known from experiments that the measured macroscopic adhesion for the contact of skin and dry glass is negligibly small (Wang et al., 2020) and cannot explain the significant reduction without any influence of electroadhesive forces.

Delhay et al. (2014) make the non-linear elastic properties of skin responsible for the area reduction. The originally coiled collagen fibers in the skin become oriented and straightened in the direction of stress. This results in a significant stiffening under a tangential force. However, we believe that the large deformation of the finger and the complex layered structure contribute to the observed effect as well. Without further investigation, this assumption was already expressed by Wang et al. (2020).

We conducted a simple two-dimensional plane strain finite element study using ABAQUS to study the origin of the area reduction further. Following Wu et al. (2006), a cross-section of the finger depicted in **Figure 6A** is modeled with layers for skin, subcutaneous tissue, bone and nail in *non-adhesive contact* with a rigid smooth plate. The friction properties of the contact are modeled by the local form of the Amontons-Coulomb law.



Both, geometrical and material parameters, were taken from Wu et al. (2006). Skin and subcutaneous tissue are modeled as hyperelastic materials using the Ogden model (Ogden, 1973). In contrast to commonly used hyperelastic models such as the Mooney-Rivlin or the Neo-Hookean model, the Ogden model may accurately describe the skin stiffening behavior (Shergold et al., 2006).

In a first step, the modeled cross-section is pressed against the rigid plate by applying a normal force per unit width to the inner bone layer (see **Figure 6B**). Then, the cross-section is moved tangentially while keeping the applied normal force constant until the whole contact length is slipping (see **Figure 6C**). At the onset of the tangential displacement, the skin layer behaves similar to a fluid filled membrane. It is generally much stiffer than the subcutaneous tissue and begins to “roll” around the bone (notice the marked node in **Figure 6** that moves into contact). As the finger is further displaced, the soft tissues begin to stiffen and slip propagates from the trailing edge. At the onset of full slip (**Figure 6C**), the geometry has drastically changed due to the large deformations and the contact length is decreased significantly.

The evolution of the contact length from the onset of tangential loading to the onset of full slip is shown in **Figure 7**. The total reduction of contact length is small for low normal forces and increases for higher normal forces. Owing to the finite geometry, the contact length and the observed reduction

are not increasing at the same rate for higher normal forces. The contact length during tangential loading can be described by a simple quadratic function of the tangential force per unit width,

$$L(f_T) = L_0 - \xi f_T^2, \quad (34)$$

where L_0 is the initial contact length prior to tangential loading. Fits of this form are included as dashed gray lines in **Figure 7A**, where ξ is a function of the normal force per unit width with $\xi \sim (f_N)^{-1.1}$.

The conducted finite element analysis supports the assumption that the observed contact area reduction under tangential loading is not caused by adhesion, but a result of large deformation and strain stiffening behavior of skin and subcutaneous tissue. However, due to the plane strain assumption and the negligence of fingerpad ridges, it is difficult to compare the results to experimental data quantitatively. Collecting the tissue layers of the epidermis and dermis in a homogeneous “skin” layer with smeared properties is a further limitation. Thus, in the present form, the model can only be used as a rough estimate. Nevertheless, similar to Equation (34) for the contact length and in accordance with Sahli et al. (2018) an empirical formula for the reduction of the ridge contact area is employed in section Sliding Friction.

Sliding Friction

Now the question arises which friction law is suitable for mapping the sliding contact between the finger and the screen under electroadhesion on the defined macroscopic scale. For this purpose the well-known approach of Bowden and Tabor is chosen as a starting point:

$$F_T = \tau \cdot A_{\text{real}}, \quad (35)$$

where τ denotes the interfacial shear strength and A_{real} is the “real” contact area. Usually, for many polymeric materials as well as human skin, the interfacial shear strength itself is composed of a constant intrinsic term τ_0 and a second term that is linearly dependent on the mean contact pressure p (Briscoe and Tabor, 1975; Adams et al., 2007).

$$\tau = \tau_0 + \tilde{\alpha}p, \quad (36)$$

where $\tilde{\alpha}$ is constant. Substituting Equation (36) in Equation (35) yields:

$$F_T = \tau_0 A_{\text{real}} + \tilde{\alpha} F_N. \quad (37)$$

Depending on whether the friction force according to Equation (37) is dominated by the first or second term, the friction is called adhesion-controlled or pressure-controlled, respectively. Note, that in this context adhesion refers to the rupture of interfacial junctions and has nothing to do with our electroadhesion. In order to include electroadhesion in the approach above, it is assumed that electroadhesion only contributes to the normal contact pressure associated with an increased contact area. Then, the following extension replaces Equation (36):

$$\tau = \tau_0 + \tilde{\alpha} (p + \sigma_{\text{el}}), \quad (38)$$

and from Equation (35) the friction force results in:

$$F_T = \tau_0 A_{\text{real}} + \tilde{\alpha} (F_N + \sigma_{\text{el}} \cdot A_{\text{real}}) = \tau_0 A_{\text{real}} + \tilde{\alpha} (F_N + F_{\text{el}}). \quad (39)$$

From a microscopic or atomic point of view, the real contact area is made up of the sum of the micro-asperity contact areas and thus only a very small fraction of the apparent contact area. Associated with this small-scale multi-asperity contact model, values of the shear strength are typically in the order of a few Megapascal (Persson, 2018). From experimental results by Adams et al. (2007), the intrinsic interfacial shear strength between wet human skin and glass lies in the range of some Kilopascal which indicates that the assumed real contact area is of the same order as the ridge contact area. By using this information, the real contact area can be replaced by the ridge contact area:

$$F_T = \tau_0 A_R + \tilde{\alpha} (F_N + \sigma_{\text{el}} \cdot A_R) = \tau_0 A_R + \tilde{\alpha} (F_N + F_{\text{el}}). \quad (40)$$

From our macroscopic point of view, Equation (40) shall represent the central law of sliding friction under

electroadhesion. In the following, focus is only on pressure-controlled friction which might be a strong assumption. An investigation considering the complete approach in Equation (40) should be part of a future work. Under the pressure-controlled assumption, Equation (40) reduces to

$$F_T = \mu_0 (F_N + \sigma_{\text{el}} \cdot A_R) = \mu_0 (F_N + F_{\text{el}}), \quad (41)$$

where $\tilde{\alpha}$ has been replaced by the “real” friction coefficient μ_0 , as it was termed by Derjaguin (1934). Equation (41) represents the generalized Amontons-Coulomb law of friction, which is often used to model friction within problems of electrovibration (see e.g., Kaczmarek et al., 2006; Shultz et al., 2018; Heß and Popov, 2019). From Equation (41) the “apparent” friction coefficient reads:

$$\mu = \frac{F_T}{F_N} = \mu_0 \left(1 + \frac{F_{\text{el}}}{F_N} \right). \quad (42)$$

Note, that the electroadhesive force per unit area σ_{el} in the sliding friction law according to Equation (41) is given by Equation (21). For pure normal contact under electroadhesion, the ridge contact area (in the following renamed to $A_{R,0}$) is calculated from the model introduced in section Approach Based on Ridge Contact Area described by Equation (9) with exponents $m_R = 0.52$ and $n_R = 1.41$. For the sliding contact, however, we further must account for the area reduction described in section Transition From Stick to Slip—the Evolution of Contact Area in the Non-adhesive Case that depends on the tangential force. Similar to the empirical formula (34) that we found for the contact length of the finite element model, the empirical quadratic law proposed by Sahli et al. (2018) for the ridge contact area:

$$A_R(F_T) = A_{R,0} - \eta F_T^2 \quad (43)$$

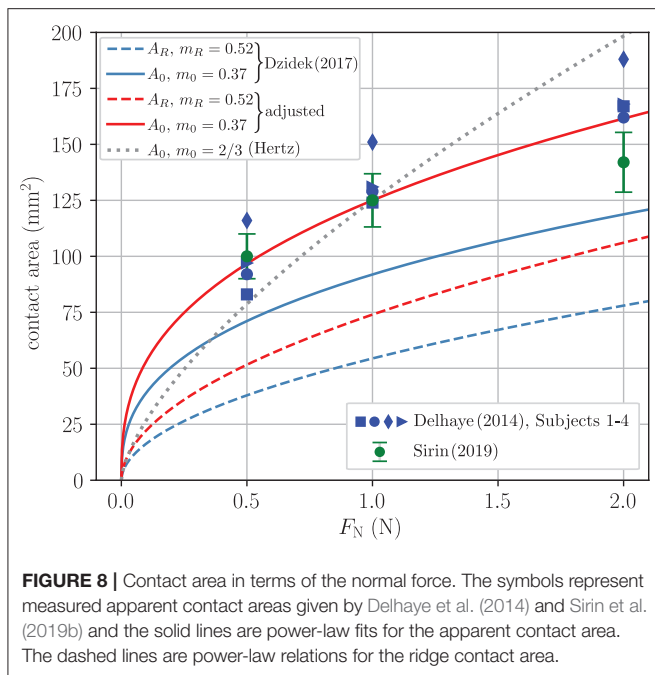
is employed, where $A_{R,0}$ is the ridge contact area prior to tangential loading and the parameter η is a function of the normal force of the form:

$$\eta = c_2 (A_{R,0})^{-1}, \quad (44)$$

where the constant c_2 is yet to be determined. Here, the exponent -1 is adopted, which is experimentally determined by Sahli et al. (2018) for the real contact area. It should be noted that in the current work of Basdogan et al. (2020) an analogous empirical approach was adopted for the apparent contact area, but η was assumed to be constant. Inserting Equation (43) into Equation (41) and solving for the tangential force yields:

$$F_T = - (2\mu_0\eta\sigma_{\text{el}})^{-1} + \sqrt{(2\mu_0\eta\sigma_{\text{el}})^{-2} + (\eta\sigma_{\text{el}})^{-1} F_N + \eta^{-1} A_{R,0}}. \quad (45)$$

In the following, the proposed model for sliding friction is verified by comparison to the experimental data presented by Sirin et al. (2019a). In this experimental study, the tangential force of a sliding fingerpad on a 3 M touchscreen was measured for different voltages applied to the conducting layer of



the touchscreen. The experimenter was trained to keep the normal force at a constant value during the measurements. Unfortunately, neither apparent nor ridge contact area were measured during this series of experiments.

In order to obtain adequate relations for the non-adhesive ridge area in the form of the power-law functions (5, 6), we again make use of the functions proposed by Dzidek et al. (2017) that are fitted to measurements of the finger of a female experimenter. In **Figure 8**, Dzidek's functions are compared to the measured apparent contact areas of four different subjects by Delhaye et al. (2014) and the averaged measurements by Sirin et al. (2019b). Obviously, Dzidek's original relations underestimate the apparent contact area of the other studies significantly. Under the assumption that the exponents remain the same, a scaling of the power-law functions for the apparent contact area as well as the ridge area was performed to obtain a better agreement with the experimental results. The exponents of the best fits for each subject vary considerably, but the exponent provided by Dzidek et al. represents a good compromise. The parameters of the adjusted relations for the non-adhesive ridge contact area are listed in **Table 1**.

It should be noted that many authors simply employ the Hertzian relation with the exponent $m = \frac{2}{3}$ to model the non-adhesive normal contact. A corresponding fit is included in **Figure 8**. As expected, the Hertzian relation is inadequate owing to the finite size and the complex layered structure of the finger.

With the non-adhesive ridge contact area, the ridge contact area for the electroadhesive contact prior to tangential loading $A_{R,0}$ can be determined by inversion of Equation (9). The work of adhesion w is given in Equation (22). The electromechanical parameters needed for the electroadhesive force per unit area σ_{el} and the work of adhesion are listed in **Table 1**. Most parameters

TABLE 1 | Parameters used for the model of sliding friction.

Symbol	Parameter name	Value and unit
μ_0	Friction coefficient	0.26
$\epsilon_{r,sc}$	Relative permittivity of stratum corneum	3,000
$\epsilon_{r,i}$	Relative permittivity of insulating layer	3.9
$\epsilon_{r,a}$	Relative permittivity of air	1
ϵ_0	Permittivity of free space	$8.854 \cdot 10^{-12} \text{ As/Vm}$
d_{sc}	Thickness of stratum corneum	250 μm
d_i	Thickness of insulating layer	1 μm
U	Applied voltage	0–200 V
F_N	Applied normal force	0–2.3 N
m_R, n_R	Exponents of power-law relationships (Equations 5, 6)	$m_R = 0.52$, $n_R = 1.41$
α_R, β_R	Factors of power-law relationships (Equations 5, 6)	$\alpha_R = 74 \text{ mm}^2/\text{N}^{m_R}$, $\beta_R = 43.5 \text{ mm}^{2-n_R}$
$d_{a,0}, c_1$	Empirical parameters for the air gap (Equation 46)	$d_{a,0} = 3.1 \mu\text{m}$, $c_1 = 0.5 \mu\text{m}/\text{N}$
c_2	Empirical parameter for the area reduction (Equation 44)	$c_2 = 5000 \text{ mm}^4/\text{N}^2$

are taken from Sirin et al. (2019a) and the thickness of the stratum corneum for the fingerpad in contact is taken from Lee et al. (2020).

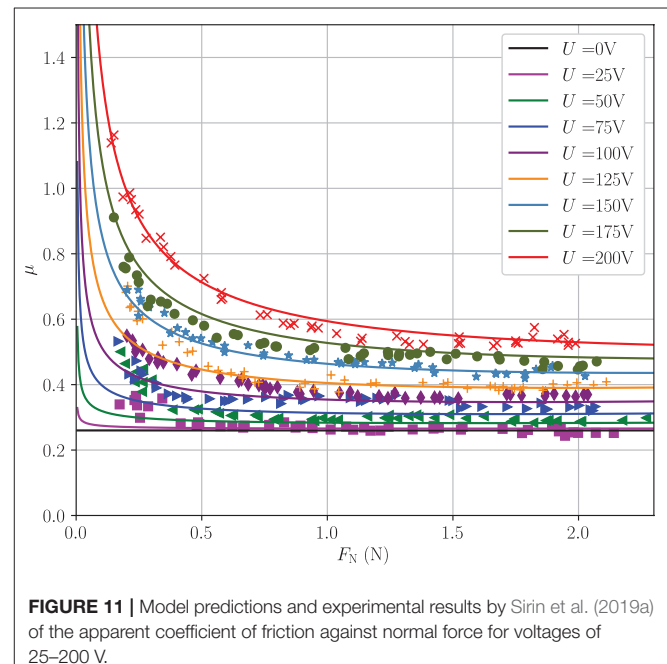
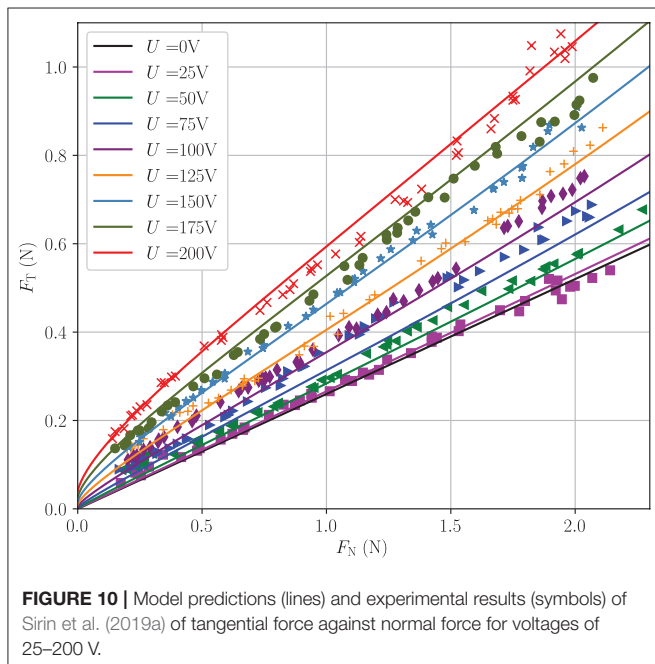
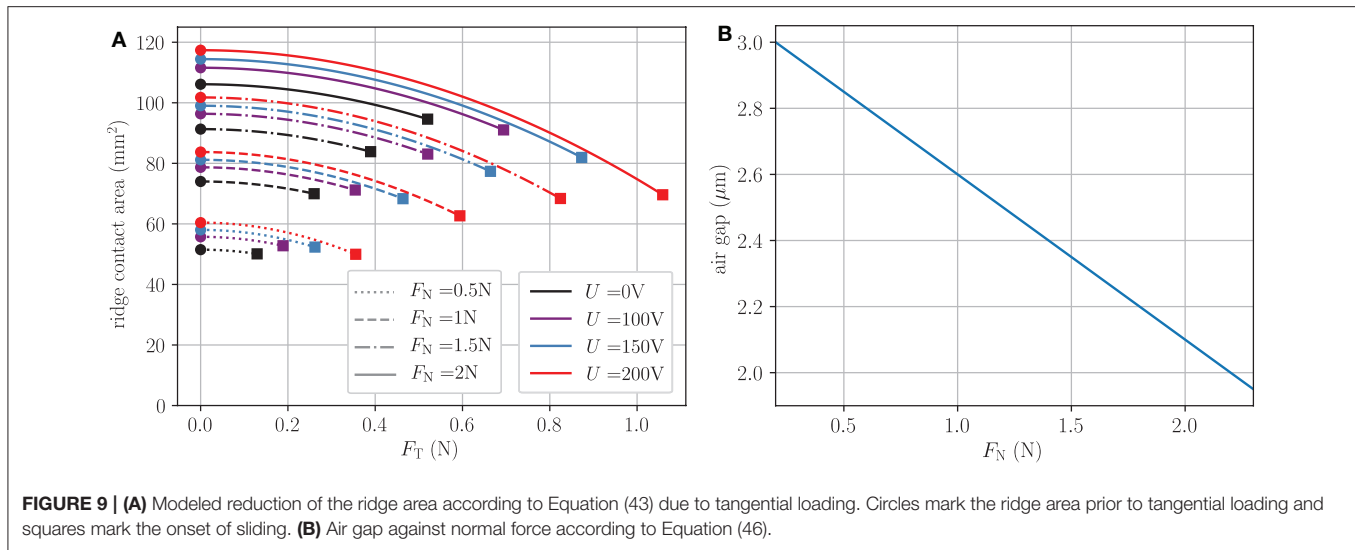
The concept of the equivalent air gap is discussed in section Theoretical Background. For the model of sliding friction, it is assumed that the thickness of the equivalent air gap d_a is a linearly decreasing function of the normal force,

$$d_a = d_{a,0} - c_1 F_N, \quad (46)$$

where the parameters $d_{a,0}$ and c_1 are yet to be determined.

The three unknown parameters for the area reduction and the air gap and in Equations (44) and (46), respectively, are found by fitting the model to the experimental results. However, the curve fitting is closely confined by the requirement that both, the area reduction and the air gap, remain within realistic ranges known from experiments.

The fitted parameters are included in **Table 1** and the ridge contact area reduction as well as the air gap thickness are shown in **Figures 9A,B**, respectively. Prior to tangential loading, the ridge contact area at constant normal loading increases up to 20% due to the electroadhesion. At the onset of sliding, the ridge contact area is reduced significantly for high initial contact areas and only marginally for smaller initial contact areas. Furthermore, the reduction is increased significantly for high voltages. Sirin et al. (2019b) investigated the reduction of apparent contact area without electroadhesion and with electroadhesion at 100 V applied voltage. They report an average reduction of 8% without and 13% with electrovibration at an applied normal force of 1 N as well as 15 and 20%, respectively, at 2 N. Our model predicts a similar ridge area reduction of approximately 5% without electrovibration and 10% with 100 V at 1 N as well as 12 and 19% at 2 N. It is valid to compare these results, because the apparent contact



area and the real contact area follow analogous reduction mechanisms and the relative reduction is very similar for both (Sahli et al., 2018). It should further be noted that Sirin et al. (2019b) measured a higher coefficient of friction presumably due to the oil they needed for image analysis. Thus, we expect a slightly increased relative area reduction for this study.

The thickness of the equivalent airgap shown in **Figure 9B** reduces from $2.85 \mu\text{m}$ at 0.5 N to $2.1 \mu\text{m}$ at 2 N . The thickness is thus in the range reported by other studies (see also Section Concept of an equivalent air gap) and agrees with $<0.5 \mu\text{m}$ difference especially well with the predictions by Guo et al. (2019).

Figure 10 shows the tangential force of the sliding fingerpad in terms of the applied normal force. The symbols represent the measurements by Sirin et al. (2019a) and the solid lines the results of the model described above. Especially for applied voltages above 50 V , the tangential force is increased significantly in the whole range of applied normal forces. Apart from small differences for low voltages, the model is in very good agreement with the experimental results. The apparent coefficient of friction (COF) $\mu = F_T/F_N$ depicted in **Figure 11** reaches a constant value for normal forces larger than 1 N . In this range, the apparent COF increases almost linearly with the applied voltage and is more than doubled for 200 V . Again, the model is in very good agreement.

The apparent COF increases significantly with decreasing normal forces in the range of normal forces lower than 0.5 N. This behavior is characteristic for adhesive contacts. Here, the model predictions are lower than the experimental results for voltages smaller than 150 V. We believe these deviations can be at least partly explained by classical van der Waals adhesion: Sirin et al. (2019a) also measured the tangential force for a sliding fingerpad without electroadhesion and show that the curves for 25 and 0 V are not distinguishable. Thus, mainly classical adhesion or capillary bridges are responsible for the small increase in apparent COF for 25 V. However, we believe the effect of electroadhesion is predominant in most cases justifying the negligence of classical van der Waals adhesion. Other inaccuracies may be found in the power-law relations for the ridge contact area at low normal forces. In this range very little experimental data is available and contact area has a very steep slope making it prone for errors.

Sirin et al. (2019a) also describe a model of sliding friction based on multiscale contact mechanics which agrees reasonably well with the experimental results. In this model, the real contact area is described by microasperities of skin in contact and is several orders of magnitude smaller than the apparent contact area. However, our much simpler macroscopic modeling approach based on the ridge contact area and the concept of an equivalent air gap is equally suited and yields, in fact, even better agreement with the experimental results.

CONCLUSION AND DISCUSSION

A new macroscopic model for sliding friction of a fingerpad over a smooth surface under electroadhesion has been developed. One of the cornerstones of the modeling is the application of Shull's compliance method to large deformations. In this way, a solution was first obtained for pure normal contact under electroadhesion, which in particular includes the dependence of the ridge contact area on the normal force as well as the applied voltage. To account for the experimentally observed reduction of the ridge contact area during transition from stick to slip, an empirical formula is employed according to which the area reduction is proportional to the square of the applied tangential force. This is consistent with experiments by Sahli et al. (2018) as well as results of our finite element simulations. After incorporating the developed solution for the pure normal contact and the empirical formula of area reduction into a pressure-based approach for the frictional force, the new macroscopic model for voltage-induced friction has been found. A comparison with recent experimental data has shown that the model adequately predicts both the frictional force and the coefficient of friction over the entire range of relevant voltages and applied normal forces.

A further important outcome with regard to the reduction of the contact area results from our finite element study

with a two-dimensional model of the fingerpad, where skin and subcutaneous tissue were considered as hyperelastic materials using the Ogden model to accurately describe skin stiffening. Although the simulations are restricted to *non-adhesive* tangential contacts, they show a significant reduction of the contact area, which is mainly caused by large deformations of the non-linear elastic material around the distal phalanx. This result challenges numerous recent studies on tangential contacts of soft materials that attempt to describe the experimentally observed area reduction caused by the frictional force using adhesion theories based on fracture mechanics concepts. Obviously, adhesion plays only a minor role for the area reduction. Hence, completely new theoretical approaches are required to model this effect in a physically meaningful way.

One uncertainty in the proposed modeling concerns the calculated ridge contact area under normal loading with electroadhesion. Since the influence of electroadhesion on the ridge area has not yet been investigated, the corresponding results obtained are currently still purely theoretical. Appropriate measurements should be part of a future task. The experimental results could as well be used to quantify the equivalent air gap more precisely. In addition, contributions to adhesion due to van der Waals forces have so far been completely ignored in the model. The same applies to capillary forces, although it is known that they cause a further increase of the coefficient of friction at low normal loads under wet conditions (Morales-Hurtado et al., 2017).

Furthermore, it should be noted that the proposed model requires the solution of the non-adhesive normal contact a priori. Although it is irrelevant for the application of Shull's method, whether this solution originates from experiments (like here), a simulation with the finite element method or something similar, the claim should be to find a suitable theoretical model also for the ridge contact area as a function of the normal force.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to Prof. Dr. Cagatay Basdogan and his colleagues for many valuable discussions during the international conference on Contact Mechanics (Berlin, Oct. 2019) and for providing us with their experimental data.

REFERENCES

- Adams, M. J., Briscoe, B. J., and Johnson, S. A. (2007). Friction and lubrication of human skin. *Tribol. Lett.* 26, 239–253. doi: 10.1007/s11249-007-9206-0
- Argatov, I. I., and Borodich, F. M. (2020). A macro model for electroadhesive contact of a soft finger with a touchscreen. *IEEE Trans. Haptics* 13, 504–510. doi: 10.1109/toh.2020.2969628
- Ayyildiz, M., Scaraggi, M., Sirin, O., Basdogan, C., and Persson, B. N. J. (2018). Contact mechanics between the human finger and a touchscreen under electroadhesion. *Proc. Natl. Acad. Sci. U.S.A.* 115, 12668–12673. doi: 10.1073/pnas.1811750115
- Basdogan, C., Sormoli, M. R., and Sirin, O. (2020). Modeling sliding friction between human finger and touchscreen under electroadhesion. *IEEE Trans. Haptics* 13, 511–521. doi: 10.1109/toh.2020.2989221
- Briscoe, B. J., and Tabor, D. (1975). The effect of pressure on the frictional properties of polymers. *Wear* 34, 29–38. doi: 10.1016/0043-1648(75)90306-3
- Ciavarella, M. (2018). An approximate JKR solution for a general contact, including rough contacts. *J. Mech. Phys. Solids* 209–218. doi: 10.1016/j.jmps.2018.03.005
- Delhay, B., Lefèvre, P., and Thonnard, J. L. (2014). Dynamics of fingertip contact during the onset of tangential slip. *J. R. Soc. Interf.* 11:20140698. doi: 10.1098/rsif.2014.0698
- Derjaguin, B. (1934). Molekulartheorie der äußeren Reibung. *Zeitschrift für Physik* 88, 661–675. doi: 10.1007/BF01333114
- Derler, S., and Gerhardt, L. C. (2012). Tribology of skin: review and analysis of experimental results for the friction coefficient of human skin. *Tribol. Lett.* 45, 1–27. doi: 10.1007/s11249-011-9854-y
- Dzidek, B., Bocherreau, S., Johnson, S., Hayward, V., and Adams, M. (2016). “Frictional dynamics of finger pads are governed by four length-scales and two time-scales,” in *IEEE Haptics Symposium (HAPTICS)* (Philadelphia, PA), 161–166. doi: 10.1109/HAPTICS.2016.7463171
- Dzidek, B. M., Adams, M. J., Andrews, J. W., Zhang, Z., and Johnson, S. A. (2017). Contact mechanics of the human finger pad under compressive loads. *J. R. Soc. Interf.* 14:127. doi: 10.1098/rsif.2016.0935
- Giraud, F., Amberg, M., and Lemaire-Semail, B. (2013). “Merging two tactile stimulation principles: electrovibration and squeeze film effect.” in *2013 World Haptics Conference (WHC)*, 199–203. doi: 10.1109/WHC.2013.6548408
- Guo, X., Zhang, Y., Wang, D., Lu, L., Jiao, J., and Xu, W. (2019). The effect of applied normal force on the electrovibration. *IEEE Trans. Haptics* 12, 571–580. doi: 10.1109/TOH.2019.2897768
- Heß, M., and Popov, V. L. (2019). Voltage-induced friction with application to electrovibration. *Lubricants* 7:102. doi: 10.3390/lubricants7120102
- Hutchinson, J. W., and Suo, Z. (1991). Mixed mode cracking in layered materials. *Adv. Appl. Mech.* 29, 63–191. doi: 10.1016/S0065-2156(08)70164-9
- Johnson, K. L. (1996). Continuum mechanics modeling of adhesion and friction. *Langmuir* 12, 4510–4513. doi: 10.1021/la950889a
- Johnson, K. L., Kendall, K., and Roberts, A. D. (1971). Surface energy and the contact of elastic solids. *Proc. R. Soc. Lond. A* 324, 301–313. doi: 10.1098/rspa.1971.0141
- Kaczmarek, K. A., Nammi, K., Agarwal, A. K., Tyler, M. E., Haase, S. J., and Beebe, D. J. (2006). Polarity effect in electrovibration for tactile display. *IEEE Trans. Biomed. Eng.* 53, 2047–2054. doi: 10.1109/TBME.2006.881804
- Lee, Z. S., Maiti, R., Carré, M. J., and Lewis, R. (2020). Morphology of a human finger pad during sliding against a grooved plate: a pilot study. *Biotribology* 21:100114. doi: 10.1016/j.biotri.2019.100114
- Lin, H. T., Hong, T. F., and Li, W. L. (2015). Grip performance affected by water-induced wrinkling of fingers. *Tribol. Lett.* 58:38. doi: 10.1007/s11249-015-0515-4
- Lin, Y. Y., Chang, C. F., and Lee, W. T. (2008). Effects of thickness on the largely-deformed JKR (Johnson-Kendall-Roberts) test of soft elastic layers. *Int. J. Solids Struct.* 45, 2220–2232. doi: 10.1016/j.ijsolstr.2007.11.025
- Lin, Y. Y., and Chen, H. Y. (2006). Effect of large deformation and material nonlinearity on the JKR (Johnson-Kendall-Roberts) test of soft elastic materials. *J. Polymer Sci. Part B Polymer Phys.* 44, 2912–2922. doi: 10.1002/polb.20914
- Liu, G., Sun, X., Wang, D., Liu, Y., and Zhang, Y. (2018). Effect of electrostatic tactile feedback on accuracy and efficiency of pan gestures on touch screens. *IEEE Trans. Haptics* 11, 51–60. doi: 10.1109/toh.2017.2742514
- Liu, X., Carré, M. J., Zhang, Q., Lu, Z., Matcher, S. J., and Lewis, R. (2018). Measuring contact area in a sliding human finger-pad contact. *Skin Res. Technol.* 24, 31–44. doi: 10.1111/srt.12387
- Liu, X., Lu, Z., Lewis, R., Carré, M. J., and Matcher, S. J. (2013). Feasibility of using optical coherence tomography to study the influence of skin structure on finger friction. *Tribol. Int.* 63, 34–44. doi: 10.1016/j.triboint.2012.08.020
- Mergel, J. C., Scheibert, J., and Sauer, R. A. (2020). Contact with coupled adhesion and friction: computational framework, applications, and new insights. *arXiv [Preprint]*. arXiv:2001.06833.
- Meyer, D. J., Peshkin, M. A., and Colgate, J. E. (2013). “Fingertip friction modulation due to electrostatic attraction,” in *IEEE World Haptics Conference (WHC)* (Daejeon), 43–48. doi: 10.1109/WHC.2013.6548382
- Morales-Hurtado, M., de Vries, E. G., Peppelman, M., Zeng, X., van Erp, P. E. J., and van der Heide, E. (2017). On the role of adhesive forces in the tribo-mechanical performance of ex vivo human skin. *Tribol. Int.* 107, 25–32. doi: 10.1016/j.triboint.2016.11.006
- Nakamura, T., and Yamamoto, A. (2016). A multi-user surface visuo-haptic display using electrostatic friction modulation and capacitive-type position sensing. *IEEE Trans. Haptics* 9, 311–322. doi: 10.1109/TOH.2016.2556660
- Nakamura, T., and Yamamoto, A. (2017). Modeling and control of electroadhesion force in DC voltage. *ROBOMECH J.* 4:18. doi: 10.1186/s40648-017-0085-3
- Ogden, R. W. (1973). Large deformation isotropic elasticity - on the correlation of theory and experiment for incompressible rubberlike solids. *Rubber Chem. Technol.* 46, 398–416. doi: 10.5254/1.3542910
- Papangelo, A., and Ciavarella, M. (2019). On mixed-mode fracture mechanics models for contact area reduction under shear load in soft materials. *J. Mech. Phys. Solids* 124, 159–171. doi: 10.1016/j.jmps.2018.10.011
- Persson, B. N. J. (2018). The dependency of adhesion and friction on electrostatic attraction. *J. Chem. Phys.* 148:144701. doi: 10.1063/1.5024038
- Popov, V. L. (2018). Solution of adhesive contact problem on the basis of the known solution for non-adhesive one. *Facta Univ. Series Mech. Eng.* 16, 93–98. doi: 10.22190/FUME180105009P
- Popov, V. L., and Heß, M. (2018). Voltage induced friction in a contact of a finger and a touchscreen with a thin dielectric coating. *arXiv [Preprint]*. arXiv:1805.08714.
- Popov, V. L., Heß, M., and Willert, E. (2019). *Handbook of Contact Mechanics, Exact Solutions of Axisymmetric Contact Problems*. Berlin Heidelberg: Springer. doi: 10.1007/978-3-662-58709-6
- Radivojevic, Z., Beecher, P., Bower, C., Cotton, D., Haque, S., Andrew, P., et al. (2012). Programmable electrostatic surface for tactile perceptions. *SID Symp. Digest Tech. Papers* 43, 407–410. doi: 10.1002/j.2168-0159.2012.tb05802.x
- Sahli, R., Pallares, G., Ducottet, C., Ben Ali, I. E., Al Akhrass, S., Guibert, M., et al. (2018). Evolution of real contact area under shear and the value of static friction of soft materials. *Proc. Natl. Acad. Sci. U.S.A.* 115, 471–476. doi: 10.1073/pnas.1706434115
- Savkoor, A. R., and Briggs, G. A. D. (1977). Effect of tangential force on the contact of elastic solids in adhesion. *Proc. R. Soc. London Ser. A* 356, 103–114. doi: 10.1098/rspa.1977.0123
- Shergold, O. A., Fleck, N. A., and Radford, D. (2006). The uniaxial stress versus strain response of pig skin and silicone rubber at low and high strain rates. *Int. J. Impact Eng.* 32, 1384–1402. doi: 10.1016/j.ijimpeng.2004.11.010
- Shull, K. R. (2002). Contact mechanics and the adhesion of soft solids. *Mater. Sci. Eng. R Rep.* 36, 1–45. doi: 10.1016/S0927-796X(01)00039-0
- Shull, K. R., Ahn, D., Chen, W., Flanagan, C. M., and Crosby, A. J. (1998). Axisymmetric adhesion tests of soft materials. *Macromol. Chem. Phys.* 199, 489–511. doi: 10.1002/(SICI)1521-3935(19980401)199:4<489::AID-MACP489>3.0.CO;2-A
- Shultz, C., Peshkin, M., and Colgate, J. E. (2018). The application of tactile, audible, and ultrasonic forces to human fingertips using broadband electroadhesion. *IEEE Trans. Haptics* 11, 279–290. doi: 10.1109/TOH.2018.2793867
- Shultz, C. D., Peshkin, M. A., and Colgate, J. E. (2015). “Surface haptics via electroadhesion: expanding electrovibration with Johnsen and Rahbek,” in *IEEE World Haptics Conference (WHC)* (Evanston, IL), 57–62. doi: 10.1109/WHC.2015.7177691

