Modeling and simulation in tribology across scales: An overview


Published in: Tribology International

Link to article, DOI: https://doi.org/10.1016/j.triboint.2018.02.005

Publication date: 2018

Document Version Peer reviewed version

Accepted Manuscript

Modeling and simulation in tribology across scales: An overview


PII: S0301-679X(18)30075-6
DOI: 10.1016/j.triboint.2018.02.005
Reference: JTRI 5104

To appear in: Tribology International

Received Date: 18 November 2017
Revised Date: 1 February 2018
Accepted Date: 4 February 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Modelling in Tribology: from Electrons to Machines

Coupling with Continuum from MD or intermediate scales: "hierarchical" or "concurrent"
Modeling and Simulation in Tribology Across Scales: an Overview


A.I. Vakis
Advanced Production Engineering, Engineering and Technology Institute Groningen, Faculty of Science and Engineering, University of Groningen, Nijenborg 4, 9747 AG Groningen, the Netherlands

V.A. Yastrebov
MINES ParisTech, PSL Research University, Centre des Matériaux, CNRS UMR 7633, BP 87, F 91003 Evry, France

J. Scheibert and C. Minfray
Univ Lyon, Ecole Centrale de Lyon, ENISE, ENTEP, CNRS, Laboratoire de Tribologie et Dynamique des Systèmes LTDS UMR5513, F-69134, ECUFFLY, France

L. Nicola
Department of Materials Science and Engineering, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands; Department of Industrial Engineering, University of Padova, via Venezia 1, 35015 Padua, Italy

D. Dini
Tribology Group, Department of Mechanical Engineering, Imperial College London, South Kensington Campus, Exhibition Road, London SW7 2AZ, UK

A. Almqvist
Division of Machine Elements, Luleå University of Technology, Luleå, Sweden

M. Paggi
IMT School for Advanced Studies Lucca, Multi-scale Analysis of Materials Research Unit, Piazza San Francesco 19, 55100 Lucca, Italy

S. Lee
Department of Mechanical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark

G. Limbert
National Centre for Advanced Tribology at Southampton (nCATS) | Bioengineering Science Research Group, Faculty of Engineering and the Environment, University of Southampton, Southampton SO17 1BJ, UK; Biomechanics and Mechanobiology Laboratory, Biomedical Engineering Division, Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Anzio Road, Observatory 7925, South Africa

J.F. Molinari and G. Anciaux
LSMS, ENAC, Swiss Federal Institute of Technology (EPFL), CH-1015, Lausanne, Switzerland

R. Aghababaei
Department of Engineering, Aarhus University, Inge Lehmanns Gade 10, 8000 Aarhus C, Denmark

S. Echeverri Restrepo
SKF Engineering & Research Centre (ERC), SKF B.V., Nieuwegein, the Netherlands; Department of Physics, King’s College London, Strand, London WC2R 2LS, England, UK

A. Papangelo
Hamburg University of Technology, Department of Mechanical Engineering, Am Schwarzenberg-Campus 1, 21073 Hamburg, Germany

A. Cammarata and P. Nicolini
Department of Control Engineering, Faculty of Electrical Engineering | Czech Technical University in Prague, Karlovo namesti 13, 12135, Prague 2, Czech Republic

C. Putignano, G. Carbone and M. Ciavarella
Politecnico di Bari. V. le Gentile 182, 70125 Bari-Italy

S. Stupkiewicz and J. Lengiewicz
Institute of Fundamental Technological Research, Polish Academy of Sciences, Pawinskiego 5B, 02-106 Warsaw, Poland

G. Costagliola and F. Bosia
Department of Physics and Nanostructured Interfaces and Surfaces Centre, University of Trento, Via Pietro Giuria 1, 10125 Turin, Italy

R. Guarino
Laboratory of Bio-Inspired & Graphene Nanomechanics, Department of Civil, Environmental and Mechanical Engineering, University of Trento, Via Mesiano 77, 38123 Trento, Italy

N.M. Pugno
Laboratory of Bio-Inspired & Graphene Nanomechanics, Department of Civil, Environmental and Mechanical Engineering, University of Trento, Via Mesiano 77, 38123 Trento, Italy; Ket Lab, Edoardo Amaldi Foundation, Italian Space Agency, Viale del Politecnico sn, 00133 Rome, Italy; School of Engineering and Materials Science, Queen Mary University of London, Mile End Road, E1–4NS London, United Kingdom

M.H. Müser
Department of Materials Science and Engineering, Saarland University, 66123 Saarbrücken, Germany

* Corresponding author. E-mail: mciava@poliba.it
Abstract
This review summarizes recent advances in the area of tribology based on the outcome of a Lorentz Center workshop surveying various physical, chemical and mechanical phenomena across scales. Among the main themes discussed were those of rough surface representations, the breakdown of continuum theories at the nano- and micro-scales, as well as multiscale and multiphysics aspects for analytical and computational models relevant to applications spanning a variety of sectors, from automotive to biotribology and nanotechnology. Significant effort is still required to account for complementary nonlinear effects of plasticity, adhesion, friction, wear, lubrication and surface chemistry in tribological models. For each topic, we propose some research directions.

Keywords: tribology, multiscale modeling, multiphysics modeling, roughness, contact, friction, adhesion, wear, lubrication, tribochemistry

1. Introduction
The word tribology introduced in the famous Jost report of 1966 [1] was apparently coined by David Tabor and Peter Jost, deriving from the root tribo- (Greek τρίβος, meaning “rubbing”) and the suffix -logy (Greek -λογία, meaning “the study of”). The Jost report suggested that problems of lubrication in engineering needed an interdisciplinary approach—including chemistry and materials science, solid mechanics and physics. At that time, Jost suggested that the British industry could have saved £500 million a year “as a result of fewer breakdowns causing lost production; lower energy consumption; reduced maintenance costs; and longer machine life.” Fifty years later, frictional losses are often evaluated as costing more than 1 per cent of GDP [2], and tribology is therefore still flourishing.

There is no doubt that tribological interactions have a profound impact on many areas of engineering and everyday life. The widespread significance of these effects has been highlighted in many articles and reports over the years, which, until recently have mainly focused on lubrication and friction and wear-related energy and material losses for “traditional” industrial applications, such as manufacturing and automotive. The reader is referred to recent reviews, which have, for example, looked at the development of solid lubricant coatings [3], lubrication [4], and the interplay between surfaces and lubricants [5]. Other works have focused on how improvements in friction reduction technologies could significantly reduce frictional energy losses in passenger cars in the short, medium and long term [6]. Reducing wear can also improve long-term efficiency and performance of moving components, as well as reducing costs of maintenance and/or improving quality of life. Accordingly, much research into means of reducing friction and wear, together with the development of new additives, lubricants and functional materials to improve the performance of interfaces, has taken place, typically in the form of experimental studies for developing improved surface materials, topography/textures or lubrication. Most of these activities have been supported and accompanied by fundamental developments in contact mechanics, e.g., [7,8], as well as surface and material science, e.g., [9]. This has in turn improved our understanding of how surface roughness and surface modifications affect the response of components in various applications [10,11].

More recently, new areas of tribology have emerged, including nanotribology, i.e. the study of friction, wear and lubrication at the nanoscale as applied, for example, to micro- and nano-electromechanical systems (MEMS/NEMS), e.g., [12,13], and magnetic storage, e.g., [14,15], and biotribology, which deals with human joint prosthetics, dental materials, skin, etc., and ecological aspects of friction, lubrication and wear (tribology of clean energy sources, green lubricants, biomimetic tribology) [2,16-19]. Studies of superlubricity, i.e. the mechanisms responsible for extremely low friction [20-23], have created great expectations of energy savings, and the creation of graphene is also greatly promising in...
In this direction [24]. Insects’ and reptile’s adhesive performance inspired numerous studies on adhesive contacts (e.g. [25-29]) and resulted in improved understanding and successful mimicking of Nature-made feet [30-37]. Massive usage of tactile interfaces triggered multiple studies in understanding sensing through contact and friction [38], and in reproducing interactive haptic feedback to moving fingers [39-42]. In keeping up with and enabling such developments, new knowledge is necessary to describe complex multiscale and multiphysical phenomena within the context of tribology, both in the modeling and experimental domains.

In this contribution, we aim to summarize the presentations and discussions that took place during a Lorentz workshop on “Micro/Nanoscale Models for Tribology” in Leiden, the Netherlands, between 30 January and 3 February 2017. It was found that one of the key issues facing the tribology community is the apparent disparity between the fields of expertise relevant to such an interdisciplinary topic, which leads to a lack of communication between engineers, material scientists, applied physicists and chemists who work to solve similar tribological problems; differences exist in notation, language, methods, the way in which problems are posed and how solutions are presented. Another finding is that new analytical models are necessary to understand the behavior at tribological interfaces, partly to avoid that numerical simulations become “black boxes” where the nuances of the phenomena involved are lost, and partly because full computational models often require prohibitively long computational times. At the same time, the industry would benefit from lightweight analytical models as long as those are sufficiently robust and able to predict critical quantities of interest with a priori known precision. Further adding to these challenges is the complexity of model validation: as the contact interface in most cases is not accessible to direct in situ observations, it is very difficult to carry out experiments aiming to access local near-surface states.

Difficulties are further enhanced by divisions between modelers and experimentalists, as well as those working on analytical versus computational methods—not only by the different theories, computational methods and tools—and depending on the research applications. Since increased visibility and cooperation between tribologists from different backgrounds is necessary, the present review aims at providing a starting point for further collaboration and possible focal points for future interdisciplinary research in tribology. Accordingly, the paper is organized as follows: various modeling methods and tools are discussed in §2; research themes in tribology, including multiphysical aspects, rough surface representations, scale effects and the breakdown of continuum theories at the nano- and microscales, material models, normal contact, friction and other phenomena, as well as interdisciplinary case studies in biotribology are addressed in §3, and conclusions are given in §4.

2. Tribological modeling methods

This section introduces the main tools currently used in tribological modeling, starting from analytical models and discussing continuous and discrete mechanical and multiphysical methods suitable for simulations characterized by different time- and length-scales (see Fig. 1 for a map of representative tribological models built across the scales), namely finite and boundary element methods, discrete dislocation dynamics and atomistic methods, as well as multiscale approaches.
2.1. Analytical Methods

2.1.1. Contact mechanics: where we stand

A full overview of the field of contact mechanics and related developments that took place over the last century or so is out of the scope of the current contribution, as this would require a devoted review. For someone approaching this scientific area for the first time, K.L. Johnson's Contact Mechanics book [43] is still a very good starting point today. Later books and review papers, e.g. [44-49], have accounted for some of the progress made, but the field continues to expand across disciplines. The purpose of this sub-section is to briefly summarize some of the important milestones in this field and provide pointers to the readers interested in its different branches.

Starting from the mechanics of nominally smooth contact problems, the Hertzian theory, which solves the problem of two non-conformal elastic bodies being subjected to frictionless contact [50], is considered as a cornerstone of contact mechanics and tribology. Many of the analytical solutions available to practitioners and scientists have been building on Hertz; as is the case, for example, for two early models that constitute seminal advances in contact mechanics focused on the issue of adhesion: the models by Johnson-Kendall-Roberts (JKR) [51] and Derjaguin-Muller-Toporov (DMT) [52]. While the JKR and DMT models, which describe the adhesive contact between compliant or hard spheres, respectively, are still very popular, the body of literature available on this topic is very substantial, especially given its relevance to, e.g., biomimetic applications; adhesion is discussed in detail in §3.7.

Remaining in the realm of smooth contact problems, but moving away from the Hertzian theory of elastic contacting bodies and its limitations (only accurate for small contact areas), progress has been made in a number of other areas: these include, for example, layered and coated systems, also in the presence of anisotropic and functionally graded materials [53-61], contacts in the presence of sharp edges [62-65] and conformal configurations [66]. Other examples of recent developments in the field are the use of asymptotic analyses to study the stress fields and sliding behavior associated with
different contact configurations [67-70], the study of contact in the presence of anisotropic and functionally graded materials, and varying friction coefficient along the interface in sliding and partial slip conditions [71]. In the case of the normal contact of inelastic solids, significant developments have been made since Johnson’s core model of elasto-plastic indentation based, for example, on the progress of instrumented nanoindentation in the last 25 years (see, e.g., [72-74]); issues of plasticity and material models are discussed further in §3.4. Some progress has also been made on tangential loading and cyclic contact with the generalized solution of contact problems characterized by time-dependent stick-slip transitions at the macroscopic scale (see, e.g. [62,75-79]).

Somewhat in parallel to the above advances and studies, many developments in the study of nominally smooth contacts in the presence of lubrication have also been made; these are discussed in §2.4 and §3.8.

On dynamic effects and impact, much work was published on the rate-and-state friction (RSF) law (also discussed in §3.6.2) and Adams’ instability [80-82], while impact remains a somewhat separate and large research area, with applications in different research area and applications including powder technology, manufacturing processes and ballistics [83-88]. Following the classical contributions by J.R. Barber on both static and sliding contact reviewed in Johnson’s book, new refined solutions and finite element formulations have appeared on thermoelastic contact (see, e.g., recent contributions [89,90] and further discussion in §2.6).

Moving on to applications strongly linked to the development of contact mechanics methodologies, various advances have been made. An example is the development of various techniques used to individually or simultaneously study various aspects of fretting fatigue, such as stress gradients, fatigue, surface damage and wear [91-96]. Progress has also been made in the study of rolling contact of elastic and inelastic (shakedown, ratchetting, etc.) bodies and rolling contact fatigue (see, e.g., [97-104]). Calendering, i.e. the elastic-plastic rolling of strips have also seen some developments [105].

On the topic of contact mechanics of rough surfaces, the seminal work by Greenwood and Williamson (GW) [7] forms the basis for a number of multi-asperity models (discussed critically in §2.1.2). Among many subsequent analytical models, some were developed based on the analysis of two or more scales, adding for example the periodic microgeometry of multi-layered elastic or viscoelastic half spaces to study normal contact and friction in the presence of coatings [106,107] or adhesion and lubrication [108,109]. Interestingly, one of the most popular theories after the GW is that of Majumdar and Bhushan [110], where Kočak’s law was used to define a power law distribution of contact spots, a “bearing area” result very much in contrast with the present understanding of the contact area being formed by “resolution-dependent” contact spot sizes. This view of “magnification-dependent” solution is not too different from the original Archard model [111] of spheres sitting on top of spheres, or work on fractal description based on a Weierstrass series within the elasticity assumption to obtain the result that the contact area decreases without limit as the resolution (or magnification) is increased [112].

The alternative to the solutions proposed in the methodologies to study rough contacts reviewed above is Persson’s theory [8], which has become the basis of another class of models, in which the stress probability distribution is considered as a function of the surface resolution under examination. The tribology community still uses both the GW and Persson approaches to model rough contact based on considerations of accuracy and simplicity which may well reflect the corresponding physics and engineering perspectives. The GW and Persson models are introduced in more detail next; a comparison between them in the context of the recent contact-mechanics challenge [113] is given in §3.5, while the topic of roughness itself is described extensively in §3.2.
2.1.2. Multi-asperity models and Persson’s theory: an introduction

The nature and various representations of surface roughness, discussed in more detail in §3.2, have been central to the prediction of tribological quantities ranging from the true area of contact—in contrast to the apparent or effective area—to the normal, friction and adhesion forces, as well as phenomena such as electrical conductance and percolation. Starting from the simplest problem definition of normal contact between two rough surfaces in the absence of other phenomena, two seminal works have formed the backbone of research in the field: the Greenwood-Williamson (GW) model [7] and Persson’s theory [8]. These are introduced below, while the results of a recent contact-mechanics challenge are summarized in §3.5, extending beyond predictions of the true contact area and into more detailed metrics of normal contact.