- Sirin, O., Ayyildiz, M., Persson, B. N. J., and Basdogan, C. (2019a). Electroadhesion with application to touchscreens. *Soft Matter*. 15, 1758–1775. doi: 10.1039/C8SM02420K
- Sirin, O., Barrea, A., Lefèvre, P., Thonnard, J. L., and Basdogan, C. (2019b). Fingerpad contact evolution under electrovibration. *J. R. Soc. Interf.* 16:156. doi: 10.1098/rsif.2019.0166
- Soneda, T., and Nakano, K. (2010). Investigation of vibrotactile sensation of human fingerpads by observation of contact zones. *Tribol. Int.* 43, 210–217. doi: 10.1016/j.triboint.2009.05.016
- Strong, R. M., and Troxel, D. E. (1970). An electrotactile display. *IEEE Trans. Man Machine Syst.* 11, 72–79. doi: 10.1109/TMMS.1970.299965
- van Kuilenburg, J., Masen, M., van Der Heide, A. E. (2013a). The role of the skin microrelief in the contact behaviour of human skin: contact between the human finger and regular surface textures. *Tribol. Int.* 65, 81–90. doi: 10.1016/j.triboint.2012.11.024
- van Kuilenburg, J., Masen, M., and van der Heide, A. E. (2013b). Contact modelling of human skin: What value to use for the modulus of elasticity? *Proc. Inst. Mech. Eng. Part J. J. Eng. Tribol.* 227, 349–361. doi: 10.1177/1350650112463307
- van Kuilenburg, J., Masen, M., and van der Heide, A. E. (2015). A review of fingerpad contact mechanics and friction and how this affects tactile perception. *Proc. Inst. Mech. Eng. Part J. J. Eng. Tribol.* 229, 243–258. doi: 10.1177/1350650113504908
- Vardar, Y., Güçlü, B., and Basdogan, C. (2017). Effect of waveform on tactile perception by electrovibration displayed on touch screens. *IEEE Trans. Haptics* 10, 488–499. doi: 10.1109/TOH.2017.2704603
- Vezzoli, E., Amberg, M., Giraud, F., and Lemaire-Semail, B. (2014). “Electrovibration modeling analysis,” in *Haptics: Neuroscience, Devices, Modeling, and Applications. EuroHaptics 2014. Lecture Notes in Computer Science*, Vol. 8619, eds M. Auvray, and C. Duriez (Berlin; Heidelberg: Springer). doi: 10.1007/978-3-662-44196-1_45
- Vodlak, T., Vidrih, Z., Vezzoli, E., Lemaire-Semail, B., and Peric, D. (2016). Multi-physics modelling and experimental validation of electrovibration based haptic devices. *Biotribology* 8, 12–25. doi: 10.1016/j.biotri.2016.09.001
- Wang, J., Tiwari, A., Sivebaek, I., and Persson, B. N. J. (2020). Sphere and cylinder contact mechanics during slip. *arXiv [Preprint]*. arXiv:2002.02226. doi: 10.1016/j.jmps.2020.104094
- Warman, P. H., and Ennos, A. R. (2009). Fingerprints are unlikely to increase the friction of primate fingerpads. *J. Exp. Biol.* 212, 2016–2022. doi: 10.1242/jeb.028977
- Waters, J. F., and Guduru, P. R. (2010). Mode-mixity-dependent adhesive contact of a sphere on a plane surface. *Proc. R. Soc. A Math. Phys. Eng. Sci.* 466, 1303–1325. doi: 10.1098/rspa.2009.0461
- Wu, J. Z., Dong, R. G., and Welcome, D. E. (2006). Analysis of the point mechanical impedance of fingerpad in vibration. *Med. Eng. Phys.* 28, 816–826. doi: 10.1016/j.medengphy.2005.11.013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Heß and Forsbach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Milestones in Natural Lubrication of Synovial Joints

Alessandro Ruggiero*

Department of Industrial Engineering, University of Salerno, Salerno, Italy

For years, the physical and mathematical modeling of tribological phenomena in natural human synovial joints has been a great challenge for researchers and scientists. This short review aims to summarize the main findings on lubrication modeling of human synovial joints over the years, starting from first insights and then underlining the evolution of the proposed theories. The complexity of natural human synovial joints, in which biological, fluid dynamic, and tribological phenomena takes place, makes this research area fascinating for scientists from several investigation fields. This manuscript underlines the necessity of deep scientific cooperation between researchers from different branches of the involved disciplines to achieve complete knowledge of these tribo-systems by taking in to account different points of view.

Keywords: biotribology, lubrication, joint, synovial fluid, cartilage

OPEN ACCESS

Edited by:

Irina Goryacheva,
Institute for Problems in Mechanics
(RAS), Russia

Reviewed by:

Vladimir Ivanovich Pakhaliuk,
Sevastopol State University, Russia
Valentin L. Popov,
Technical University of
Berlin, Germany

*Correspondence:

Alessandro Ruggiero
ruggiero@unisa.it

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 08 May 2020

Accepted: 08 June 2020

Published: 31 July 2020

Citation:

Ruggiero A (2020) Milestones in
Natural Lubrication of Synovial Joints.
Front. Mech. Eng. 6:52.
doi: 10.3389/fmech.2020.00052

INTRODUCTION

The framework of biotribology, the science devoted to the study of “those aspects of tribology concerned with biological systems” (Dowson, 2012), offers the possibility to understand and to model the complex lubrication phenomena acting in natural synovial joints. This field has attracted interest from researchers and scientists for years, and assumes particular significance this year, as Duncan Dowson passed away in January 2020 (Jin et al., 2020). Synovial joints, also called diarthrosis, are freely movable joints that possess a cavity bounded by a synovial membrane (Khan et al., 2007). From a kinetics point of view, synovial joints can be viewed as a sophisticated bio-bearing which allows for wide movements and supports high loads, of up to 10 times the body weight, in the presence of very low friction due to the biological lubricant, named synovial fluid. From an anatomical point of view, the constitution of a synovial joint is complex: the joined bones are covered by the hyaline articular cartilage which plays a key role in load supporting and joint lubricating. These are lined by a fibrous capsule, which also provides stabilization to the joint and is covered by a synovial membrane devoted to synovial fluid secretion. Setting aside the complexity of the biological phenomena acting in the synovial joints, the possibility of a complete understanding of their tribological performances, accounting for the possibility of a detailed mathematical description of the contact and lubrication phenomena, could be useful to the scientists involved in both the medical or pharmacological treatment of joint diseases and also in the optimal biomechanical and tribological design of artificial joints (Popov, 2019; Ruggiero and Zhang, 2020).

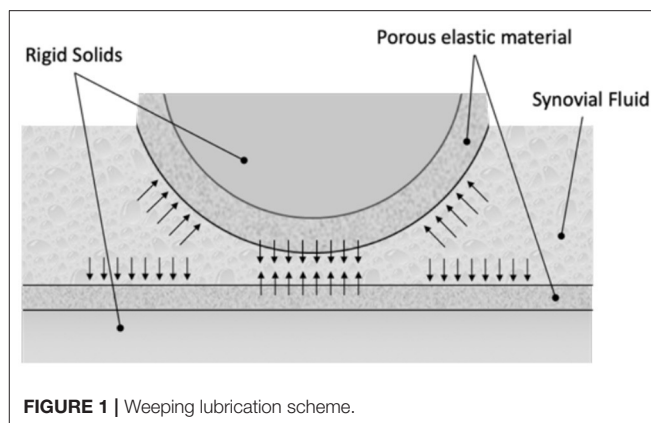
Even if this topic remains under investigation, this short review aimed to summarize and highlight the scientific progress made through the years in the understanding and modeling of lubrication mechanisms in natural human synovial joints, covering more popular and relevant theories from earlier research in this field up to the latest ones, introducing briefly some milestones from the 1930's to today.

LUBRICATION MODELS THROUGH THE YEARS

The first coherent theory about human joint lubrication dates to 1932, when MacConaill stated that human joint comportment was regulated by a hydrodynamic action (MacConaill, 1932). Before his studies, Walmsley (1928) recognized and investigated the similarities between articulating surfaces in human joints and the physical wedge shown by Reynolds to be necessary for hydrodynamic or fluid-film lubrication. Between 1932 and 1960, MacConaill (1932, 1956, 1960) suggested that the intra-articular cartilage (for example, the meniscus in the knee joint) assumes a slight inclination of the coupled surfaces in order to create wedge-shaped lubricant films analogous to those observed in Michell tilting pads. Four years after MacConaill's earliest paper, Jones (1936) concluded that the joints can be viewed as hydrodynamically lubricated systems. Moreover, he performed one of the first detailed experiments on friction in natural joints, observing a horse stifle mounted as the fulcrum in a pendulum machine. He noticed that if the joint was lubricated by synovial fluid, a very low coefficient of sliding friction $f = 0.02$ arose. Furthermore, an unlubricated joint evidenced an audible creaking noise after 8,000 cycles when a constant load of 445 N was applied. After 4 h, Jones noticed effects such as joint heating, steam rising, debris being thrown out, bone grating, etc. In his investigation, Jones highlighted the essential role of lubricant in natural joints. Moreover, according to a study on human finger joints dating back to 1936 (Jones, 1936), Jones observed that decay in amplitude of the pendulum swing was exponential. From this, he concluded that viscous damping is present in the joint and the mode of lubrication was fluid film. In 1959, Charnley (1959) measured, with reference to the human knee joints, the friction in lubricated conditions at a very low rubbing speed and found a value no larger than 0.02. Hydrodynamic theory was at that time the first concept chosen in the study of natural human joint lubrication.

Boundary Lubrication

It was not until 1959 that both Charnley (1959) and McCutchen (1962) started to question MacConaill's concept. According to Charnley, hydrodynamic action could not exist due to low sliding velocities under the heavy loads acting in the human joints. This conclusion was based on several pendulum tests with dissected ankle joints, in which a linear decay in amplitude was observed. Charnley imputed Jones' exponential decay to the lack of congruity in the joint and to the contribution, at high amplitudes, from the capsule and the ligaments which Jones left intact in his study on the finger joint. Charnley proposed, as an alternative to MacConaill's theory, a boundary lubrication action. From several experiments on articular cartilage, Charnley (1959) recorded friction coefficients values, finding values between $f = 0.005 \div 0.023$ in dissected ankle joints. In 1962, Barnett and Cobbald (1962) commented upon Charnley's theory. They proved that a linear decay in amplitude was attributable to the dissection of the joint. However, when they replaced the joint in the pendulum fulcrum with a hydrostatic bearing, it was found that the decay still showed an essentially linear relationship with



time. In 1967 and 1968, Linn (1967, 1968) and Linn and Radin (1968) performed experiments on animal joints at a constant load and the results suggested that an extraneous dynamic force component, generated by the eccentricity of the joint, had to be added to the friction force. Linn concluded that animal joints operate within the mixed film region of lubrication.

Weeping Lubrication

An in-between solution to the problem of joint lubrication was proposed in 1966 by McCutchen (1966). For the first time, he considered the porosity and elasticity of articular cartilage. McCutchen stated that, due to the load, the pressurized synovial fluid flows through the porous cartilage which behaves as sponge-like material, in a similar way to a self-pressurized hydrostatic bearing. The term weeping suggested that the lubricant film was sweated into the high-pressure region between opposing cartilages, while the boundary lubrication effect between the contact surfaces was still present. McCutchen used the term “weeping lubrication” because bearing materials which perform it weep liquid when compressed.

In several manuscripts (McCutchen, 1966, 1973, 1983) both weeping and boundary lubrication were discussed. A simple experiment was performed by using the sawn-off shoulder end of a pig which was pressed with a known force against a smooth glass surface, measuring the required force to make it slide. The friction of cartilage against glass increased over time and consequent squashing down of the cartilage was observed. If the cartilage was left in a fluid for 1 s, a lower friction force was observed for a brief period, while for a period of 10 s the cartilage swelled visibly and friction decrease was more marked, with the conclusion drawn that the cartilage is able to lose and regain water, like a sponge, through its fine pores. Friction is very low at the beginning and rises as the wringing out of the water permits the cartilage to be squeezed down (Figure 1).

Later, in 1994, Ateshian et al. (1994) and Ateshian and Hung (2006) proved that cartilage interstitial fluid plays a key role in the load support during the first 100–200 s after contact loading.

Elastohydrodynamic Lubrication

In 1963, a similar time to which weeping lubrication was hypothesized, Dintenfuss (1963) provided new insights on

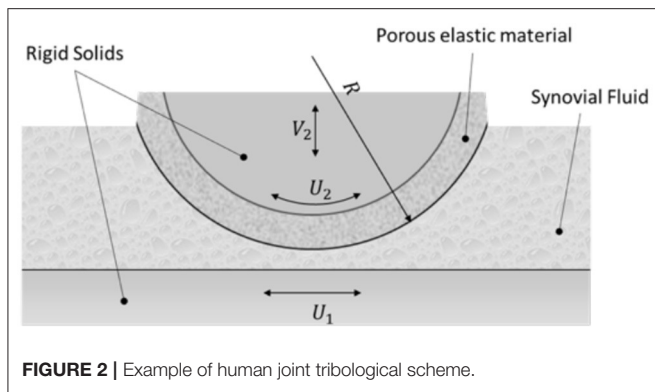


FIGURE 2 | Example of human joint tribological scheme.

human joint lubrication. Firstly, he showed the failure of the hydrodynamic lubrication mode, in which deformations were completely neglected. His studies took into account the deformability of articular cartilage and led to the elastohydrodynamic lubrication theory (EHD, see Popova and Popov, 2014). The main result was that in highly loaded lubricated contacts the operating film thickness may be up to 100 times greater than those predicted by conventional lubrication theory, since the human joint requires the correct operating conditions for the cartilage to be separated by a synovial film, and the boundary lubrication therefore does not exist. In 1966, Tanner (1966) calculated the film thickness in hip joints at normal walking speeds to be $h = 10\text{--}5\text{ cm}$. Even though the correct values of the loading pattern during gait were not yet available at the time, a mathematical incorporation of non-Newtonian lubricant properties was omitted. Tanner's theoretical values suggested that in the joints a lubricating film is possible when the lubricated surfaces are subject to relative motion. However, the question remained over what the lubricating mechanism is after a period of time under constant load with no motion (e.g., in human standing position).

Later, Medley et al. (1984), Dowson and Jin (1986), and Dowson (1995) confirmed EHD lubrication as the dominant mode of lubrication of many highly deformable or "soft" bearing systems, such as synovial joints.

Squeeze-Film Lubrication

In 1966, Dowson (1966) and Popova and Popov (2014) suggested a squeeze-film mechanism of lubrication. He based this assumption on the analysis concerning a loaded rigid cylinder approaching a rigid plate. A schematic representation of a generic model of a human joint is provided in Figure 2, in which a rigid plane opposes a rigid bone, covered by a layer of porous, elastic articular cartilage lubricated by synovial fluid. The considered relative velocities are both sliding and squeezing.

Dowson applied for the first time in tribology's history the main results of hydrodynamic and EHD theories in a simplified model for human joints.

The study demonstrated that EHD lubrication, caused by sliding and/or squeezing motions of the porous surfaces, seems to be the most common lubricating mechanism during usual

body movement, while classical hydrodynamic lubrication is inadequate in knee and hip joints.

Similar conclusions were obtained by Fein (1966), Higginson and Norman (1974), and Higginson (1977), who considered a pair of compliant surfaces, between which the lubricant was squeezed out. However, the calculations did not consider the non-Newtonian behavior of the synovial fluid.

In 1966, Fein (1966) investigated and supported the squeeze-film lubrication between contact lubricated surfaces and performed optical experiments to validate his theory. Excellent agreement between theoretical and experimental values of film thickness was found.

In 1992, Hou et al. (1992) performed an asymptotic analysis of a lubrication problem for a model of articular cartilage and synovial fluid under the squeeze-film conditions, while more recently in 2011 and in 2013, Ruggiero et al. (2011, 2013) proposed an original analytical approximate model for the synovial pressure field determination in the ankle joint in a pure squeeze motion, accounting for the non-Newtonian behavior of synovial fluid and porosity of the cartilage. In 2000, very interesting research was published by Hlaváček (2000, 2001) on the squeeze-film lubrication of the human ankle joint with synovial fluid filtrated by articular cartilage.

Boosted Lubrication, Ultrafiltration, and Hydration Lubrication

Maroudas' ultrafiltration theory (Maroudas, 1968; Maroudas et al., 1968), as well as the boosted lubrication theory from Walker et al. (1968, 1969) introduced the attractive idea that the lubricant is retained between the loaded surfaces due to some specific properties of both synovial fluid and the articular cartilage. In 1967–69, Maroudas removed, by ultrafiltration, the water from the synovial solution, obtaining a jelly-like gel; this prompted them to consider the joint cartilage as a "filter," permeable to water but not to the macro-molecules, allowing the gel formation under specific conditions. However, the stable thickness of the gel was estimated to be $h = 0.01\text{ }\mu\text{m}$, which is too small to provide separation between the surfaces. In 1968, Walker et al. (1968) noticed that during the loading the liquid component of the fluid becomes preferentially squeezed out, leaving the lubricant film enriched by the macromolecular components. In 1970, Dowson et al. (1970) observed that, due to the rising viscosity of the enriched synovial fluid, the squeezing times were greatly increased. Moreover, it was observed that macromolecules from the lubricant showed an affinity for the cartilage surface, allowing the formation of a skin-like protective gel. Investigations by Radin and Paul (1969), Radin et al. (1970), Swann and Radin (1972), Swann and Mintz (1979), and Swann et al. (1981) demonstrated the presence of an adsorption mechanism onto articular cartilage and identified the component in synovial fluid which was responsible for it. In 1975, new evidence on human joint lubrication was published by Unsworth et al. (1975). The authors studied friction coefficients in human joints using a pendulum machine. They examined six cadaveric human joints, one of which came from a patient affected by rheumatoid arthritis. They tested them both in dry conditions

and with synovial lubricant. Used loads were between 135 and 1,500 N. The initial amplitude of swing was 0.0785 rad. Authors increased loads by 220 N after each test cycle. The main result was that, in human joints during the walking cycle, squeeze phenomena can be found, which occur under high load and high squeeze velocity without sliding. Moreover, it was observed that elastohydrodynamic lubrication takes place when the sliding velocity is quite high, so the formation of an EHD fluid film is possible. The results from this research enabled a description of the phenomena which occurs during the walking cycle in human joints.

More recently, Raviv et al. (2001, 2003), Briscoe et al. (2006), and Klein (2013) introduced the novel concept of a hydration lubrication mechanism as a new framework for understanding boundary lubrication processes in aqueous media. Schmidt and Sah (2007) investigated the connection between synovial fluid and the articular cartilage by means of the boundary lubrication mode, while Greene et al. (2011) stated that the tribology of synovial joints needs to be investigated considering the synergistic effect of several modes of lubrication. Other recent investigations (Hui et al., 2012) were also devoted to the connections between synovial lubrication and the system biology.

THE RHEOLOGY OF SYNOVIAL FLUID

Synovial fluid's (SF) rheological properties play a key role in the lubrication modeling of the joints. SF contains the molecules hyaluronan, proteoglycan 4 (proteins also known as lubricin, superficial zone protein, and megakaryocyte-stimulating factor), and surface-active phospholipids, each of which interacts with and adsorbs to the articular surface (Schmidt et al., 2007).

The non-Newtonian character of the viscosity was proposed, among others, by Ropes et al. (1947). Ogston et al. (1950) related changes in viscosity to the variations of both the concentration and conformation of the hyaluronic acid molecules, and in 1953 Ogston and Stanier (1953) qualitatively stated that synovial fluid also possesses elastic properties.

Until the year 1966, research on rheological properties of synovial fluid did not make any noticeable progress. From 1967, new interesting results succeeded each other very quickly, probably due to the improvement of experimental equipment.

Through the years several non-Newtonian models were proposed (Lai et al., 1978). The most common are:

- **Power Law Model:** The so-called “power-law equation,” or Ostwald–de Waele relationship, was mentioned. It relates the viscosity μ to the shear rate $\dot{\gamma}$ in a steady shear flow:

$$\mu = K\dot{\gamma}^{n-1} = K\left(\frac{\partial u}{\partial y}\right)^{n-1}$$

with K and n representing two coefficients obtained by the process of curve fitting.

- **Generalized Newtonian fluid**

The shear-thinning fluid described by Ostwald–de Waele relationship is a type of generalized Newtonian fluid that, in

general, satisfies the rheological equation:

$$\tau = \mu(\dot{\gamma}) \dot{\gamma}$$

A generalized Newtonian fluid is an idealized fluid for which the shear stress is a function of shear rate at a particular time, but not dependent upon the history of deformation. For the power-law fluid, the rheological equation becomes:

$$\tau = K\dot{\gamma}^n = K\left(\frac{\partial u}{\partial y}\right)^n$$

- **Cross-WLF Model**

The empirical equation to describe the shear thinning behavior of synovial fluid gained wide acceptance in the literature and it can be written as:

$$\frac{\mu\dot{\gamma} - \mu_\infty}{\mu_0 - \mu_\infty} = \frac{1}{1 + (K\dot{\gamma})^{1-n}}$$

where:

μ_0 is the zero shear rate viscosity

μ_∞ is the infinite shear rate viscosity

K is a time constant

n is the Power Law index

- **Stokes Couple stress fluid** (Stokes, 1966)

According to this theory used in some cases to model synovia, the momentum equation and the continuity equation of synovial fluid are:

$$\begin{aligned} \rho \frac{D\mathbf{V}}{Dt} &= -\nabla p + \rho \mathbf{F} + \frac{1}{2} \rho \nabla \times \mathbf{C} + \mu \nabla^2 \mathbf{V} - \eta \nabla^4 \mathbf{V} \\ \nabla \cdot \mathbf{V} &= 0 \end{aligned}$$

The vectors \mathbf{V} , \mathbf{F} , and \mathbf{C} represent, respectively, the velocity, the body force, and body couple per unit mass while ρ is the density of the oil, μ is the viscosity, and η is a “couple stress constant.”

CONCLUSIONS

This short review was focused on the development of understanding and modeling lubrication phenomena in natural synovial joints, highlighting some relevant research through the years. The topic is particularly interesting since it represents a fascinating research field connects both tribological and biological issues which are necessary for a deep understanding of the complex phenomena acting in the investigated bio-tribosystems. It is acknowledged that the topic is very wide and hence difficult to summarize in a short manuscript, highlighting the “milestones” may be an interesting read and also a useful support for researches and scientists who are approaching this research field for the first time. It could stimulate the scientific community toward stronger cooperation between researchers from different scientific areas, resulting in new insights, and also allow for insights in the optimal tribological design of modern artificial joints, which requires more accurate models to be used in their *in-silico* pre-clinical testing.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

This research was funded by MIUR, PRIN 2017- BIONIC.

REFERENCES

- Ateshian, G. A., and Hung, C. T. (2006). The natural synovial joint: properties of cartilage. *Proc. Inst. Mech. Eng. Pt. J J. Eng. Tribol.* 220, 657–670. doi: 10.1243/13506501JET86
- Ateshian, G. A., Lai, W. M., Zhu, W. B., and Mow, V. C. (1994). An asymptotic solution for the contact of two biphasic cartilage layers. *J. Biomech.* 27, 1347–1360. doi: 10.1016/0021-9290(94)90044-2
- Barnett, C. H., and Cobbald, A. F. (1962). Lubrication within living joints. *J. Bone Joint Surg.* 44, 662–674. doi: 10.1302/0301-620X.44B.3.662
- Briscoe, W. H., Titmuss, S., Tibergh, F., Thomas, R. K., McGillivray, D. J., and Klein, J. (2006). Boundary lubrication under water. *Nature* 444, 191–194. doi: 10.1038/nature05196
- Charnley, J. (1959). “The lubrication of animal joints,” in *Proceedings of the Symposium on Biomechanics* (London), 12–19.
- Dintenfuss, L. (1963). Lubrication in synovial joints. *Nature* 197, 496–497. doi: 10.1038/197496b0
- Dowson, D. (1966). “Paper 12: modes of lubrication in human joints,” in *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings* (London: SAGE Publications).
- Dowson, D. (1995). Elastohydrodynamic and micro-elastohydrodynamic lubrication. *Wear* 190, 125–138. doi: 10.1016/0043-1648(95)06660-8
- Dowson, D. (2012). Bio-tribology. *Faraday Discuss.* 156, 9–30. doi: 10.1039/c2fd20103h
- Dowson, D., and Jin, Z. M. (1986). Micro-elastohydrodynamic lubrication of synovial joints. *Eng. Med.* 15, 63–65. doi: 10.1243/EMED_JOUR_1986_015_019_02
- Dowson, D., Unsworth, A., and Wright, V. (1970). Analysis of ‘boosted lubrication’ in human joints. *J. Mech. Eng. Sci.* 12, 364–369. doi: 10.1243/JMES_JOUR_1970_012_060_02
- Fein, R. S. (1966). “Research report 3: are synovial joints squeeze-film lubricated?” in *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings* (London: SAGE Publications), 125–128.
- Greene, G. W., Banquy, X., Lee, D. W., Lowrey, D. D., Yu, J., and Israelachvili, J. N. (2011). Adaptive mechanically controlled lubrication mechanism found in articular joints. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5255–5259. doi: 10.1073/pnas.1101002108
- Higginson, G. R. (1977). Elastohydrodynamic lubrication in human joints. *Proc. Inst. Mech. Eng.* 191, 217–223. doi: 10.1243/PIME_PROC_1977_191_028_02
- Higginson, G. R., and Norman, R. (1974). The lubrication of porous elastic solids with reference to the functioning of human joints. *J. Mech. Eng. Sci.* 16, 250–257. doi: 10.1243/JMES_JOUR_1974_016_045_02
- Hlaváček, M. (2000). Squeeze-film lubrication of the human ankle joint with synovial fluid filtrated by articular cartilage with the superficial zone worn out. *J. Biomech.* 33, 1415–1422. doi: 10.1016/S0021-9290(00)00109-3
- Hlaváček, M. (2001). The thixotropic effect of the synovial fluid in squeeze-film lubrication of the human hip joint. *Biorheology* 38, 319–334.
- Hou, J. S., Mow, V. C., Lai, W. M., and Holmes, M. H. (1992). An analysis of the squeeze-film lubrication mechanism for articular cartilage. *J. Biomech.* 25, 247–259. doi: 10.1016/0021-9290(92)90024-U
- Hui, A. Y., McCarty, W. J., Masuda, K., Firestein, G. S., and Sah, R. L. (2012). A systems biology approach to synovial joint lubrication in health, injury, and disease. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 4, 15–37. doi: 10.1002/wsbm.157
- Jin, Z. H., Meng, Y., Hu, Y., and Luo, J. (2020). In memoriam: Duncan Dowson (1928–2020). *Friction* 8, 1–3. doi: 10.1007/s40544-020-0360-9
- Jones, E. S. (1936). Joint lubrication. *Lancet.* 227:1043–5. doi: 10.1016/S0140-6736(01)37157-X
- Khan, I. M., Redman, S. N., Williams, R., Dowthwaite, G. P., Oldfield, S. F., and Archer, C. W. (2007). The development of synovial joints. *Curr. Top. Dev. Biol.* 79, 1–36. doi: 10.1016/S0070-2153(06)79001-9
- Klein, J. (2013). Hydration lubrication. *Friction* 1, 1–23. doi: 10.1007/s40544-013-0001-7
- Lai, W. M., Kuei, S. C., and Mow, V. C. (1978). Rheological equations for synovial fluids. *J. Biomech. Eng.* 100, 169–186. doi: 10.1115/1.3426208
- Linn, F. C. (1967). Lubrication of animal joints: I. The arthrotripsometer. *JBJS* 49, 1079–1098. doi: 10.2106/00004623-196749060-00005
- Linn, F. C. (1968). Lubrication of animal joints: II the mechanism. *J. Biomech.* 1, 193–205. doi: 10.1016/0021-9290(68)90004-3
- Linn, F. C., and Radin, E. L. (1968). Lubrication of animal joints. III. The effect of certain chemical alterations of the cartilage and lubricant. *Arthritis Rheumat.* 11, 674–682. doi: 10.1002/art.1780110510
- MacConaill, M. A. (1932). The function of intra-articular fibrocartilages, with special reference to the knee and inferior radio-ulnar joints. *J. Anat.* 66(Pt. 2):210.
- MacConaill, M. A. (1956). Studies in the mechanics of synovial joints. *Irish J. Med. Sci.* 31, 353–364. doi: 10.1007/BF02951170
- MacConaill, M. A. (1960). Lubrication of mammalian joints. *Nature* 185, 920–920. doi: 10.1038/185920a0
- Maroudas, A. (1968). Physicochemical properties of cartilage in the light of ion exchange theory. *Biophys. J.* 8, 575–595. doi: 10.1016/S0006-3495(68)86509-9
- Maroudas, A., Bullough, P., Swanson, S. A. V., and Freeman, M. A. R. (1968). The permeability of articular cartilage. *J. Bone Jt. Surg.* 50, 166–177. doi: 10.1302/0301-620X.50B1.166
- McCutchen, C. W. (1962). The frictional properties of animal joints. *Wear* 5, 1–17. doi: 10.1016/0043-1648(62)90176-X
- McCutchen, C. W. (1966). “Paper 1: physiological lubrication,” in *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings* (London: SAGE Publications), 55–62.
- McCutchen, C. W. (1973). A note on weeping lubrication. In: Kenedi RM, editor. *Perspectives in Biomedical Engineering*. London: Palgrave Macmillan. doi: 10.1007/978-1-349-01604-4_18
- McCutchen, C. W. (1983). Lubrication of and by articular cartilage. *Cartilage* 1983, 87–107.
- Medley, J. B., Dowson, D., and Wright, V. (1984). Transient elastohydrodynamic lubrication models for the human ankle joint. *Eng. Med.* 13, 137–151. doi: 10.1243/EMED_JOUR_1984_013_035_02
- Ogston, A. G., and Stanier, J. E. (1953). The physiological function of hyaluronic acid in synovial fluid; viscous, elastic and lubricant properties. *J. Physiol.* 119, 244–252. doi: 10.1113/jphysiol.1953.sp004842
- Ogston, A. G., Stanier, J. E., Toms, B. A., and Strawbridge, D. J. (1950). Elastic properties of ox synovial fluid. *Nature* 165, 571–571. doi: 10.1038/165571b0
- Popov, V. L. (2019). Active bio contact mechanics: concepts of active control of wear and growth of the cartilage in natural joints. *AIP Conf. Proc.* 2167:020285. doi: 10.1063/1.5132152
- Popova, E., and Popov, V. L. (2014). On the history of elastohydrodynamics: the dramatic destiny of Alexander Mohnstein-Ertel and his contribution to the theory and practice of lubrication. *ZAMM J. Appl. Math. Mech.* 95, 652–663. doi: 10.1002/zamm.201400050
- Radin, E. L., and Paul, I. L. (1969). Failure of synovial fluid to cushion. *Nature* 222, 999–1000. doi: 10.1038/222999a0
- Radin, E. L., Swann, D. A., and Weissner, P. A. (1970). Separation of a hyaluronate-free lubricating fraction from synovial fluid. *Nature* 228, 377–378. doi: 10.1038/228377a0
- Raviv, U., Giasson, S., Kampf, N., Gohy, J. F., Jérôme, R., and Klein, J. (2003). Lubrication by charged polymers. *Nature* 425, 163–165. doi: 10.1038/nature01970

- Raviv, U., Laurat, P., and Klein, J. (2001). Fluidity of water confined to sub-nanometre films. *Nature* 413, 51–54. doi: 10.1038/35092523
- Ropes, M. W., Robertson, W. V. B., Rossmeisl, E. C., Peabody, R. B., and Bauer, W. (1947). Synovial fluid mucin 1, 2. *Acta Med. Scand.* 128, 700–744. doi: 10.1111/j.0954-6820.1947.tb14705.x
- Ruggiero, A., Gomez, E., and D'Amato, R. (2011). Approximate analytical model for the squeeze-film lubrication of the human ankle joint with synovial fluid filtrated by articular cartilage. *Tribol. Lett.* 41, 337–343. doi: 10.1007/s11249-010-9710-5
- Ruggiero, A., Gómez, E., and Roberto, D. (2013). Approximate closed-form solution of the synovial fluid film force in the human ankle joint with non-Newtonian lubricant. *Tribol. Int.* 57, 156–161. doi: 10.1016/j.triboint.2012.06.024
- Ruggiero, A., and Zhang, H. (2020). Editorial: biotribology and biotribocorrosion properties of implantable biomaterials. *Front. Mech. Eng.* 6:17. doi: 10.3389/fmech.2020.00017
- Schmidt, T. A., Gastelum, N. S., Nguyen, Q. T., Schumacher, B. L., and Sah, R. L. (2007). Boundary lubrication of articular cartilage: role of synovial fluid constituents. *Arthritis Rheum.* 56, 882–891. doi: 10.1002/art.22446
- Schmidt, T. A., and Sah, R. L. (2007). Effect of synovial fluid on boundary lubrication of articular cartilage. *Osteoarthr. Cartil.* 15, 35–47. doi: 10.1016/j.joca.2006.06.005
- Stokes, V. K. (1966). Couple stresses in fluids. *Physics Fluids* 9, 1709–1715. doi: 10.1063/1.1761925
- Swann, D. A., Hendren, R. B., Radin, E. L., and Sotman, S. L. (1981). The lubricating activity of synovial fluid glycoproteins. *Arthritis Rheumat.* 24, 22–30. doi: 10.1002/art.1780240104
- Swann, D. A., and Mintz, G. (1979). The isolation and properties of a second glycoprotein (LGP-II) from the articular lubricating fraction from bovine synovial fluid. *Biochem. J.* 179, 465–471. doi: 10.1042/bj1790465
- Swann, D. A., and Radin, E. L. (1972). The molecular basis of articular lubrication I. Purification and properties of a lubricating fraction from bovine synovial fluid. *J. Biol. Chem.* 247, 8069–8073.
- Tanner, R. I. (1966). An alternative mechanism for the lubrication of synovial joints. *Phys. Med. Biol.* 11:119. doi: 10.1088/0031-9155/11/1/312
- Unsworth, A., Dowson, D., and Wright, V. (1975). The frictional behavior of human synovial joints—part I: natural joints. *J. Tribol.* 97, 369–376.
- Walker, P. S., Dowson, D., Longfield, M. D., and Wright, V. (1968). “Boosted lubrication” in synovial joints by fluid entrapment and enrichment. *Ann. Rheum. Dis.* 27:512. doi: 10.1136/ard.27.6.512
- Walker, P. S., Sikorski, J., Dowson, D., Longfield, M. D., Wright, V., and Buckley, T. (1969). Behaviour of synovial fluid on surfaces of articular cartilage. A scanning electron microscope study. *Ann. Rheumat. Dis.* 28:1. doi: 10.1136/ard.28.1.1
- Walmsley, T. (1928). The articular mechanism of the diarthroses. *JBS* 10, 40–45.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Ruggiero. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Review on the Bubble Dynamics Based Cavitation Dynamics for the Negative Squeeze Motion in Lubricated Contacts

Thomas Geike*

Beuth Hochschule für Technik Berlin, Berlin, Germany

OPEN ACCESS

Edited by:

Irina Goryacheva,
Institute for Problems in Mechanics
(RAS), Russia

Reviewed by:

Qiang Li,
Technical University of Berlin,
Germany

Mihai M. Arghir,
University of Poitiers, France

Xuesong Han,
Tianjin University, China

*Correspondence:

Thomas Geike
thomas.geike@beuth-hochschule.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 28 February 2020

Accepted: 04 May 2020

Published: 03 June 2020

Citation:

Geike T (2020) Review on the Bubble Dynamics Based Cavitation Dynamics for the Negative Squeeze Motion in Lubricated Contacts. *Front. Mech. Eng.* 6:33. doi: 10.3389/fmech.2020.00033

Simulation models for the cavitation dynamics in lubricated contacts can be roughly clustered into two groups: either without or with bubble dynamics, the first one being the standard case for most fluid film bearing calculations. The approach with bubble dynamics has been introduced to the lubrication community about 20 years ago by Someya, and it is based on the coupling of Reynolds equation and Rayleigh–Plesset equation. It has been used for journal bearings, squeeze film dampers, and it is essentially required for correct numerical calculations of the negative squeeze motion (i.e., the separation of two plates) or the oil stiction problem. More than a decade ago, in 2009, the first paper on the negative squeeze motion with bubble dynamics—allowing numerical calculations of tensile stresses in the lubricant—had been published. The application in mind is the simulation of mixed lubrication for rough surfaces. The negative squeeze motion is then understood as the motion of asperities (on smaller length scales). The paper at hand summarizes some of the research on the dynamics of cavitation in lubricated contacts from different research groups from the last 10–15 years and sketches key topics for further research on the topic. The roadmap is centered around the three key issues that remained from the previous research of the author: (a) numerical stability of the calculations for curved plates, (b) characteristic time scale for separation of plates, and (c) experimental evidence for validating the calculation results.

Keywords: cavitation, mixed lubrication, oil stiction, negative squeeze motion, bubble dynamics, negative pressure

1. INTRODUCTION

Cavitation is a typical phenomenon in lubricated contacts and appears in different forms. For journal bearings, film rupture and reformation is typically connected to air ingestion respectively feeding new oil. In squeeze film dampers, oil stiction problems or mixed lubrication on the other hand, pressure induced growth and collapse of bubbles is most relevant.

In most simulations of lubricated contacts (e.g., journal bearings) tensile stresses (negative pressure) are neglected. With respect to pressure p cavitation is then characterized by $p = p_v$ or even simpler $p = 0$ (p_v being the vapor pressure at the given temperature).

Nonetheless, negative pressure appears even in journal bearings. To the authors knowledge, Someya (2003) was the first to do numerical calculations for journal bearings that included

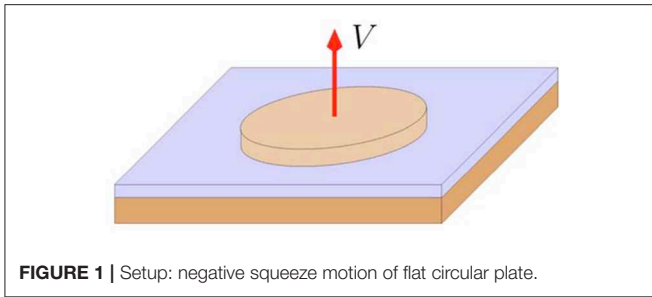


FIGURE 1 | Setup: negative squeeze motion of flat circular plate.

negative pressure¹. He used a bubble dynamics approach including a dilatational viscosity. For discretization, he used the Finite Difference method. Someya reports a good agreement between experiment and calculations based on bubble dynamics.

Tensile stresses cannot be neglected when simulating the negative squeeze motion of lubricated contacts (i.e., the separation of two plates, see **Figure 1**). At first sight, the negative squeeze motion of a macroscopic plate does not seem relevant for practical purposes. Far from it! An important practical application is the oil stiction, which is highly relevant for the performance of reed valves in compressors or of switching valves in hydraulic applications (e.g., Resch and Scheidl, 2014; Yoshizumi et al., 2018). Switching times as low as 1 ms require sophisticated models for the lubricant's behavior in valves.

Certainly the negative squeeze motion is also relevant when looking at smaller length scales. Then we look at asperities moving relative to each other and thereby again and again separating from each other. Tensile stresses on the microscopic level resulting from negative squeeze motion can contribute to fatigue and wear in mixed lubrication. Indeed, the simulation of rough surfaces including negative pressure was the application in mind when the author started his research work on lubricated contacts.

There is experimental evidence for tensile stresses in lubrication films for the setup under study for more than five decades. But it was only in 2009 that calculations with tensile stresses have been reported for the negative squeeze motion Geike and Popov (2009a). To reproduce tensile stresses in numerical simulations bubble dynamics needs to be taken into account. Traditional methods based on Reynolds equation and *static* cavitation conditions do not yield tensile stresses. Bubble dynamics is also required for getting the relevant time scale for the separation of plates.

The key question is which system of equations shall constitute the numerical simulation model. The answer was and still is the coupling of Reynolds equation for compressible fluid flow and Rayleigh–Plesset equation for bubble dynamics.

In order to identify the key topics for further research, it is worthwhile to look at the three questions that remained when the author's research came to a temporary end in early 2008.

1. How can the numerical stability problems be solved for the case of spherical plates (finite radius of curvature)?
2. What is the correct time scale for the plate separation?
3. What experimental evidence is available for a quantitative validation of numerical calculations?

The paper at hand looks at relevant contributions to the topic, particularly in the last 12 years², and suggests potential corner stones for future research on the topic. The paper is organized along the three questions from above.

2. SOLVING THE NUMERICAL STABILITY PROBLEMS

2.1. The Authors Approach to the Cavitation Dynamics for the Negative Squeeze Motion

To the author's knowledge, the papers from January 2009 Geike and Popov (2009a,b) are the first to report on numerical calculations of the negative squeeze motion based on Reynolds equation and bubble dynamics, thereby allowing tensile stresses to develop in the lubricant.

The author used Reynolds equation

$$\frac{1}{r} \frac{\partial}{\partial r} \left(\frac{r \rho h^3}{\eta} \frac{\partial p}{\partial r} \right) = 12 \left(h \frac{\partial \rho}{\partial t} + \rho V \right) \quad (1)$$

and Rayleigh–Plesset-equation

$$R \frac{d^2 R}{dt^2} + \frac{3}{2} \left(\frac{dR}{dt} \right)^2 = \frac{p_v - p}{\rho_{liq}} - \frac{2\gamma}{\rho_{liq} R} - \frac{4\eta_{liq}}{\rho_{liq} R} \frac{dR}{dt} \quad (2)$$

as the basis of the simulation model. p_v is the vapor pressure, γ is the surface tension constant, η_{liq} is the viscosity, and ρ_{liq} the density of the lubricant (liquid).

Reynolds equation is the partial differential equation for the pressure in the lubricant and can be obtained from Navier–Stokes equation and continuity equation on assumptions such as negligible inertia, Newtonian fluid of constant viscosity and thin film geometry. Rayleigh–Plesset equation describes the growth in radius of a spherical bubble in an incompressible fluid of infinite extent. It is used to approximately calculate the vapor fraction and thus the density, which is needed in the Reynolds equation. It is important to understand that the simulation does not consider single bubbles in the lubricant. Instead, the simulation is based on a characteristic bubble radius at each location r that is connected to the vapor fraction. Rayleigh–Plesset equation is used to derive a partial differential equation for vapor fraction, which is then used in the simulation model.

For the published results the first term with the second derivative of R had been neglected. Further simulations with the full Equation (2) yielded almost the same results.

²Many papers that cite the author's previous work study journal bearings and do not take the bubble dynamics into account. These contributions will not be discussed here. However, papers on journal bearings as the ones by Someya or Snyder et al. are considered as they provide helpful insight for the mixed lubrication problem at hand.

¹First published in 2000 in Japanese only. Unfortunately the authors had not been aware of Someya's work when they started to work on the topic.

The partial differential Equation (1) had been transformed into a system of ordinary differential equations based on the Differential Quadrature method (Shu, 2000). The final system of equations is a differential-algebraic system of differentiation index 1. One of the challenges in solving the problem numerically is the strong transient behavior. Within a short time interval significant negative pressure appears and disappears again, while vapor fraction increases quickly.

The numerical solution converged for flat plates (infinite radius of curvature) for a variation of grid types. The results are—at least qualitatively—correct. Husen et al. (2016) came in their paper from 2015 to the same results using a Finite Element method, thereby validating our discretization method and computer implementation of the numerical calculations.

For spherical plates (finite radius of curvature) numerical simulations did not converge for no obvious reason. The numerical integration of the initial value problem (that results from the spatial discretization) fails to converge. However, for the mixed lubrication of rough surfaces it is key to master the spherical plate problem.

From first-hand experience, the author knows that in the Differential Quadrature method stability of numerical calculations sometimes depend on the grid type used (e.g., uniform grid vs. Chebyshev–Gauss–Lobatto grid), but the stability problems remained no matter which grid had been used. The stability issue's root cause and appropriate countermeasures have not been identified yet.

2.2. Calculations for Squeeze Film Dampers With Dilatational Viscosity Term

Few months later, in April 2009, a paper by Gehannin et al. (2009) was published on a quite similar topic—cavitation modeling based on the coupling of Reynolds equation and Rayleigh–Plesset equation. The setup and the application in mind is however quite different: they study squeeze film dampers, i.e., they look at a tangential motion instead of the vertical motion (separation). Bubbles contain vapor and undissolved gas thus the pressure inside the bubble is the sum of the partial pressure of the gas and the vapor pressure. Consequently, they write p_B instead of p_v in equation (2), with

$$p_B = p_0 \left(\frac{R_0}{R} \right)^3 + p_v. \quad (3)$$

In addition, following Someya (2003), Gehannin et al. used an additional term on the right-hand side of the Rayleigh–Plesset Equation (2) with the dilatational viscosity,

$$-\frac{4\kappa}{\rho_{\text{liq}} R^2} \frac{dR}{dt}. \quad (4)$$

They finally solve the system of equations using a Finite Volume approach and conclude that for the squeeze film damper the bubble pressure term and the dilatational viscosity term are most relevant. The agreement between their calculations and the experimental results from Adiletta and Pietra (2006) will be discussed below.

Gehannin et al. do not report on stability issues in their numerical approach. This might be caused by one of the three major differences.

- The setup is different (tangential instead of vertical motion).
- Additional terms have been used in the Rayleigh–Plesset equation.
- A different method for discretization has been used (Finite volume instead of Differential Quadrature).

In future research, the relevance of the additional terms for the negative squeeze motion may be studied first.

2.3. Calculations for Bearings With Dilatational Viscosity Term

Snyder et al. (2015) apply the coupling of Reynolds equation and Rayleigh–Plesset equation to bearings. In accordance with the work of Someya (2003) and Gehannin et al. (2009) they include the dilatational viscosity term in the Rayleigh–Plesset equation and emphasize its importance.

A later paper from the same group (Braun et al., 2017) reports on a simulation model for journal bearings that is extended by an energy equation to model temperature effects and heat transfer.

2.4. Mathematical Studies on the Well-Posedness of Reynolds–Rayleigh–Plesset Coupling for Journal Bearings

Jaramillo et al. (2018) study the coupled equations from the mathematicians perspective, i.e., they study the existence and stability of stationary solutions. They study the case with zero vertical relative velocity (e.g., the standard procedure for journal bearing simulation) and take the dilatational viscosity into account. For journal bearings they perform numerical experiments based on the Finite Volume method and a backward Euler scheme. They conclude that for the eccentricity $\epsilon < 0.41$ the transient solution converges toward the stationary solution; for higher values of ϵ time-convergence is no longer obtained. One of the hypothesis for stability identified by Jaramillo et al. is indeed violated for $\epsilon > 0.41$. This hypothesis is centered around the term

$$f_1(R) = \frac{1}{\rho_{\text{liq}}} \left(p_B - \frac{2\gamma}{\rho_{\text{liq}} R} \right). \quad (5)$$

Even though the setup under study is different, this result is understood as an additional hint to look at the additional terms in the Rayleigh–Plesset equation when doing further research on the the negative squeeze motion.

2.5. The Way Forward: Extend the Model Equations

For further numerical simulations, it seems most promising to include the dilatational viscosity term and extend the pressure term to allow for other pressure than the vapor pressure inside the bubble. The stability issue can be revisited based on the extended model. In case that stability issues remain, the method

of discretization can be changed, from Differential Quadrature method to Finite Volume method or Finite Element method.

In addition to numerical experiments it would be beneficial to partner up with mathematicians to study the negative squeeze motion from the mathematical perspective. This could result in statements on the parameter values for which stable numerical calculations can be expected.

3. GETTING THE TIME SCALE RIGHT

3.1. The Authors Results Based on Bubble Dynamics

The numerical results published in 2009 are qualitatively reasonable. The pressure distribution starts at the onset of motion with the distribution that is expected without cavitation. Thus, significant tensile stresses can be observed for a very short time. Tensile stresses quickly drop while vapor fraction increases. However, the characteristic time scale for plate separation seems rather short and needs attention.

3.2. Calculations of the Oil Stiction Force Without Bubble Dynamics

As mentioned above, an important practical application for the plate separation is the oil stiction in technical devices as valves in hydraulic applications or reed valves in compressors. Oil stiction influences the opening time and therefore the performance of valves in hydraulic or pneumatic applications.

Roemer et al. (2015) state that for small initial distances the assumption of negligible tensile strength of the liquid is not applicable. They propose a model with a fluid tensile strength as additional material property to deal with the tensile stresses that develop when two plates quickly separate. Also, a distinct cavitation zone is modeled. The cavitation zone appears in the event of the tension exceeding the threshold. Roemer et al. report of one example where a time of about 12 ms is required before a cavitation zone is formed.

Scheidl and Gradl (2016) study the problem of two separating plates based on Reynolds equation and cavitation. Their calculation is based on *static* cavitation conditions, i.e., is without bubble dynamics, and disregards tensile stresses³. Another paper from the research group (Resch and Scheidl, 2014) reports tensile stresses for short periods of time—in the range of tenth of milliseconds.

3.3. Calculations With Bubble Dynamics and Experiments for the Oil Stiction of Reed Valves

The paper by Yoshizumi et al. (2018) from 2018 reports on experiments and numerical calculations for the opening of a reed valve. As reed and valve seat are flat and the motion is vertical this situation is relatively close to the negative squeeze motion. However, reed and valve seat are parallel only at the beginning.

As the reed bends while opening, the motion of the two parts is not parallel.

The authors include the elastic deformation of the reed as well as the dynamics of the oil film. For the latter, they use a coupling of Reynolds equation and Rayleigh–Plesset equation. Again, the dilatational viscosity is included and the inertia terms are neglected. The Finite Volume method is used for discretization. For comparison, not only the dynamic cavitation model with bubble dynamics is used for calculation but also two other models (one with static cavitation, one without cavitation). They do not report any stability issues in their calculations.

For the reed valve delay time (a very relevant performance indicator) Yoshizumi et al. got similar results from experiments and from numerical calculations with the dynamic cavitation model⁴. As one would expect, the delay time is far too high for the model without cavitation and far too small for the static model. This is a wonderful hint that the dynamic cavitation model with bubble dynamics is just the right choice—it yields correct results, and simpler choices for modeling are insufficient for a full understanding.

3.4. The Way Forward: Extend the Model Equations

The results for reed valves from Yoshizumi et al. indicate a dependency between delay time and the cavitation model used for calculations (see above). Even more important and not as obvious: the delay time significantly depends on the dilatational viscosity. In particular, the lower the value of the dilatational viscosity the lower the delay time. Therefore, a model without dilatational viscosity might be insufficient for a correct modeling of the time scale.

Using the above mentioned extensions to the simulation model, first of all including the dilatational viscosity, the time scale needs to be looked at closely again. Based on the results from Yoshizumi et al. it is quite save to assume an increase in the characteristic time for the negative squeeze motion.

Having the simulation of mixed lubrication of rough surfaces in mind, Roemer et al.'s (2015) approach with a tensile stress should also be considered. The advantage of this approach is the lower computational effort.

4. GETTING FURTHER EXPERIMENTAL EVIDENCE

Geike and Popov (2009a) give an overview on experimental evidence for tensile stresses in the negative squeeze film motion and discuss why the published data are not sufficient for model validation. In particular, they point to the work of Hays and Feiten (1964), Parkins and May-Miller (1984), Chen et al. (2004), and Wang et al. (2005), who had studied the time-dependent cavitation experimentally in a simple parallel-plate squeeze film configuration.

³Having a design optimization in mind, the authors focus on relatively simple models to speed up the optimization process.

⁴For the reed valve under study, the delay time is 25 ms.

It seems that no experimental results for the negative squeeze motion of spherical plates—sufficient for a quantitative validation of simulation models—have been published yet. The work of Sun et al. (2008) seems promising for the model validation for flat plates in oscillatory motion. The experiments from Yoshizumi et al. (2018) are also relevant as their setup is somewhat close to the negative squeeze motion and a characteristic time scale is studied. It is also worthwhile to study other contributions on the oil stiction topic to find further experimental data.

Most experimental studies, however, focus on other setups⁵. The focus is mostly on journal bearings or squeeze film dampers. As an example, the paper by Adiletta and Pietra (2006) from 2006 studies the squeeze film damper. Their results show that negative values for the absolute pressure appear in squeeze film dampers too. There seems to be still a difference between numerical results from Gehannin et al. (2009) and the experimental results from Adiletta and Della Pietra with respect to the absolute pressure inside the cavitation zone. Gehannin et al. conclude that it is essential to include the dilatational viscosity to get negative pressure in the studied squeeze film damper. They also emphasize that a model based on Reynolds and Rayleigh–Plesset equations relies on many parameters, temperature dependent lubricant properties and also initial conditions on bubble size and vapor fraction.

For the future, a joint research between partners who on the one side undertake the necessary experimental studies and on the other side work on the numerical simulation model seems beneficial to answer the open questions.

⁵Someya's paper on journal bearings Someya (2003) shows a diagram for the negative squeeze but does not give any background information, which would be needed for the validation of computational results.

REFERENCES

- Adiletta, G., and Pietra, L. D. (2006). Experimental study of a squeeze film damper with eccentric circular orbits. *ASME J. Tribol.* 128, 365–377. doi: 10.1115/1.2162555
- Braun, M., Pierson, K., and Snyder, T. (2017). Two-way coupled reynolds, Rayleigh-Plesset-Scriven and energy equations for fully transient cavitation and heat transfer modeling. *IOP Conf. Ser. Mater. Sci. Eng.* 174:012030. doi: 10.1088/1757-899X/174/1/012030
- Chen, X., Sun, M., Wang, W., Sun, D. C., Zhang, Z., and Wang, X. (2004). Experimental investigation of time dependent cavitation in an oscillatory squeeze film. *Sci. China Ser. G* 47, 107–112. doi: 10.1360/03yb0237
- Gehannin, J., Arghir, M., and Bonneau, O. (2009). Evaluation of Rayleigh-Plesset equation based cavitation models for squeeze film dampers. *ASME J. Tribol.* 131:024501. doi: 10.1115/1.3063819
- Geike, T., and Popov, V. (2009a). A bubble dynamics based approach to the simulation of cavitation in lubricated contacts. *ASME J. Tribol.* 131:011704. doi: 10.1115/1.2991290
- Geike, T., and Popov, V. (2009b). Cavitation within the framework of reduced description of mixed lubrication. *Tribol. Int.* 42, 93–98. doi: 10.1016/j.triboint.2008.05.002
- Hays, D. F., and Feiten, J. B. (1964). "Cavities between moving parallel plates," in *Cavitation in Real Liquids*, ed R. Davies (Amsterdam: Elsevier), 122–127.
- Husen, L. V., Müller, S., Reinartz, B., and Knoll, G. (2016). Berechnung von Kavitationsvorgängen in hydrodynamischen Gleitlagern auf der Basis

5. WHAT IS NEXT?

In the author's opinion the research on the cavitation dynamics for mixed lubrication should be centered around the above mentioned key topics—stability of numerical calculation, clarifying time scale and finding further experimental evidence.

As mentioned above, it seems most promising to set the focus on extending the Rayleigh–Plesset equation (bubble dynamics). Hopefully, the questions around the stability issues and the characteristic time scale can then be answered. Involving partners for (i) the mathematical study on the stability issue and (ii) for further experiments would be beneficial for future research.

From today's perspective, the next leap forward would be the simulation of the elasto-hydrodynamic problem of rough surfaces including the cavitation dynamics. For this the boundary element method (BEM) seems the right choice for modeling the elastic part. Only the surface is discretized in the BEM. Hence the method allows to model surface roughness with very fine meshes and it is often more efficient for contact problems than methods that require the discretization of the entire volume. The tool and experience of Popov's research group Pohrt and Li (2014) could be used here.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author thanks Valentin Popov for proposing the idea to resume the exciting research work on the cavitation dynamics.