Greenwood and Williamson conducted a pioneering study targeted towards predicting the link between the approach of nominally flat but rough surfaces (quantified as the distance between their mean planes) and the resulting force and true contact area [7]. The GW and subsequent multi-asperity models are based on the following assumptions: 1) the effective rough surface (a superposition of two rough profiles or surfaces) can be represented by an ensemble of asperities (surface summits), characterized by the vertical coordinate of the tip and its curvature(s); 2) these characteristics are known in the statistical sense, for example, via the probability density of the asperities' vertical position; 3) the relation between penetration, force and the contact area follows the Hertzian theory of contact; 4) the asperities of the effective rough surfaces coming into contact are separated in the plane by distances at which their mutual influence can be neglected. In the original GW, all asperities are approximated as parabolic with the same curvature radius, and an arbitrary height distribution is assumed, contrary to numerous references in the literature erroneously stating that the GW model is based on Gaussian distribution of asperity heights: both Gaussian and exponential tails are considered in the original paper.

Subsequent progress in statistical multi-asperity models was triggered by the seminal paper of Nayak [114], which was in turn inspired by the works of Longuet-Higgins who was the first to apply the random process model for analysis of random surfaces in the ocean [115,116]. Based on the same assumption, i.e. that a rough surface can be represented as a two-dimensional isotropic Gaussian process, Nayak obtained the relation between the spectral moments of the surface and the distribution of asperities, their density, curvature, ellipticity, etc. He also introduced a central quantity for roughness description, a dimensionless combination of the zeroth, second and fourth momenta, subsequently referred to as the Nayak parameter that characterizes spectral breadth. Based on Nayak's statistical results, Bush, Gibson and Thomas (BGT) [117] obtained a new approximation for the dependence of the force density and contact area fraction taking into account, among other of Nayak's results, the ellipticity of asperity tips. Much later, Greenwood [118] demonstrated that, according to Nayak's theory, the ellipticity of asperities is rather mild, and thus an approximate Hertzian equation for the elliptic contact can be employed, which makes use of the geometric mean value of two principal asperity curvatures. This “simplified elliptic model” yields relatively simple equations for force and area dependence as functions of the approach (or separation). Among other interesting results, Greenwood demonstrated that according to the random process model, the probability of finding a spherical asperity is strictly zero.

Multi-asperity models predict asymptotic linearity between the contact area and the load with a factor containing a proportionality coefficient $\kappa$ and, in the denominator, a product of the effective elastic modulus and the root mean squared roughness gradient (or equivalently, a square root of the doubled second spectral moment). However, it is important to remark here that this proportionality holds only for vanishingly small contact area intervals, which depend on the Nayak parameter: the higher this
parameter is, the smaller the region of validity [118-120]. In this light, the proportionality predicted between the load and the area remains a mathematical abstraction and cannot be used directly in engineering practice. However, the usage of multi-asperity models is not restricted to vanishingly small areas, but can also be used for higher loads at which the area evolves nonlinearly with the load and strongly depends again on the Nayak parameter [117,120]: the higher the Nayak parameter, the smaller the contact area. Comparison of multi-asperity models with full numerical simulations of rough contact (free of the multiple assumptions of multi-asperity models) demonstrated that, indeed, the Nayak parameter plays an important role in contact area evolution, but its effect in multi-asperity models is strongly exaggerated [121].

Further improvements in multi-asperity models attempted to incorporate elastic interaction between asperities, based on the following motivation: if one asperity comes into contact and produces a force, then the vertical position of all surrounding asperities needs to be changed by, approximately, a value proportional to this force and inversely proportional to the distance to its point of application (for the precise formulations refer to [43]). Note that because of such a slow decay, this long-range elastic interaction cannot be cutoff without considerable loss in accuracy. Approximately, this interaction can be included in a statistical framework by assuming a zero-order approximation, i.e. the vertical positions of all asperities are decreased by a value proportional to the product of a nominal pressure and the contact area [122-124]. A further improvement in terms of elastic interaction relied on the rejection of a purely statistical model and the resorting to deterministic models instead, taking into account the in-plane positions of all asperities. In this deterministic framework, not only elastic interactions can be accurately accounted for [119,125], but so can the merging of contact areas related to distinct close asperities [126].

In 2001, B.N.J. Persson suggested another analytical model for predicting the contact area and other related quantities [8] that relies on completely different considerations and, therefore, does not suffer from the multiple assumptions inherent in multi-asperity models (even though it introduces its own). Persson’s theory is based on the following consideration: let us assume contact between two flat surfaces squeezed together by a nominal pressure \( p_0 \) such that the probability density of interfacial pressure is simply a Dirac delta-function centered at \( p_0 \). When new modes are progressively injected into the spectrum of contacting surfaces, the corresponding pressure distribution function spreads out as a Gaussian distribution. If the full contact is preserved, the link between the statistical characteristics of the height distribution and interfacial pressure distribution can be easily established: the variance of the contact pressure is proportional to the product of the variance of the surface gradient and squared effective elastic modulus. Based on these considerations, a diffusion-type equation was formulated for the contact pressure distribution (acting as the concentration quantity), with the pressure variance acting as the time and the local pressure acting as the space coordinates [8,127,128] considering, up to this point, only full contact. Since Gaussian support is infinite, tensile stresses will occur in the contact interface for an arbitrary finite external pressure. Persson introduced a boundary condition stating that the probability density function of contact pressures vanishes at zero pressure. Indeed, this boundary condition seems very reasonable if one thinks about the fact that, for Hertzian contact, the pressure drops to zero at the contact edges with an infinite slope, thus resulting in the linear growth of probability density near zero pressure. The main remaining assumption of Persson’s theory is the validity of the diffusion equation for partial contact accounting for the fact that it was derived for full contact.

Apart from other quantities of interest, Persson's theory predicts that the contact area evolves as an error function, from zero to full contact, which is reached for infinite nominal pressure. Since the Taylor expansion of the error function in the vicinity of zero contains only odd powers, the contact
area can be approximated with a high degree of confidence by a linear function of nominal pressure with a factor given by a proportionality coefficient divided by the product of a root mean squared roughness gradient and the effective elastic modulus. The first difference between this prediction and those of multi-asperity models is the proportionality factor $\kappa$, which is approximately 1.60 in Persson's theory and approximately 2.51 in multi-asperity models. The second crucial difference is that, contrary to multi-asperity models, Persson's linearity is valid for realistic area/pressure intervals. Finally, the third difference is that the sole roughness parameter needed for Persson's theory is the root mean squared roughness gradient so that, contrary to multi-asperity models, this theory has no dependence on the Nayak parameter.

Numerous comparisons between complete numerical simulations, multi-asperity models and Persson's theory can be found in the literature [119-121,127-141]. The rough conclusion of all these studies with respect to the contact area evolution can be formulated as follows: Persson's model nicely predicts the qualitative growth of the contact area with increasing nominal pressure up to full contact [137]. For moderate loads, the true contact area evolves slightly nonlinearly and is below the asymptotic prediction of multi-asperity models and above the prediction of Persson's theory. Meanwhile, an improvement in Persson's theory was introduced to take into account partial contacts in a more rigorous way [142], yielding results that are much closer to numerical solutions. Very recent findings demonstrate that the contact area growth is dependent not only on the root mean squared gradient but also weakly on the Nayak parameter [121] which is absent in Persson's theory, but is inherent to multi-asperity models that, however, strongly overestimate its effect.

2.2. Finite and Boundary Element Methods

Two major families of methods can be distinguished in continuum mechanics: the Finite Element Method (FEM) [143] and the Boundary Element Method (BEM) [144]. The FEM is a versatile method for solving boundary value problems in many fields of science and technology [143,145,146]).

In the FEM, an explicit relation between the strain (and possibly strain rate and its history) and the stress can be prescribed, either within infinitesimal or finite strain formulations, enabling this method to consider arbitrary constitutive material models starting from simple linear elasticity up to complex crystal plasticity. The BEM uses in its formulation a fundamental solution for the normal and tangential point forces, which enables linking surface tractions with surface displacements. Equivalently, to formulate a spectral version of the BEM, a fundamental solution linking pressure and vertical displacement for a combination of harmonics in two orthogonal directions should be used [147,148]. Such solutions exist for a limited number of cases and mainly under the assumption that the solid can be locally considered as a flat half-space. These limitations imply a more restrictive field of application for the BEM compared to the FEM, which is a versatile numerical method. It is worth mentioning that, in general, contact problems are nonlinear even if frictionless and non-adhesive contact is considered between linearly elastic solids. This is because the contact area is a priori unknown, apart from simple cases such as the rigid flat stamp problem or the case of full contact. In analogy, a full stick frictional condition (infinite friction) makes the frictional problem much easier to handle than a problem with a finite friction.

Detailed descriptions of numerical methods within the FEM formulation can be found in the literature, e.g., [149-151], while details on the application of the BEM in rough surface contact mechanics can be found in a comparative analysis of BEM formulations [152]. There are also many instances in which FEM and BEM can be coupled into FEM-BEM solvers for the solution of three-dimensional contact problems [153] or can be combined to achieve different levels of refinement in the solution to the problem under investigation (see, e.g., [154]).
The application of the FEM to tribological problems involves the discretization of the volumes of contacting bodies and an appropriate treatment of their contact interaction. The arbitrariness of material models, as well as the geometries of contacting solids and their heterogeneity that can be reached in the treatment of contact interfaces, make this method a multipurpose engineering tool. However, this is all at the cost of a higher computational complexity than in the BEM, which has less versatility but a much higher efficiency in the treatment of interfacial problems, since it requires solving the problem only for surface degrees of freedom and does not require any discretization in the volume. On the other hand, the BEM results in dense systems of linear algebraic equations, contrary to the FEM, which renders sparse systems of equations. Thus, the BEM has to rely on iterative solvers, whereas the FEM can successfully use either iterative or direct solvers based on the sparse matrix storage.

When interested in near-surface stress fields, which are crucial in the reliable analysis of surface deterioration (e.g., fretting fatigue and wear) and microscopic contact at the roughness scale, imprecise integration and/or discretization may result in huge errors in local fields and, thus, in realistic estimations. To properly capture the stress field in the vicinity of a contact zone, and especially near its edges (which, in most problems, is unknown), requires a very dense spatial discretization. The accuracy of the integration technique is especially crucial when a conformal mesh cannot be ensured on the contacting parts (e.g., large-deformation or large-sliding contact systems) and if two deformable solids of comparable stiffness are brought into contact, i.e., when one of the solids cannot be considered as rigid. In addition, the path-dependence of frictional problems requires that the load increment should be chosen properly, as the temporal discretization plays a crucial role even in quasi-static problems: as an example, for the shear tractions in normal Hertzian cylindrical contact with friction in the interface, the self-similar character of the solution, as argued by Spence [155], can be obtained with one hundred load steps with the displacement increment proportional to the time squared, but not within one single load step.

In tribology, due to its computational cost, application of the FEM is justified if the problem at hand cannot be solved within the assumptions of the BEM, namely the existence of a fundamental solution and the local flatness of the surface (small slope). A broad family of systems falls within this context: large-deformation, large-sliding contact of soft bodies, which can be observed in various biological systems (oral food processing, contact of skin, etc.), but also in engineering applications (contact of tires, polymeric seals and many others) or contacts involving strongly nonlinear material behavior which is hard to represent within the BEM framework such as indentation involving strong finite-strain plastic deformations or fracture in the interface.

Concerning the applications to microcontacts and microtribology, both FE and BE methods are used extensively. At the scale of roughness, the macroscopic shape of the contacting solids can be usually neglected and, since the roughness slope is in general rather small, the problem satisfies the main assumption of the BEM, which can be successfully used for its solution. The evolution of the true contact area, interface permeability, electric and thermal contact resistance can all be resolved in the framework of the BEM for linear material laws. Regarding material nonlinearities, elasto-plastic [156-158] and viscoelastic [159,160] material behavior can be incorporated in the BEM framework by assuming that deformations and slopes remain small, otherwise an FEM would be needed [161,162]. It should remarked that most contact systems involving elasto-plastic materials operate mainly in the elastic regime both at the micro- and macroscales; hence, depending on the level of stress and the type of loading, considering plastic deformation may be important during the first loading cycles but may not be needed in subsequent ones. Furthermore, severe plasticity is associated with wear and must
therefore be incorporated in the simulations, but how can one explicitly model wear numerically (e.g., using both BEM and FEM)? The issue of wear is partly discussed in §3.9.1.

The BEM framework can consider homogeneous nonlinear material behavior, but can also account for heterogeneous inclusions in the bulk, see e.g., [163], which is computationally much more expensive. Accounting for heterogeneous materials is often critical for microscale analyses in which the material’s microstructure might play an important role. This, for example, is the case in contact problems involving functionally graded interfaces [164], metallic polycrystalline [165] or monocrystraline [166] microstructures, whose accurate treatment requires the FEM. Concerning multiphysical (multi-field) problems, both methods are comparable at the scale of roughness, with the same limitations and advantages: simple but fast BEM versus slow FEM but with capabilities to account for arbitrary complexity. Examples of applications include: lubrication problems [167-169], electro-elastic contact modeling [170,171], thermo-mechanical coupling [172], and many others. Using BEM-type formulations has also been used to treat elasto-dynamic frictional problems [173,174], whereas complex geometries and boundary conditions would still require usage of FEM or equivalent formulations [175,176].

In summary, both the FE and BE methods are well developed and able to solve most micro-tribological problems involving both material nonlinearities and multiphysical couplings with the FEM being more versatile and more easily accessible for a general researcher and engineer (numerous commercial and open software are available) but computationally costly, and the BEM being less available and versatile, but still capable of solving most problems under reasonable assumptions and for very moderate computational costs. The main challenge here for the researchers and engineers would be to promote both methods within the homologue communities and to enable them to use one or the other based on the needs of the target application.

2.3. Crystal plasticity and Discrete Dislocation Dynamics

Crystal plasticity is a well-established constitutive framework for the modeling of elasto-plastic deformations of metal crystals [177-180]. The essential feature of crystal plasticity is that plastic deformation is assumed to result from plastic slip on specified crystallographic slip systems. An individual slip system is active when the shear stress acting on it (called the resolved shear stress) exceeds the corresponding critical resolved shear stress, the latter being governed by an evolution (hardening) law that is expressed in terms of slip rates for all active slip systems. By considering the crystallographic features of plastic deformation, crystal plasticity provides a physics-based continuum description of single crystals as well as of individual grains in polycrystalline aggregates [181,182].

Once combined with a suitable scale transition scheme (mean-field homogenization, Representative Volume Element (RVE)-based computational homogenization, etc.), crystal plasticity has proven to be highly successful in predicting the effective elasto-plastic behavior of polycrystalline aggregates, e.g., [183-185]. A notable example is the visco-plastic self-consistent (VPSC) model [186], which is widely used for predicting hardening and texture evolution in plastic forming processes. The crystal plasticity framework has also been extended to include, in a simplified manner, other deformation mechanisms, such as deformation twinning [187,188] and martensitic phase transformations [189,190].

Being a continuum theory, crystal plasticity is not applicable at very small scales at which discrete events, e.g., those related to the nucleation and propagation of dislocations, become important, and other approaches, such as discrete dislocation dynamics (see below) and molecular dynamics (see §2.4), are then more appropriate. Even at higher scales, important phenomena that accompany plastic deformation, e.g., the formation of dislocation structures, deformation banding and grain refinement,
are not captured by the available crystal plasticity models, even though attempts in that direction have been made [191-194]. In general, plastic deformation is inhomogeneous at multiple scales, and crystal plasticity is not capable of describing many of the related phenomena.

Discrete Dislocation Dynamics (DDD) is a modeling technique to study plasticity at the microscale [195-200]. In DDD, the solid is modeled as a linear elastic continuum, and the dislocations by means of their linear elastic fields, which are accurate outside of the dislocation core. Atomic aspects are included by means of constitutive rules that govern dislocation nucleation/annihilation, glide, and interaction with obstacles and dislocations. Given that both the dislocations and the solid are described using linear elasticity, it is possible to solve boundary value problems relying on the principle of superposition. The solution to the boundary value problem is given at each time increment and at every material point as the sum of the dislocation fields and their image fields. The image fields can be calculated using finite elements, although, for contact problems, where rough surfaces need to be described using a fine discretization, it is computationally more efficient to use other techniques, such as, for instance, Green’s Function Molecular Dynamics (GFMD) [201].