- eines blasendynamischen Modells. *Tribologie und Schmierungstechnik* 63, 25–31. Available online at: <https://elibrary.narr.digital/article/99.125005/tus201650025>
- Jaramillo, A., Bayada, G., Ciuperca, I., and Jaib, M. (2018). On the well-posedness of Reynolds-Rayleigh-Plesset coupling. *arXiv [Pre-print]*. arXiv:1805.07219v1 [math-ph].
- Parkins, D. W., and May-Miller, R. (1984). Cavitation in an oscillatory oil squeeze film. *ASME J. Tribol.* 106, 360–367. doi: 10.1115/1.3260934
- Pohrt, R., and Li, Q. (2014). Complete boundary element formulation for normal and tangential contact problems. *Phys. Mesomech.* 17, 334–340. doi: 10.1134/S1029959914040109
- Resch, M., and Scheidl, R. (2014). A model for fluid stiction of quickly separating circular plates. *Proc. IMechE C J. Mech. Eng. Sci.* 228, 1540–1556. doi: 10.1177/0954406213509613
- Roemer, D., Johansen, P., Pedersen, H., and Andersen, T. (2015). Fluid stiction modeling for quickly separating plates considering the liquid tensile strength. *J. Fluids Eng.* 13:061205. doi: 10.1115/1.4029683
- Scheidl, R., and Gradl, C. (2016). An approximate computational method for the fluid stiction problem of two separating parallel plates with cavitation. *J. Fluids Eng.* 138:061301. doi: 10.1115/1.4032299
- Shu, C. (2000). *Differential Quadrature and Its Applications in Engineering*. London: Springer. doi: 10.1007/978-1-4471-0407-0
- Snyder, T., Braun, M., and Pierson, K. (2015). On Rayleigh-Plesset based cavitation modeling of fluid film bearings using the Reynolds

- equation. *J. Phys. Conf. Ser.* 656:12072. doi: 10.1088/1742-6596/656/1/012072
- Someya, T. (2003). "Negative pressure in the oil-film of journal bearing," in *Rotrib 03 National Tribology Conference* (Galați: University of Galati), 215–220.
- Sun, M., Zhang, Z., Chen, X., Wang, W., Meng, K., and Sun, D. C. (2008). Experimental study of cavitation in an oscillatory oil squeeze film. *Tribol. Trans.* 51, 341–350. doi: 10.1080/10402000701793989
- Wang, W., Zhang, Z., Chen, X., Sun, M., and Sun, D. C. (2005). "Investigation of cavitation phenomenon in an oscillatory oil squeeze film," in *World Tribology Congress III* (Washington, DC: WTC), 64167. doi: 10.1115/WTC2005-64167
- Yoshizumi, F., Kondoh, Y., Moroi, T., Tamano, S., and Morinishi, Y. (2018). Model study on the oil stiction of a discharge reed valve in compressors. *ASME J. Tribol.* 140:021103. doi: 10.1115/1.4037539
- Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Geike. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Strengthening and Weakening Effects in Bilayer Coated Spherical Contact

Kurien S. Parel, Zhou Chen and Izhak Etsion*

Department of Mechanical Engineering, Technion, Haifa, Israel

The effect of a middle-layer of thickness t_1 with intermediate Young's modulus and yield strength in between a hard coating of thickness t_2 and a softer spherical substrate of radius R on the yield inception of a coated system is investigated. Finite element method is used to model bilayer-coated spheres flattened by a rigid flat. It is found that the addition of the middle-layer can enhance or reduce the resistance to yield inception depending on its dimensionless thickness t_1/t and t/R where t is the total thickness of the two coating layers. Some practical results are presented, to enable optimum selection of bilayer coated system and prevention of undesired weakening effect.

Keywords: contact mechanics, yield inception, coatings, bilayers, strengthening and weakening

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Antonio Papangelo,
Politecnico di Bari, Italy
Sina Mobasher Moghaddam,
Apple, United States

*Correspondence:

Izhak Etsion
etsion@technion.ac.il

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 27 February 2020

Accepted: 09 April 2020

Published: 29 April 2020

Citation:

Parel KS, Chen Z and Etsion I (2020)
Strengthening and Weakening Effects
in Bilayer Coated Spherical Contact.
Front. Mech. Eng. 6:23.
doi: 10.3389/fmech.2020.00023

INTRODUCTION

Stiff hard coatings, such as TiN or CrN on metallic substrates, are often used to enhance tribological properties of components (Kot, 2012). Goltsberg et al. (2011) studied the yield inception of a coated sphere pressed by a rigid flat. They found that a coated sphere can be more resistant to yielding than a homogenous sphere made of the hard coating material. However, ultrathin hard coatings can cause a weakening effect (Goltsberg et al., 2011; Huang et al., 2012; Goltsberg and Etsion, 2013), by which the yield resistance of the coated sphere is lower than that of a homogenous sphere made of the soft substrate material. (Komvopoulos, 1988, 1989) and Sun et al. (1995) also demonstrated this weakening effect of ultrathin hard coatings in the case of a coated half-space subject to indentation.

A mismatch of the Young's moduli at the coating/substrate interface can lead to additional stresses in the coated system (van der Zwaag and Field, 1982; Komvopoulos, 1988, 1989; Chai et al., 1999; Piao et al., 2010; Goltsberg et al., 2011; Goltsberg and Etsion, 2013), which can reduce the resistance to yield inception of the system. To alleviate this effect, the mismatch of the Young's moduli at the interface should be reduced. This can be achieved by using functionally graded material (Stephens et al., 2000; Liu et al., 2016) or by applying multilayers of coatings (Djabella and Arnell, 1994). A simpler solution is to insert a single middle-layer, which has an intermediate Young's modulus between the substrate and coating.

Fontalvo et al. (2010) showed experimentally that such bilayer coatings can enhance wear resistance. Finite element studies (van der Zwaag et al., 1986; Djabella and Arnell, 1993a,b) have modeled the spherical indentation of an elastic half-space with a bilayer coating, assuming a Hertzian pressure distribution. These studies show a reduction of adverse stresses within the coating material compared to single-layer cases. However, the assumption of Hertzian pressure loading is valid only for small mismatch values (Gupta and Walowit, 1974). More accurately, Lardner et al. (1992) and Guo and Zhao (2019) modeled a

bilayer coated half-space indented by a rigid sphere, relaxing the Hertzian pressure distribution assumption. Lardner et al. (1992) identified which interface (the contact interface, the outer-coating/middle-layer interface or the middle-layer/substrate interface) had the greatest shear and tensile stresses for different total coating thicknesses and the Young's moduli ratio of the two coating layers. Guo and Zhao (2019) showed that a middle-layer of intermediate Young's modulus can reduce the stress discontinuities in the coated system (and lower the chances of delamination at the coating/substrate interface). Yu et al. (2016) recently identified the location of the maximum von Mises stress for the bilayer coated half-space during indentation, for different coating thicknesses and material properties. The half-space was subjected to both normal and tangential loading with a spherical indenter. However, little attention has been paid to the yield resistance of the bilayer coated system. The aim of the present study is to investigate the yield inception of a bilayer coated sphere flattened by a rigid flat.

Modeling with the flattening approach is chosen over the indentation one, because it is more relevant for good tribological designs associated with mild adhesive friction and wear, when asperities indentation is avoided (Goltsberg et al., 2011). The results from this study are expected to be relevant in the design of mechanical components that involve stiff-hard coatings (such as TiN or CrN on metallic substrates) to enhance load carrying capacity and component lifetime (Kot, 2012). These include applications such as bearings, mechanical tools, electromechanical switches and bio-implants.

THEORETICAL BACKGROUND

The yield resistance of a single-layer coated sphere flattened by a rigid frictionless flat was intensively investigated in Goltsberg et al. (2011) and Goltsberg and Etsion (2013). The relevant results are summarized in **Figure 1**, which schematically presents the dimensionless critical load, P_c/P_{c_co} , as a function of dimensionless coating thickness, t/R , for the case of $P_{c_co} > P_{c_su}$. Here t is the coating thickness and R is the substrate radius. P_c is the critical load at yield inception for the coated sphere, and P_{c_co} and P_{c_su} are the critical loads for a homogenous sphere of radius $R+t$ made of coating or substrate materials, respectively. Values for P_{c_co} and P_{c_su} are given by the following expression for the critical load for the flattening of a homogenous sphere of radius R (Brizmer et al., 2006).

$$P_c = \frac{\pi^3}{6} C_v^3 Y [R(1 - \nu^2) \frac{Y}{E}]^2 \quad (1)$$

where E , Y , and ν are the Young's modulus, yield strength, and Poisson's ratio of the relevant material of the homogenous sphere, respectively, and $C_v = 1.234 + 1.256\nu$. Yield inception was noted to occur always along the central axis of symmetry.

In the figure at $t/R = 0$, $P_c = P_{c_su}$ and at very large t/R , P_c approaches P_{c_co} as expected. At ultralow t/R , a weakening effect exists such that P_c is lower than P_{c_su} . Maximum weakening occurs at $t/R = (t/R)_{MW}$, and the location of yield inception is at the substrate side of the coating/substrate interface (Goltsberg

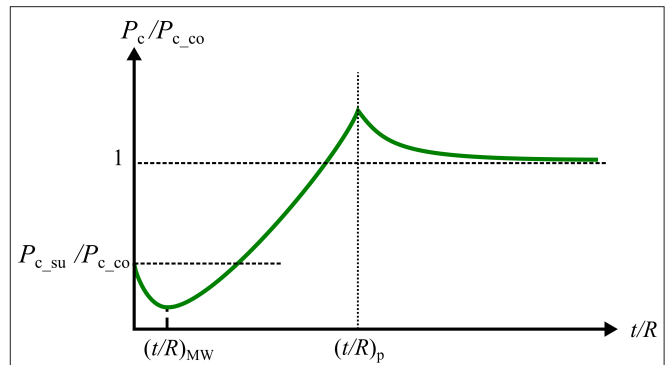


FIGURE 1 | Schematic of the dimensionless critical load, P_c/P_{c_co} , vs. dimensionless coating thickness, t/R , for a single-layer coated sphere flattened by a rigid flat.

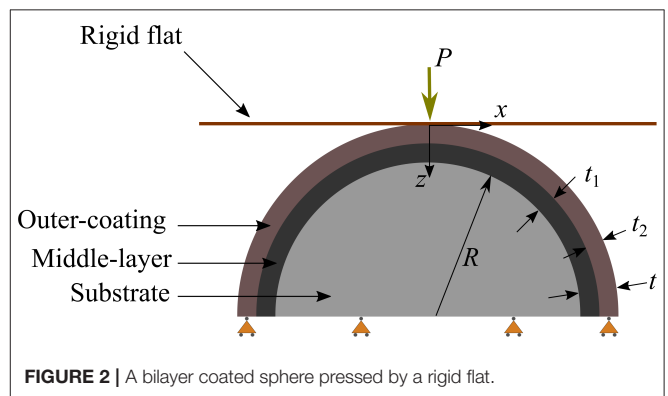
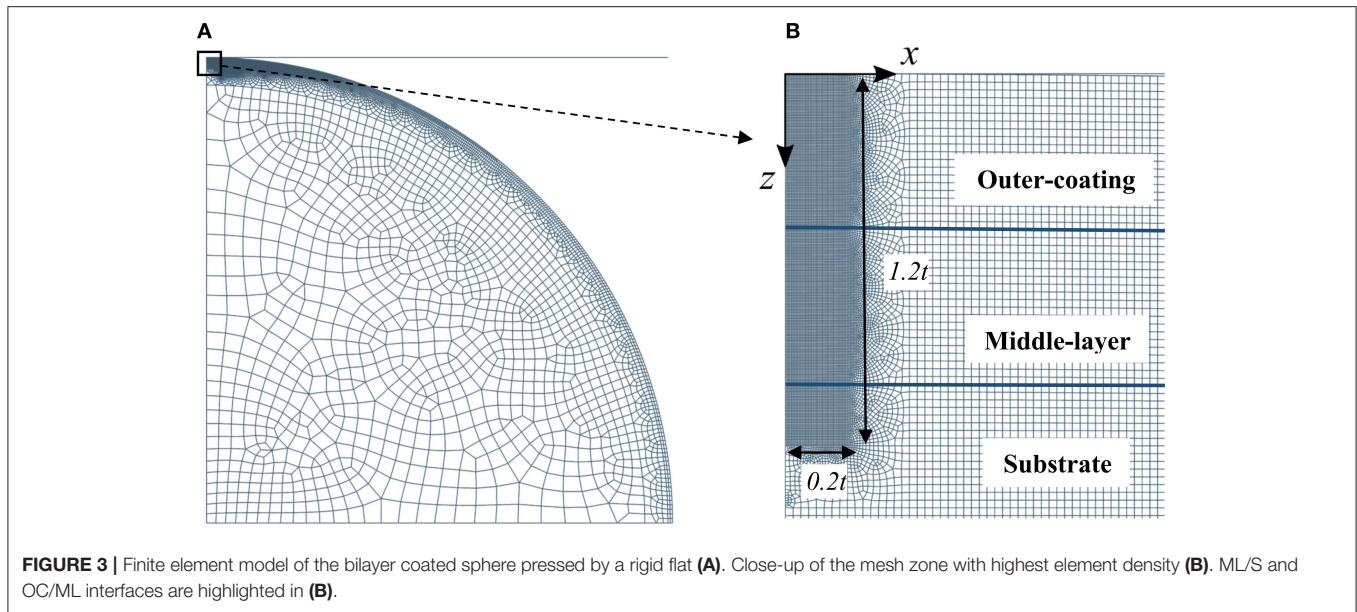


FIGURE 2 | A bilayer coated sphere pressed by a rigid flat.

and Etsion, 2013). As argued by Goltsberg and Etsion (2013), the contribution to the equivalent von Mises stress, σ_{eq} , due to the external normal loading is maximum slightly below the contact interface. The contribution due to the Young's moduli mismatch is maximum at the coating/substrate interface. At $t/R = (t/R)_{MW}$ the above two locations coincide and the maximum weakening occurs. Under a given normal load P , when t/R is different than $(t/R)_{MW}$, the coating/substrate interface moves away from the location of maximum contribution from the external loading. This results in lower equivalent von Mises stresses and a higher P_c is needed for yielding as seen in **Figure 1**. The critical load P_c/P_{c_co} reaches a maximum at $t/R = (t/R)_p$, when the location of yield inception moves to within the coating. Near $(t/R)_p$, a strengthening effect is observed as P_c is greater than P_{c_co} .

METHODOLOGY

Figure 2 schematically presents a bilayer coated sphere compressed by a rigid flat under normal load P . The radius of the spherical substrate is R . The bilayer coating has a total thickness t , and consists of a middle-layer and outer-coating of thicknesses t_1 and t_2 , respectively ($t = t_1 + t_2$). The bottom surface of the sphere is restricted from normal displacement [further restriction of tangential displacement has negligible



effect on the results of the contact problem (Goltsberg et al., 2011)]. To simplify the problem, the following assumptions are adopted:

1. Perfect slip contact condition (frictionless) is assumed at the contact interface to avoid shear stresses there.
2. Perfect bonding is assumed at the outer-coating/middle-layer (OC/ML) and middle-layer/substrate (ML/S) interfaces.
3. The outer-coating, middle-layer and substrate materials are isotropic, free of residual stresses and obey linear elastic constitutive law prior to yielding.
4. The Poisson's ratio for outer-coating, middle-layer and substrate materials are equal, i.e. $\nu_{co} = \nu_m = \nu_{su} = \nu = 0.32$. Note that the subscripts 'co', 'm', and 'su' refer to the outer-coating, middle-layer and the substrate materials, respectively.

These simplifying assumptions help in providing a good physical insight of the problem but, if needed, can be relaxed in future work.

The contact problem shown in **Figure 2** was solved by using the commercial package ABAQUS CAE 2017, with a 2D axisymmetric finite element model presented in **Figure 3A**. The bilayer coated sphere was modeled with a quarter circle and the rigid flat as a line. The quarter circle was meshed with 4-node bilinear quadrilateral elements (CAX4R) and the rigid flat was modeled with an 'analytical rigid' element. The densest mesh was applied in a zone of $0.2t$ width and $1.2t$ depth at the sphere summit (see **Figure 3B**), where yield inception occurs. The element length in this zone is $0.005t$. Outside this zone the mesh density is decreased gradually as seen in **Figure 3A**. In total, around 30,000 to 40,000 elements were used depending on the thicknesses of the outer-coating and middle-layer. To define the contact pair, the rigid flat surface was chosen as the master surface and the outer-coating surface as the slave surface, with "small sliding" formulation as the tracking approach [see section 38.1.1 of (Dassault Systèmes, 2016) for details]. Loading of the

coated sphere is accomplished by displacement control of the rigid flat, in increments of $4 \times 10^{-7} R$. Yield inception occurs in the bilayer coated sphere once the equivalent von Mises stress σ_{eq} (see Equation 2) at a certain location reaches the yield strength of the relevant material.

$$\sigma_{eq} = \{0.5 [(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_1 - \sigma_3)^2]\}^{0.5} \quad (2)$$

where σ_1 , σ_2 , and σ_3 are the principal stresses. For all the cases studied here, yield inception was noted to occur along the axis of symmetry ($x = 0$), at which this expression becomes (Goltsberg et al., 2011)

$$\sigma_{eq} = |\sigma_1 - \sigma_2| \quad (3)$$

where σ_1 and σ_2 are the normal and radial principal stresses, respectively.

In order to validate the accuracy of the numerical model, mesh convergence was tested by running models with greater mesh density in all zones until no significant change ($<1\%$) in the critical load at yield inception in the coated system is observed (see **Appendix A** for details). To validate the accuracy of the numerical model, identical material properties were applied to both the outer-coating and the middle-layer, corresponding to a single-layer coated sphere. The critical load as well as the load-displacement relation from such cases showed good agreement (within 10%) with the results reported by Goltsberg et al. (2011) and Goltsberg and Etsion (2015).

RESULTS AND DISCUSSION

For studying the yield inception of a bilayer coated sphere, models were constructed such that $E_{co} > E_m > E_{su}$ and $Y_{co} > Y_m > Y_{su}$. In the models, R was fixed at 100 mm and E_{su} at 1 GPa. However, changing these dimensional values do not affect the

TABLE 1 | Material property ratios to study the strengthening effect in bilayer coated spheres.

E_{co}/E_{su}	E_m/E_{su}	E_{su}/Y_{su}	Y_{co}/Y_{su}	Y_m/Y_{su}	ν
10	5.5	1500	10	5.5	0.32

results when presented in non-dimensional form. Simulations to study both the weakening and strengthening zones were run for t_1/t from 0 to 1 in increments of 0.1, and t/R from 0.001 to 0.02 in increments of 0.0005. Since the value of peak critical load near the maximum strengthening is sensitive to t/R (see **Figure 1**), additional simulations were run near the peak strengthening with smaller t/R increments of 2×10^{-5} . Note that the coating thickness in the present study is normalized by the substrate radius R , which is different from that normalized by half-width of the contact area, e.g., Komvopoulos (1988). Since R is known a priori, contrary to the unknown half width of contact area, which depends on load and coating thickness, the current normalization approach enables easy selection of the optimum coating thickness. The practicality and validity of this normalization approach was demonstrated experimentally in Bar-Hen and Etsion (2017). Moreover, this normalization approach was also used successfully in the study on the electrical conductance of a bilayer coated spherical contact (Korchevnik et al., 2018) and enabled a concise interpretation of the results for the application of electromechanical switches.

Strengthening Effect of Bilayer Coated Spheres

The material property ratios used in the models to study the strengthening effect are shown in **Table 1**. An extreme $E_{co}/E_{su} = 10$ was chosen to allow a substantial reduction in the Young's moduli mismatch by adding the middle-layer with $E_{co}/E_m = 1.82$ and $E_m/E_{su} = 5.5$. **Figure 4** shows the dimensionless critical load $P_c/P_{c_{co}}$ vs. the dimensionless coating thickness t/R for two typical bilayer cases of relatively small and large dimensionless middle-layer thicknesses $t_1/t = 0.4$ and $t_1/t = 0.8$, respectively. For comparison, results are also shown for $t_1/t = 0$ corresponding to a single-layer case with coating made of the same outer-coating material ($E_{co}/E_{su} = 10$).

For the small $t_1/t = 0.4$, $P_c/P_{c_{co}}$ increases monotonically with t/R until a maximum value $(P_c/P_{c_{co}})_p$ at a certain $t/R = (t/R)_p$ and then reduces approaching unity, like the single-layer case. A similar behavior is shown for the large $t_1/t = 0.8$, but this case involves an additional small peak at relatively small t/R . From **Figure 4**, it is noted that compared to the single-layer case with $t_1/t = 0$, $(P_c/P_{c_{co}})_p$ is greater for the bilayer case with $t_1/t = 0.4$ but lower for the bilayer case with $t_1/t = 0.8$. This suggests that the dimensionless middle-layer thickness, t_1/t , can enhance or reduce the maximum strengthening of a coated sphere. Hence, there is an optimum t_1/t at which the ultimate maximum strengthening effect is obtained.

Further, there is a transition thickness, $(t/R)_{tr,S}$, for any t_1/t at which the bilayer coated sphere has the same value of $P_c/P_{c_{co}}$ as the single-layer coated sphere (see vertical dashed lines in

Figure 4 for the cases with $t_1/t = 0.4$ and $t_1/t = 0.8$). While for t/R below its $(t/R)_{tr,S}$, the bilayer coated sphere experiences lower critical loads than the single-layer coated sphere, for t/R above $(t/R)_{tr,S}$ the bilayer coated sphere experiences higher critical loads.

Figure 5 shows, for the strengthening zone, the effect of dimensionless middle-layer thickness, t_1/t , on the peak critical load, $(P_c/P_{c_{co}})_p$, the dimensionless coating thickness at peak critical load, $(t/R)_p$, and the dimensionless transition coating thickness, $(t/R)_{tr,S}$. In **Figure 5A**, at $t_1/t = 0$, which corresponds to a single-layer case, the peak dimensionless critical load is around 1.8 (see also **Figure 4**). As can be seen from **Figure 5A**, $(P_c/P_{c_{co}})_p$ hardly changes up to $t_1/t = 0.3$. Therefore, the bilayer effect in this range is negligible. An increase is noted for $t_1/t \geq 0.3$ and the ultimate peak critical load is at an optimum dimensionless middle-layer thickness $t_1/t = 0.5$. The optimum t_1/t value may be different for other material properties and this will be explored in a future study. Beyond the optimum t_1/t , $(P_c/P_{c_{co}})_p$ decreases and can even become lower than that for the single-layer case.

Further, from **Figures 5B,C** both $(t/R)_p$ and $(t/R)_{tr,S}$ become significantly large for t_1/t larger than the optimum value. Hence, bilayer coatings with large values of t_1/t are undesirable since the required total coating thickness, t , to obtain peak strengthening becomes very large, while the resulting peak strengthening is lower than the ultimate peak critical load at optimum t_1/t .

In the present model, yield inception always occurs along the symmetry axis ($x = 0$ in **Figure 2**). From the many simulations performed in this study it was found that the location of yield inception in the bilayer cases depends on t_1/t and t/R . With t_1/t up to the optimum value (≤ 0.5) the yield inception is at the substrate side of the ML/S interface when $t/R < (t/R)_p$, and within the outer-coating when $t/R \geq (t/R)_p$. This is similar to the location of yield inception for single-layer coated cases as described in Goltsberg et al. (2011).

With t_1/t greater than the optimum value (> 0.5), the yield inception is also at the substrate side of the ML/S interface for low t/R . However, when t/R starts increasing the yield inception location jumps to the middle-layer side of the OC/ML interface. This jump corresponds to the first peak for the case with $t_1/t = 0.8$ in **Figure 4**. When t/R further increases to its $(t/R)_p$ the yield inception location jumps to within the outer-coating. This corresponds to the second peak for the case with $t_1/t = 0.8$ in **Figure 4**.

As seen in **Figure 5A**, for t_1/t up to the optimum $t_1/t (\leq 0.5)$, an increase in t_1/t leads to an increase in the dimensionless critical load $(P_c/P_{c_{co}})_p$, obtained at $(t/R)_p$, when the yield inception location jumps to within the outer-coating. As explained in Goltsberg et al. (2011), at a given load, Young's moduli mismatch between the outer-coating and the substrate causes additional stresses within the outer-coating that is stretched by the substrate. As found from the present simulations, these additional stresses are reduced by increasing the thickness of the middle-layer and hence, $(P_c/P_{c_{co}})_p$ is increased.

The two main sources for maximum contribution to the von Mises equivalent stress (Goltsberg and Etsion, 2013) were explained in the discussion of **Figure 1** at the end of the

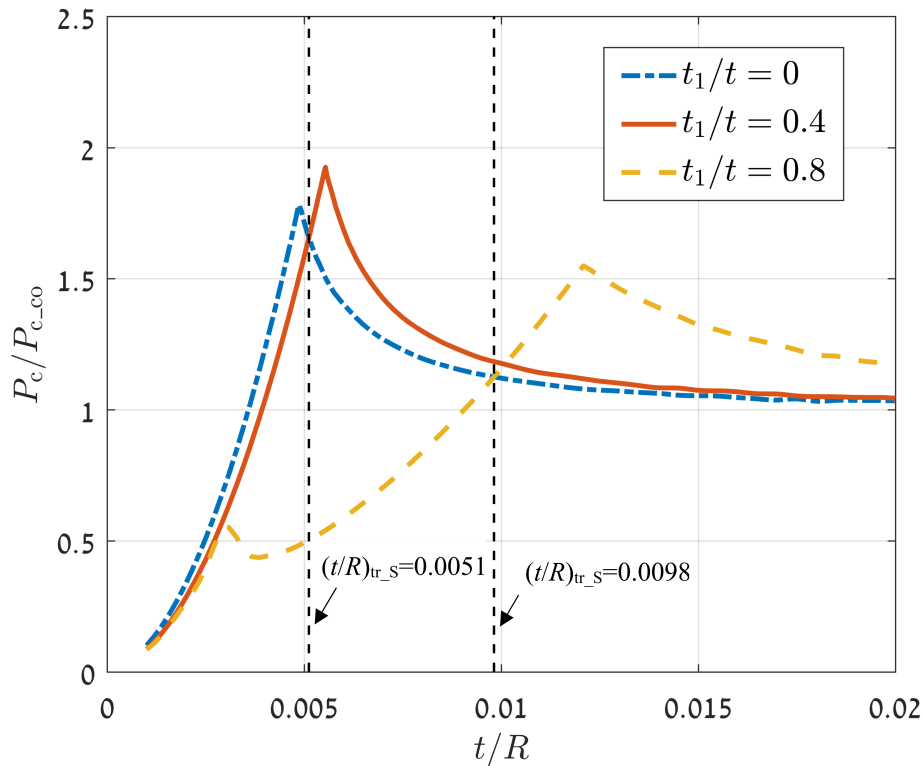


FIGURE 4 | Dimensionless critical load, P_c/P_{c-co} , vs. dimensionless coating thickness, t/R , for bilayer coated spheres. Material properties as in **Table 1**.

Theoretical Background section. In the present study of bilayer coated spherical contact under a normal load P , the center point ($x = 0$) of the OC/ML interface is the location of maximum contribution to the von Mises equivalent stress due to the Young's moduli mismatch at this interface. For any given t/R , when t_1/t increases, this location becomes closer to the contact area (see **Figure 2**) and the location of maximum contribution due to the normal loading. Hence, if the increasing von Mises equivalent stress at the OC/ML interface reaches the lower yield strength of the middle-layer, yield inception will occur at this point and the load P will become the critical one P_c . As we found from our simulations this occurs for a range of t/R values below $(t/R)_p$ whenever t_1/t is greater than the optimum value (>0.5). For these values of t_1/t , yield inception for a range of t/R is at relatively low critical loads, compared to the case with optimum t_1/t . This results in a reduction of the $(P_c/P_{c-co})_p$ value as seen in **Figure 5A**.

As discussed in relation to **Figure 4**, a bilayer coated case has lower critical loads than the corresponding single-layer case with same outer-coating ($E_{co}/E_{su} = 10$) and t/R , when $t/R < (t/R)_{tr-S}$ and greater critical loads when $t/R > (t/R)_{tr-S}$. To investigate the reasons for this, comparisons were made between the dimensionless von Mises equivalent stresses, σ_{eq}/Y , along the axis of symmetry (at $x = 0$) in bilayer cases and corresponding single-layer cases. Y is the yield strength of the relevant material. The comparisons were done for each bilayer case and its corresponding

single-layer case subjected to the lower critical load of the two cases.

The comparisons showed that the maximum σ_{eq}/Y in the outer-coating is typically lower in a bilayer case than in its corresponding single-layer case (except for locations next to the OC/ML interface at $z/t = t_2/t$). The parameter z is the axial distance from the contact interface of the coated sphere as shown in **Figure 2**. However, the σ_{eq}/Y at the middle-layer side of the $z/t = t_2/t$ interface is typically significantly higher in a bilayer case than the σ_{eq}/Y at the same z/t location in its corresponding single-layer case. Further, the σ_{eq}/Y at the substrate side of the $z/t = 1$ interface is also typically higher in a bilayer case than in its corresponding single-layer case. This suggests in the strengthening zone, reducing the mismatch at $z/t = 1$ interface, by inserting a middle-layer, increases the σ_{eq}/Y at the substrate side of this interface.

The lower σ_{eq}/Y within the outer coating in a bilayer case compared to the corresponding single-layer case is due to the reduction of stresses within the outer-coating with the presence of the middle-layer, as discussed earlier. The bilayer case experiences additional stresses at $z/t = t_2/t$ due to the mismatch at the OC/ML interface at this location, which is not present in the corresponding single-layer case. Further, the material at the middle-layer side of the OC/ML interface in the bilayer case has lower yield strength than the outer-coating material at this z/t location in the single-layer case. Due to both of these reasons, σ_{eq}/Y at the middle-layer side of $z/t = t_2/t$ interface is higher in

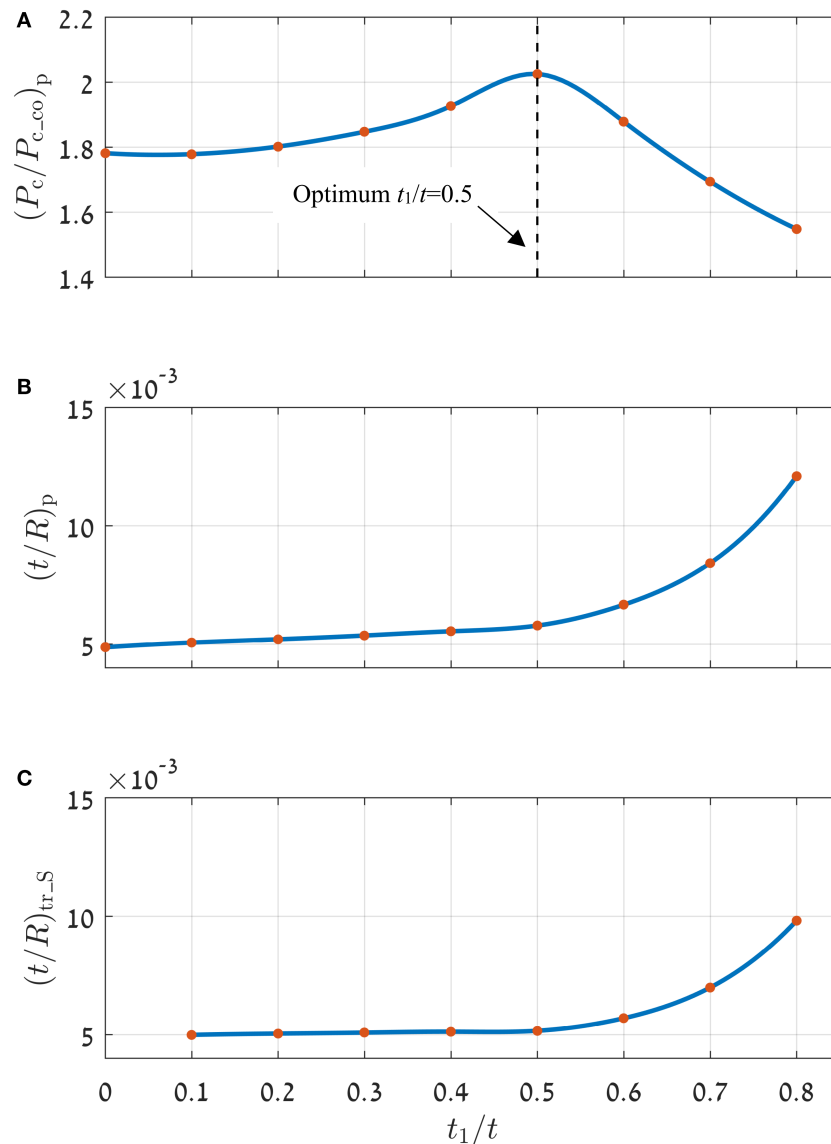


FIGURE 5 | Peak dimensionless critical load, $(P_c/P_{c-co})_p$ **(A)**, dimensionless coating thickness at peak critical load, $(t/R)_p$ **(B)**, and dimensionless transition coating thickness, $(t/R)_{tr-s}$ **(C)**, for different middle-layer dimensionless thickness, t_1/t . Material properties as in **Table 1**.

the bilayer case than the σ_{eq}/Y at the same z/t location in the corresponding single-layer case.

From the many simulations, it was observed that small mismatch at the $z/t = 1$ interface increases (makes more compressive) σ_1 more than σ_2 at the substrate side of the interface. From Equation (3) we can see that this increases σ_{eq} in the bilayer case compared to the single-layer case.

When $t/R < (t/R)_{tr-s}$, high σ_{eq}/Y at either $z/t = t_2/t$ or $z/t = 1$ results in lower critical loads for a bilayer case compared to the corresponding single-layer case. When $t/R > (t/R)_{tr-s}$, high σ_{eq}/Y within the outer-coating of the corresponding single-layer case results in smaller critical loads for this case compared to the bilayer case as seen in **Figure 4**.

TABLE 2 | Material property ratios to study the weakening effect in bilayer coated spheres.

E_{co}/E_{su}	E_m/E_{su}	E_{su}/Y_{su}	Y_{co}/Y_{su}	Y_m/Y_{su}	ν
10	2	350	10	2	0.32

Weakening Effect in Bilayer Coated Spheres

Weakening occurs for coated spheres when $P_c/P_{c-su} < 1$. As per Equation (12) in Goltsberg and Etsion (2013), weakening for single-layer coated spheres occurs for small coating

thickness when

$$\frac{t}{R} < \frac{1.417}{(E_{su}/Y_{su})^{0.9587}} \quad (4)$$

Hence, to avoid constructing models with very low t/R , which would require a very dense mesh, models were constructed with a lower value of E_{su}/Y_{su} to study the weakening effect in bilayer coatings. **Table 2** shows the material property ratios used in these models. The value of E_{su}/Y_{su} was chosen to be within the range analyzed by Goltsberg and Etsion (2013) for the weakening effect

in single-layer coated spheres. E_{su} was kept the same as the value used in strengthening (1 GPa) and the E_{su}/Y_{su} ratio was reduced by increasing Y_{su} .

As noted by Goltsberg and Etsion (2013), the weakening effect in the single-layer case is due to additional stresses in the substrate due to the mismatch at the $z/t = 1$ interface. Hence, for more effective bilayer, $E_m/E_{su} = 2$ was selected in **Table 2** instead of 5.5 in **Table 1**. The yield strengths values of the outer-coating and the middle-layer materials were changed so that $Y_{co}/Y_{su} = E_{co}/E_{su}$, and $Y_m/Y_{su} = E_m/E_{su}$ as in **Table 1**.

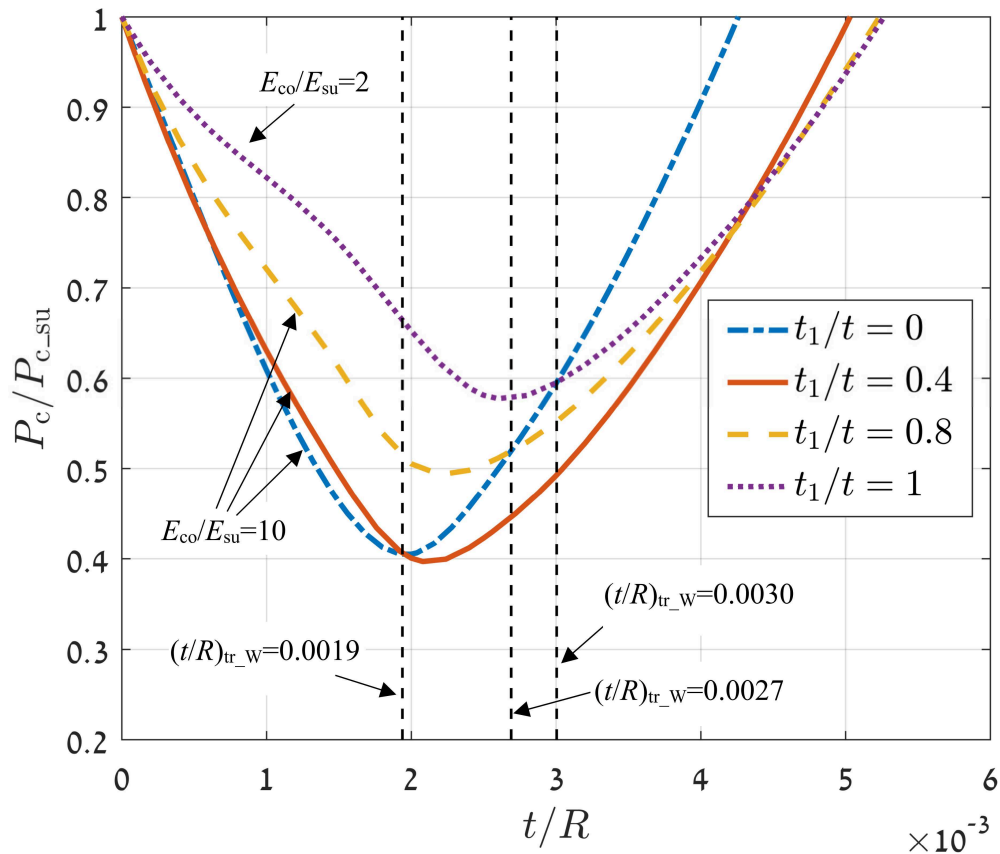


FIGURE 6 | Dimensionless critical load, $P_c/P_{c,su}$, vs. dimensionless coating thickness, t/R . Material properties as in **Table 2**.

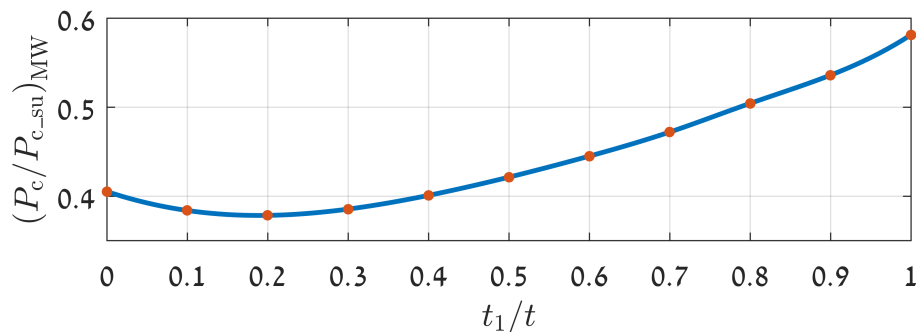


FIGURE 7 | Dimensionless critical load at maximum weakening, $(P_c/P_{c,su})_{MW}$, for different t_1/t . Material properties as in **Table 2**.

Figure 6 presents $P_c/P_{c_{su}}$ vs. t/R in the weakening zone for the same two typical cases of bilayer coated spheres, $t_1/t = 0.4$ and $t_1/t = 0.8$, as in **Figure 4**. Plots are also provided for single-layer coated cases with $t_1/t = 0$ and $t_1/t = 1$, which correspond to single-layer coatings with $E_{co}/E_{su} = 10$ and $E_{co}/E_{su} = 2$, respectively.

As shown in **Figure 6**, the single-layer case with higher mismatch ($E_{co}/E_{su} = 10$) has greater maximum weakening than the case with less mismatch ($E_{co}/E_{su} = 2$), consistent with Goltsberg and Etsion (2013). However, it is interesting to note that beyond a transition thickness $(t/R)_{tr_W} = 0.0030$ the single-layer case with high mismatch $E_{co}/E_{su} = 10$ shows less weakening than the single-layer case with small mismatch $E_{co}/E_{su} = 2$. The two bilayer cases with $t_1/t = 0.4$ and $t_1/t = 0.8$ also exhibit greater weakening than the single-layer case $t_1/t = 0$ for t/R beyond respective $(t/R)_{tr_W}$ values.

Further, in **Figure 6** it is noted that the extension of the weakening zone, in terms of t/R , is smaller in the single-layer case with greater mismatch ($E_{co}/E_{su} = 10$) than for the two bilayer cases and the single-layer case with less mismatch ($E_{co}/E_{su} = 2$). This suggests that while a lower mismatch at the $z/t = 1$ interface is beneficial in reducing the maximum weakening, it adversely extends the weakening zone to higher t/R values.

Considering the results in the Strengthening Effect of Bilayer Coated Spheres section, a bilayer coated sphere has a higher critical load than a single-layer case with the same outer-coating when t/R is either above $(t/R)_{tr_S}$ or below $(t/R)_{tr_W}$. Since the second case has to be avoided as being inside the weakening zone, the preferred selection for the total coating thickness of a bilayer case would be $t/R > (t/R)_{tr_S}$.

Figure 7 shows dimensionless critical load at maximum weakening, $(P_c/P_{c_{su}})_{MW}$, for bilayer coated spheres with different t_1/t . As can be seen from **Figure 7** $(P_c/P_{c_{su}})_{MW}$ hardly changes up to $t_1/t = 0.4$. Therefore, the bilayer effect in this range is negligible. For $t_1/t > 0.4$, the bilayer cases show less maximum weakening than the single-layer case with $t_1/t = 0$ but greater weakening than the single-layer case with $t_1/t = 1$.

In the weakening zone, at very low t/R , yield inception is within the substrate. For $t/R \geq (t/R)_{MW}$, yield inception is at the substrate side of the ML/S interface. This behavior was noted for all bilayer coated spheres with different t_1/t , and is similar to the behavior for single-layer coated spheres as described in Goltsberg and Etsion (2013).

In a previous study (Goltsberg et al., 2011) it was assumed that the Young's moduli mismatch is detrimental, in the weakening zone, due to a reduction in the compressive stresses at the substrate side of the coating/substrate interface. However, as seen in **Figure 6**, the single-layer case with high mismatch $E_{co}/E_{su} = 10$ exhibits less weakening than the other cases (that have less mismatch) at t/R beyond each relevant $(t/R)_{tr_W}$. From the many simulations, it was observed that above $(t/R)_{tr_W}$ reducing the mismatch at the $z/t = 1$ interface increases σ_1 more than σ_2 at the substrate side of the interface (at $x = 0$). From Equation (3) we can see that this increases σ_{eq} in the bilayer case compared to the single-layer case.

This suggests reducing the mismatch at the $z/t = 1$ interface reduces the σ_{eq}/Y at the substrate side of the interface only for low t/R (when t/R is less than the relevant $(t/R)_{tr_W}$).

TABLE 3 | Range of parameters for which similar behavior as described in the Strengthening Effect of Bilayer Coated Spheres section was observed ($Y_{co}/Y_{su} = E_{co}/E_{su}$ and $Y_m/Y_{su} = E_m/E_{su}$).

E_{co}/E_{su}	$(E_m/E_{su}-1)/(E_{co}/E_{su}-1)$	E_{su}/Y_{su}	ν
3 to 8	0.3 to 0.6	600 to 1200	0.32

Effect of Different Material Properties

Additional simulations were run for a range of material properties described in **Table 3**. Within this range, it is noted

- 1) The bilayer coated sphere exhibits greater strengthening than the corresponding single layer case only when t/R is greater than a transition thickness $(t/R)_{tr_S}$.
- 2) There exists optimum values of t_1/t and t/R for ultimate peak strengthening.

Hence, the general observations made here are likely to be applicable for the design of real coating systems, particularly those with E_{co}/E_{su} ratios within this range. For example, E_{co}/E_{su} for TiN outer-coating on Ti substrate is 3.33, and TiN outer-coating on Al substrate is 5.71 (Sun et al., 1995). Likewise, the ratio E/Y for most engineering metallic materials ranges from 100 to 2000, so the cases covered by **Table 3** are in line with practical applications.

Due to the large number of material and dimensional parameters, finding quantitative expressions between the material properties and the optimum values of t_1/t and t/R is beyond the scope of this paper. It shall be attempted in a future study.

CONCLUSION

Yield inception of bilayer coated spheres, with a middle-layer of intermediate Young's modulus and yield strength, flattened by a rigid flat was studied for a range of dimensionless thicknesses of the two layers. Both the strengthening and weakening effects were observed for bilayer coated spheres. It was expected that the gradual reduction of the Young's moduli mismatch in the bilayer coating would enhance the critical loads for yield inception over a single-layer coated sphere with the same outer-coating. However, the results show that within the strengthening zone, the presence of a middle-layer leads to an increase in the critical load only when the total dimensionless coating thickness is above a certain transition value, $(t/R)_{tr_S}$, which depends on the material properties and the dimensionless thickness of the middle-layer, t_1/t . Further, there is an optimum dimensionless middle-layer thickness t_1/t , which maximizes the peak critical load $(P_c/P_{c_{co}})_p$ of a bilayer coated sphere.

In addition, it is shown that there exists a $(t/R)_{tr_W}$ within the weakening zone above which the bilayer coated case experiences lower critical loads than the single-layer case with the same outer-coating. This implies that bilayer coatings are undesirable when $(t/R)_{tr_W} < t/R < (t/R)_{tr_S}$.

It is therefore recommended to select $t/R > (t/R)_{tr_S}$ for beneficial effect of bilayer coated spheres. Finding the expressions

for the optimum t_1/t and optimum t/R as functions of material properties requires a complex parametric analysis, which is out of the scope of the present paper but will be attempted in a future study.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

KP, ZC, and IE conceptualized, wrote, reviewed, and edited the manuscript. KP wrote the original draft. IE supervised.

REFERENCES

- Bar-Hen, M., and Etsion, I. (2017). Experimental study of the effect of coating thickness and substrate roughness on tool wear during turning. *Tribol. Int.* 110, 341–347. doi: 10.1016/j.triboint.2016.11.011
- Brizmer, V., Kligerman, Y., and Etsion, I. (2006). The effect of contact conditions and material properties on the elasticity terminus of a spherical contact. *Int. J. Solids Struct.* 43, 5736–5749. doi: 10.1016/j.ijsolstr.2005.07.034
- Chai, H., Lawn, B. R., and Wuttiphon, S. (1999). Fracture modes in brittle coatings with large interlayer modulus mismatch. *J. Mater. Res.* 14, 3805–3817. doi: 10.1557/JMR.1999.0514
- Dassault Systèmes (2016). *Abaqus. Analysis User's Guide*. Available online at: <http://ivt-abaqusdoc.ivt.ntnu.no:2080/v2016/books/usb/default.htm?startat=pt06ch32s05ael30.html>
- Djabella, H., and Arnell, R. D. (1993a). Finite element analysis of contact stresses in elastic double-layer systems under normal load. *Thin Solid Films* 223, 98–108. doi: 10.1016/0040-6090(93)90732-5
- Djabella, H., and Arnell, R. D. (1993b). Finite element comparative study of elastic stresses in single, double layer and multilayered coated systems. *Thin Solid Films* 235, 156–162. doi: 10.1016/0040-6090(93)90259-R
- Djabella, H., and Arnell, R. D. (1994). Finite element analysis of elastic stresses in multilayered systems. *Thin Solid Films* 245, 27–33. doi: 10.1016/0040-6090(94)90872-9
- Fontalvo, G. A., Daniel, R., and Mitterer, C. (2010). Interlayer thickness influence on the tribological response of bi-layer coatings. *Tribol. Int.* 43, 108–112. doi: 10.1016/j.triboint.2009.05.002
- Goltsberg, R., and Etsion, I. (2013). A model for the weakening effect of very thin hard coatings. *Wear* 308, 10–16. doi: 10.1016/j.wear.2013.09.019
- Goltsberg, R., and Etsion, I. (2015). A universal model for the load-displacement relation in an elastic coated spherical contact. *Wear* 322–323, 126–132. doi: 10.1016/j.wear.2014.11.002
- Goltsberg, R., Etsion, I., and Davidi, G. (2011). The onset of plastic yielding in a coated sphere compressed by a rigid flat. *Wear* 271, 2968–2977. doi: 10.1016/j.wear.2011.06.025
- Guo, Y. X., and Zhao, Y. W. (2019). Effect of interlayer on the elastic-plastic deformation of coating systems. *J. Mech.* 35, 1–8. doi: 10.1017/jmech.2018.46
- Gupta, P. K., and Walowit, J. A. (1974). Contact stresses between an elastic cylinder and a layered elastic solid. *J. Lubr. Technol.* 96, 250–257. doi: 10.1115/1.3451940
- Huang, X., Kasem, H., Shang, H. F., Shao, T. M., and Etsion, I. (2012). Experimental study of a potential weakening effect in spheres with thin hard coatings. *Wear* 296, 590–597. doi: 10.1016/j.wear.2012.08.018
- Komvopoulos, K. (1988). Finite Element Analysis of a layered elastic solid in normal contact with a rigid surface. *J. Tribol.* 110, 477–485. doi: 10.1115/1.3261653
- Komvopoulos, K. (1989). Elastic-plastic finite-element analysis of indented layered media. *J. Tribol.* 111, 430–439. doi: 10.1115/1.3261943
- Korchevnik, O., Goltsberg, R., Kligerman, Y., and Etsion, I. (2018). Electrical resistance model of a bilayer-coated spherical contact. *IEEE Trans. Compon. Packag. Manuf. Technol.* 8, 1614–1620. doi: 10.1109/TCPMT.2018.2854755
- Kot, M. (2012). Contact mechanics of coating-substrate systems: monolayer and multilayer coatings. *Arch. Civ. Mech. Eng.* 12, 464–470. doi: 10.1016/j.acme.2012.07.004
- Lardner, T. J., Giovino, R. J., and Ritter, J. E. (1992). The effect of high modulus single and double layer coatings on contact stresses. *Philos. Mag. A* 66, 437–455. doi: 10.1080/01418619208201568
- Liu, T. J., Xing, Y. M., and Wang, Y. S. (2016). The axisymmetric contact problem of a coating/substrate system with a graded interfacial layer under a rigid spherical punch. *Math. Mech. Solids* 21, 383–399. doi: 10.1177/1081286514555563
- Piao, Z. Y., Xu, B. S., Wang, H. D., and Pu, C. H. (2010). Effects of thickness and elastic modulus on stress condition of fatigue-resistant coating under rolling contact. *J. Cent. South Univ. Technol.* 17, 899–905. doi: 10.1007/s11771-010-0574-5
- Stephens, L. S., Liu, Y., and Meletis, E. I. (2000). Finite Element Analysis of the initial yielding behavior of a hard coating/substrate system with functionally graded interface under indentation and friction. *J. Tribol.* 122, 381–387. doi: 10.1115/1.555373
- Sun, Y., Bloyce, A., and Bell, T. (1995). Finite element analysis of plastic deformation of various TiN coating/substrate systems under normal contact with a rigid sphere. *Thin Solid Films* 271, 122–131. doi: 10.1016/0040-6090(95)06942-9
- van der Zwaag, S., Dear, J. P., and Field, J. E. (1986). The effect of double layer coatings of high modulus on contact stresses. *Philos. Mag. A* 53, 101–111. doi: 10.1080/01418618608242810
- van der Zwaag, S., and Field, J. E. (1982). The effect of thin hard coatings on the Hertzian stress field. *Philos. Mag. A* 46, 133–150. doi: 10.1080/01418618208236213
- Yu, C., Wang, Z., Liu, G., Keer, L. M., and Wang, Q. J. (2016). Maximum von Mises stress and its location in trilayer materials in contact. *J. Tribol.* 138:041402. doi: 10.1115/1.4032888

All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGMENTS

This paper is part of IEA AMT IA technical activities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmech.2020.00023/full#supplementary-material>

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Parel, Chen and Etsion. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

NOMENCLATURE

E	Young's modulus
P	Normal load applied to the rigid flat
R	Radius of the substrate
t	Total thickness of the two layers combined
t_1	Thickness of the middle-layer
t_2	Thickness of the outer-coating
$(t/R)_{tr_S}$	Transition t/R value in the strengthening zone at which two different cases have the same dimensionless critical load, P_C/P_{C_CO}
$(t/R)_{tr_W}$	Transition t/R value in the weakening zone at which two different cases have the same dimensionless critical load, P_C/P_{C_SU}
Y	Yield strength
z	Distance along the axis of symmetry from the contact interface
ν	Poisson's ratio
σ_{eq}	Equivalent von Mises stress
$\sigma_1, \sigma_2, \sigma_3$	Principal stresses.
<i>Subscripts</i>	
co	Outer-coating material
c_co	Value at critical load for homogenous sphere made of the outer-coating material
c_su	Value at critical load for homogenous sphere made of the substrate material
m	Middle-layer material
MW	Value at maximum weakening
p	Value at peak critical load
su	Substrate material



Studies on the Influence of Residual Stresses on the Fatigue Life of Rolling Bearings in Dependence on the Production Processes

Florian Pape*, Timm Coors and Gerhard Poll

Institute of Machine Design and Tribology, Leibniz University Hannover, Hanover, Germany

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Jose Daniel Biasoli De Mello,
Federal University of Uberlandia, Brazil
Árpád Czifra,
Óbuda University, Hungary

*Correspondence:

Florian Pape
pape@imkt.uni-hannover.de

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 16 March 2020

Accepted: 16 June 2020

Published: 29 July 2020

Citation:

Pape F, Coors T and Poll G (2020)
Studies on the Influence of Residual
Stresses on the Fatigue Life of Rolling
Bearings in Dependence on the
Production Processes.
Front. Mech. Eng. 6:56.
doi: 10.3389/fmech.2020.00056

The production process significantly influences the surface properties of rolling element bearings raceways. Deep rolling can induce a depth dependent residual stress state. Previous numerical and experimental studies have shown that rolling bearings fatigue life can be positively influenced by high compressive residual stress to a depth of around 300 μm from the surface. By extending the components life, the resource efficiency of machine components can be increased. In order to determine the influence of the residual stress state in bearing fatigue life, a calculation method was developed for predicting the bearing fatigue life. This method was validated for hard-turning and subsequent deep rolling by experiments on a test rig in four-bearing configuration under radial load. An increase of the L_{10} bearing life by a factor of 2.5 has been achieved by inducing residual stresses on the bearing's inner ring. Due to similar process control, the manufacturing steps turning and deep rolling were combined. Bearings were manufactured combining the processes hard-turning and deep rolling (called *turn-rolling*). The heat from the hard machining has an effect on the residual stresses in the bearing subsurface, thus further altering the magnitude and maximum depth of the residual stress influencing the microstructure. With these bearings, the additional fatigue life was determined experimentally and compared to the results of the bearings produced by hard-turning and subsequent deep rolling. It could be shown, that the process of hard turning and subsequent deep rolling has highest potential to achieve improved bearing fatigue life. These findings were transferred to a "Tailored Forming" shaft with integrated raceway in a second step. In this case, a shaft made of mild steel is combined with a cladding layer of high strength bearing steel to be used as a bearing raceway.

Keywords: cylindrical roller bearing, fatigue life, residual stresses, deep rolling, turn-rolling

INTRODUCTION

A high reliability of machine elements under relative motion is of main importance. The dimensions of machine elements are calculated for reliability. Bearings properties as typical highly loaded machine elements are based on proven conventional production techniques. During operation bearings have to resist very high cyclic stresses. The fatigue life of rolling bearings is influenced in particular by the raceway properties. Böhmer provides an overview of the load conditions and material behavior of rolling contacts in Böhmer (1998). The author defines a large

number of influencing variables on component fatigue life: surface micro topography, especially under mixed lubrication, lubricant and its contamination, residual stresses, and temperature effects as well as the material and the heat treatment processes. The cyclical rolling of the rolling elements over the raceways of inner and outer ring causes alternating stresses in the material volume of the contact partners. This can finally result in the nucleation of fatigue cracks in the very high cycle regime, which is known as rolling contact fatigue (RCF). If the alternating stress continues, crack growth processes occur on the boundary layer resulting in the separation of particles from the surface. Under ideal lubrication conditions the resulting chips, known as pittings, are one of the main causes of rolling bearing failure in practical applications (Glaeser and Shaffer, 1996).

Zwirlein and Schlicht (1989) describe the development of residual stresses due to the actual over-rolling process. They also describe the resulting equivalent stresses in the Hertzian contact zone as a function of existing residual stresses and friction occurring at the transition from full film lubrication to mixed lubrication. Under the influence of friction, the maximum of the equivalent stress of the multi-dimensional stress state below the running surface shifts increasingly in the rolling direction and toward the surface.

According to Voskamp, the material zone beneath the overrolled raceway goes through three phases over the period of the bearing's fatigue life in which a change of the microstructure takes place (Voskamp, 1985). In the first phase, named "shakedown phase," which is occurring within the first 1,000 revolutions, plastic micro-deformations take place in the area below the raceway surfaces, leading to work hardening of the bearing material. It is characterized by the rapid reduction of residual austenite trapped in the martensitic structure of the bearing material and the build-up of residual compressive stresses. The extent of the changes depends on the load acting on the bearing. In the subsequent stationary second phase, also called steady-state phase, practically no plastic deformation and no structural change takes place. The duration of the stationary phase depends on the level of the load. For loads exceeding the "shakedown limit," there is a direct transition from the first to the third phase. For small loads, the fatigue process remains in the second phase and no transition to the third phase takes place, and thus no component failure due to RCF occurs. In the third phase, the so called "instability phase," the component subsurface can no longer absorb the acting loads in the form of elastic deformation. As a result of the cyclic contact conditions, a further change in the complex residual stress state of the microstructure occurs. The micro plastic deformations evolve to macro plastic deformations and result in component failure. Hollox et al. were able to detect the course and magnitude of the residual stresses below the raceways as a function of the running times for radially loaded deep groove ball bearings by X-ray diffraction (XRD) (Hollox et al., 1987). The alignment of the crystal structure as a result of the rolling stress and thus the formation of a texture in the subsurface of the rolling bearing was also determined and recognized as a characteristic feature of the third phase of the fatigue process. The development of residual stresses in near-subsurface areas of the raceways during

cyclic over rolling was also proven by theoretical investigations by (Hills, 1983).

Voskamp later experimentally determined an increase in bearing fatigue life by subjecting the inner rings to a running-in process (Voskamp, 1996). During this running-in process, the bearings were subjected to high loads for a short time of over rolling. In the subsequent fatigue life experiments under normal load, the fatigue life was increased by a factor of three in comparison to standard bearing inner rings (which were not run-in) could be determined. Voskamp carried out these investigations under full film lubrication.

Poll and Hacke investigated the effect of a short time of overload during the running in on cylindrical roller bearings of the size NU206 and NJ206 (Poll and Hacke, 2010). The bearings were subjected to two different loads: a short phase of high load ($C/P = 2.5$ Hertzian pressure $p_m = 3,000 \text{ N/mm}^2$) as an overloading and subsequent operation under normal load ($C/P = 4$, $p_m = 2,500 \text{ N/mm}^2$). In the statistical evaluation of the tests, it was determined that the running times of the collective tests with run-in at increased load were higher than those of the pure single-stage tests. The increased bearing fatigue life is assumed to be the result of residual compressive stresses introduced in the running-in phase.

Besides a preloading of the bearings, there is also the possibility of applying pre-induced residual stresses by the manufacturing process. A relatively new method is laser shot peening (Montros et al., 2002). By usage of a high energy laser pulse, a plasma shock wave is generated on a metallic surface. When the shock wave is transmitted into the steel surface, a plastic deformation results, which leads to residual compressive stresses.

Altenberger emphasizes the positive influence of temperature on SAE 1045 and AISI 304 during deep rolling (Altenberger, 2005). In case of AISI 304 a much higher fatigue lifetime improvement compared to ultrasonic shot peened, deep rolled, and laser shock peened samples could be achieved. In similar investigations it could be shown that by means of mechanical surface treatments, especially deep rolling, also for non-steel materials such as aluminum-7075-T6 (Majzoobi et al., 2009) and titanium Ti-6Al-4V (Nalla et al., 2003) an increase in fatigue life could be achieved.

Pabst was also able to prove an increase in the bearing fatigue life of rolling bearings which had been modified by deep-rolling in order to induce compressive residual stresses (Pabst, 2018). Additional shot peening was investigated. Pabst assumed that the residual stress state did not increase the bearing fatigue life as much. Rather, the subsequent induction hardening process and its influence on material microstructure increased the fatigue life. He mentioned that the magnitude of the measured residual stresses can serve as an indirect assessment criterion. The machining processes results in a finer-grained microstructure with positive properties with regards to deformation and fracture.

The effect of an enhanced induced residual stress state by deep rolling of cylindrical roller bearings (CRB) was investigated by Neubauer (2016). The deep rolling process can be enhanced in the depth of the achieved residual stresses by using the effect of heat. In this case the heat from the turning process can be used

if the rolling ball is placed directly behind the turning tool (Pape et al., 2017).

METHODOLOGY

The aim of this study is to provide guidelines for optimized manufacturing processes for so called “Tailored Forming” machine elements such as multi-material shafts with integrated raceways for CRBs, as used in Coors et al. (2018). Thus, the here presented results based on a former project on the influence of production processes on the fatigue performance of machine elements are elaborated and transferred. The scheme of a Tailored Forming shaft is shown in **Figure 1**. In this case, a novel process chain for the production of load-adapted machine elements using steel-steel multi-material systems has been developed. A bearing steel with a high resistance to wear and fatigue like 100Cr6 (material number: 1.3505/SAE 52100) or better, is applied by plasma powder deposition welding on low-cost steel shafts, c.f. (Golovin et al., 2019). This allows for the local deposition of high strength steel, which serves as the raceway for a rolling element bearing and covers the fatigue life-determining material volume regarding rolling contact fatigue, c.f. (Behrens et al., 2019). The base material allows for basic requirements regarding structural load-bearing capacity, thus serving as a support structure and for the mechanical connection to other components. This hybrid workpiece is then formed by cross wedge rolling to achieve a fine grained microstructure, c.f. (Kruse et al., 2019). The finishing process is carried out by local induction hardening and hard-turning, and will be extended by deep rolling, c.f. (Denkena et al., 2019). In order to optimize the finishing process, the production steps and their effect on the bearing service life will be examined more closely.