Important recent advances in this area include, for example, the development of a formulation that incorporates elastodynamic effects in the description of the interactions between dislocations. The resulting methodology, Dynamic Discrete Dislocation Plasticity (D3P; see, e.g., [202]), allows the treatment of problems characterized by high strain rate deformation such as shock waves [203] and could be used to perform concurrent coupling (see §2.6) with atomistic simulations in order to avoid issues with the transition between the atomistic-continuum boundaries. Furthermore, concurrent methodologies (also see §2.6) to directly couple crystal plasticity and DDD have been also developed [204,205] to take advantage of the fact that the DDD formulation is only required in very small regions in the presence of stress concentrations, such as cracks and indentation of asperity-to-asperity interactions.

2.4. Modelling Methods for Lubrication, Solid/Fluid Interactions and Particle Dynamics

The computational methods introduced in the previous two sections mainly cover formulations and methodologies adopted to model individual dry contact problems and focus on detailed descriptions of solid deformations and stresses. However, other techniques must be adopted when modeling lubrication and solid/fluid interactions in the presence of a fluid film interposed between contacting bodies and when multiple contacts are generated simultaneously through the complex interactions between many particles. In this sub-section, we give an overview of standard and advanced methods developed over the last century to predict film thickness, friction, rheological response of fluids and interactions between surfaces in lubricated conditions. A brief summary of the techniques developed to study interactions between particles in different environments and the dynamics of systems involving multiple contacts is then provided.

Hydrodynamic Lubrication (HL) and Elasto-Hydrodynamic Lubrication (EHL) are lubrication regimes where a thin lubricant film is formed between two surfaces in relative motion. HL takes place in conformal contacts, when low pressures are established between the two surfaces, while EHL takes place when pressures are significant enough to cause considerable elastic deformation of the surfaces. EHL usually occurs in non-conformal contacts and many machine elements, including rolling bearings and gears, rely on EHL in their operation. Existence of a fluid film sufficient to separate two surfaces under hydrodynamic conditions, such as in a journal bearing, has been known since the work of Tower in 1883 [206]; however, it was not until 1949 that Grubin predicted that a thin fluid film can also separate surfaces in high pressure, non-conformal contacts [207]. Formation of such a film is possible due to high pressure having two beneficial effects: firstly, it increases lubricant viscosity in the contact
inlet and, secondly, it elastically deforms and flattens the contacting surfaces, hence the term elasto-hydrodynamic lubrication.

Classical solutions of HL and EHL contact problems use the Reynolds’ equation [208] to describe the behavior of the lubricant, while elastic deformation is traditionally calculated using Hertz theory of elastic contact, although nowadays BEM or FEM solvers are also routinely used. Reynolds's equation is a simplification of full Navier-Stokes equations, derived by assuming a Newtonian lubricant with constant density and constant pressure and viscosity across the film thickness. Cameron et al. [209] developed the first Reynolds-based computerized numerical solutions for hydrodynamic lubrication and in 1959 Dowson and Higginson [210] produced the first full numerical solution for EHL. Subsequently, Dowson and co-workers, also proposed regression equations for prediction of the EHL film thickness based on their numerical solutions and a number of other improvements including the consideration of material properties and thermal effects (e.g., [211-213]. In the last fifty years, many numerical approaches [214-218] have been developed to address the solution of this set of equations: nowadays, it is possible to account for a variety of non-Newtonian effects, ranging from piezo-viscosity to shear thinning. The majority of these approaches uses a Finite Difference (FD) scheme, although the use of the FEM and Finite Volume (FV) methodologies has recently been proposed especially to overcome some of the limitations of FD when dealing with complex domains in the presence of micro-textured surfaces and cavitation using mass-conserving algorithms [219-221], but also to extend a Reynolds-type solver to full Computational Fluid Dynamics (CFD) studies looking at the fluid flow outside the contact, overcoming the limitations of the Reynolds’ assumptions in specific extreme contact conditions [222-225]. The development of fully-coupled Solid/Fluid Interactions (SFI) solvers [226] constitutes the new frontier of this particular area of research, with the promise that advances in computational power may lead to a more comprehensive study of the multiphysics phenomena governing three-dimensional contact problems considering full field deformations, thermal and multi-field effect, and the complex rheologies of the fluids and the solids under investigation. Hybrid techniques (e.g., the element-based finite volume method – EbFVM [227,228]) have also been recently developed to combine the flexibility of finite elements in terms of studying complex domains and using unstructured meshes and the use of finite volumes to accurately solve the fluid-dynamic problem at hand.

Another important area of interest, often to industrial applications, is the solution of problems involving particle interactions and multi-body contacts, as many industrial and natural processes involve granular systems. Diverse phenomena such as avalanches, fluidized beds and asthma inhalers all depend on assemblies of particles. The understanding of such systems is therefore of interest to a number of scientific disciplines, as well as industry. Due to their complexity, it is often very difficult to study such systems, in which large numbers of particles interact, and macroscopic behavior depends both on the physical properties of individual particles, and the interactions between them. The Discrete Element Method (DEM) is ideally placed to tackle these contact configurations, as it allows the description of the physical state of a system using a large number of discrete elements. This approach shares many similarities with atomistic simulations (see §2.5) where atoms are replaced by particles that interact via constitutive equations rather than interaction potentials; however, depending on the problem under investigation, the DEM requires constitutive laws to describe individual interactions, which often are obtained by adopting hierarchical multi-scale approaches (see §2.6). Noticeable examples are studies of particle-particle interactions to derive elastic, viscoelastic and plastic constitutive laws that capture the right kinematics during particle collisions [229-232] and the integration of the effect of adhesion [233,234], particle shape [235,236] and roughness [237,238] into
DEM codes. Recently this method has been also used to study wear involving complex fragmentation, but also problems affected by complex rheological and/or multi-physics behavior [239,240].

2.5. Atomistic methods

Molecular Dynamics (MD) was first developed to study the interaction of hard spheres [241] and, in the following decades, has been expanded into methods and tools suitable for investigations in a number of physical, chemical and mechanical phenomena both for diagnostic [242-247] and predictive purposes [248-257]. Classical MD essentially calculates the kinematics of atoms (or representative “particles”) by solving their Newtonian (or Langevin) equations of motion based on potentials that describe the interactions between them. This tool was applied to the study of tribological interfaces especially in the high speed regime, which lends itself to the length and timescales of MD [258-261]. Other examples of studies include: elementary phenomena such as the mechanical mixing between two surfaces in contact [262]; different wear regimes [263], plastic deformation [264,265]; the tribology of Diamond-Like Carbon (DLC) coatings [266]; the frictional behavior of self-assembled layers formed from additives [267,268]; the rheology of lubricant films in contact in the EHL regime [269,270]; and other tribological phenomena including friction, adhesion, and wear [271].

The classical MD framework can provide a description of the dynamics at atomistic level, but without explicitly modelling individual interactions in terms of surface reactivity, bond formation and evolution of electronic structures, which can be dealt with using first principles or ab initio MD techniques (examples of this include Car-Parrinello MD [272] and Tight-Binding Quantum Chemical MD (TB-QCMD) [273] and will be discussed in more detail at the end of this sub-section); hence, the key ingredient of any classical MD simulation is the interaction potential (also referred to as the Force Field, FF). Even though the availability of suitable interaction potentials is still a limiting factor for the study of complex systems, several families of FFs have been presented in the literature (along with their explicit parameterization), each of them designed to capture the essential features of a different type of material. The simpler functional forms of FF are represented by pairwise interactions that generally account for an attractive (describing London dispersion forces) and a repulsive term (originating from core-core repulsion). Probably the most popular examples are the Lennard-Jones (LJ) [274] and Morse potentials [275]. The number of (free) empirical parameters is kept at a minimum (for each atomic species, this number is two and three for the LJ and Morse potentials, respectively), as is the computational cost of simulations based on these FFs. It turns out that the LJ and Morse potentials are not able to realistically describe the behavior of many materials (for example, the LJ potential can accurately model noble gases only). Nevertheless, the usage of the LJ potential has produced fundamental results over the years, as evinced, for example, in the prediction of the breakdown of continuum contact mechanics at the nanoscale [276,277], discussed in more detail in §3.3, and in Non-Equilibrium MD (NEMD) simulations to shed light on the phase behavior of fluids in confinement [278-280].

A class of potentials routinely used in tribology can be grouped into the family of non-reactive FFs (see, e.g. [281-284]). This class of potentials is often employed to model intramolecular interactions in organic molecules and contains several two-, three- and four-body terms (usually including LJ, electrostatics, bond stretching, angle bending and torsional parts). As already mentioned, despite the simplicity and relatively low computational cost of such non-reactive FFs, a fixed topology has to be provided as an input for an MD simulation, thus preventing the possibility of investigating tribochemical reactions or events that require the breaking/formation of chemical bonds in general.

Metallic systems are more often (and more accurately) described by the family of the Embedded Atom Method (EAM) potentials [285]. EAM potentials comprise a pairwise repulsive term modeling the
core-core interaction and a cohesive contribution representing the energy that an ion core experiences when it is “embedded” in the electron density originating from neighboring atoms. The use of these many-body potentials overcomes intrinsic limitations of two-body potentials, which, for example, are bound to satisfy the Cauchy relation, or lead to defect energies that correlate with cohesion energies much more strongly than in real materials. Examples of the application of the aforementioned potentials in tribology are studies of the frictional behavior of an indenter tip against different metallic surfaces [286-289], or the interfacial friction characteristics of different metal pairs [290]. For carbon-based (e.g., diamond, graphite/graphene, diamond-like coatings, nanotubes) and other covalent systems, a series of FFs has been developed, all based on the bond order concept originally formulated by Pauling [291]. Examples include the Finnis-Sinclair [292], Tersoff [293] and Brenner [294] potentials, as well as more recent derivations such as the Adaptive Intermolecular Reactive Empirical Bond Order (AIREBO) [295] and ReaxFF [296] FFs. These all share the common assumption that it is possible to properly model the strength of a chemical bond on the basis of the bonding environment, thus considering the number of bonds and, if necessary, bond lengths and bending angles. Such kinds of potentials have been successfully used to investigate the tribological properties of different systems, including the interaction between diamond samples [297-299], the frictional behavior of corrugated nano-structured surfaces [300], the wear mechanisms of tungsten-carbon systems [301], friction and adhesion properties of carbon nanotubes and polymers [302,303], and tribochemical reactions on silicon/ silicon oxide interfaces [304,305].

Classical MD –especially when calculating and tracking the kinematics of all atoms (all-atom MD) as opposed to aggregates of these (united-atom or coarse-grained MD)– require significant computational resources, meaning that the method is usually reserved for systems of relatively small sizes (less than a cubic micrometer) studied for a short time (less than a microsecond), even with today’s increased capabilities. In what is essentially a boundary element method, Green’s Function MD (GFMD) [306] integrates out “all internal (harmonic) modes of an elastic body, […] leading to effective interactions of those atoms whose degrees of freedom couple to an external force.” In this manner, “the full elastic response of semi-infinite solids is incorporated so that only the surface atoms have to be considered in molecular dynamics simulations” [307]. GFMD is being used extensively in the study of tribological systems, including in the recent contact-mechanics challenge summarized in §3.5.

When modeling tribochemistry, MD techniques [284,308,309] or quantum calculations (using Density Functional Theory, DFT) [310] are used to study atom motion during friction or chemical reactivity, respectively. To combine both types of information, reactive force-field MD [311], ab initio MD techniques [272] or tight-binding coupled with MD [312] techniques have also been used to extract in situ information of interfacial material behavior. A deeper insight of the local electronic and geometric characteristics is required to capture subtleties that a molecular mechanical description cannot represent. Indeed, quantum mechanical approaches have been used toward this aim, e.g., [313], focusing on the theoretical modeling of a specific stoichiometry and chemical composition. Tribochemistry is discussed in more detail in §3.9.2.

2.6. Multiscale modeling: concurrent and hierarchical schemes

By multiscale modeling, one refers to a technique in which two (or more) different models related to different scales (or different matter descriptions) interact, i.e. exchange data, in a way that enhances the information that can be obtained about the modelled phenomenon. Contact between rough surfaces with geometrical features present on multiple scales, starting from the shape of contacting solids down to the atomic fluctuating nature of the “surface” at the nanoscale, is an example of a spatially multiscale problem. Earthquakes, on the other hand, constitute the most characteristic example of a temporally multiscale problem, in which the stresses building up in the earth’s crust for many years
are released within seconds inside the fault zone, giving rise to seismic waves. In general, spatially multiscale problems are much more complicated to model than temporally multiscale ones, as time is only a one-dimensional quantity. Consider, for example, a multiscale contact problem between rough surfaces: this can be solved using either a classical model (FEM, BEM and so on, as discussed in §2.2), e.g., as in [119], or a multiscale model, e.g., as in [314]. In such a multiscale model of rough contact, the upper scale model (e.g., treated with the FEM) determines the state (for example, the contact pressure) for the microscale (e.g., treated with the BEM), whereas the microscale provides the upper state with some properties of the contact interface such as, for example, the contact stiffness, contact area, friction, etc.

Having been generalized by many authors, the problem of multiscale rough contact inspired numerous theoretical and computational studies aimed at understanding the role of roughness at different scales of observation; see, e.g., [112,314-317] among many others. Recently, the topic has gained renewed interest with the increased potential of MD in studying nanoscale contact problems [263,276,318,319] that unveil interesting mechanisms of contact interactions occurring at the nanoscale. While the advent of MD opened new challenges due to the still limited time and size scales of the simulations that can be performed with the aid of supercomputers, it has also revealed new opportunities for the use of various multiscale approaches.

An important question in multiscale modeling is the following: how to identify which spatial and temporal scales and mechanisms are relevant for understanding the phenomena to be modelled? A simple recipe would be to start with a simpler model, based on a single scale and uncoupled physical processes, and then adaptively introduce additional scales to permit coupled multiscale-multiphysics considerations, whenever and wherever these are needed, until the simplest possible model is obtained. Scale, in this context, does not only refer to the spatial and temporal dimensions, but also to the different computational models relevant to different scales. Inevitably, some multiscale coupling also implies multiphysical coupling as, for example, in the case of coupling mechanical FEM with classical MD in which thermal oscillations are inherent to the model [320]. However, this simple recipe can be often ineffective as it depends on the ability of the “user” to add the right details at the right scale and may lead to the neglect of important information flow across the scales.

In tribological models, key processes are usually localized in a thin interface layer, but have important implications or can even fully control the macroscopic behavior of the system. In this light, the interfacial laws of friction, wear, heat and electrical transfer, as well as other relevant phenomena can be obtained with microscale models for use in macroscale ones. In terms of accuracy, one can determine two levels in this hierarchical approach: 1) the microscale model is assumed to not affect the macroscale state, in which case the microscale data can be obtained by simply post-processing the macroscale results; 2) the microscale model affects the macroscale state and, thus, the constitutive interface model has to be directly included in the latter scale. For most applications in which scale separation between the micro- and the macroscales exists, a hierarchical multiscale model is acceptable and the relevant question would be: when would a finer and truly multiscale model –i.e., one which requires stronger scale coupling– be needed? Normally, a finer model is required when no scale separation exists, as is normally the case for surface roughness. Such models, dealing with concurrent multiscale coupling are in general much more complex and can hardly be used to obtain statistically meaningful results; see, e.g., [204,320-322]. At the same time, finer models can be used for rare-event simulations and are of high importance in understanding the physics of certain phenomena happening in contact interfaces such as dislocation interaction with free surface in contact interfaces [323,324], ballistic heat diffusion through small contact spots [320,325], partial slip
conditions in lubrication at the molecular level [326], and so on. Most such phenomena can be studied at a single relevant scale and integrated at a bigger scale in a hierarchical manner.