TEST SPECIMEN PREPARATION AND TEST EQUIPMENT

For the investigations, CRB type NU206 were used as exemplary object of investigation for fatigue life effects of residual compressive stresses regarding RCF. Bearing inner rings were machined at the Institute of Production Engineering and Machine Tools, Leibniz University Hannover, Germany as shown

in Denkena et al. (2016a). The bearings are made of 100Cr6 steel, and feature a hardness of 62 HRC. For the hard-turning process, carbide inserts with $\text{Al}_2\text{O}_3 + \text{Ti}(\text{C},\text{N})$ coating were chosen. A hydrostatic rolling tool was used for the deep rolling process. In this case the bearings were prepared in a first step by hard-turning and subsequently the bearings were deep rolled. The combination of hard turning and deep rolling leads to an improved surface quality (Maiß et al., 2017). Due to the high hardness of the bearing material, the surface roughness from the turning process is not completely eliminated by the deep rolling process. Both processes influence each other and must be coordinated. In the deep rolling process, the roughness peaks are plastically deformed. This reduces the maximum profile height of the roughness, which in turn is advantageous for surfaces subject to tribological load. For the process, rolling pressure, ball diameter, and rolling feed are taken into account. The main process control variable is the degree of coverage, i.e., partial rolling over the same material area several times.

Additionally, a combination of the turning and deep rolling processes was developed, as shown by Denkena et al. (2016b) and Maiß et al. (2016). In this case, a deep rolling tool is combined with a cutting insert in one tool, merging the aforementioned sequential steps in one to increase efficiency (Maiß et al., 2016). The process was compared to hard turning and subsequent deep rolling by Denkena et al. illustrating the feasibility of the approach (Denkena et al., 2016b). The tool additionally allows a more precisely alignment of the deep rolling ball on the surface toward the cutting insert. The so called turn-rolling concept for machining of the inner rings is shown in **Figure 2**. The ball diameters ($d_k = 3.175$ or 6.35 mm) and the insert corner radius ($r_\beta = 1.6$ mm) have been adjusted to achieve the positioning of the deep rolling ball in or on the turning grooves in the feed direction. The pressure on the rolling ball (rolling pressure p_w) is supplied by oil. The rolling tool allows the induction of subsurface residual stresses. In combination with the heat generated by the cutting insert, the depth of the residual stresses can be increased.

To study the influence of the production processes on fatigue life, bearing inner rings of type NU206 were manufactured with the aforementioned turn-rolling process. These were mounted in a test rig for rolling bearings in four-bearing configuration and operated until failure, see **Figure 3**. The load is applied

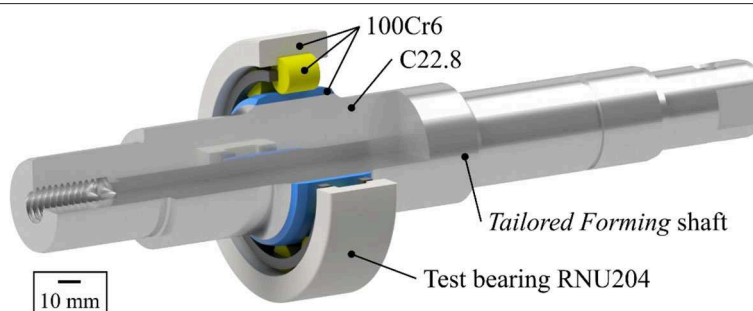


FIGURE 1 | Tailored Forming shaft.

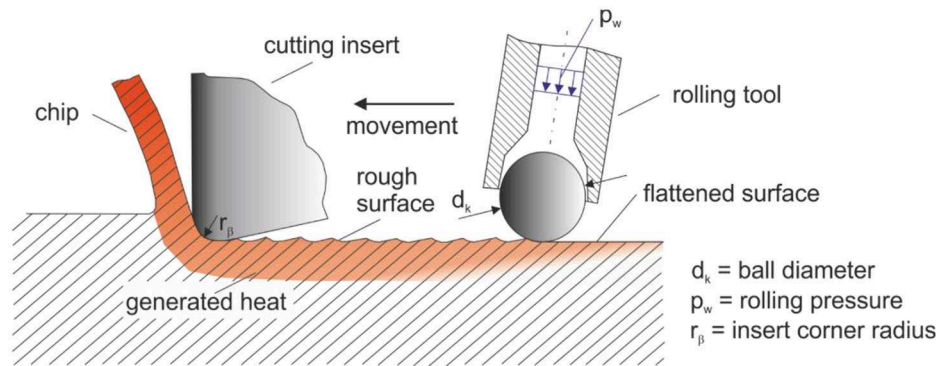


FIGURE 2 | Hybrid process tool for turn-rolling.

via disk springs, which act on a hollow shaft onto the both inner bearings. The symmetrical arranged outer bearings transfer the load toward the housing. The applied load is measured with a load cell during the tests. The radial load was chosen with $C/P = 4$, which is equivalent to a Hertzian pressure of $p_m = 2,500$ MPa. The bearings are directly supplied with oil. The oil is tempered by a process thermostat to achieve permanently constant lubrication conditions. The tests were carried out under fully flooded lubrication conditions at 60°C . Synthetic oil of the viscosity $\nu_{40} = 68$ mm²/s was used. The bearing ran with a rotational speed of $n = 4,050$ min⁻¹. The lubrication parameter is calculated as $\lambda = 3$. Piezoelectric vibration transducers measure the vibrations and structure-borne sound in the load direction. By means of a self-developed condition monitoring system, an envelope curve analysis is carried out during the runtime of the experiment. If a fatigue damage in the rolling bearing develops to the point of surface chipping, the evaluation of the shock pulse repetition frequencies in the envelope curve spectrum allows conclusions to be drawn about bearing damage that has occurred. The test is aborted by a previously defined switch-off criterion of exceeding the steady state signal by 150%.

The bearings were investigated regarding the subsurface residual stress state. Doing so, X-ray diffraction measurements were performed at the Institute of Production Engineering and Machine Tools, Leibniz University Hannover, Germany. A two circle Bragg-Brentano-Diffractometer System (Type GE XRD 3003TT and Seifert XRD 3000P) with CrK α radiation was used for the measurement. For the depth profiling, the material on the surfaces was removed by electrolytical polishing.

RESULTING RESIDUAL STRESS FOR DIFFERENT FINISHING PROCESSES

Part of the current investigation is focused on how favorable boundary zone properties can be adjusted with respect to the residual stress state. For this purpose, the methods hard-turning, hard-turning with subsequent deep rolling, and turn-rolling were in the scope. To quantify changes of the boundary zone properties as a function of the test duration,

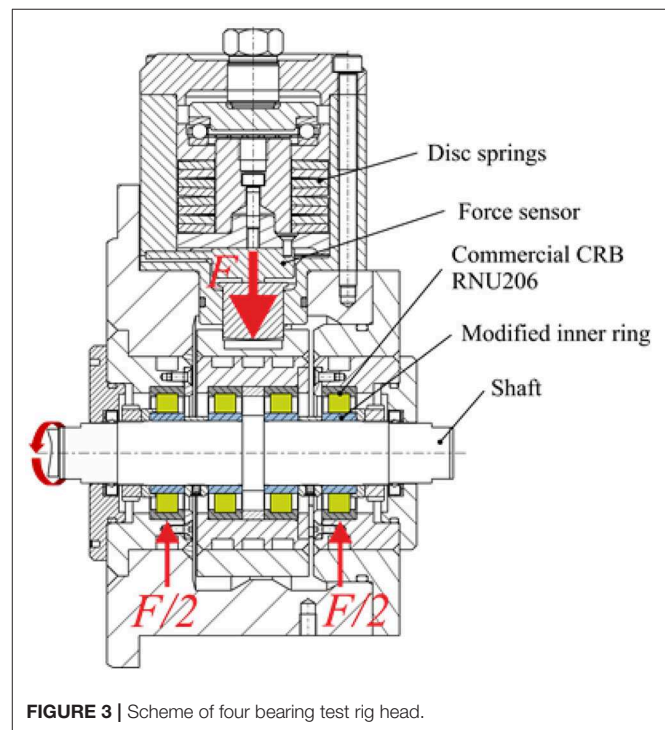


FIGURE 3 | Scheme of four bearing test rig head.

the bearings were investigated in terms of XRD after defined running times.

In order to form a reference, standard bearings of the type NU206 C3 with increased internal clearance were examined. The raceways of the bearings were honed in the final machining process; the inner rings have a crowned profile. The maximum pressure in crowned roller contact between inner ring and rolling element is $p_m = 2,500$ MPa. In new condition, residual compressive stresses of about $\sigma_{\text{circumference}} = -500$ MPa are present on the surface, which are almost eliminated at a depth of about $z = 20$ μm as shown in Figure 4A. Within the first 3.9 million revolutions, the rolling bearings investigated show significant changes in the surface and subsurface properties during operation, the roughness peaks on the surfaces were

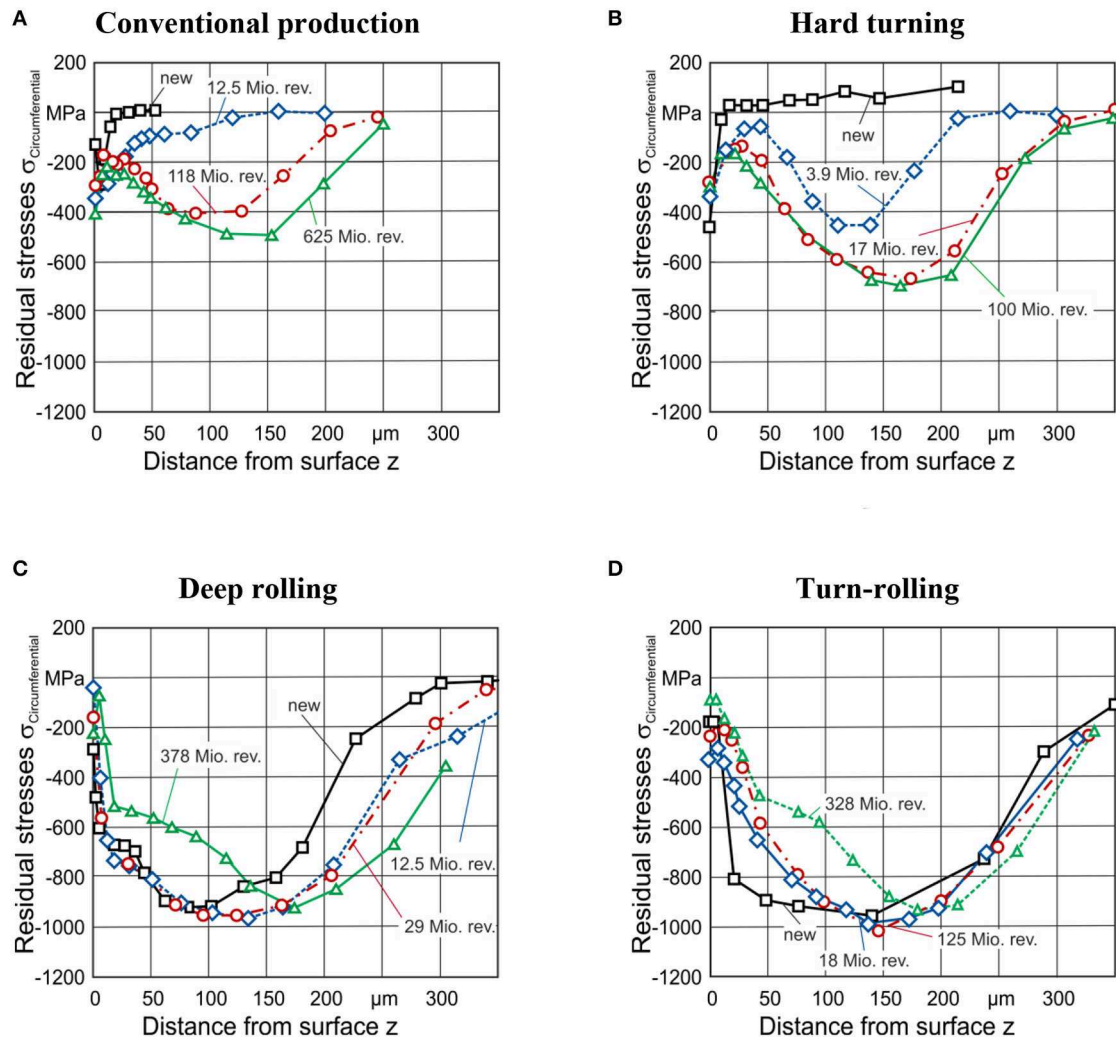


FIGURE 4 | Comparison of residual stress depth profiles at different running times depending on the respective production strategy for NU206 inner rings and a bearing load of $C/P = 4$. **(A)** Reference from conventional large series production, **(B)** Hard turning, **(C)** Hard turning with subsequent deep rolling, **(D)** Combined turn-rolling.

plastically deformed and residual stresses were built up in the depth. **Figure 4** also depicts the development of the residual stress state over the bearing test duration. The residual stresses have developed at the depth of the maximum shear stresses and then remain constant at this level after 10^6 to 10^8 revolutions (Pape et al., 2017).

For the hard-turned bearings, the process had to be advanced to achieve bearings of a sufficient low surface roughness to allow for the bearing tests (Neubauer et al., 2013). It was shown that a low surface roughness is required in order to avoid surface-induced early bearing failures. With bearings manufactured with a sufficiently low roughness, results comparable to standard bearings could be achieved. It could be shown that in the running-in period the roughness has been reduced and residual stresses have increased. The changes of the residual stress

over test duration are shown in **Figure 4B**. Due to hard turning process, the surfaces of the bearing inner rings do not have a crowned surface, so that a lower maximum Hertzian pressure of $p_{\text{max}} = 2,300$ MPa is present. As a result, slightly lower residual stresses were induced in the bearings (Denkena et al., 2016b; Pape et al., 2017).

In comparison, hard-turned and subsequent deep rolled bearing rings were examined; the residual stress depth curves after production and after varied running times were compared (**Figure 4C**). The manufacturing process induced a higher residual stress state than a running in phase would have. For the first 29 million revolutions, the residual stresses up to a depth of $z = 120 \mu\text{m}$ are comparable to those directly after deep rolling. At greater depths, there is a slight increase in residual stresses. After 378 million revolutions, a reduction of the residual stresses can be observed in the

near subsurface area (Denkena et al., 2016b; Pape et al., 2017).

The development of the residual stresses for the turn-rolled bearings are presented in **Figure 4D**. After manufacturing the residual stress state is shifted to a higher depth and, compared to the deep rolled bearings, the magnitude of the residual stresses could be increased. It can be seen that, within the first 18 million and 125 million revolutions, the residual stresses remain stable. After 328 million revolutions, a reduction of the residual stresses can be observed in the near-surface area.

BEARING FATIGUE LIFE REGARDING MODIFIED RESIDUAL STRESS STATES

The bearings underwent endurance tests. For the evaluation the slightly different contact pressure due to the hard-turning process has to be regarded. A bearing fatigue life of $L_{10} = 274$ h and a Weibull slope of 0.6 could be determined for the standard bearings. In this case the bearing inner rings feature a crowned profile and thus higher Hertzian pressure in the contact between inner ring and rollers. In case of the hard-turned bearings a lifetime of $L_{10} = 771$ h was determined. The hard-turned and subsequently deep rolled bearings exhibit a bearing fatigue life $L_{10} = 1,924$ h. Due to the deep rolling process, the service life could be increased by a factor of 2.5 (Neubauer, 2016). This increase can be explained by the induced residual stresses, which reach their maximum approximately at the depth of the highest load induced stresses. The turn-rolled bearing feature a $L_{10} = 471$ h which is even below the fatigue life of the hard-turned bearings. Although in a first view, the pre-induced residual stress state should allow an increased bearing fatigue life, so the process was not beneficial. The Weibull distribution at a load of $C/P = 4$ are shown for the hard-turned, deep rolled, and turn-rolled bearings in **Figure 5**.

The results of the bearing fatigue tests show that the bearings can be significantly influenced by changes of the production step properties. The test results of the hard-turned bearings prove that this alternative machining process to the standard honing process allows the achievement of comparable fatigue lives. However, it should be noted that the hard-turned bearing rings have a different bearing clearance and profile than the reference bearings. A direct comparison can be done to the deep-rolled bearings as well as the turn-rolled bearings due to the comparable turning process with the same bearing clearance and the same profile. The steel rods of the hard-turned, deep-rolled, and turn-rolled bearings, in contrast to the reference bearing rings, come from the same material batch.

To obtain influences on the fatigue of the surface of the bearings, the development of the direction of the residual stresses before and after testing was inspected as shown in **Figure 6**. In case of the hard-turned bearing after production, only the first $10\ \mu\text{m}$ experience a measurable residual stress condition. Due to RCF loading, after 18 million revolutions a residual stress state is induced to the material. For the first $25\ \mu\text{m}$, the circumferential and axial residual stress state differs, for higher depths both values deviate slightly. In comparison, the stress state for a deep rolled

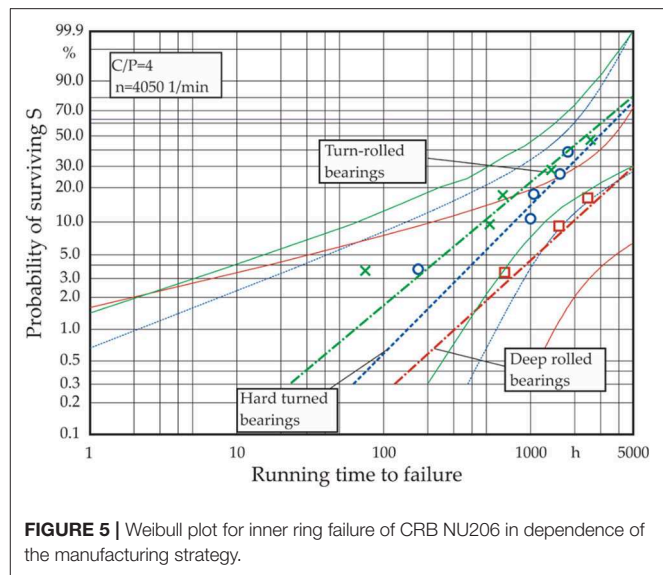


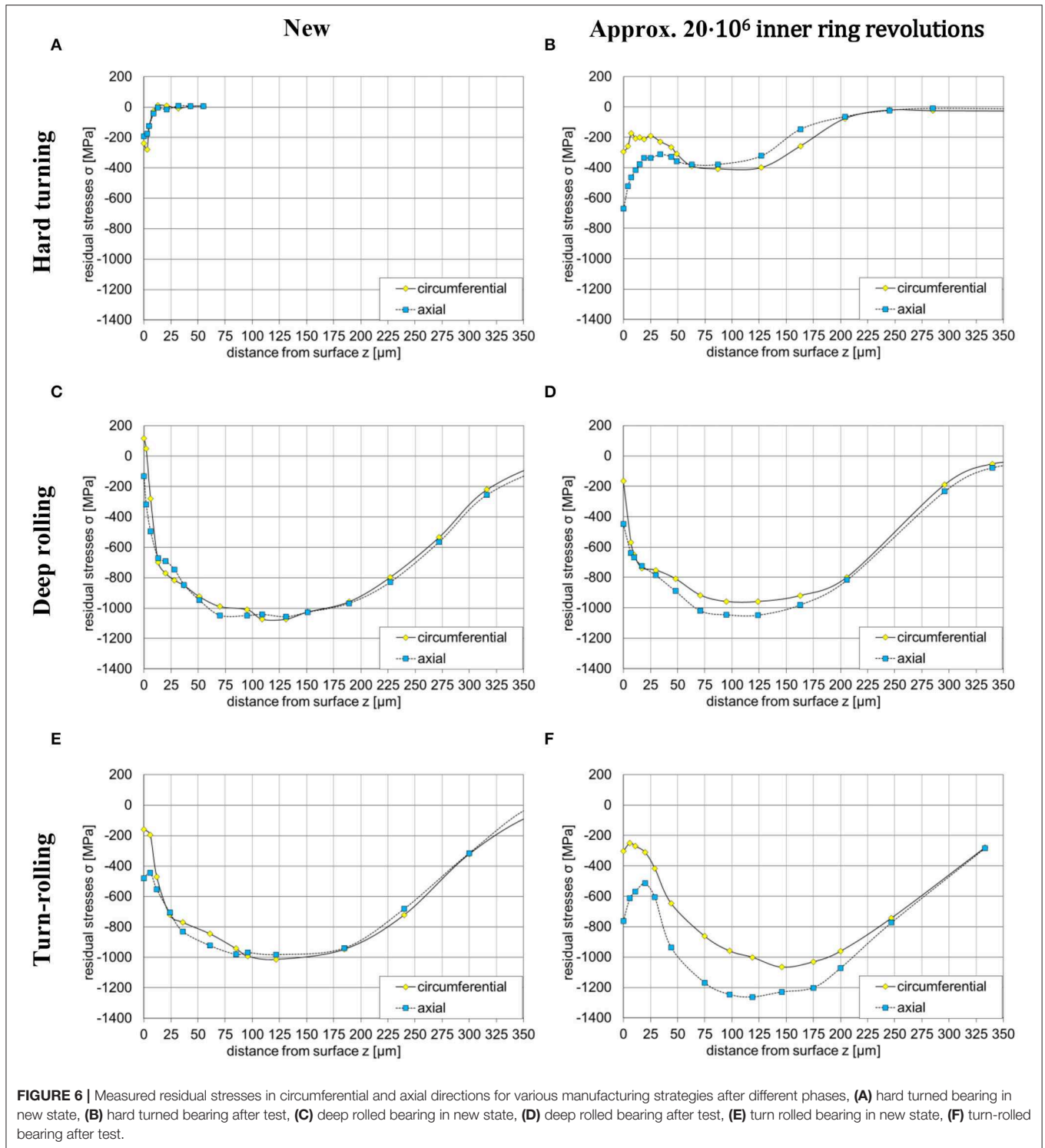
FIGURE 5 | Weibull plot for inner ring failure of CRB NU206 in dependence of the manufacturing strategy.

bearing in circumferential and axial direction remained equal even after 290 million revolutions. Even for very long running times (595 million revolutions) the circumferential and axial residual stress state remained equal.

For the turn-rolled bearing, the circumferential and the axial residual stress states before test are nearly equal. For a bearing inner ring removed at an early stage of 18 million revolutions, the circumferential and axial residual stress states differ significantly. Also the induced residual stress state in axial direction is increased up to magnitudes of more than $-1,200$ MPa. The distribution and difference between circumferential and axial residual stresses could also be proven for longer bearing running times. It could be shown that the material of the turn-rolled bearing inner ring exhibits anisotropic behavior regarding the orientation of the residual stress state. It is assumed that the heat influence causes an alignment of the surface structure which compensates the positive effect of the rotary rolling process. Additional a cross section for the turn rolled bearing was investigated (**Figure 7**). The cross section of a turn rolled bearing inner ring before test shows martensitic texture (**Figure 7A**). After test (125 million revolutions) the textures features darker spots due to carbon redistribution (**Figure 7B**).

FATIGUE LIFE CALCULATION REGARDING RESIDUAL STRESSES

To consider the influence of residual stresses on the fatigue life of bearings, a finite element model of a bearing inner ring segment was prepared and combined with a calculation routine. The external bearing load is added to the bearing ring as three-dimensional Hertzian pressure. The profile of the bearing inner ring is measured and taken into account in the modeling. It allows the three-dimensional stress state under the raceway surface to be determined. In doing so, the external loads



are superimposed with an additional residual stress state. The procedure is described by Neubauer (2016). The determines the von Mises equivalent stresses, the orthogonal shear stresses, and the hydrostatic stresses as a result. The resulting stresses are used as input to the lifetime model based on Ioannides and Harris (1985). This approach regards a stress fatigue limit as well as

a fatigue stress criterion. The probability of survival S can be obtained by the following equation (Ioannides et al., 1999):

$$\ln\left(\frac{1}{S}\right) \approx N^e \int_V \frac{(\tau_i - \tau_u)^c}{z'^h} dV$$

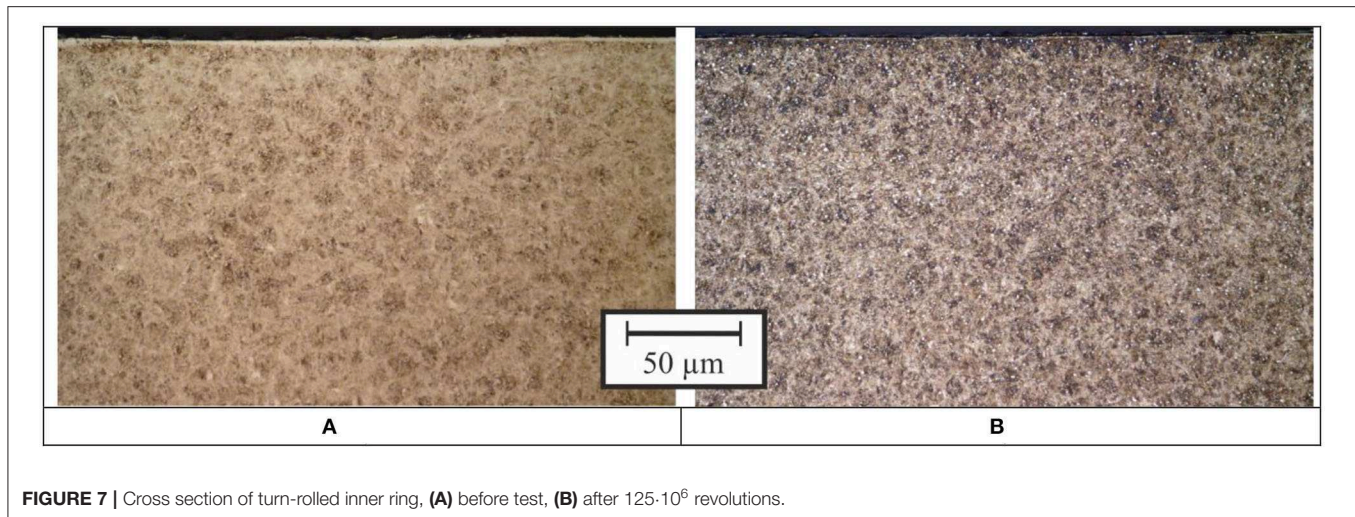


FIGURE 7 | Cross section of turn-rolled inner ring, (A) before test, (B) after $125 \cdot 10^6$ revolutions.

The concept considers the magnitude and depth of the load induced stresses as well as the stressed volume V . In the formula S is the probability of survival: N stands for the number of load cycles, and z' is the stress-weighted depth from the surface. The exponents e , c , and h can be obtained from the literature (Ioannides et al., 1999): $e = 9/8$ (for line contact), $c = 1/3$ (exponents of Lundberg and Palmgren Lundberg and Palmgren, 1947 and $h = 7/3$. T_u represents the stress fatigue limit of the raceway material. For the stress related fatigue criterion, T_i , the Dang Van criterion Dang Van et al., 1989 is inserted: $\tau_i = \tau_{Omax} - k_{hyd} \cdot p'_{hyd}$. This can be computed with the max. shear stress τ_{Omax} , k_{hyd} as a weighting factor for the hydrostatic stress p'_{hyd} , which includes compressive stresses. This model was evaluated with tests and allows the calculation of the bearing fatigue life with consideration of the residual stresses (Denkena et al., 2016b; Pape et al., 2018). The calculated bearing fatigue life $L_{10,mod}$ are nearly equal for the deep-rolled and turn-rolled bearing rings or the same rolling tool ($d_k = 6.35$ mm). In both cases similar residual stress states are induced below the surface of the bearing inner ring. As in the Dang Van criterion, the orthogonal shear stresses are corrected by the residual stress state and, in both cases, an increase in the bearing fatigue life can be expected. The calculation does not regard any material anisotropy of the subsurface zone due to the turn-rolling process. Due to the modified process of turn-rolling involving a subsurface heat profile, it can be expected that the subsurface features anisotropic material behavior which should be regarded in future investigations.

CONCLUSION

Pre inducing a subsurface residual stress state in rolling bearings allows achieving an increased bearing fatigue life. This could be shown by experimental validation and in terms of a calculation approach based on Ioannides and Harris regarding the Dang Van fatigue criterion. By hard-turning and subsequent deep rolling a beneficial residual stress state in the depth of the highest load

induced stresses was achieved. The residual stress state remains constant for very long bearing running times and the bearing fatigue life could be extended.

Despite this, a combined tool with the combination of hard-turning and deep rolling resulted in disadvantageous subsurface structure and slightly reduced bearing fatigue life compared to hard-turned bearings. It can be concluded that, due to the anisotropic heating during turn-rolling, the surface fine structure features anisotropic material behavior, resulting in an increased modification of the residual stresses. On the one hand, the residual stress state with a value higher than $-1,200$ MPa results in surface degradation. On the other hand, the differences in the residual stress direction can result in undesirable stress superposition in case of cyclic over rolling. There is demand to investigate this modification by ongoing studies and investigations on grain size.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/supplementary material.

AUTHOR CONTRIBUTIONS

FP and TC performed tests and researched on the influences of ESP on RCF and on the topic Tailored Forming. The contribution was written by FP and TC under supervision of GP. All authors contributed to the article and approved the submitted version.

FUNDING

Funding by the DFG (German research Foundation) within the research program Resource efficient Machine Elements (SPP1551) and the SFB 1153 Tailored Forming—Grant No. 252662854.

ACKNOWLEDGMENTS

The authors thank the DFG (German Research Foundation) for supporting this project in the context of the research

program the SFB Tailored Forming in the subproject C3. The authors would also like to thank apl. Prof. Breidenstein's team for carrying out the residual stress measurements using XRD.

REFERENCES

- Altenberger, I. (2005). "Deep rolling—the past, the present and the future," in *9th International Conference on Shot Peening (ICSP9)* (Paris), 144–155.
- Behrens, B.-A., Chugreev, A., Matthias, T., Poll, G., Pape, F., Coors, T., et al. (2019). Manufacturing and evaluation of multi-material axial-bearing washers by tailored forming. *Metals* 9:232. doi: 10.3390/met9020232
- Böhmer, H. J. (1998). Wälzverschleiß und -ermüdung von Bauteilen und Maßnahmen zu ihrer Einschränkung. *Materialwiss. Werkstofftech.* 29, 697–713. doi: 10.1002/mawe.19980291203
- Coors, T., Pape, F., and Poll, G. (2018). Bearing fatigue life of a multi-material shaft with an integrated raceway. *Bear. World J.* 3, 23–30.
- Dang Van, K., Griveau, B., and Message, O. (1989). *On a New Multiaxial Fatigue Criterion: Theory and Application*. London: Mechanical Engineering Publications, 479–496.
- Denkena, B., Bergmann, B., Breidenstein, B., Prasanthan, V., and Witt, M. (2019). Analysis of potentials to improve the machining of hybrid workpieces. *Prod. Eng.* 13, 11–19. doi: 10.1007/s11740-018-00870-3
- Denkena, B., Grove, T., and Maiß, O. (2016a). Influence of hard turned roller bearings surface on surface integrity after deep rolling. *Proc. CIRP* 45, 359–362. doi: 10.1016/j.procir.2016.02.047
- Denkena, B., Poll, G., Maiß, O., Pape, F., and Neubauer, T. (2016b). Enhanced boundary zone rolling contact fatigue strength through hybrid machining by hard turn-rolling. *Bear. World J.* 1, 87–102.
- Glaeser, W. A., and Shaffer, S. J. (1996). "Contact fatigue," in *ASM Handbook, Vol. 19* (Materials Park, OH: ASM Handbook Committee), 331–336.
- Golovin, P. A., Mildebrath, M., and Zemlyakov, E. V. (2019). Effects of PTA deposition parameters on coating quality of 41Cr4, 50CrV4 and 100Cr6 steel. *Key Eng. Mater.* 822, 801–808. doi: 10.4028/www.scientific.net/KEM.822.801
- Hills, D. A. (1983). Some aspects of post-yield contact problems. *Wear* 85, 107–119. doi: 10.1016/0043-1648(83)90339-3
- Hollox, G. E., Voskamp, A. P., and Ioannides, E. (1987). Ermüdungssichere Auslegung von Wälzlagerkomponenten. *Kugellagerzeitschrift* 62, 20–27.
- Ioannides, E., Bergling, G., and Gabelli, A. (1999). "An analytical formulation for the life of rolling bearings," in *Acta Polytechnica Scandinavica*. Mechanical engineering series (Espoo; Stockholm: Finnish Acad. of Technology).
- Ioannides, E., and Harris, T. A. (1985). A new fatigue life model for rolling bearings. *ASME J. Tribol.* 107, 367–337. doi: 10.1115/1.3261081
- Kruse, J., Mildebrath, M., Behrens, B.-A., Stonis, M., and Hassel, T. (2019). Cross-wedge rolling of PTA-welded hybrid steel billets with rolling bearing steel and hard material coatings. *AIP Conf. Proc.* 2113:040019. doi: 10.1063/1.5112553
- Lundberg, G., and Palmgren, A. (1947). "Dynamic capacity of roller bearings," in *Acta Polytechnica, Mechanical Engineering Series, Vol. 1, No. 3, Royal Swedish Academy of Engineering Sciences* (Stockholm).
- Maiß, O., Denkena, B., and Grove, T. (2016). Hybrid machining of roller bearing inner rings by hard turning and deep rolling. *J. Mater. Process. Technol.* 230, 211–216. doi: 10.1016/j.jmatprotec.2015.11.029
- Maiß, O., Grove, T., and Denkena, B. (2017). Influence of asymmetric cutting edge roundings on surface topography. *Prod. Eng. Res. Dev.* 11, 383–388. doi: 10.1007/s11740-017-0742-7
- Majzoubi, G. H., Azadikhah, K., and Nemati, J. (2009). The effects of deep rolling and shot peening on fretting fatigue resistance of aluminum-7075-T6. *Mater. Sci. Eng. A* 516, 235–247. doi: 10.1016/j.msea.2009.03.020
- Montros, C. S., Wei, T., Ye, L., Clark, G., and Mai, Y.-W. (2002). Laser shock processing and its effects on microstructure and properties of metal alloys: a review. *Int. J. Fatigue* 24, 1021–1036. doi: 10.1016/S0142-1123(02)0022-1
- Nalla, R. K., Altenberger, I., Noster, U., Liu, G. Y., Scholtes, B., and Ritchie, R. O. (2003). On the influence of mechanical surface treatments—deep rolling and laser shock peening—on the fatigue behavior of Ti–6Al–4V at ambient and elevated temperatures. *Mater. Sci. Eng. A* 355, 216–230. doi: 10.1016/S0921-5093(03)00069-8
- Neubauer, T. (2016). *Betriebs- und Lebensdauerverhalten hartgedrehter und festgewalzter Zylinderrollenlager* (dissertation). Leibniz Universität Hannover, Hannover, Germany.
- Neubauer, T., Poll, G., Denkena, B., and Maiß, O. (2013). "Fatigue life extension of rolling element bearings by residual stresses induced through surface machining," in *5th World Tribology Congress* (Turin).
- Pabst, A. (2018). *Steigerung der Ermüdungslebensdauer von Wälzlagern durch gezielt eingebrachte Druckeigenspannungen* (dissertation). Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
- Pape, F., Maiß, O., Denkena, B., and Poll, G. (2018). Computational approach to improve bearings by residual stresses based on their required bearing fatigue, life. *Int. J. Comput. Methods Exp. Meas.* 6, 656–666. doi: 10.2495/CMEM-V6-N4-656-666
- Pape, F., Neubauer, T., Maiß, O., Denkena, B., and Poll, G. (2017). Influence of residual stresses introduced by manufacturing processes on bearing endurance time. *Tribol. Lett.* 65:70. doi: 10.1007/s11249-017-0855-3
- Poll, G., and Hacke, B. (2010). "Berücksichtigung von Betriebszuständen, Sonderereignissen und Überlasten bei der Berechnung der Wälzlager-Lebensdauer in Windenergieanlagen und Großgetrieben," in *Informationstagung 2010 der Forschungsvereinigung Antriebstechnik*. Würzburg: FVA Forschungsreport.
- Voskamp, A. (1996). *Microstructural changes during rolling contact fatigue—metal fatigue in the subsurface region of deep groove ball bearing inner rings* (dissertation). Technical University Delft, Delft, Netherlands.
- Voskamp, A. P. (1985). Werkstoffverhalten bei Wälzbeanspruchung. *Kugellagerzeitschrift* 60, 2–10.
- Zwirlein, O., and Schlicht, H. (1989). Werkstoffanstrengung bei Wälzbeanspruchung - Einfluss von Reibung und Eigenspannungen. *Werkstofftechnik* 11, 1–14. doi: 10.1002/mawe.19800110104

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Pape, Coors and Poll. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mechanical Aspects of Deformation-Induced Surface Roughening in the Presence of Inclusions in a Subsurface Layer. Numerical Modeling

Varvara Romanova*, Ekaterina Dymnich, Ruslan Balokhonov and Olga Zinovieva

Institute of Strength Physics and Materials Science, Siberian Branch of the Russian Academy of Sciences, Tomsk, Russia

OPEN ACCESS

Edited by:

Valentin L. Popov,
Technical University of
Berlin, Germany

Reviewed by:

Vladimir Skripnyak,
Tomsk State University, Russia
Shingo Ozaki,
Yokohama National University, Japan
Vera Petrova,
University of Stuttgart, Germany

*Correspondence:

Varvara Romanova
varvara@ispm.s.tsc.ru

Specialty section:

This article was submitted to
Tribology,
a section of the journal
Frontiers in Mechanical Engineering

Received: 03 June 2020

Accepted: 07 July 2020

Published: 14 August 2020

Citation:

Romanova V, Dymnich E,
Balokhonov R and Zinovieva O (2020)
Mechanical Aspects of
Deformation-Induced Surface
Roughening in the Presence of
Inclusions in a Subsurface Layer.
Numerical Modeling.
Front. Mech. Eng. 6:66.
doi: 10.3389/fmech.2020.00066

The mechanical aspects of deformation-induced surface roughening inherent in microstructural inhomogeneity are studied numerically using single inclusion models. Three-dimensional finite-element calculations of uniaxial tension are performed for a set of single inclusion models where a cubic-shaped inclusion is embedded into a homogeneous matrix. The inclusion-to-surface distance, tilt angle about the axis of tension, and the ratio between the matrix and inclusion elastic-plastic properties are varied in different combinations to study the effects which these parameters have on the development of out-of-plane surface displacements under uniaxial tension. It has been shown that all stress and strain tensor components in the vicinity of inclusions take on non-zero values, including those directed across the load axis. Thus, the free surface becomes rough under the action of internal forces originated from the inhomogeneous stress-strain fields. Some illustrative examples of surface roughening under uniaxial tension are shown for multiple ellipsoidal inclusions periodically arranged in a subsurface layer of an elastic-plastic material.

Keywords: deformation-induced surface roughening, stress-strain analysis, numerical simulation, microstructure, uniaxial tension

INTRODUCTION

Deformation-induced surface roughening is a common feature for many materials (Raabe et al., 2003; Stoudt et al., 2011; Yoshida, 2014; Shi et al., 2017; Ma et al., 2019a,b). Being flat in the undeformed state, the material free surface roughens under deformation in the absence of external forces applied across its plane. Generally, roughening develops throughout all length scales from micro to macro (Anongba et al., 1993; Raabe et al., 2003) with roughness characteristics depending on many factors (crystal lattice, grain size and shape, texture, interface geometry, loading conditions, mechanical properties, etc.) (Raabe et al., 2003; Stoudt et al., 2011; Yoshida, 2014; Shi et al., 2017; Ma et al., 2019a,b; Solhjo et al., 2020).

For the most part, deformation-induced roughening is thought to be undesirable feature impairing cosmetic and functional material properties, such as reflectivity, lubricant transport, wear resistance, yield strength and the like (Raabe et al., 2003; Stoudt et al., 2011; Yoshida, 2014; Shi et al., 2017; Ma et al., 2019a,b; Cinat et al., 2020; Joe et al., 2020; Ozaki et al., 2020).

On the other hand, the roughness pattern can be used for non-destructive testing of the material stress-strain state. In our recent studies (Romanova et al., 2017a, 2019, 2020), we gave a special consideration to the mesoscale processes where roughness patterns were formed by grain groups involved in cooperative out-of-plane surface displacements. Experimental and numerical results obtained for a titanium alloy (Romanova et al., 2017a, 2019, 2020) suggested us that it is the mesoscale roughness which might be used as an early precursor of macroscale plastic strain localization and fracture.

In tribological applications, surface roughening acquires a special significance. Being one of essential factors influencing friction and adhesion [see, e.g., Popov and Li, 2018; Popov et al. (2017)], this phenomenon can be treated as positive or negative effect depending on the requirements to operative conditions. A comprehensive review of the related phenomena is provided by Popov (2010). Along with other methods to control adhesion and friction behavior (see, e.g., Popov and Li, 2018), the idea of controllable deformation-induced surface roughening sounds attractive.

The development of efficient methods to suppress or control deformation-induced surface roughening requires a deep understanding of the roughening mechanisms operative at different scales. While abundant experimental evidence has been produced and numerous theoretical studies have been performed, the problem of identifying the mechanisms involved and the factors responsible for the surface roughening in plastically deformed materials is the subject of considerable controversy and further investigations along these lines are clearly necessary.

Experimental and numerical studies have shown that it is structural inhomogeneity which is responsible for free surface roughening in the absence of external forces. Classical mechanics considering a homogeneous solid fails to provide an adequate description of this event. On a flat free surface of a homogeneous material there are no stresses likely to cause out-of-plane displacements. It is, therefore, reasonable to describe the related phenomena in terms of micro- and mesomechanics where material microstructure is taken into account.

In our earlier studies of micro- and mesoscale deformation-induced surface roughening in polycrystalline and coated materials (Romanova et al., 2017a,b, 2019, 2020), three-dimensional finite difference and finite element calculations were performed for microstructural models with realistic morphological features. The calculation results well-agreed with the experimental evidence and provided new insights into the roughening mechanisms, but the amount of the information gained with the use of the complex models was too vast and nearly as difficult to analyze as the data obtained from real experiments. Therefore, it is thought to be reasonable to examine the roughening mechanisms using simplified models to avoid the effect of the set of parameters used in the complex models. In this contribution, the mechanical aspects of deformation-induced surface roughening under uniaxial tension are studied numerically, using single inclusion models. While a large number of papers offer analytical and numerical solutions to a single inclusion mechanical problem (see, e.g., Eshelby, 1959;

Mori and Tanaka, 1973), the discussion of the related free surface out-of-plane displacements has but a little attention.

In this paper, out-of-plane surface displacements developing in a homogeneous material under uniaxial tension in the presence of a cubic-shaped inclusion in a subsurface layer are studied numerically. The inclusion orientation and position and the matrix and inclusion mechanical properties are varied in different combinations to study the effects which these parameters have on the surface geometrical characteristics. Some calculation results are presented for deformation-induced roughening in a multiple inclusion model where ellipsoidal inclusions are arranged in a subsurface layer in a regular manner.

SIMULATION OF SURFACE ROUGHENING IN SINGLE INCLUSION MODELS

Geometry, Constitutive Description, and Loading Conditions

Three-dimensional calculations for uniaxial tension were performed for a set of single inclusion models where a cubic-shaped inclusion was embedded into a homogeneous matrix in a subsurface layer. The inclusion-to-surface distance, inclusion tilt angle about the axis of tension, and the ratio between the matrix and inclusion elastic-plastic properties were varied in different combinations to investigate their effects on the free surface out-of-planeness.

Six representative geometrical models with the inclusions tilted at an angle of 0, 22, and 45 degrees about the axis of tension are shown in **Figure 1**. The depth of the inclusion position below the free surface was varied from 0 (**Figures 1A–C**) to a value equal to the inclusion linear size (**Figures 1D–F**). The models measuring $400 \times 200 \times 400 \mu\text{m}$ were approximated by tetrahedron finite-element meshes.

Calculations were performed for two sets of the matrix and inclusion properties. In the first case referred to as a hard inclusion model the matrix was aluminum and inclusion was ceramic. In the second case referred to as a soft inclusion model the inclusion was aluminum and the matrix was ceramic. The elastic-plastic model with a linear strain hardening was used to describe the constitutive behavior of the aluminum alloy. Ceramics was calculated in terms of elasticity. A perfect mechanical contact at the matrix-inclusion interface was kept throughout the entire loading process. The material constants applied in the calculations are given in **Table 1**.

Three-dimensional boundary-value problems in a quasistatic formulation were solved numerically by the finite element method using ABAQUS/Standard (Dassault Systèmes Simulia, 2011). The boundary conditions in all calculations were set to simulate uniaxial tension along the X-axis, as shown schematically in **Figure 1D**. For doing so, the displacements were assigned in the nodal points of the two opposite faces perpendicular to the X-axis to reach 1% tensile strain. This degree of strain is reasonable to study the mechanical factors responsible for the roughness formation. The other two lateral faces and the top surface, which was the main object of our interest, were

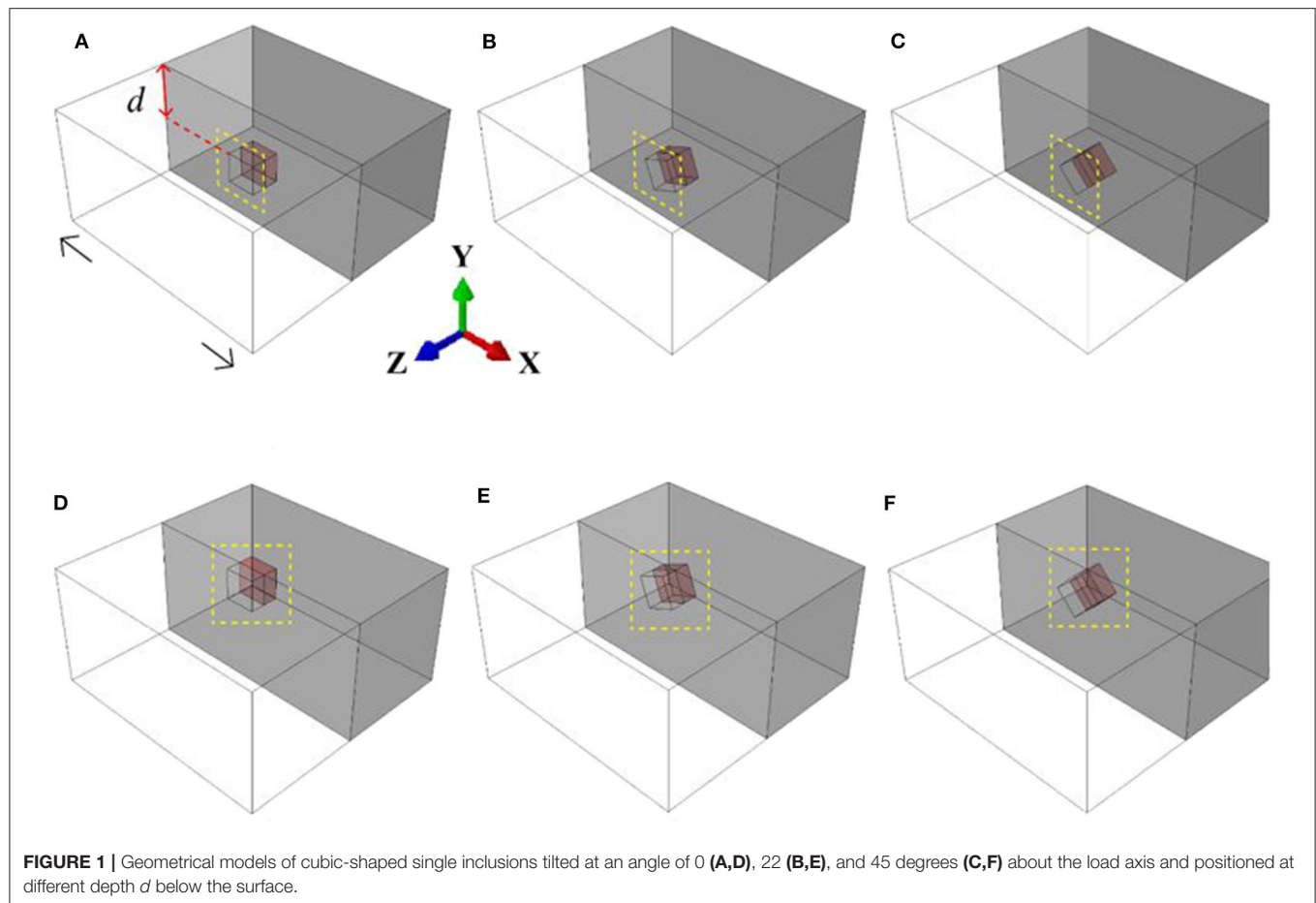


TABLE 1 | Material constants used in the calculations.

Material	Young module [GPa]	Poisson's ratio	Yield strength [MPa]	Strain hardening module [MPa]
Aluminum	70	0.25	100	50
Ceramics	400	0.2	–	–

free of external forces, and the bottom surface was taken to be a symmetry plane about the Y-direction.

Stress and Strain Fields

In order to identify the mechanisms of deformation-induced surface roughening in single inclusion models, let us analyze the relationship between the processes operative on the free surface and in the bulk of the material. For the sake of illustration, calculation results are presented for six models where inclusions are positioned near the surface (Figures 1A–C) and at a distance below the surface (Figures 1D–F), with the general conclusion being supported by the whole set of numerical data including intermediate inclusion positions. For each geometrical model, the mechanical properties were assigned in two combinations: the elastic inclusion - elastic-plastic matrix and vice versa.

The stress and equivalent plastic strain fields developing in the vicinity of hard inclusions embedded in the elastic-plastic matrices are presented in Figures 2, 3 for the regions bordered by dashed lines in Figure 1. Similar results for elastic-plastic inclusions embedded in the elastic matrix are given in Figures 4, 5.

Inside the hard inclusions, the von Mises stress exhibits strong inhomogeneity with the highest values developing near the inclusion boundaries (Figures 2A–C, 3A–C). Correspondingly, the elastic-plastic matrix demonstrates strain localization near the matrix-inclusion interfaces (Figures 2G–I, 3G–I). Common features of plastic strain fields in the vicinity of the inclusions are the appearance of cross-shaped regions of lower strains (blue-colored regions) and spherically-shaped regions of higher strains (red-colored regions) where the strains deviate from the average strain level (green-colored regions in Figures 2G–I, 3G–I). While the position of the low and high strain localization regions is mainly controlled by the loading conditions, the strain values in the interfacial regions are affected by the inclusion orientations. The most pronounced stress concentration develops in the inclusions with faces parallel to the axis of tension (Figures 3A, 4A) which have a higher ability to resist deformation. Accordingly, extensive regions of low plastic strains are formed in the matrix near the inclusion faces (Figures 2G–I, 3G–I). At a short distance from the faces perpendicular to the axis of tension,

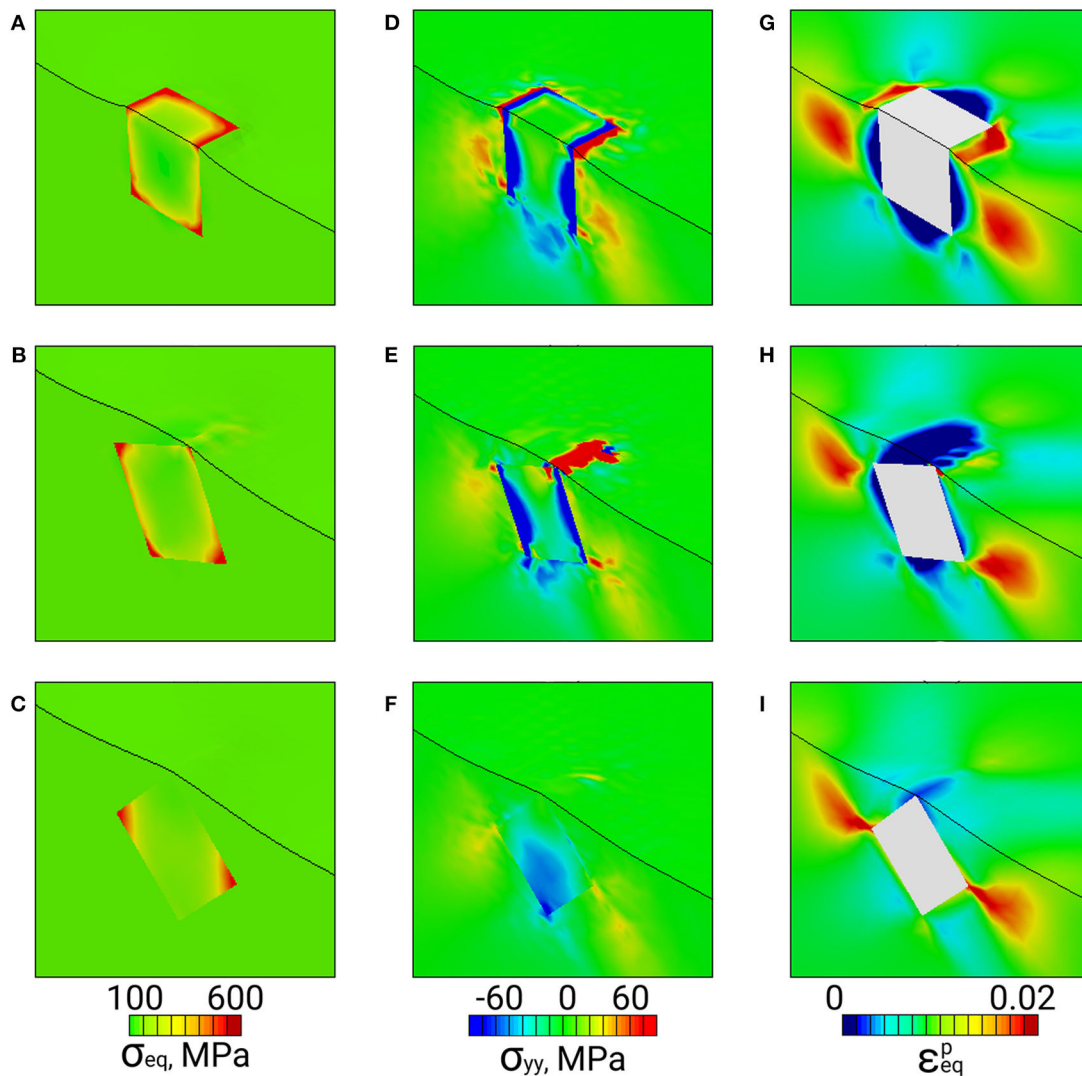


FIGURE 2 | Von Mises stress (A–C), normal stress σ_{yy} (D–F), and equivalent plastic strain fields (G–I) in the vicinity of a hard inclusion positioned in an elastic-plastic matrix near the surface (dash-bordered regions in **Figures 1A–C**).

the regions of high plastic strains are formed to compensate the low strains in the interfacial regions. Different plastic strain pattern is formed in the vicinity of the inclusions whose faces are tilted at an angle of 45 degrees to the load axis where high plastic strains localize along the whole matrix-inclusion interface (**Figures 2I, 3I**).

In the soft inclusion models (**Figures 4, 5**), the von Mises stresses developing in the elastically deformed matrix take on the highest values near the inclusion top and bottom boundaries and low values in the regions symmetrical about the inclusion vertical centerlines (**Figures 4A–C, 5A–C**). In the plastically deformed inclusions, the strains mainly localize near boundaries; the plastic strain patterns are controlled by the inclusion orientation relative to the axis of tension. It is a common tendency for all orientations that the inclusions positioned close to the surface exhibit higher strains than those located at a distance (cf. **Figures 4G–I, 5G–I**).

It is interesting to note that the inclusion whose faces are parallel to the load axis demonstrates less pronounced localization among all models when it is positioned at a distance from the surface (**Figure 5G**) and the highest strain level when it is exposed on the surface (**Figure 4G**).

The calculation results exemplify the fact that it is the material inhomogeneity which is responsible for the surface out-of-plane displacements. From the mechanical viewpoint, in a homogeneous material subjected to uniaxial tension, all components of the stress and strain tensors are equal zero except for those associated with the load direction. In this case, the free surface undergoes uniform deformation along the loading axis and remains flat throughout the entire loading process. The material inhomogeneity, however, gives rise to a complex stress-strain state with all stress and strain tensor components being non-zero in the near-interface regions.

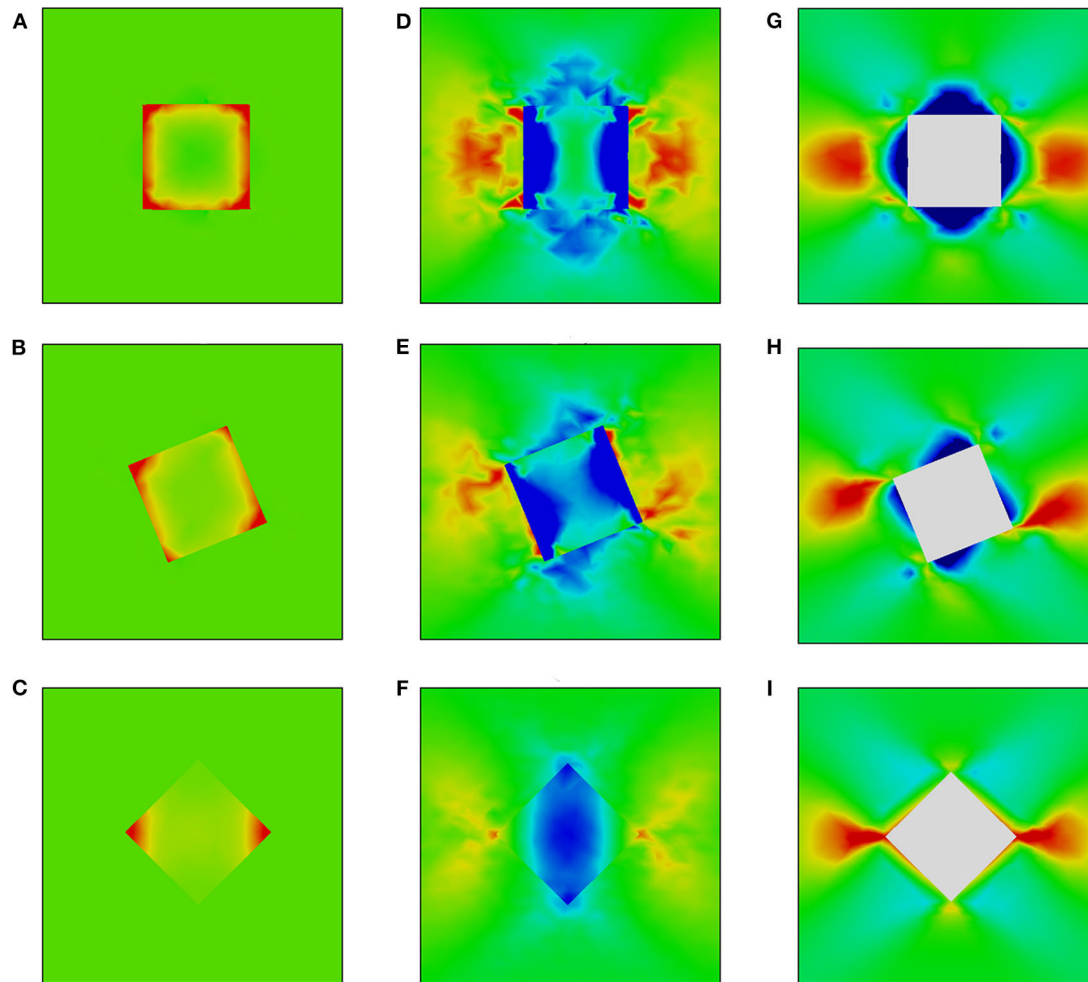


FIGURE 3 | Von Mises stress (A–C), σ_{yy} (D–F), and equivalent plastic strain fields (G–I) in the vicinity of a hard inclusion embedded in an elastic-plastic matrix at a distance below the surface (dash-bordered regions in **Figures 1D–F**). Color legends are given in **Figure 2**.

Out-of-Plane Surface Displacements

From the viewpoint of surface roughening, particular attention should be paid to the stress tensor components associated with the direction normal to the free surface (hereinafter referred to as normal stresses). The normal stress fields σ_{yy} are shown in **Figures 2D–F**, **3D–F** for the hard inclusion models and in **Figures 4D–F**, **5D–F** for the soft inclusion models. Again, these stresses in the case of a homogeneous isotropic material must be zero due to the absence of external forces directed normal to the surface. Due to the presence of inclusions, the normal stresses take on non-zero values near the matrix-inclusion interfaces with the positive and negative stress regions bordering each other. The negative normal stresses develop inside the hard inclusions and in the matrix above and below the inclusion boundaries to resist compression across the load direction (**Figures 3D–F**). As dictated by the stress equilibrium condition, the regions of positive normal stresses are formed in the matrix symmetrically about the inclusion vertical centerline. The positive and negative stresses compensate each other so that the average normal stress

values are zero. A similar conclusion is held for the soft inclusion models with the positive and negative normal stress regions reversing their positions (**Figures 4D–F**, **5D–F**). The stresses σ_{yy} are associated with the forces directed normal to the free surface. Acting from the inside, these forces give rise to the out-of-plane surface displacements in the form of the humps above the hard inclusion (**Figures 6A–C**) and dimples above the soft inclusion (**Figures 6D–F**). These conclusions are supported experimentally. For instance, Solhjoo et al. (2020) have reported formation of surface humps in a stainless steel under uniaxial tension due to the presence of hard particles in a subsurface layer.