In the case of plastic deformations occurring, for instance, during sliding motion between two metallic surfaces, many dislocations are nucleated at the surfaces and under maintained load may travel long distances. In an MD simulation, the small size of the domain will artificially trap them and create artificial hardening, which should occur in very thin coatings. In order to address this issue, advanced concurrent coupling strategies are being developed where dislocations can be passed to a continuum representation [327,328]. In three dimensions, dislocations are line networks, so that a dislocation may cross the coupling interface. Such hybrid dislocations should behave as single dislocation structure, which requires the use of reciprocal boundary conditions and may significantly increase the complexity of coupling strategies.

Another important aspect to consider is the possibility to perform concurrent coupled simulations where atomistic and molecular details need to be captured near the wall in lubricated contacts when the fluid film is larger than the Root Mean Square (RMS) composite surface roughness; this is particularly useful, for example, when slip at the wall or atomistic details of the surface topography must be explicitly modelled. In this case, MD-continuum coupling strategies involve the transfer of information between MD and CFD, and particular care must be taken when the two descriptions merge [329,330]; a number of schemes exist to achieve this [331-333].

Finally, comparison with experimental data is of crucial importance for all types of models, and multiscale ones are not an exception. Difficulties here arise from the fact that it is not always possible to reproduce the relevant scales for the application/model in the lab. For example, the friction of rocks (as well as their fracture) is a very scale-dependent phenomenon [334] that is intimately linked to the probability of presence of critical defects in a given volume. The related key question in this example would be: what are the features of real earthquakes, which can be reproduced in the lab? Also, can multiscale models tuned at the lab scale, e.g. [335], be used at earthquake scales? Further research on scale separation in contact interactions is required to guide the choice of the most appropriate computational method preserving the accuracy of the description of a given physical problem while considering the effect of inherent uncertainties.

3. Research themes in tribology

The problem of normal contact between rough surfaces has been studied extensively—for example, the reader is referred to a recent paper on a contact-mechanics challenge whose results are summarized in §3.5 and can be considered to be well understood, but almost all other issues in tribology remain open for future research. While different theories, techniques and models used to investigate these issues were reviewed in §2, this section introduces active topics for modeling research in tribology. As a foreword, let us emphasize that, since the global forces acting on an interface are integral quantities along the interface (for example, the friction force is the integral of the shear stress over the contact area), various models can predict rather similar forces using different assumptions. Comparisons of models to experiments are therefore necessary, not only in terms of global forces but also in terms of local measurements, for instance, of temperature, strains or the real area of contact. Multiple successful examples of such comparisons can be found in the literature [336-346]. Local measurements become increasingly accessible due to the miniaturization of local probes and the development of full-field evaluation techniques like Digital Image Correlation (DIC) [347] or infrared imaging [348]. Imaging techniques are especially interesting for performing local measurements at a contact interface in a non-invasive way, but the choice of possible materials is limited as they must be
transparent to the radiation used (e.g., visible or infrared light). In this context, wherever relevant, we will also present experimental results that are amenable to direct comparison with models.

### 3.1. Multiphysical phenomena in tribology

All tribological phenomena happening near interfaces between solids are determined by the atomic interactions within and between solids, as well as those between atoms of the substances present at the interface. Since these interactions give rise to various physics described at the macroscale by different theories and models, the tribological interface can be considered a “paradise” of Multiphysics (coupled multiple fields; see Fig. 2). The following types of phenomena may take place in such an interface or in its immediate vicinity: mechanical (solid and fluid), thermal, electro-magnetic, metallurgical, quantum and others.

![Multiphysical phenomena in tribology](image)

Mechanical phenomena can refer to the mechanical deformation of solids and their contact interaction including adhesion and friction. The process of material removal or surface deterioration (microcracking, abrasive and adhesive wear) can be also included within this type. Thermal phenomena are related to heat transfer from one solid to another, as well as to heat generation due to interfacial friction or due to dissipation in the bulk (viscoelasticity, viscoelastoplasticity, damage accumulation or micro-fractures): heat exchange can be either ballistic or diffusive depending on the size of contact spots [349-351], while radiative and convective heat exchange also contribute considerably to the overall heat conductance [352]. The local heating of contacting asperities up to the point of local melting, recognized in early tribological studies [353] and known as flash-heating, has important implications for friction, especially in dry contacts [354,355]. Metallurgical phenomena happening in near-interface layers span various microstructural changes that are, either, triggered by changes in temperature (e.g. because of Joule or frictional heating) or by severe deformations, and include dynamic recrystallization and various phase transformations; an example is the formation of the so-called “white layer,” a fine-grained and rather brittle martensitic layer [356].

For materials experiencing glass transition, the local rise in temperature can be critical for their mechanical performance [357]: in general, mechanical properties are strongly dependent on the temperature, thus making the thermo-mechanical problem one of the most natural and strongly coupled multiphysical problems in tribology, especially in dry contact or in the mixed lubrication regime. Because of excessive local heating, the solids can reach their melting or sublimation point and...
experience phase transition [353]; thus, melting, evaporation and sublimation appear to be important phenomena in dry and lubricated micromechanical interactions. More complicated physics emerge for composite and porous materials; examples of the latter are rocks experiencing chemical decomposition, water evaporation, pressurization, and so on [358,359]. A complex interaction of the aforementioned physics with a fluid present in the interface is another strongly coupled multiphysical problem, especially for EHL (see §2.4 and §3.8), sealing applications and saturated fractured media [360-362]. In most situations, the interfacial fluid flow can be considered as a thin flow that can thus be properly described by the Reynolds equation but, in the case of the fluid viscosity depending on the pressure (piezoviscosity) or temperature, a consistent development of the Navier-Stokes equations for thin flow should be performed with a priori included pressure dependence in the original equation and not directly into the Reynolds equation [363].

In addition, tribofilm formation and various tribochemical phenomena taking place at tribological interfaces make them very challenging objects for multiphysical research [310,364]. At the same time, to understand and model such a complex multiphysical problem as a tribological interface, one needs to construct reliable multiphysical models and design appropriate multiphysical tools. Some recent examples of tribology-related modeling applications involving multiphysical coupling include, for example, excitable biological cells (see §3.9.5), weakly coupled modeling of creeping fluid flow through the contact interface between rough solids [365], and electro-mechanical coupling in contact problems [170]. Because of the complexity of direct experimental measurements and the inseparability of various multiphysical mechanisms in real interfaces, a big challenge is to construct reliable and precise multiphysical models having predictive power while, at the same time, being verifiable and sufficiently comprehensive.

3.2. Surface roughness

Real (engineering) surfaces brought into mechanical contact touch only over a number of discrete contact spots forming the real or true contact area, which, in general, is much smaller than the nominal contact area that can be computed for the case of perfectly smooth surfaces. Under increasing pressure, the true contact area grows towards the limit of the nominal one that can be reached under relatively high squeezing pressures. The integral true contact area, as well as the localization and morphology of the clusters of true contact, affect numerous tribological mechanisms and thus present a topic of intensive engineering and scientific research. In particular, the following quantities are dependent on the true contact area: 1) the stress state near the contact interface, which is proportional to the applied stress and inversely proportional to the true contact area; 2) friction, adhesion and adhesive wear; 3) the transport of electric charge and/or heat through the contact interfaces; and, finally, 4) the fluid flow through the free volume of contact interfaces in sealing problems. Apart from the phenomena affected by the contact area, roughness is responsible for the additional interface stiffness of contact interfaces, which can be related to heat/electrical conductivity [366]. To understand the effect of roughness on all aforementioned phenomena, accurate mechanical models are needed.

One of the fundamental issues in the modeling of contact between rough surfaces is the realistic representation of roughness. As the roughness of real engineering surfaces spans multiple length scales—whether measured experimentally or created using numerical methods, for example, via simulations of sandblasting and shot peening [367], or through surface randomization algorithms [121,135,368,369]—, the question is essentially which length scales are relevant to a specific tribological system or, alternatively, to what extent should one implement accurate roughness representations in a tribological model? The wealth of parameters used in roughness characterization—amplitude (\(S_a\), \(S_q\), \(S_{sk}\), \(S_{ku}\)), spatial (\(S_{al}\), \(S_{tr}\), \(S_{td}\)) and hybrid parameters (\(S_{dq}\), \(S_{dtr}\)), or Abbott-
Firestone (bearing area) curve-based parameters ($S_k, S_{pk}, S_{sk}$, material ratios, and volume parameters for 3D measurements)– demonstrate the complexity of reaching a universal description of surface roughness; see, e.g. [370,371]. Indeed, most models use only a small subset of those parameters, the ones deemed necessary to describe a specific function.

Representations based on concepts of self-affinity were apparently introduced to tribology much more recently, although Archard first introduced a concept of fractals already in 1957 [111] with his model of spheres upon larger spheres upon larger spheres applied to contact and friction. A key point is what was recognized into tribology with Whitehouse and Archard [372]: they first introduced the topography’s Autocorrelation Function (ACF), and noted that the Fourier transform of the ACF, i.e. the Power Spectrum Density (PSD), of their topographies was a power law at large wavevectors, as Sayles and Thomas [373] would later confirm for a number of surfaces. One implication of their work was that between one-third and one-quarter of all the sample points of their topography would be a peak, regardless of the sampling interval they chose, while the mean peak curvature depended strongly on the sampling interval. The tribology community still debates on the effect of the upper wavevector truncation in the PSD, which significantly affects contact area, rubber friction dissipation, and many other physical properties. On the contrary, the fact that the lower wavevector determines the RMS amplitude for non-stationary roughness has been neglected in later literature, since the time of highly influential works on stationary roughness by Longuet-Higgins [116] and later by Nayak [114] on whose basis most multi-asperity models are constructed (see §2.1.2).

A very interesting finding of Whitehouse and Archard came when they measured the profile of a rough surface along the same track, before and after a single passage of a lubricated slider. They found that, while the main scale roughness was still present, all the fine scale roughness had been removed [372], a finding which also tends to be neglected in the literature. Keeping in mind the limited metrology of the time, one could ask to what extent we should measure or worry about the initial roughness when irreversible deformations might remove it? On the other hand it is known that, if a metallic sample is heated after mechanical polishing, the initial surface roughness might reappear on its surface [374].

Much emphasis in modeling is placed today on nominally flat stationary self-affine fractals, while very little work was performed on the macroscopic “shape" of surfaces –particularly in the presence of adhesion–, where the basic contact problem of a rough sphere remains incompletely understood. One exception is a rather special case of roughness for the sphere (axisymmetric waviness) which can be solved analytically [375]. Otherwise, numerical calculations are necessary and in this case it may be of little interest to argue a priori on models describing shape and roughness assuming they consist of very separate scales. Summarizing, most of the real practical problems remain unanswered: what is the real contact area? How can it be estimated quantitatively form “scale/magnification-dependent” quantities? Which mechanisms (plasticity, failure processes, adhesion at small scales) does one need to incorporate to converge to a well-defined value?

Following the introduction of fractal roughness, numerical models began to utilize the PSD to fully define surface roughness. However, one has to keep in mind that the PSD does not represent the full information about topography: different realizations of surfaces in real space are possible for the same PSD, depending on the phase associated to each spectral component [376]. While the effect of deviation from Gaussianity has limited effect on some quantities, it can be crucial for others. For example, even small deviations from the ideal Gaussian random roughness case seem to lead to a dramatic increase in adhesion for rough surfaces due to a finite number of asperities or a finite tail (unlike the infinite nominal Gaussian tail) in the asperities’ height distribution [377-381]. Furthermore,
as modern fractal parameters do not include such traditional ones as skewness, there might be an advantage in using traditional characterizations, perhaps to augment fractal ones for non-Gaussian surfaces, e.g. [382].

The perceived universality of the PSD in fully describing surface roughness was demonstrated by Persson who showed that a 1D line scan, a 2D Atomic Force Microscopy (AFM) scan and a 2D Scanning Tunnel Microscopy (STM) scan all lie on the same PSD plot for a grinded steel surface with the fractal dimension being $D_f = 2.15 \pm 0.15$ for many engineering surfaces [367]. At the same time, however, and in the absence of random phases, a profile PSD with a slope of $-2$ (as in the work of Whitehouse and Archard) does not necessarily represent a rough surface, but can also be a square wave (that has all phases equal to zero), while a slope of $-3$ may well correspond to semi-circles nestling together. Also, having a Gaussian distribution of heights does not automatically suggest uncorrelated spectra. Higher order autocorrelation functions may be needed but the topic of non-Gaussian fractal surfaces is not very developed at present. It is worth mentioning here that, in many practical applications, the surfaces in contact are actually non-Gaussian; road surfaces, worn-out or polished surfaces, blasted surfaces, etc. The class of anisotropic rough surfaces, also very frequent in engineering, is also relatively unrepresented in modern modeling.

On the critical issue of the definition of the low- and high-frequency cutoff values of the roughness PSD, some macroscopic quantities, such as stiffness, electrical and thermal conductance, are well known to depend principally on the RMS amplitude of roughness, i.e. on the lower frequency contents of the PSD, as demonstrated by Barber [366]. Other quantities, like the real contact area or the RMS slope of the topography depend on the higher frequency part of the PSD. This suggests that attempts to measure the real contact area with indirect methods, e.g., measuring conductance, have the intrinsic difficulty of measuring two quantities which depend very differently on the PSD content. The reader should keep in mind that the high-frequency cutoff or, more realistically, the requirement of surface smoothness necessary in analytical and numerical models, is not a well-defined quantity. The roughness persists down to the atomistic scale [383,384], which is usually challenging to measure and goes beyond the continuum description of matter [276].

The metrology of surface roughness measurements plays a crucial role in our understanding of roughness as well. Abbott and Firestone measured surface roughness by using a pen-recorder to draw an amplified version of the motion of a “stylus” (a broken razor blade) over a surface [385]. Since then, a multitude of techniques have been developed or adapted for measuring roughness: contact and optical profilometry, stripe projection scanning, Scanning Probe Microscopy (SPM), Transmission Electron Microscopy (TEM), etc. The scope here is not to give an extensive overview of those various methods, which the reader can find for instance in [372,386,387]. The main message to be conveyed here is that these techniques, whether contacting or non-contacting, present a number of limitations and artefacts that should be carefully taken into account when interpreting the data (see e.g. [388] for white light interferometry and [389,390] for scanning force microscopy). Knowledge of those artefacts is particularly important when using contact mechanics or lubrication models based on topographical features [391]. It is well known, for example, that the stylus tip geometry filters the measured signal, while high contact stresses at the stylus tip can lead to significant deformations [392]. Post-processing is also critical in extracting roughness information from raw data with a number of aspects –shape removal (tilt), the restoration of missing data (“perforated” surface data) using built-in triangulation or grid-fit routines, and the filter type and cut-off length (Gaussian versus Robust Gaussian Regressive Filter, RGRF)– affecting the end result. Furthermore, artefacts may occur due to diffraction effects around sharp edges caused by calibration grid height steps. In certain cases, results differ across measurement methods: comparisons of contacting and non-contacting measurement techniques show
large differences in predicted bearing curves, for example, with confocal microscopy typically yielding higher roughness values than atomic force microscopy [393].

3.3. Scale effects and the breakdown of continuum theories
Contact between two bodies –perceived as continua– is well-defined and occurs when the distance between them is zero; however, the same reasoning cannot be applied to the atomistic scale. Also, experimentally atomistic contact area cannot be measured, even in the hypothetical case that the instrument would allow such a resolution, since atoms have no well-defined boundaries. Luan and Robbins numerically studied the contact between a flat surface and nanoscale indenters of different structures (spherical crystalline, amorphous and stepped crystalline) and showed that the details of the atomic structure matter in the contact pressure distribution in adhesive versus non-adhesive contact conditions [276,277]. Subsequent work by other research groups showed that the accurate calculation of the contact area at a given length scale could yield reliable results [140,244,394,395], but this requires careful post-processing and interpretation of atomistic results with appropriate definitions of criteria for contacting atoms and the “area of contact for an atom.” For the latter, one method of calculation involves the assumption that the real contact area is the sum of the contact areas of each atom determined to be in contact [244,396]. But is the concept of contact area really meaningful for atomistic models? Similarly to the notion of contact itself, the contact area is a well-defined quantity only at low magnifications, i.e. at scales where the discrete nature of atoms is not relevant. Perhaps extracting the pressure distribution over the interface by looking at the distribution of forces [276,277] may be more meaningful than attempting to measure the real contact area with indirect methods; furthermore, the contact area is difficult to measure experimentally [343] since transparent materials need to be used to image the interface, while no information can be obtained at scales below the pixel size, which may yield errors in the real area of contact of the order of 10% [397].

The concept of contact distance, defined as the distance between atoms at which contact occurs, is a well-defined quantity only at low magnification. To begin with, at the atomic scale the thermal fluctuations of atoms need to be averaged over time to estimate the contact area [318]. Even with averaging, the distance between atoms at which contact “occurs” is not straightforward to calculate. Researchers have used various methods in atomistic simulations using idealized materials and introducing, for instance, potential energy- or distance-based cutoffs for specific crystal or amorphous material structures [396], but the situation is far from clear when real materials with multiple elements or alloys, inhomogeneities, impurities, and so on, are considered. Even in the ideal case where a Lennard-Jones-type potential can be used to define repulsion and adhesion between two particles (or atoms) [398], contact and friction are actually described to occur when the contact distance is nonzero.

Mapping roughness parameters from continuum models to discrete atomic systems is also challenging. For instance, given a continuum function of position, one can calculate the mean contact slope used, for example, in Persson’s theory (see review in §2.1.2), but how should one proceed when the surface is discrete? A viable option is to turn the discrete surface into a continuous surface, for instance by means of bi-cubic splines to ensure that the RMS gradient and curvature are finite. Contact behavior in atomistic simulations is known to depend on the specific realizations of the system under study (see, e.g., [399]). Questions then arise as to which extent local differences in atomistic structures might affect the macroscopic picture. They seem to be relevant already at the microscale for percolation problems, while statistical fluctuations seem to be important in cyclic loading (hysteresis). It appears that a roll-off or other robust sampling strategies are required to model representative rough surfaces at the various scales as well as a proper way to map quantities from one scale to another, both for crystalline and amorphous surfaces.
The breakdown of continuum at the atomistic scale can also be observed in other phenomena. When referring to Density Functional Theory (DFT), for example, the work function of transition metals (TM) becomes non-scalable when particle clusters decrease in size, and the continuum model by Smalley [400] breaks down. The transition between the scalable and non-scalable regimes is at around 100 atoms in the case of gold. An anti-correlation is found between the binding energy and the vertical detachment energy, which may have important implications in relation to catalysis: e.g., while bulk gold is inert, small gold clusters are reactive [401]. The question that arises is whether rough metal surfaces are more reactive than atomically smooth surfaces and, also, whether amorphous surfaces are more reactive than crystalline surfaces, given that they contain more imperfections. To tackle these questions there is a need for accurate tight-binding and/or empirical models at the atomistic scale.

In the case of fluid lubricants, the breakdown of continuum is related to an increase in viscosity and a transition towards a solid-like state, accompanied by stick-slip behavior. The increased viscosity is non-scalable: when the lubricant film thickness decreases down to a few nanometers, i.e. the size of the lubricant molecules, there is a deviation from typical bulk behavior as was observed in Surface Force Apparatus (SFA) studies [402-404]; this transition from ultra-thin lubrication to dry friction under high pressure and shear has been studied using MD [405]. The presence of nanoscale roughness frustrates the ordering of the fluid molecules, leading to high friction states. Experimentally measured viscosities were reported, for example, for perfluoropolyethylene (PFPE) molecularly thin films deposited on the atomically rough substrates used in hard disk drives [406,407] and used in subsequent analytical models to predict the tribological behavior at the head-disk interface [408]. In the case of SFA-type experiments, analytical expressions for the normal (e.g., Kapitza’s solution [409]) and shear forces acting on a spherical probe sliding on a substrate with a fluid film [410] should only hold up to the point where the film can be viewed as a continuum; however, these are routinely used to extract the complex viscosity from amplitude and phase information of the probe vibrations even in cases when very few lubricant molecules exist at the interface [411]. After all, how many lubricant molecules can be said to constitute a continuum?

Additional scale effects related to material models and plasticity are discussed in the next section.

3.4. Material models and plasticity
Crystal plasticity is the relevant constitutive framework when modeling rough surface contact and whenever the size of contact spots is comparable to the grain size in a polycrystalline material. This, of course, includes single crystals. It may seem surprising that only very few tribology-related applications of crystal plasticity can be found in the literature, apparently limited to the analysis of asperity flattening [412,413] and indentation hardness [166,414,415]. Although plasticity of crystals exhibits strong anisotropy (captured by crystal plasticity), the elasto-plastic normal compliance of a rough crystal surface is expected to only weakly depend on crystal orientation as demonstrated by instrumented spherical indentation and crystal-plasticity simulations, e.g., [416,417]. At the same time, plastic anisotropy manifests itself in complex, orientation-dependent pile-up and sink-in patterns [166,417,418]. The related effects may influence the evolution of real contact area in rough contacts, but seem not to have been studied yet.

Nano-indentation tests have revealed another important effect, namely the increase of hardness with decreasing indentation depth, which is referred to as the indentation size effect [419,420]. Several gradient crystal plasticity models have been developed with the aim to describe the related size effects, e.g., [421-424], accompanied by much more scarce three-dimensional crystal-plasticity simulations of the indentation size effect [425,426]. The related effects may also impact the elasto-plastic contact of
rough surfaces. This has been illustrated using a conventional strain gradient plasticity model [427], but the corresponding gradient crystal plasticity studies have not been reported so far.

An important cause of the indentation size-effect in metals is that the dislocations, which are the carriers of plastic deformation, are discrete. Continuum models, including crystal plasticity, are based on the assumption that plasticity can always occur at any location, as long as a critical strength is exceeded; however, in reality, dislocation availability is limited at the small scale. Upon contact, even a very high local pressure might not induce sufficient dislocation nucleation to sustain plastic deformation. Thus, continuum plasticity models for contact and friction are expected to break down at the (sub)micron scale, since they miss a length scale capable of capturing size-dependence. Neglecting the size-dependence of plasticity would lead to the prediction of an earlier onset of plasticity and underestimate the amount of work hardening during plastic deformation. This would have consequences in the estimation of the evolution of the contact area. Size-dependent plasticity can, however, be captured by DDD simulations (see §2.3) [200,428], which can be coupled to MD simulations to accurately capture the nucleation of dislocation loops [324].

Contact between bodies with simple geometries has been studied using two-dimensional dislocation dynamics, where edge dislocations glide on three sets of slip systems, e.g., [323]. Contact results in highly fragmented contact areas due to the exit of dislocations from free surfaces. This leads to a serrated contact area and a peaky contact pressure profile, with high localized pressure, very different from what a continuum model would predict. A comparison between contact pressure profiles obtained using dislocation dynamics and crystal plasticity is presented in [429]. Komvopoulos et al. [430] used two-dimensional DDD to model the indentation of a flat crystal by means of a rigid rough surface with multiscale roughness. Surface asperities were treated as a collection of Hertzian contacts and dislocations could glide only on a single crystallographic slip system. An interesting outcome of this study is that, as the load increases, asperity interactions emerge at different length scales, and so does the interaction between plastic zones. The onset of static friction for a flat contact was presented by Deshpande et al. [431], whose work points to the competition between plastic deformation – dominant for larger contact areas– and loss of adhesion –dominant when the contact is so small that plasticity is limited. There is much room for additional friction studies in the framework of discrete dislocation plasticity.

A way to incorporate microscale size-dependent plasticity into contact models could be to fit the dislocation dynamics results for the deformation of a non-local plasticity theory, such as strain gradient plasticity or even include such effects in a statistical model. The advantage of statistical models, like the one recently developed by Song et al. [432], is their extremely low computational cost, which would make them attractive for use by the industry. However, a statistical approach based on the GW model, for example, would suffer from the same limiting assumptions discussed earlier (see §2.1.2) and may not be directly applicable to realistic representations of roughness (see §3.2).

Plasticity is not only limited to dislocations, as it can also appear in the form of grain boundary sliding [433-435] when high strain rates are involved. In this case, even the material crystallographic structure can change. During dry sliding, grain coarsening [265] as well as grain refinement and amorphization have been observed [436]. As an example, Stoyanov et al. [301] show that tungsten carbide (WC) in a frictional contact with tungsten (W) causes the crystalline WC structure to turn into amorphous WC with a dispersion of nano-diamonds. Some interfacial phenomena in metal sliding are related to near-surface austenization induced by frictional heat and subsequent formation of fine-grained martensite known as a white layer [437-439].
3.5. Normal contact between rough surfaces: the contact-mechanics challenge

One of the few tribological problems that are relatively well understood (at least in theory) is normal contact between rough surfaces. A comparison of various modeling approaches in their ability to properly solve a well-defined normal contact problem has been tackled in the recent contact-mechanics challenge [113]. A surface height spectrum was generated [367] featuring a roll-off and power-law decay region, as was a realization of this randomly rough surface in real space. The following approximations were made: small surface slopes, linear elasticity, short-range adhesion (based on the value of the local Tabor parameter $\mu_T = 3$, which was close to the JKR limit; see §3.7), periodic boundary conditions, and a hard-wall contact constraint. The problem setup results in insignificant adhesive hysteresis up to moderate contact pressures. This information was made available to researchers who were asked to compute any well-defined property ranging from spatially resolved information, via distribution functions of stress and gaps, to compound properties like contact area as a function of load. Specific metrics used in the subsequent analysis included the gap and stress along a reference line; stress and contact patch histograms; and relative contact area and mean gap values. Submitted solution methods could be categorized into brute-force computing, where errors could come from the discretization, and models mapping onto simpler equations using uncontrolled approximations. More specifically, results utilized exact (boundary-value) methods, Persson theory without adhesion, multi-asperity models that assume local constitutive relations without interaction between contact patches ("bearing models"), as well as all-atom MD simulations, where the surface size was scaled down by a factor of 100, and experiments, where the surface size was scaled up by a factor of 1000. The reference solution was calculated using GFMD (see §2 for a review of computational methods and models).

Good agreement with the reference solution was found for both experiments and all-atom MD; when comparing the gap across the reference line, the effect of removing the small-slope approximation gave excellent agreement for all-atom MD. Expectedly, multi-asperity models were found to overestimate the gap, while exact methods agreed almost exactly at the greatest magnification; however, results of the stress across the reference line (local zoom-in) showed great scatter. Stress distribution histograms were almost Gaussian at compressive contacts, featuring a high adhesive peak at zero pressures and a rapid decay to tensile tractions. Multi-asperity models were found to overestimate the stress while, in the presence of adhesion, when small patches become unlikely, these models produced very similar trends for the patch-size distribution. Most solutions showed reasonable agreement for the contact area as a function of load, as well as for the mean gap as a function of load. The exception is models based on the geometrical overlap of rough surfaces whose results strongly deviate from more accurate models (see [113] for details).

In summary, very close agreement was observed between all systematic approaches with differences becoming visible when quantities required high resolution. At the same time, these approaches showed good agreement with experiments and all-atom MD, suggesting that common approximations might be less problematic than believed. Reasonable agreement was found between the reference solution and the non-adhesive Persson’s theory on all reported properties, while multi-asperity methods agreed with each other but deviated from the reference solution (it is worth mentioning that more recent asperity models accounting, for example, for asperity interaction were not compared in this study). Adhesive Persson’s theory is compared with the contact-challenge’s results in [440]. It could therefore be argued that the suitability of modeling methods and tools can be determined based on the properties one would need to extract: for example, predicting contact area versus load or mean gap versus load seems to be consistent across methods and, arguably, the most suitable model would be the simplest.
one. On the other hand, extracting local quantities at higher resolution would require numerical methods able to achieve sufficient discretization.

As soon as the contact is not only compressed but is also sheared, the real contact area of real (frictional) rough interfaces has been measured to evolve significantly. In particular, recent experiments [397] showed that for rough elastomers in contact with smooth rigid substrates, the area of real contact significantly drops during incipient shearing, well before macroscopic sliding. Similar behavior has been suggested, from indirect measurements, to happen in sheared rough-on-rough rock contacts [441]. We believe that such frictional contacts under shear have not yet received sufficient attention in the modeling literature. It should be noted that this elastic behavior is different from the shape change observed in viscoelastic contact in sliding [159].

3.6. Friction

Although normal contact between rough surfaces can serve as a reference situation in many tribological systems, it is not a priori sufficient to address issues related to moving surfaces. Lateral motion does involve fundamentally new phenomena, related, for example, to those occurring at continuum scales: frictional heating, wear, and third body or shear-rate-induced dissipation (through fluid lubrication or bulk viscoelasticity). Those effects need to be understood in order to assess the origin of macroscopic friction and quantify it in various tribological systems. The breadth of the field of friction is too large to attempt an extensive summary here. Instead, the reader is referred to reference books for an overview of the field, e.g., [46,442]. While the question of the microscopic origin of the friction force has already been addressed elsewhere, e.g., [176,443-445], in this section we will take for granted that, at the micrometer scale and above, i.e. at continuum scales involving a large number of atoms or molecules, a friction force exists. Furthermore, we will only address a few recent advances made in the understanding of macroscopic friction, from its onset and transition from static to kinetic values, to rubber friction in relation to viscous bulk dissipation, through to patterned surfaces.

3.6.1. Friction laws

As soon as any motion occurs at a macroscopic interface, a transition from stick to slip takes place, either sharp [446,447] or smooth [448,449], and models need to incorporate a friction law accordingly. The most classical and widely known friction law is the one of Amontons-Coulomb (AC) [450], which states that no sliding occurs as long as the ratio of the shear force \( F_s \) to the normal load \( F_n \) remains below a certain threshold defined as the static friction coefficient, \( \mu_s \). Maintaining a constant sliding speed requires the application of a kinetic friction force, \( F_k = \mu_k F_n \), with \( \mu_k \) being the kinetic friction coefficient which usually cannot exceed \( \mu_s \), with some exceptions reported in frictional interfaces which are strengthening at high sliding velocities [451]. Note that, in the current acceptation of the AC friction law, the friction coefficients are considered constants for given materials in contact. Coulomb actually had already found that \( \mu_s \) increases logarithmically with the contact time, and \( \mu_k \) depends logarithmically on the sliding velocity [450,452]. Today, laws incorporating those dependencies are denoted as rate-and-state friction laws, as further described below. The AC law, which has been defined here from the global forces acting on the interface, is commonly used locally along extended interfaces. In those cases, the friction coefficients are to be compared to the local ratio of shear to normal stress \( q(x)/p(x) \), where \( x \) is the coordinate along the interface. Practically, a fundamental question arises about the value to be used for the local friction coefficients: should one use the values of the corresponding global coefficients or should these be different at the local contacts? As demonstrated, e.g., in [176,453], assuming a simple AC friction at the interface may result in a velocity-dependent macroscopic friction coefficient.
Whereas the global and local kinetic friction coefficients are expected to be equal (in the quasi-static case), the situation is very different for static friction coefficients. It has been shown experimentally that the static friction coefficient depends on the stress distribution at the interface prior to the onset of sliding [454], and that $\mu(x)/p(x)$ can exceed the macroscopic friction coefficient by a factor of two [455]; these results have been reproduced in models of heterogeneous frictional interfaces [456,457]. The fundamental reason behind this behavior is that the global and local static friction coefficients are equal only if all points of the interface reach their slipping threshold at the very same instant. This situation corresponds, for instance, to an ideally homogeneous interface submitted to homogeneous loading. In practice, this never happens: when slip at the interface becomes unstable, a large portion of the interface is loaded below its threshold, so that the total tangential load born by the interface is smaller than its theoretical maximum value. The consequence is that, in general, the global static friction coefficient is smaller than its local counterpart [447,458], and it is thus challenging to infer a local static friction coefficient from macroscopic measurements.

Although practically useful and rather easy to implement in models, AC’s friction law, in its limited current acceptation (see two paragraphs above), cannot capture a series of effects repeatedly observed in rough contacts (see, e.g., [459] or [460] for reviews). First, the static friction coefficient, $\mu_s$, slowly increases with the time the interface spends at rest. This effect is interpreted as an increase of the area of real contact over time through asperity creep, an effect denoted as geometrical aging. Depending on the material, creep can be of viscoplastic [339] or viscoelastic in nature [461]. Another cause for the increase of $\mu_s$ is the strengthening of the contact with time, presumably due to relaxation of the glass-like material forming the very interface [462], or to the formation of chemical bonds [463], an effect denoted as structural aging [459]. Secondly, the kinetic friction coefficient in steady sliding is velocity-dependent, typically with a logarithmic velocity-weakening. This effect is partly due to an intrinsic velocity-dependence of the interface’s shear strength, and partly to the time-dependence of the real area of contact: slower sliding gives more time to the micro-contacts to grow in size before they break and are replaced by fresh, smaller micro-contacts. Those effects are taken into account in rate-and-state friction laws, and apply to various fields related to friction, in particular earthquake and landslide science.

Despite its many successes, the rate-and-state friction law must also be used with caution. The logarithmic velocity-weakening is based on observations at low slip-velocity, smaller than about $100 \mu m/s$. At higher slip rates, a velocity strengthening regime due to viscous effects is also expected, and is indeed generally observed beyond some crossover velocity [451]. Note that, in AFM experiments, velocity-strengthening can also be observed due to thermally activated breaking of nanoscale junctions (see, e.g., [464]). At even higher velocities, in the range typical to unstable slip up to a few $m/s$, sliding is accompanied by significant temperature rise, possibly by several hundred degrees. Such heating can induce transient phase changes in the vicinity of the contact interface [455]. In these conditions, friction may not be controlled only by a critical length scale (the average micro-contact size) but also by time scales [447,455]. Heat can also favor chemical reactions, in particular in tectonic faults with fluids and high pressure. Such reactions tend to self-lubricate the interface, with low friction resistance at the highest slipping rates [465]. Such systems remain challenging to model, due to the strong multiphysics coupling required to capture the most salient controlling phenomena.

3.6.2. The relevance of space and time scales on the onset of sliding

Apart from identifying and understanding new and specific mechanisms occurring at or close to the contact interface, tribological models can be used as quantitative tools to reproduce and interpret experimental observations: this is especially true for friction. Since most contact and friction measurements are made at the system-size level (e.g., total normal and friction forces), models
predicting system-size quantities could be denoted as “macroscale models”, irrespective of the actual length scale considered. As a provocative example, a model of atomic force microscopy experiments is a macroscale model if its aim is to predict the total friction force that the tip experiences. But what are the properties of models actually enabling such quantitative comparisons?

A frictional interface can be modelled using a homogeneously loaded contact between elastic half-spaces only in very specific instances; instead, most real contacts have complex geometries, boundary conditions, and loading configurations leading to unavoidable pressure and shear stress heterogeneities along the contact interface. Since friction laws need to couple both normal and shear stresses to predict where and when slip will occur, the stress distribution along the interface needs to be accurately modelled. Although a large portion of friction-related works deals with static or quasi-static situations, most realistic contacts also experience transient phenomena: either the loading is unsteady (oscillating contacts, impacts) or the interfacial response is itself transient (instabilities). This is why, in order to offer improved quantitative predictions of the tribological behavior of an interface, macroscale models need to account for the elasto-dynamics of the bodies in contact: the incorporation of temporal phenomena, together with realistic boundary conditions, into frictional models is essential.

As a practical example, one can consider how macroscale models were progressively improved to reproduce some aspects of the experimental results reported by the group of Fineberg about the onset of sliding of extended interfaces [340,446,454,455,466-469]. Their main observation is that the transition from static to kinetic friction is mediated by the dynamic propagation of micro-slip fronts along the interface: ahead of the front, the interface is still in its stuck state, while it is already slipping behind it. Macroscopic sliding only occurs when the front has spanned the whole interface [446]. In this context, not all fronts lead to macroscopic sliding. Precursors to sliding are sometimes observed, which correspond to fronts spanning only a fraction of the contact interface. These precursors manifest themselves at macroscale as a series of dents in the loading curve, indicating partial load relaxation [466]. Note that it is still an open question whether those observations of slip fronts, which have been made on polymers, might also be made on other materials like metals. Part of the answer may be related to the concept of elastic coherence length [46,470], i.e. the length scale below which a contact interface can be considered as rigid. In the case of metals, the elastic coherence length is expected to be excessively large [46], which may prevent the observation of front propagation, at least on decimetric samples such as those used for polymers.

The first models for the length of precursors were one-dimensional [456,471-475]. Although the ad-hoc introduction of an initial shear stress field was improving the results [473], none of these models could be compared quantitatively with Fineberg’s experiments, in which the height of the slider was not negligible. Only with two-dimensional models based on spring-block or FEM representations of the elasto-dynamics of the slider [457,476-478] could the predictions quantitatively match the observations. While the aforementioned models were based on the AC description of the frictional interactions at the interface with static and kinetic friction coefficients, a recent fracture-based description appears to provide equally good predictions of the precursor length [340,341], strengthening the idea of an equivalence between the friction and fracture descriptions of the onset of sliding, often used in earthquake science [479]. In particular, the fracture-like stress field around the tip of micro-slip fronts, measured through an array of miniature strain gauges was captured by analytical [468] and FEM models [469].

Although a velocity-independent AC friction law is sufficient to predict the precursor length and the fact that front speed depends on the local pressure to shear stress ratio [467], such a law fails to explain the unexpectedly large range of front speeds observed [476,480]. While the fastest fronts,
propagating at about the speed of sound in the contacting materials, were expected from standard shear fracture theory, abnormally slow fronts—slower by orders of magnitude—were observed but remained unexplained, while a single front could alternate between both types in a single event [446]. It should be noted that slow fronts here are distinct from quasi-static fronts like those involved in the onset of sliding of sphere-on-plane contacts, the propagation speed of which is proportional to the external driving velocity [43,448,481]. Dynamic slow fronts have been obtained theoretically within a one-dimensional model of the interface using an improved rate-and-state friction law featuring a velocity-weakening-then-strengthening behavior. In this model, the slow front speed is related to the velocity at which the steady-state friction coefficient is minimum [482,483], which is supported by observations of slow rock friction [484].

Unfortunately, such an approach does not explain the possible transition from fast to slow front regimes observed within a single event; this was achieved using a multi-scale model [447,485] in which a 2D model [476] is complemented by a micro-junction based description of the interface [486] in which the loading/breaking/reformation cycle of each junction is controlled by a time scale. This time scale is inspired by the one identified experimentally in [455], which was observed to control the transition from fast slip to slow slip when the interface starts to slide, and was argued to correspond to the cooling time of the interface after the rapid heat deposition as the micro-junctions break upon front passage. Such heating is presumably responsible for local melting of the interface, a phenomenon which is also clearly involved in seismology where sliding rocks melt and reform leaving fault veins.

The main implication of this time scale is that, in the model, after a slip phase, the interface does not re-stick perfectly, but transiently allows for some further, slow slipping. Thus, slow fronts are fronts that would arrest in the absence of this slow slip mechanism, but can continue to propagate, much more slowly, due to the slow slipping occurring in the broken part of the interface. It was also found that the selection of the front type (fast or slow) is not only dependent on the shear to normal stress ratio, but also on the local disorder in shear forces sustained by the micro-junctions [447]. As a result, local static friction is history-dependent, with potentially a factor of two in the variation of the coefficient of static friction due to the rupture history of the interface [486]. All these results suggest that friction features multiscale aspects both in the spatial and time domains that must be considered in models.

3.6.3. Rubber friction: Some open issues from mesoscale experiments on elastomers

Rubber friction has received much attention in the literature, both because of its practical relevance, for instance to tire/road contact, and because of the particular way energy is dissipated through friction. The seminal work of Grosch [487] has shown that the temperature- and sliding velocity- (or equivalently, frequency)-dependence of the friction coefficient closely follows that of the viscoelastic moduli of the rubber. His results suggest that both the surface and bulk dissipation during rubber friction are of viscoelastic origin. As for the bulk, each spatial frequency present in the surface roughness is expected, through the sliding velocity, to correspond to a temporal frequency for the excitation of the viscoelastic material. Persson’s 2001 multiscale theory of contact [8] was aimed at clarifying the relationship between the continuum of frequencies within the roughness and the dissipation caused by them. For a review of this issue, the reader is referred to the following review paper [488]. In the rest of the section, the focus is mainly on the recent use of elastomers to gain insights into specific frictional phenomena.

It has already been argued that new insights into friction can be reached by comparing model predictions to experimental measurements made not only at the system-sized scale (macroscopic loads) but also at local scales (ideally full field evaluations). In several aspects, elastomers are good model materials with which comparisons can be performed. Due to their low elastic modulus, the
amplitude of the interfacial displacements under tribological solicitations is typically large enough to be routinely monitored optically, using contact imaging techniques (see e.g., [489,490] for tire rubber). In particular, polydimethylsiloxane (PDMS) is increasingly used for in situ measurements of displacement fields (see e.g., [448,481,491-493]). PDMS has the further advantages to have a low loss modulus, and to fracture at extremely high strains, well beyond those associated with frictional solicitations. Thus, its behavior can be compared to elastic models, sometimes incorporating nonlinear elasticity at high strains [342].

Access to local displacement and stress at such rubber interfaces enabled the identification of some phenomena that are not yet satisfactorily incorporated into friction models. As a first example, rough interfaces have finite normal and shear stiffness (compared to the infinite stiffness of a complete contact between smooth bodies) due to the compliance of each individual micro-contact forming the multi-contact interface. Although those stiffness values affect the behavior of contact interfaces (see e.g., [494] for the role of the normal stiffness and [448] for that of tangential stiffness on rough sphere-on-plane contacts), most models consider, for the sake of simplicity, perfectly smooth interfaces. Such models could be improved by including the effect of roughness through effective boundary conditions on smooth interfaces (as done, for example, in [495,496]). As a second example, the contact mechanics and frictional properties of elastomer contacts are found to be affected by the value of a pre-stretching applied to the rubber (see, e.g., [497,498]), due to a stretching-induced anisotropy of the interface. Keeping in mind that any contact loading leads to a non-vanishing field of in-plane tensile strain, in particular near the contact edges, stretching effects are expected to be involved in virtually all tribological situations. Improved friction models should aim at incorporating those effects.

3.6.4. Dry friction between patterned surfaces

In many practical applications, the emergent frictional behavior is not only determined by microscopic degrees of freedom or surface roughness, but also by other mesoscopic or macroscopic length scales characterizing the material surfaces. The hierarchical structure of the gecko paw is one of the most cited examples to illustrate the role of a complex contact structure, and many research efforts have been devoted to understanding the origin of its properties of adhesion and friction [13,499-502] (biotribology is further discussed in §3.9.3-5). In general, many biological materials are characterized by a non-uniform complex surface structure, e.g., insect legs [503], lotus leaves [504,505], nacre [506], as well as animal [507-509] and human skin [510-512], whose hierarchical scheme of contact splitting has been described as a way to optimize surface adaptability, self-cleaning abilities, and to avoid self-bunching [500], and are therefore difficult to model in a single framework. The exceptional mechanical properties of these systems have attracted a lot of interest, and led to attempts to reproduce their behaviors artificially with specific geometric features of the surfaces. The main focus of research in bio-inspired materials is to design new materials by mimicking nature, aiming to manipulate the mechanical properties of a system through a complex organization of microscopic components rather than introducing new chemical and physical features [31,513-517]. Understanding and optimizing friction in these bio-inspired complex surfaces is an open challenge.

Recently, experimental results have been obtained for the friction of specific textured surfaces, e.g., honeycomb structures [518,519], periodic regular grooves both in dry and wet conditions [38,520-523], as well as pillars and dimples [524-527]. MD simulations (see §2.5) have been adopted to investigate the effect of patterning in the presence of lubricants [528], but the theoretical and numerical modeling of dry friction in these systems shares the difficulties inherent to that of the friction of rough surfaces: how to take into account within a unified framework concurrent length scales spanning orders of magnitude and involving many physical mechanisms. For this reason, much work remains to be done on this topic. Some results have been obtained by means of a simplified
approach based on numerical simulations of the spring-block model [529], aiming to investigate the qualitative frictional behavior of patterned surfaces [530-533]. In order to study the role of specific surface structures, it is not necessary to include into a model the details of all microscopic interactions, since they can be taken into account with an effective description at the mesoscale, where the system is discretized into elementary components whose interactions are described in terms of forces within the framework of classical mechanics. Thus, surface structures are introduced by means of the arrangement of elementary components, and the effects on the macroscopic friction coefficient are deduced from the numerical solution of the overall equations of motion of the system. With this procedure, some versions of the spring-block model have been successfully used to model and understand the existence of slow detachment fronts [447,471,480,485], crack-like precursors of sliding [472-474,476], and stick-slip sliding [534,535], consistent with experimental observations [446,455,466-468].

Thus, despite the approximations and apparent simplicity of the model, the spring-block approach can provide a qualitative understanding of relevant phenomena with computationally inexpensive numerical simulations. The results of these studies show how static friction can be tuned and optimized by means of a specific arrangement of surface structures. In particular, it has been demonstrated that the static friction coefficient is reduced by means of large surface grooves [530] and that a hierarchical organization of grooves with different length scales can be used to tune it to a desired value [531]. Also, it has been proved that a remarkable reduction of the global static friction of a surface can be obtained by means of a hierarchical organization of regions with different local static friction coefficients [532]. Recently, a two-dimensional version of the spring-block model has been adopted to simulate the effect of surface patterns like pillars or cavities [533]. A natural development based on this research is to improve the spring-block model by relaxing some of its approximations, for example, by simulating more realistic three-dimensional surfaces; furthermore, variations of the surface roughness after the onset of sliding or other long-term effects during the dynamic phase can be incorporated.

3.7. Adhesion

Research on adhesion in the field of contact mechanics saw significant progress only in the 1970s. Any review of the literature on adhesive contacts will start with the two analytical models developed in this period, the JKR model [51] and the DMT model [52]. These models considered adhesive contact between a smooth sphere and a flat body, but with different approaches and making significantly different assumptions. They were shown to apply equally well to different contact conditions by Tabor [536] who identified a characteristic parameter, now known as the Tabor parameter, which can be systematically used to identify whether short-range or long-range adhesion dominates the contact interactions; in particular, the JKR model captures mainly short-range interactions, representative only for contacts with a large value for the Tabor parameter (>2, soft solids, small curvature, large adhesion), while the DMT model is valid for contacts with a small value (<0.01, rigid solids, large curvature, weak adhesion) [49]. Muller et al. [537] attempted to bridge the two models by removing the assumption that the Hertz profile is not affected by adhesion and developing a self-consistent analysis of adhesive contact between a sphere and a flat. Similar analyses to a higher level of accuracy were later performed by Greenwood [538]. It should be noted that, while the variation of the contact area and approach with load are well predicted for large values of the Tabor parameter, the hysteretic energy loss during a contact separation cycle can be significantly overestimated by the JKR theory [539] and a corrective theory is needed to model the precise event of “jump-in” and therefore the hysteresis loss. However, advances have been made in this respect thanks to the development of accurate numerical simulations, as discussed below.
Whilst later analyses by Muller et al. [537] and Greenwood [540] seem to provide the solution to contact mechanics of smooth adhesive contacts, their complexity and numerical basis hindered exploitation until more recently, when alternative models were developed. Maugis applied a Dugdale-type analysis (from fracture mechanics to contact mechanics) to the problem [541], replacing the true adhesive forces with a constant adhesive force acting between the surfaces at all points separated less than a critical distance. Greenwood and Johnson used a “double-Hertz” analysis to similarly simplify the solution and provide results suitable for analytical manipulation [540]. These methods, while offering a significant step forward in analytical capabilities, do not provide the same accuracy as the Muller and Greenwood analyses, which are therefore usually selected as the starting point for the development of newer deterministic formulations. More recently, finite element models for adhesive contact problems have also been developed, where the contact description obtained using the Lennard-Jones potential is incorporated into the framework of nonlinear continuum mechanics, e.g. [542] and [543], also in the presence of plasticity [544] and within the context of multi-scale simulations, e.g., [277,545]. Alternative approaches have also been developed based on the BEM, which incorporates adhesion through energy minimization; see, e.g., the work presented in [546,547]. Attempts have also been made to study the effect of indenter geometry on the macroscopic shape of the contacting region; for example, Popov et al. [548] have provided numerical and experimental results for contacts of rigid punches characterized by different shapes and contacting a soft, adhesive counterpart, showing that, in some cases, pull-off may not be instantaneous and detachment fronts can propagate from sharp corners and travel inwards, until the final configuration (circular for regular geometries) is reached.

Most of the models discussed above were developed for or applied to smooth surface contact, nominally between a sphere and a flat. A common justification for neglecting adhesive forces is the existence of surface roughness and, starting from this point, an early and significant analysis was carried out by Fuller and Tabor [549], who showed that the adhesive influence could be described by an “adhesion parameter,” which is, in effect, a ratio of the adhesive force of “lower” asperities to the elastic push of “higher” asperities. The theory was found to show reasonable agreement when fitted to experimental results. Fuller and Tabor had used the JKR model on an asperity level; Maugis repeated the analysis using the DMT model and found that an additional load would be caused by adhesive forces around each asperity [550]. Further advancements were made through the inclusion of an elastic–plastic representation of the asperities based on the DMT model, e.g., [551]. Other attempts have been recently made to incorporate the effect of thin films [552], and to extend the validity of the maps proposed by Johnson and Greenwood [553] to account for the strength limit [554].

Adhesion for rough surfaces is obviously an extremely rich problem. Simple theories, such as those proposed by Rumpf [555] and Rabinowicz [556] have been demonstrated to work well when studying nanoscales effects for hard solids for the spherical geometry. These show large reduction with RMS amplitude of roughness and a limited dependence on slopes or curvatures, as confirmed by extensive experimental measurements performed by Jacobs et al. [557]. When looking at the influence of roughness on the system response from atomic corrugation up to a few nanometers, the latter showed that the measured work of adhesion decreases by more than an order of magnitude when the RMS increases. Successful attempts have also been recently made to estimate the effect of adhesion between elastic (hard) rough solids with Gaussian multiple scales of roughness [558] and to study the effect of adhesion for sinusoidal contacts, e.g., [559].

Looking at other theoretical and non-deterministic models of multi-asperity contacts, in some of the early contributions, Persson and Tosatti considered adhesion through a fractal representation of surface roughness and showed that adhesion dropped significantly at higher fractal dimensions [560]. They suggested that the simpler analysis of Fuller and Tabor and their adhesion parameter adequately
described the full detachment stage of a particle. More recently, Persson and Scaraggi [561] used Persson’s theory and a power spectrum representation of the contact roughness to introduce a Tabor number that depends on the length scale or magnification, and which gives information about the nature of the adhesion at different length scales. They proposed the analytical study of the two limiting cases (JKR –see also Persson [562]– and DMT) for randomly rough surfaces using the Persson contact mechanics theory (see §2.1.2); it was shown that adhesion problems that are “JKR-like” for large length scales and “DMT-like” for short length scales can be approximately treated using the theory with different levels of approximations, which depend on how quickly the behavior transitions between the two limits across the scales. While these rough surface models (or asperity models) are limited to a stochastic description of the surfaces and thus cannot provide a complete contact mechanics solution for all surfaces, they may constitute a good approximation and provide a useful design tool, especially when numerical simulations may struggle or fail to produce fast and reliable results. Extensions to include hysteretic effects would be a very useful addition to the literature.

Deterministic adhesion models of contact in the presence of roughness are expected to provide an accurate representation of the response of real bodies in contact. MD simulations of contacts (see §2.5) can potentially provide an extremely accurate deterministic description of adhesive forces in a contact (see, e.g., [277,563,564]); however, the limitation in terms of the number of atoms and system sizes that can be included in MD simulations (at least when classical approaches are used and the simulated degrees-of-freedom are atoms rather than coarse-grained entities, as is the case, for example, in GFMD; see below) reduces the applicability of this method to large-scale contacts. Given the advent of new and improved numerical methodologies and increased computational power, there has been a recent resurgence in the development of contact mechanics models able to address contact between surfaces of arbitrary shape and roughness, of small and large scale, and capable of providing accurate information for contact forces, surface displacements and hysteretic effects (where present) throughout the contact. Many of these methodologies can be seen as BE methods (discussed in §2.2) relying on different discretizations and numerical techniques to solve the contact problem using “brute force” [113], and include GFMD [138,565], FFT-based (e.g., [566,567]), and Multi-Level Multi-Integration (MLMI)-based techniques [568]. These methods have been shown to capture the response of rough contact surfaces in the presence of adhesion in a number of configurations and can be used successfully to predict the scales and regimes at which roughness will play a significant role in adhesive contacts, as well as computing hysteretic losses. These models can also be applied all the way down to the nanoscale as long as the surface interactions are well captured and can be approximated using simple Lennard-Jones potential interactions [568]. Other examples of implementation within the BEM framework include the incorporation of JKR adhesion, e.g., [569,570]. Recently, Rey et al. [571] suggested an alternative approach to the adhesive BEM, which is based on the minimization of the total energy.

An open question is whether or not adhesion depends on the topography’s RMS amplitude: while asperity theories predicted a strong influence of RMS amplitude, Pastewka and Robbins [565] formulated a criterion for “stickiness” by numerical observation of the slope of the (repulsive) area-load, which appears to be independent of the RMS amplitude. This is not necessarily a contradiction as the proposed criterion studies only the existence of instabilities at the small scale, while it is the magnitude of such instabilities that is heavily influenced by long-wavelength undulations. Discussion about this issue is currently still active [377,378,572]. Future perspectives also include the need for detailed investigations of the interplay between adhesion and shear stresses/friction (see e.g. recent contributions on this topic [397,573]) and the integration of realistic adhesive interactions, which describe the surface behavior accounting for chemical interactions and bonding energies that go
beyond van der Waals forces, into multiscale roughness simulations via MD-continuum coupling strategies, which in principle allow for chemo-mechanical interactions to be more accurately captured.

3.8. Lubrication and viscoelasticity
Everyday experience shows that interposing a fluid between two contacting bodies dramatically drops the friction force. Lubrication has, then, a paramount importance in engineering and applied science research since it is clearly related to an improved energy efficiency, a better durability of components and systems, and, ultimately, to economic savings. In this section, only full film lubrication regimes are addressed. The effect of lubricant additives under boundary lubrication conditions are discussed in §3.9.2.

Theoretical investigations take their origin in the pioneering studies made by Reynolds in the 19th century [208]: Reynolds equations enable the analysis, in terms of velocity and pressure distribution, of a flow in the lubrication channel. In the last fifty years, a lot of approaches, mainly numerical [214], have been developed to address the solution of this set of equations: nowadays, it is even possible to account for a variety of non-Newtonian effects, ranging from piezo-viscosity to shear thinning. For a more comprehensive overview the reader is also referred to Hamrock’s classical book [216], while modeling approaches are discussed in §2.4.

In recent years, textured surfaces for the optimization of hydrodynamically lubricated contacts have been developed (see, e.g., [574], also inspired by nature [575]). The main effect of the presence of dimples, pockets or asperities is an increase in the load-carrying capacity of the bearing and eventually a reduction in the coefficient of friction. The main challenge in modeling the hydrodynamic lubrication between textured surfaces remains the description of the cavitation, for which many models have been proposed (e.g., finite difference algorithms [576,577] based on the well-accepted JFO boundary conditions [578,579]. In addition, multiphase CFD simulations have been used to model cavitation but, given the complexity of the problem and the coupling with the appropriate turbulence models, it is still a challenging task [580]. Multiscale approaches should be developed in order to capture both the macroscopic tribological characteristics of a lubricated contact and the micro-hydrodynamics, with the related phenomena of roughness-induced cavitation and turbulence.

Furthermore, in order to completely assess the problem, the solution of the lubricant fluid dynamics has to be coupled with the analysis of the contacting solids’ mechanics: in the so-called EHL regime (also see §2.4), the fluid pressure is high enough to entail an elastic deformation of the lubricated bodies. Consequently, the pressure field has to satisfy, at the same time, the Reynolds equations and the elasticity constitutive relations. The intricacy of the problem surges when the roughness of the contacting solids is accounted for. Indeed, the mathematical form of the problem does not change, but the number of elements required to find a numerical solution and, in particular, to explicitly resolve the effects of rough contact cannot be handled with the computational resources currently available. Consequently, a deterministic approach which accounts for the contact interactions at all relevant roughness scales is unfeasible; instead, various homogenization methods have been developed to overcome these limitations. The most commonly used approach solves the Reynolds equation as if the surfaces were smooth and uses “flow factors” as statistically corrective terms for the surface roughness [581]. This approach was pioneered by Patir and Cheng in [582], and then further developed by Elrod [583] and Tripp [584] to account for anisotropic effects. Furthermore, recent investigations have shown that more accurate estimations may be performed by employing, instead of scalar coefficients, flow factor tensors, which are functions of the surface roughness and, specifically, of the anisotropy roughness tensor [585].
When contact or environmental conditions do not permit fluid film lubrication, e.g., when extreme temperatures and/or pressures are present, as in aerospace applications [586], solid lubricants are generally employed. It should be noted that, in the literature, a distinction is made between powder and granular lubricants, on the basis of the particle characteristics and the load-carrying capacity generation mechanisms [587]. Many analytical models of solid lubrication have been developed over the years, starting from analogies with fluid mechanics and the conservation laws for mass, momentum and energy [588,589]. The kinetic theory of gases, instead, has been the basis for the development of the granular kinetic lubrication theory [590,591]. Both continuum and discrete models are available for the description of solid lubrication or, more in general, of third body friction [592]. Continuum modeling approaches are based on rheological laws describing the third body, originally introduced by Heshmat [593]. Discrete simulations, instead, allow the precise computation of particle dynamics and taking into account individual particle-particle and particle-wall interactions [594]. Solid lubrication is intrinsically a multiscale and multiphysics problem. Therefore, an effective modeling approach should be able to include the microscopic physical (e.g., surface roughness), chemical (e.g., tribo-corrosion [595]) and thermal interactions, and to link them to the frictional characteristics of the tribo-contact. Hence, discrete approaches and particle-based methods seem more promising, despite necessitating further efforts to make the micro-to-macro correlation. Novel lubricants have been successfully developed, e.g., using additives to improve anti-wear properties, allowing to extend the life of tribo-contacts. Nanolubricants, for instance, display exceptional thermal and tribological properties and are obtained by adding nanometer-sized particles to a base fluid. The development and study of the response that additive molecules and nanoparticles, and the effect they have on friction reduction and boundary lubrication, is usually achieved through detailed modelling at the atomistic scale, as discussed in §2.5. A detailed overview of modelling methods used in this area is provided in [596].

The lubrication problem becomes even more complicated when it involves the wide class of soft materials. Given its practical interest—related to the continuously increasing demand for new polymers [597,598], soft tissues [599], biomedical implants [600], biomimetic solutions [565,601] and smart materials [602]—soft matter lubrication is a field which is currently attracting a variety of research contributions. The main challenge in these investigations is in dealing with the lubricated bodies’ rheology, which is usually not perfectly elastic, and, on the contrary, is marked by nonlinear time-dependent stress-strain constitutive laws. Indeed, hyper-elasticity has been embedded in a number of models (see e.g. [169]) and was shown to be responsible for significant quantitative deviations from the classical EHL theory. However, such a step has not been sufficient to explain a variety of experimental observations involving soft materials. These include, for example, film thickness maps and contact patches whose shapes and values show, depending on the flow speed, a marked shrinkage at the flow outlet, thus looking very different from conventional Hertzian-like contact configurations [603]. Another surprising experimental finding linked to the interplay between solids and fluids in soft contact problems can be found in [604], where it is shown that the rupture of the fluid film occurs at the flow inlet in lubricated interfaces in the presence of strongly viscoelastic solids: this is very hard to explain in the absence of strong time-independent deformations, and is unexpected in classical lubrication. For these reasons, recently, new models for two-dimensional [605] and full three-dimensional interfaces [606] have been developed to account for the viscoelasticity of lubricated solids. Specifically, in the latter case, when considering a viscoelastic rheology, it is possible to appreciate a dramatic deviation from classical EHL theory, both in terms of fluid pressure and film thickness. Indeed, the film thickness has a marked shrinkage at the fluid outlet, so that the absolute minimum of the film thickness can move from the flow outlet to the inlet and the pressure distribution is peaked accordingly. All this has paramount importance when focusing on the friction developed in
tribo-systems involving viscoelastic soft materials. Indeed, the viscoelastic material hysteresis has to be added to the fluid viscous losses, a trend which is far from the classical EHL friction-speed dependence and is consistent with very recent experimental observations [607].

Beyond lubrication, the contact mechanics and tribology of viscoelastic soft matter itself can be studied via the BEM, which is significantly more cost-effective in modeling rough surfaces than FEM (see §2.2). In general, viscoelasticity causes shrinkage of the contact area for increasing speed [159]. For example, the contact behavior of a rigid sphere in reciprocating sliding contact with a viscoelastic half-space ranges from the steady-state viscoelastic solution, with traction forces always opposing the direction of the sliding rigid punch, to a multi-peaked pressure distribution with tangential forces in the direction of the sliding punch. This behavior is controlled by the size of the contact, the frequency and amplitude of the reciprocating motion, and the relaxation time of the viscoelastic body [608].

The development of comprehensive tools is necessary to simultaneously manage surface roughness, the complex rheology of lubricants and contacting bodies, surface effects linked to adhesion or the presence of surface-active molecules, and the geometry of the contacting bodies.

3.9. Other tribological phenomena and applications

3.9.1. Wear

Despite three centuries of scientific investigations on wear mechanisms [609], which led to the emergence of a myriad of empirical models (amongst which the ubiquitous Archard’s wear law), by and large, the dots remain unconnected and our macroscopic engineering-scale understanding of wear remains limited [610]. Wear processes emerge from a rich variety of complex physical and chemical mechanisms at disparate time and length scales. Due to the vastness of the literature, this brief and incomplete overview is limited to dry adhesive sliding wear focusing only on a few recent works in the literature. A fairly complete synthesis of the existing empirical models can be found in [611].

Starting in the eighties with the advancement of AFM, tribology has taken a turn towards identifying molecular mechanisms behind friction [444,612], bringing about the era of nanotribology. This has naturally lead to uncovering three fundamental asperity-level mechanisms behind wear: atom-by-atom attrition [613-616], gradual smoothening by dislocation plasticity [617-620] and amorphization [621], as well as fracture-induced third body formation [436,622,623].

Beside theoretical studies [624-627], numerical modeling of wear processes has appealed to many as it opens the possibility to zoom in on an otherwise buried contact interface; however, numerical modeling comes with its share of difficulties. This is due, on one hand, to the challenge of the length scales of wear processes (engineering wear debris are often orders of magnitude larger than the scale of molecular processes that lead to them) and, on the other hand, to the diversity of underlying mechanisms (including plasticity, third body interactions, formation and propagation of cracks, chemistry). For instance, third bodies can have a significant effect on the frictional properties of the tribo-contact [628], sometimes even reducing the coefficient of friction [629].

Wear modeling approaches can be decomposed into continuum and discrete types. Continuum models, which include the popular finite element (FE) approach (see §2.2), have the advantage of being comparatively computationally affordable, while it is also fairly easy to introduce material parameters within macroscopic constitutive laws [630-636]. Correspondingly, DDD (see §2.3) has been recently used as a mesoscale approach to investigate plasticity upon asperity collision [429,637,638]. Both approaches are commonly used to study the onset of wear only, as they suffer in performance and require adaptive meshing when intense deformation due to shearing occurs. In general, when debris
are formed, it is best to use a discrete description of matter. The most prominent discrete modeling technique to model wear is classical MD (see §2.4). This is a very useful approach in particular because it is relevant in scale to a large body of experimental work in nanotribology [254,273,639-646]. The quality of the results is very much influenced by the care put into the choice of atomistic potentials [263,647-649]. Naturally, classical MD are limited to sizes below microns, which are relevant to nanotribology but not to a vast category of engineering wear scenarios, i.e. with debris sizes of the order of or above micrometers. At a scale above, an interesting approach is the DEM (see §2.4) [594,650,651]. In this method, numerical points aim to represent an ensemble of particles, or a grain, and the physical sizes of the model can be much larger. Of course, this is at the expense of material modeling accuracy, and the artificial length scale introduced when specifying a distance between particles can influence the wear mechanisms, and has to be carefully chosen.

A recent intermediate approach aims at coarse-graining simple atomistic potentials. In particular, a recently-developed coarse-grained atomistic potential [263] (i.e. discrete particles are meant to represent an ensemble of atoms) permits one to capture the formation of a steady-state debris particle generated during an adhesive wear process. Steady state implies here that the debris reaches a size that becomes eventually independent of time, and, in fact, that can be predicted at the asperity level [652], following a local Archard's law [653] (i.e. the debris size is dictated by the junction size) and a local Reye’s law [654] (i.e. the debris volume scales with frictional work). Numerical evidence [263] shows that there exists a critical length scale for junction size, above which surface asperities lead to “fracture” and thus produce wear debris particles, while smaller junctions exhibit “plastic” deformation [263]. This concept might be applied to contact wear maps to analyze which micro contacts lead to debris, and using probabilistic arguments to deduce wear coefficients from first principles, which to-date remain fully empirical parameters.

Due to the complex multiscale and multiphysics nature of wear processes, there is need of more systematic and multidisciplinary research to better understand the origins of wear at different scales. The recent advances summarized above give new hope at revisiting empirical engineering wear models and promoting physics-based mechanistic wear models at both the single and multiple-asperity levels.

3.9.2. Tribocchemistry

The control of friction and wear in a tribological contact is known to be related to several parameters such as the nature of the rubbing surfaces (roughness, physico-chemical composition, mechanical properties), contact conditions (pressure, shear stress), temperature, environment, etc. In particular cases, chemical reactions occurring during sliding will strongly influence the tribological behavior of the interface through the generation of new compounds. These phenomena are studied in the field of tribocchemistry and are often observed in boundary lubricated contacts [442]; a characteristic example is molybdenum dialkyldithiocarbamate (MoDTC) which is a well-known friction modifier additive used in engine oil that is able to significantly reduce friction through the generation of molybdenum disulfide (MoS₂) lamellar flakes in the contact [655,656]. Another typical example of lubricant additives is zinc dialkyl dithiophosphate (ZDDP) which is well known for its anti-wear properties thanks to the generation of a sacrificial phosphate-based tribofilm on steel contact surfaces [657]. The classical approach to study such phenomena is to characterize surfaces by identifying new compounds after tribological tests (post-mortem characterization). The thickness of the tribofilms usually ranges from few to several hundreds of nanometers. Surface-sensitive tools are so needed to physico-chemically characterize surfaces over a depth of a few nanometers. The analyzed area should also be as small as possible in order to spatially resolve nanoscale features. Recently, more and more in-situ experimental tools, coupling friction testing and in-situ characterization, have been used to gain access
into interfacial material modifications during rubbing [658-662]. Alternatively, tribochemistry is studied with MD and quantum calculation tools, as discussed in §2.5.

The activation of tribochemical reactions cannot be described with a universal mechanism but depends on conditions at the interface. During severe contact, for example, a "new" (nascent) surface is revealed, which reacts differently with the additives or the chemical environment from the initial one [663]. In the presence of insulating materials –mostly under dry conditions–, studies suggest that electrons and particles are emitted during sliding that could influence tribochemical reactions [664,665]. In general, the interface is at thermodynamical equilibrium when the temperature stays constant in the contact, either, at very low sliding speeds when no significant increase of temperature is found, or at high sliding speeds when the melting point of the contacting material has been reached. In all other cases, the interface is not at thermodynamical equilibrium and its behavior becomes significantly more complex [666]: For instance, under high-speed contact, the increase of temperature could be important with the thermal energy pushing through the energy barriers of chemical reactions. In such a case, the tribochemical reaction mainly occurs because of thermal energy generated in the contact. Furthermore, in some cases, normal and shear stresses applied on the "interfacial material" could promote a tribochemical reaction [659,667,668]. In this case, tribochemical reactions are promoted by the mechanical energy, which helps decrease the energy barriers of the chemical reaction pathway. Relevant models about these topics have been reviewed by Spikes and Tysoe [669].

3.9.3. Contact scale issues in experimental biotribology
Nanotribological experimental approaches have been employed for contact mechanics and friction studies of biological tissues. Concerning the synovial joint system, for example, the use of AFM has given new insight on the frictional properties of cartilage tissues [670-672] –including in the study of synovial joints [673]–, allowed for the detection of different elasticity (stiffness) on the proximal versus distal areas [674,675] and the identification of more compliant characteristics of the pericellular matrix than territorial/interterritorial matrices of cartilage [676]. The distinction between healthy areas and enzymatically defected areas of cartilage is possible exclusively with very sharp (nanometer-sized) AFM probes [677], which led to the development of AFM-based arthroscopy [678].

A common observation is that the excellent lubricating capabilities of cartilage tissues, reported by many macroscale experimental studies [679], were not found at the small scale, not even on experiments preformed on thin films prepared with the individual constituents of cartilage [680-683]. In studies with sharp AFM tips the very small contact area achieved by the AFM probe on the cartilage surface is likely to inhibit the activation of interstitial fluid pressurization. This may indicate an intrinsic hurdle or, alternatively, a fundamental challenge in the usage of AFM for nanotribological studies of cartilage. When it comes to the frictional properties of cartilage tissues and model thin films for small scale contact, computational modeling studies have been relatively scarce to date. Multiscale and multiphysical tribological models are necessary to fill this gap.

3.9.4. Skin tribology
The skin controls many types of exchanges between our inner and outside worlds which take the form of mechanical, thermal, biological, chemical and electromagnetic processes [684]. These processes concurrently operate as parts of a very dynamic system featuring highly non-linear feedback mechanisms [512,685,686] where mechanics is pivotal. As mounting evidence suggests, skin microstructure can play a critical role in how macroscopic deformations are modulated at the microscopic level [687]. These structural mechanisms are also at the heart of skin tribology by constituting and conditioning mechanical load transmission [38,510-512,688-691].
It is widely accepted that skin friction is made of deformation-induced and adhesion components [511,692-695] but, up to now [512], adhesion-induced friction has been deemed to be the dominant contributor to macroscopic friction. Applying a computational homogenization procedure to a 2D anatomically-based finite element multilayer model of the skin, Leyva-Mendivil et al. [512] recently showed that deformation-induced friction can be significant when the skin surface is subjected to the action of a single rigid indenter of sub-millimeter size. It was shown that the macroscopic coefficient of friction between the skin and a rigid slider moving across its surface is noticeably higher that the local coefficient of friction applied as an input parameter to the finite element analyses [512]. Similar observations were reported in a 3D computational contact homogenization study [696]: geometrical effects alone can have a significant impact on the macroscopic frictional response of elastic contacts. These results support the idea that accounting for the microstructure of biological tissues and the heterogeneous nature of their mechanical properties could be critical in determining their biotribological properties. Using their computational contact homogenization modelling framework [512], Leyva-Mendivil et al. [697] recently demonstrated the pressure sensitivity of skin friction which is strongly modulated by finite deformations of skin surface asperities, Similar observations were made in an experimental context by Wolfram [695]. This has important tribological consequences in combination with the effect of relative humidity on the mechanics of the epidermis, particularly when considering mechanically-induced skin wrinkles [698].

To date, despite many experimental and modeling studies investigating shear stress at the surface of the skin in relation to skin injuries and pressure ulcers [699-701], very little effort has been devoted to develop methodologies to gain a more quantitative and mechanistic understanding of how shear stresses are induced at the level of skin micro-relief asperities, and how they propagate from the skin surface to the deeper layers where they are likely to mechanically stress living cells [38]. Ultimately, excessive stress or strain can lead to cell damage and death, which, at a meso/macroscopic level translates into tissue damage and loss of biological structural integrity. If one considers that, notwithstanding the strong sensitivity of the skin to fluctuations in environmental conditions, (finite strain) mechanics is typically coupled to biochemistry and other physical processes such as thermal transfer, it is clear that the formulation of any type of sufficiently descriptive contact theory of the skin is going to require substantial integrative efforts. Due to the fibrous nature of their cytoskeleton, cells also feature strongly anisotropic properties, which, combined with their extreme deformability, calls for new contact theories of biological soft matter. This presents numerous challenges at a theoretical, computational and experimental level but also provides outstanding opportunities to establish an ambitious research roadmap to push further the boundaries of our current knowledge and capabilities, in biotribology and biological soft matter in general, and in skin tribology in particular.

3.9.5. Cardiac dynamics: multiphysical biotribology
In the last few years, new perspectives for contact mechanics research in biotribology are emerging as far as the problem of contact interactions between biological cells is concerned; see, for example, a wide overview in [702-706]. In cardiac dynamics, myocytes, which are the fundamental cells composing the cardiac tissue, interact in a very complex way across their boundaries, transferring physiological quantities, electric current, and also mechanical tractions [707]. Moreover, their boundaries evolve in time, as a result of growth, remodeling and aging effects [708]. From the mathematical point of view, the complex myocyte dynamics and its electrophysiological behavior can be described by a set of reaction-diffusion partial differential equations for the diffusive membrane voltage and for the local electrophysiological gating fields [709,710]. The nonlinear coupling between electrophysiology and the hyperelastic material response induced by the excitation-contraction mechanisms is typically modelled via the multiplicative decomposition of the deformation gradient
into elastic and anelastic parts; see, for example, [711-713] for more details on theoretical and computational aspects related to this modeling strategy. Specifically, the inelastic active deformation gradient can be provided by the subcellular calcium/voltage dynamics, while the elastic deformation gradient is computed as customary [711].

Complementing these continuum mechanics formulations with suitable interface constitutive relations to address the problem of myocyte-myocyte interaction is an open problem, with preliminary attempts to solve having already been proposed in [707,714]. Mechanical interactions should account for adhesion and contact tractions dependent on the local cell-cell separation, to reproduce the experimental evidence. Finally, as a further model improvement, the roughness of cell-cell interfaces should be accounted for, leading to a distribution of partially insulated but still conductive spots rather than a fully conductive interface. In this regard, the fundamental discoveries in the field of electric and thermal contact problems in the presence of roughness are expected to be applicable and extendable also to myocyte contacts. As proposed in [714], the myocyte interface can be modelled as an imperfect zero-thickness boundary layer, whose response can be governed by nonlinear constitutive relations generalizing the popular cohesive zone models used in fracture mechanics for pure mechanical interactions. The mechanical field has to be coupled with other fields, such as the electric one, to be transferred across the interface. Notably, the results established in [366,715] are expected to play an important role regarding the relation between electric current and voltage.

3.9.6. Industrial case studies: Steel forming processes, wafer lithography and roller bearings

Controlling tribological properties in steel-making processes is necessary to improve quality and increase the production rate. Undesirable phenomena include temperature-dependent adhesive wear, flaking and galling. The industry currently uses tribological models that are based on continuum theories and incorporate limited microscale aspects and simplified roughness representations, or phenomenological models that strongly rely on experience: e.g., the friction coefficient is varied within a known range to predict process parameters. Philips Drachten, for example, currently uses a micromechanics-based numerical model to predict friction coefficients that vary with local pressure, strain and temperature [716]. Such models calculate the load-carrying capacity of lubricant-filled cavities, where the Young’s modulus and flow stress are modelled as temperature-dependent. There is a need for numerical models that satisfy certain criteria: they should use computationally-efficient simulation strategies, be usable in automated control systems to allow in-line adjustment of process settings based on (meta)data, and they should be robust across various processes and demonstrable results at both ends of the dimensional range. Hence, there is need for simple (perhaps, even, analytical) but comprehensive predictive models of friction as well as system-level simulations that can incorporate tribological aspects into the modeling of multi-stage deformation processes.

While unanswered questions remain and improved models are needed in the “classical” manufacturing world, tribological issues persist also for semiconductor companies such as ASML that use fast extreme ultraviolet (EUV) lithography on large tens-of-micrometers-thick wafers to manufacture integrated circuits with positioning accuracies of the order of nanometers. Physics and chemistry questions are relevant for such processes, focusing on EUV source, scanner, metrology and process attributes. Current positioning methods involve electrostatic forces used to fix the wafers onto burls on the substrate; improving and optimizing positioning accuracy requires multiphysics modeling across scales since wafer-support forces lead to wafer distortions and, in turn, to overlay and height (out of focus) errors. Adhesion and friction play an important role in wafer support as does the contact and clamping history: the order in which contact with individual burls is established is different every time. Furthermore, positioning is a dynamical contact phenomenon that, at such small scales, results in
accelerations of about 50g. One major advance for the industry would be to realize switchable friction without wear.

A final example of an industrial case study is the reduction of friction in roller bearings. Having this as the ultimate goal, researchers at SKF with academic partners performed non-equilibrium molecular dynamics simulations of stearic acid adsorbed on iron surfaces with nanoscale roughness [717]. The stearic acid films were found to be able to maintain separation of asperities on opposing surfaces due to strong adsorption of the head groups, thereby decreasing the friction coefficients and Derjaguin offsets. These effects were negligibly affected by an increase in surface roughness. To tackle larger size and time scales, multiscale methods are likely candidates for future research. Of particular interest are the quasi-continuum method [718], and the CPL library (http://cpl-library.org/) [719], a recently developed communication and topology management system for coupling continuum fluid dynamics to molecular dynamics. Other possible avenues for further research are accelerated molecular dynamics techniques.

4. Conclusions

One of the main outcomes of the Lorentz workshop on “Micro/Nanoscale Models for Tribology” was the realization that, despite the modeling community’s ability to address elastic problems of great complexity at various scales, significant effort is still required to account for effects like plasticity, adhesion, friction, wear, lubrication and surface chemistry in tribological models. Although many systems do involve two or more of those phenomena at various scales, multiscale and multiphysics models are still challenging to develop and use as they require multidisciplinary expertise and collaborative effort. Nevertheless, a few successful examples are provided in the text. Breakthroughs are thus expected from the future development of versatile and efficient multiscale/physics tools dedicated to tribology. On the other hand, tribologists still need to identify key elementary processes specific to rough contacts under shear, and associated, for example, to crack nucleation and propagation, chemical reactions, or fluid-solid interactions. In order to keep a clear physical understanding of the outcome of complex models, those processes will preferably be first studied on their own, before introducing the related behavior laws in more comprehensive tools. Only by pursuing simultaneously both research avenues will the tribology community have a chance to (i) advance on the fundamental understanding of frictional interfaces and (ii) propose simple but comprehensive models useful to optimize and control industrial processes.

As a good way of improving existing models and testing new ones, one agreement that was reached among the