In correspondence with the stress and strain patterns, the surface undulations are affected by the inclusion geometry, position below the surface and mechanical properties of the matrix and inclusion. Let us evaluate the effects which these parameters have on the surface displacement fields in the models with hard and soft inclusions (**Figure 6**). In both cases, the tilt angle of inclusions positioned below the surface at a distance equal or longer than the inclusion linear size has no significant

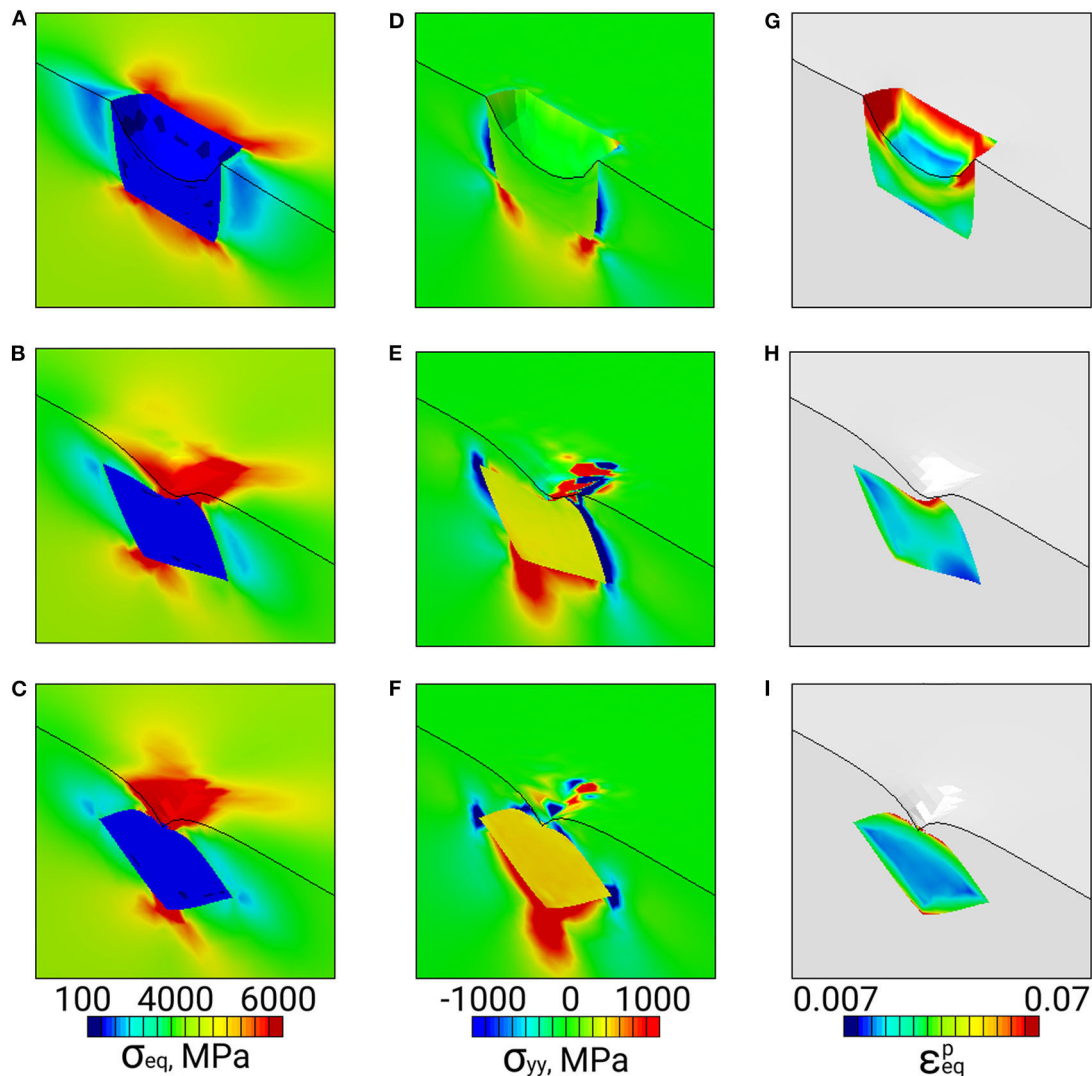


FIGURE 4 | Von Mises stress (A–C), normal stress σ_{yy} (D–F), and equivalent plastic strain fields (G–I) in the vicinity of a soft inclusion positioned in an elastic matrix near the surface (dash-bordered regions in **Figures 1A–C**).

impact on the shape and height of the surface humps and valleys. Contrastingly, calculation results for the inclusions located close to the free surface reveal a profound effect of the inclusion orientation on the surface roughness characteristics. The hard and soft inclusions with edges parallel to the free surface give rise to the formation of rectangular-shaped humps (**Figure 6A**) and dimples (**Figure 6D**), respectively. When the inclusions are tilted at an angle of 22 degrees about the loading axis, the hump and dimple shapes become asymmetric (**Figures 6B,E**). The sharpest humps and dimples are caused by the inclusions tilted at an angle of 45 degrees about the loading axis (**Figures 6C,F**).

Figure 7 demonstrates a combined effect of the inclusion orientation and inclusion-to-surface distance on the surface humps and dimples. In these plots, the height, depth, and inclusion-to-surface distance values are normalized to their

largest values. A common tendency for all models is that the heights and depths of the humps and dimples non-linearly increase with decreasing the inclusion-to-surface distance, which is judged from the curves plotted in **Figure 7**. It is interesting to note that the curves for hard inclusions demonstrate a specific behavior within a narrow subsurface region (**Figure 7A**). Particularly, an inverse dependence of the hump height on the inclusion-to-surface distance is observed for the inclusions with faces parallel to or tilted at an angle of 22 degrees about the load axis and a horizontal portion for that oriented at 45 degrees. In order to explain the mechanical cause of this non-linear behavior, let us analyze the normal stress fields in the vicinity of the inclusion top parts (**Figures 3D–F**). The non-zero normal stress regions formed in the matrix above the inclusion top strongly depend on the inclusion orientations. The larger the

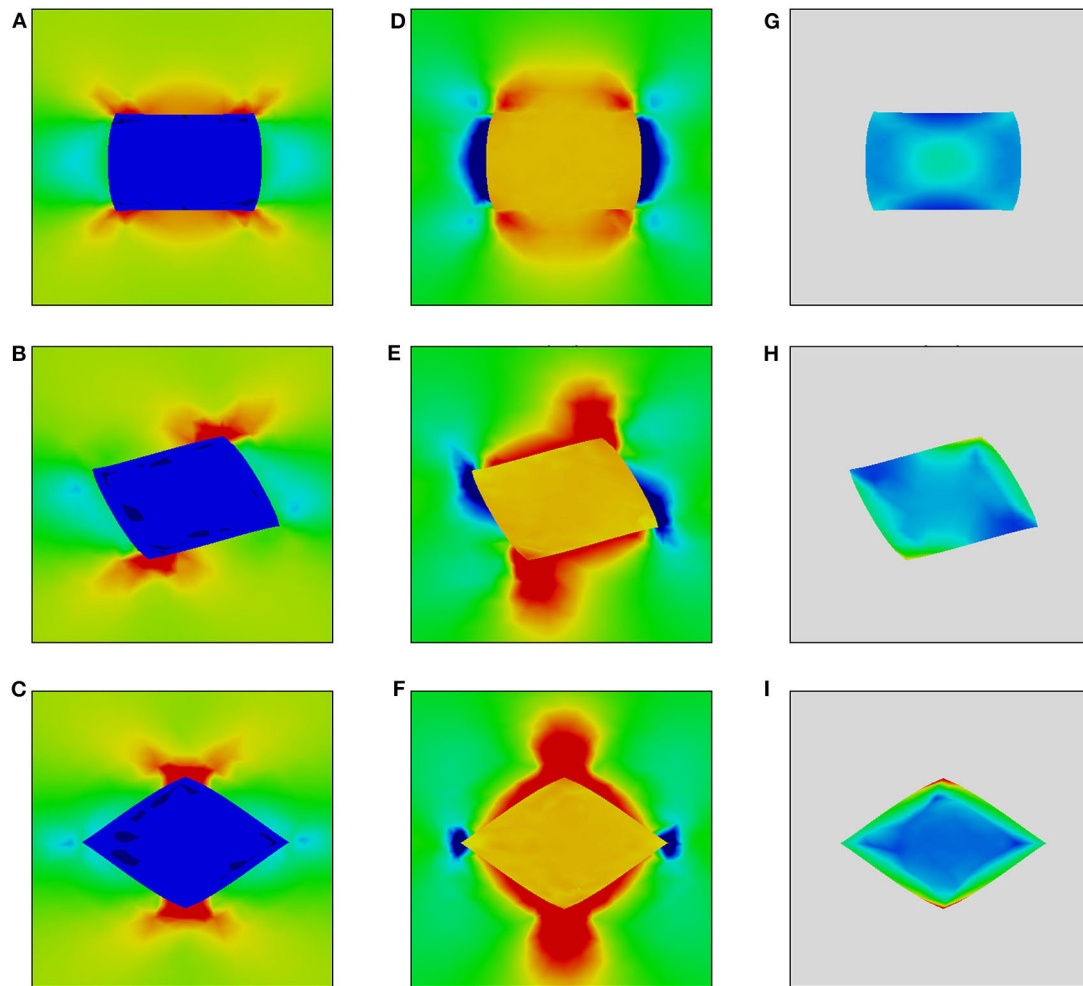


FIGURE 5 | Von Mises stress (A–C), normal stress σ_{yy} (D–F), and equivalent plastic strain fields (G–I) in the vicinity of a soft inclusion embedded in an elastic matrix at a distance below the surface (dash-bordered regions in **Figures 1D–F**). Color legends are given in **Figure 4**.

inclusion tilt angle about the loading axis, the wider the matrix region exhibiting negative normal stresses. The smallest region of compressive normal stresses is formed near the top of the inclusion whose faces are tilted at an angle of 45 degrees to the load axis (**Figure 3F**). As the inclusion moves close to the surface, the out-of-plane surface displacements become higher until the region of non-zero normal stresses reaches the surface. At a shorter distance, the stress concentration area is partially terminated by the surface and the humps are getting smaller.

DEFORMATION-INDUCED SURFACE ROUGHENING IN THE PRESENCE OF MULTIPLE INCLUSIONS IN A SUBSURFACE LAYER

While the stresses acting across the free surface are much lower than those associated with the load direction and, correspondingly, the out-of-plane surface displacements are

much smaller than the displacements in the direction of tension, the deformation-induced surface roughness is thought to be a useful characteristic for evaluating the material stress-strain state. In contrast to calculating the stress and strain fields from the in-plane surface displacements or using methods of indirect estimations (e.g., measurements of dislocation density, X-ray diffraction, etc.), the out-of-plane surface displacements are a measurable quantity which can be directly compared with the reference (undeformed) surface configuration.

As an example, calculations for deformation-induced surface roughening in the presence of multiple hard inclusions in a subsurface layer are presented in this section. The geometrical model measuring $2 \times 1 \times 2$ mm with ellipsoidal inclusions in a subsurface layer is shown in **Figure 8A**. The major and minor axes of the inclusions are 0.26 and 0.13 mm, respectively. The inclusions are arranged in a periodical manner at a depth of 0.11 mm below the surface. The shortest distance between inclusions is 0.37 mm, which reduces their interinfluence to a minimum. Thus, the roughness is not affected by the inclusion

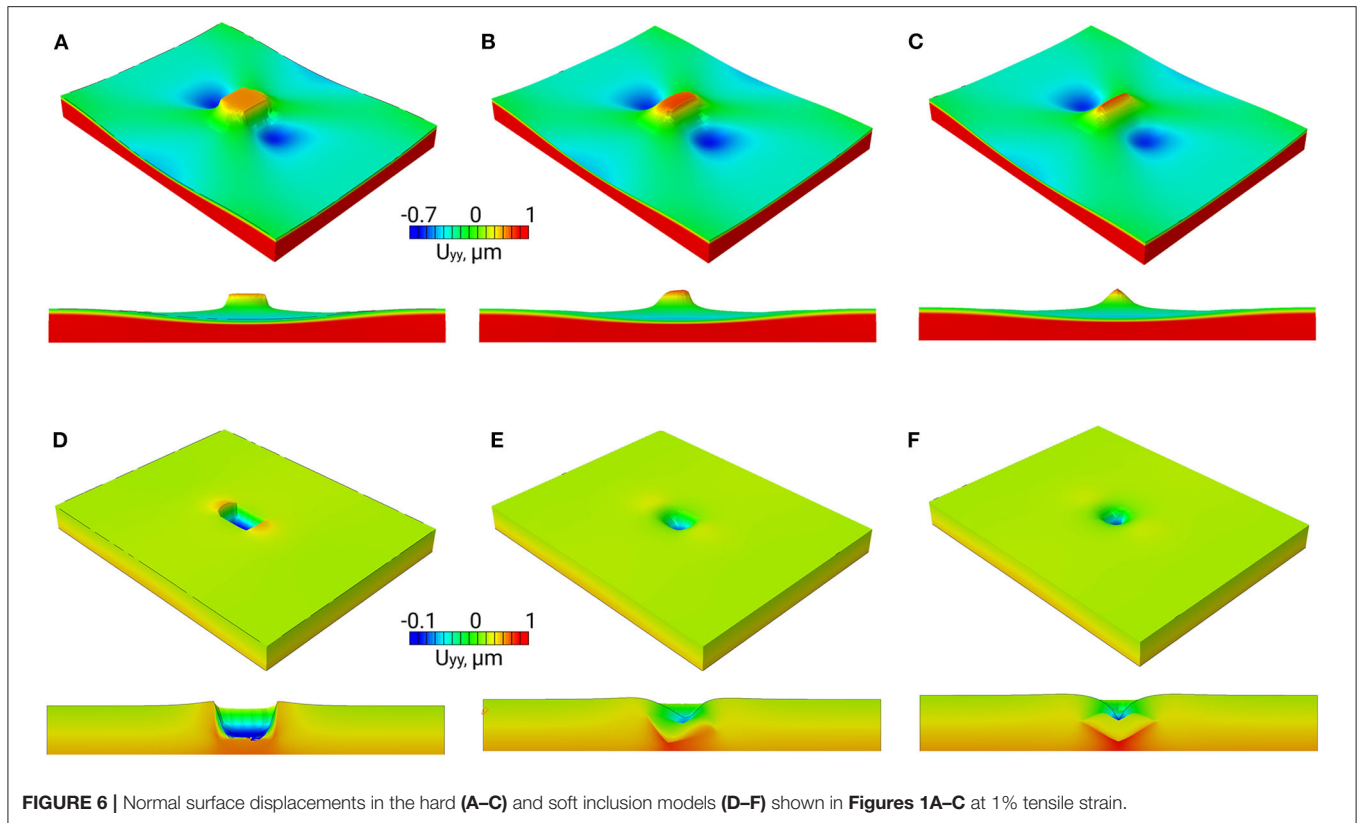


FIGURE 6 | Normal surface displacements in the hard (A–C) and soft inclusion models (D–F) shown in Figures 1A–C at 1% tensile strain.

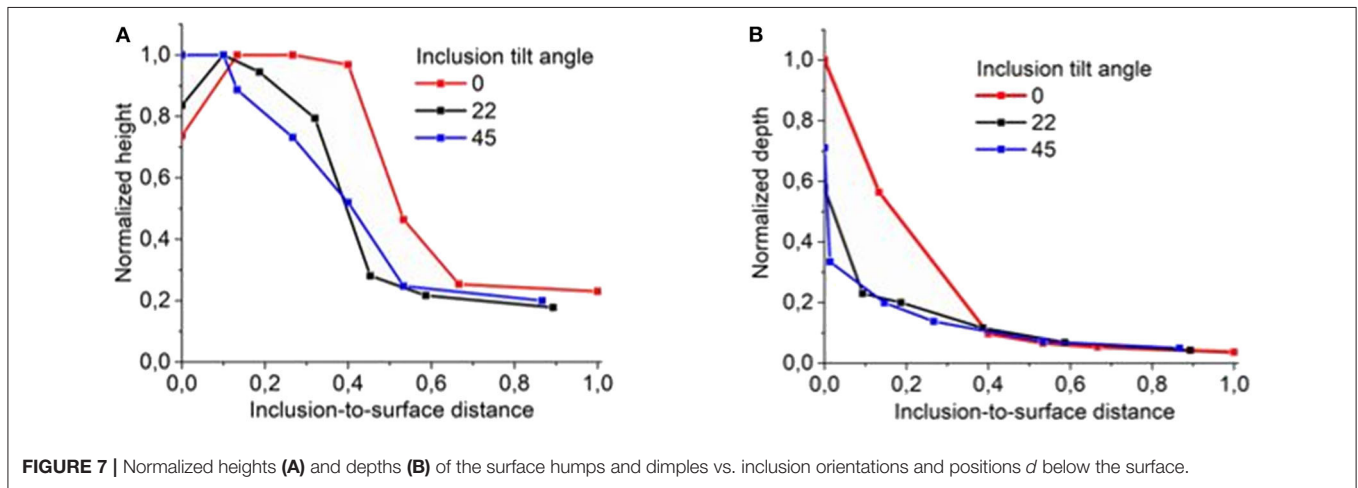


FIGURE 7 | Normalized heights (A) and depths (B) of the surface humps and dimples vs. inclusion orientations and positions d below the surface.

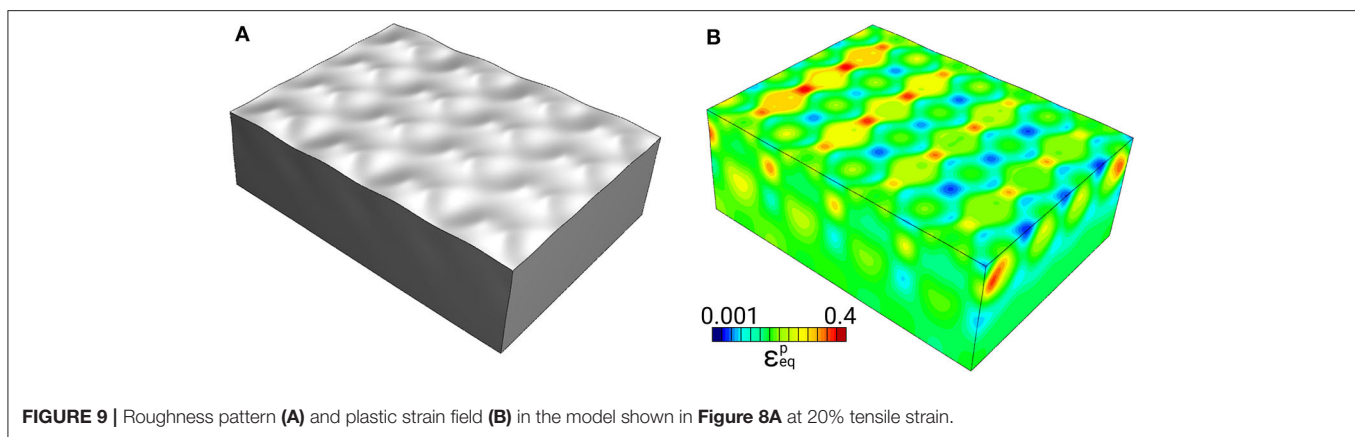
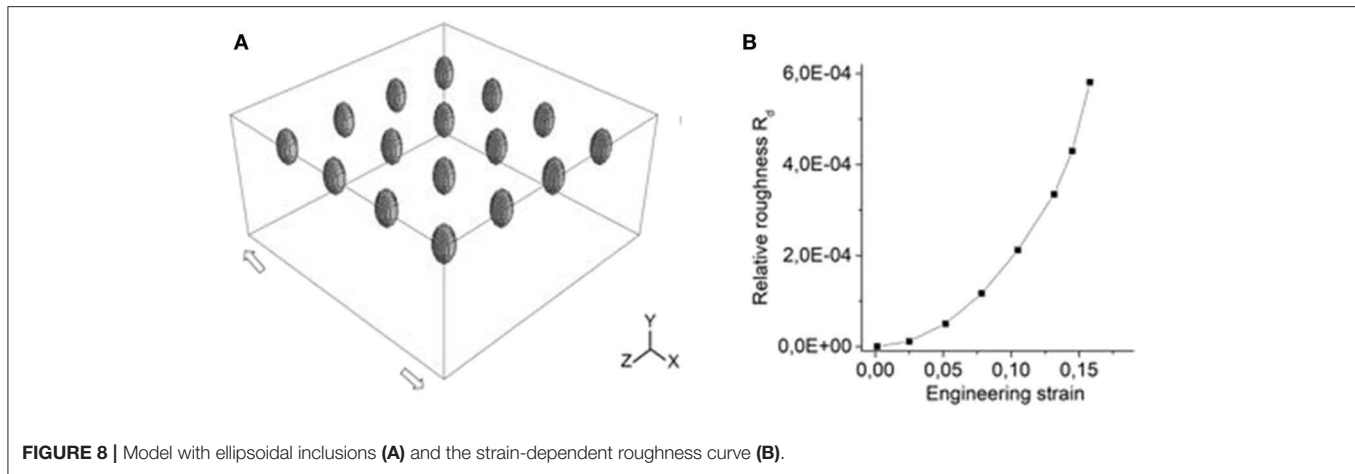
interaction but controlled by their individual contributions. The model is approximated by a regular $200 \times 100 \times 200$ mesh with hexahedral elements. The matrix and inclusions are assigned aluminum and ceramic properties, respectively; the material constants are given in **Table 1**. The boundary conditions are set in a similar way as for the single inclusion models to simulate uniaxial tension along the X-axis up to a strain of 20% to examine the roughness evolution in a wide range of plastic strains.

Similar to the case of single inclusions, all components of the stress and strain tensors are inhomogeneous, with the highest non-zero values taking place in the interfacial regions. The surface roughness develops in the form of periodically

arranged extruded and intruded regions (**Figure 9A**) and, correspondingly, plastic strain localizes in periodical bands going across the surface perpendicular to the axis of tension (**Figure 9B**).

Let us estimate the roughness evolution in terms of a dimensionless roughness parameter proposed in Romanova et al. (2017a, 2019) to quantify surface roughening in plastically deformed materials. By analogy with the strain quantification, the dimensionless roughness R_d is calculated as

$$R_d = \frac{L_r}{L_e} - 1, \quad (1)$$



where L_r is the rough profile length and L_e is the profile evaluation length. Expressed in this way, the roughness parameter is clearly interpreted: the larger is the R_d value, the rougher is the surface and the more inhomogeneous are the stress and strain fields developing in the material. Note again, in an idealized case of a homogeneous isotropic material subjected to uniaxial tension, the material free surface remains flat as long as no external forces act perpendicular to the tensile axis. In real materials, the displacement fields, inhomogeneous due to the present microstructure, give rise to the formation of rough patterns on the material free surface. Thus, the dimensionless parameter R_d shows to which extent the material is deformed in comparison with an ideal homogeneous material and can be treated as a relative roughness characteristic.

The dimensionless roughness parameter was calculated for a set of surface profiles measured along the model centerline in the range of tensile strains up to 20%. The strain-dependent roughness curve plotted in Figure 8B demonstrates non-linear growth in the course of plastic deformation, which agrees with previous experimental and numerical results obtained in (Romanova et al., 2017a, 2019, 2020). The dimensionless roughness parameter estimated for roughness profiles in plastically-deformed titanium alloy appeared to be rather

sensitive to the local plastic straining. What is more important, the R_d value grew non-linearly with the degree of plastic deformation in contrast to the commonly used roughness characteristics linearly dependent on the strain value. Thus, the R_d estimations would be promising tool for non-destructive material testing.

CONCLUSION

The mechanical cause of deformation-induced surface roughening under uniaxial tension has been analyzed numerically using single inclusion models. The effects of the inclusion orientation, position below the surface, and the matrix and inclusion elastic-plastic properties on the out-of-plane surface displacements were studied under uniaxial tension.

It has been shown that all stress and strain tensor components take on non-zero values in the vicinity of inclusions, including those directed across the load axis. Under the action of internal forces originated from the inhomogeneous stress-strain fields, the free surface becomes rough with humps and dimples formed above the hard and soft inclusions, respectively. The height of the surface undulations is affected by the

inclusion-to-surface distance while their shape (extrusion or intrusion) is controlled by the inclusion-to-matrix mechanical property ratio.

The roughness evolution in a multiple inclusion model was analyzed in terms of relative roughness estimations. It was found that the dimensionless roughness parameter calculated for the multiple inclusion model non-linearly increases in the course of plastic deformation, which agrees with earlier estimations for polycrystalline and coated materials.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

REFERENCES

- Anongba, P. N. B., Bonneville, J., and Martin, J. L. (1993). Hardening stages of [112] copper single crystals at intermediate and high temperatures—II. Slip systems and microstructures. *Acta Metallurg. Mater.* 41, 2907–2922. doi: 10.1016/0956-7151(93)90105-2
- Cinat, P., Gnecco, G., and Paggi, M. (2020). Multi-scale surface roughness optimization through genetic algorithms. *Front. Mech. Eng.* 6:29. doi: 10.3389/fmech.2020.00029
- Dassault Systèmes Simulia (2011). *Abaqus 6.11 Theory Manual*. Providence, RI: DS SIMULIA Corp.
- Eshelby, J. D. (1959). The elastic field outside an ellipsoidal inclusion. *Proc. R. Soc. A* 252, 561–569. doi: 10.1098/rspa.1959.0173
- Joe, J., Thouless, M. D., and Barber, J. R. (2020). Effect of surface roughness on adhesive instabilities for the elastic layer. *Front. Mech. Eng.* 6:31. doi: 10.3389/fmech.2020.00031
- Ma, X., Zhao, J., Du, W., Zhang, X., and Jiang, Z. H. (2019a). Analysis of surface roughness evolution of ferritic stainless steel using crystal plasticity finite element method. *J. Mater. Res. Technol.* 8, 3175–3187. doi: 10.1016/j.jmrt.2019.03.017
- Ma, X., Zhao, J., Du, W., Zhang, X., and Jiang, Z. H. (2019b). Analysis of texture-induced ridging in ferritic stainless steels 430 and 430LR during tensile deformation. *J. Mater. Res. Technol.* 8, 2041–2051. doi: 10.1016/j.jmrt.2018.12.019
- Mori, T., and Tanaka, K. (1973). Average stress in matrix and average elastic energy of materials with misfitting inclusions. *Acta Metall.* 21, 571–574. doi: 10.1016/0001-6160(73)90064-3
- Ozaki, S. H., Mieda, K., Maegawa, S., and Nakano, K. (2020). Meso–macro coupled analysis of pressure-dependent friction of rubber. *Front. Mech. Eng.* 6:24. doi: 10.3389/fmech.2020.00024
- Popov, M., and Li, Q. (2018). Multimode active control of friction, dynamic ratchets and actuators. *Phys. Mesomech.* 21, 24–31. doi: 10.1134/S1029959918010046
- Popov, V. (2010). *Contact Mechanics and Friction*. Heidelberg: Springer. doi: 10.1007/978-3-642-10803-7
- Popov, V., Pohrt, R., and Li, Q. (2017). Strength of adhesive contacts: influence of contact geometry and material gradients. *Friction* 5, 308–325. doi: 10.1007/s40544-017-0177-3
- Raabe, D., Sachtler, M., Weiland, H., Scheele, G., and Zhao, Z. (2003). Grain-scale micromechanics of polycrystal surfaces during plastic straining. *Acta Mater.* 51, 1539–1560. doi: 10.1016/S1359-6454(02)00557-8

AUTHOR CONTRIBUTIONS

VR wrote the first draft of the manuscript, performed calculations for multiple inclusions, and analyzed roughness. ED carried out calculations for single inclusions and analyzed the results. RB contributed to the discussion of the results and wrote section stress and strain fields. OZ wrote introduction and analyzed calculations for single inclusion. All authors contributed to conception and structure of the paper and approved the submitted version.

FUNDING

This work is supported by Russian Science Foundation (Project No. 20-19-00600).

- Romanova, V., Balokhonov, R., Emelianova, E., Pisarev, M., and Dymnich, E. (2020). Numerical study of the texture effect on deformation-induced surface roughening in titanium polycrystals. *Eng. Fail. Anal.* 110:104437. doi: 10.1016/j.engfailanal.2020.104437
- Romanova, V., Balokhonov, R., Emelianova, E., Sinyakova, E., and Kazachenok, M. (2019). Early prediction of macroscale plastic strain localization in titanium from observation of mesoscale surface roughening. *Int. J. Mech. Sci.* 121–162. doi: 10.1016/j.ijmecsci.2019.105047
- Romanova, V., Balokhonov, R., Panin, A., Kazachenok, M., and Kozelskaya, A. (2017a). Micro- and mesomechanical aspects of deformation-induced surface roughening in polycrystalline titanium. *Mater. Sci. Eng. A* 697, 248–258. doi: 10.1016/j.msea.2017.05.029
- Romanova, V. A., Balokhonov, R. R., Panin, A. V., Batukhtina, E. E., Kazachenok, M. S., and Shakhijanov, V. S. (2017b). Micromechanical model of deformation-induced surface roughening in polycrystalline materials. *Phys. Mesomech.* 20, 324–333. doi: 10.1134/S1029959917030080
- Shi, Y., Jin, H., Wu, P., and Lloyd, D. (2017). Analysis of roping in an AA6111 T4P automotive sheet in 3D deformation states. *Acta Mater.* 124, 598–607. doi: 10.1016/j.actamat.2016.11.028
- Solhjoo, S., Halbertsma, P. J., Veldhuis, M., Toljaga, R., and Pei, Y. (2020). Effects of loading conditions on free surface roughening of AISI 420 martensitic stainless steel. *Mater. Proc. Tech.* 275:116311. doi: 10.1016/j.jmatprotec.2019.116311
- Stoudt, M. R., Levine, L. E., Creuziger, A., and Hubbard, J. B. (2011). The fundamental relationships between grain orientation, deformation-induced surface roughness and strain localization in an aluminum alloy. *Mater. Sci. Eng. A* 530, 107–116. doi: 10.1016/j.msea.2011.09.050
- Yoshida, K. (2014). Effects of grain-scale heterogeneity on surface roughness and sheet metal necking. *Int. J. Mech. Sci.* 83, 48–56. doi: 10.1016/j.ijmecsci.2014.03.018

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Romanova, Dymnich, Balokhonov and Zinovieva. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership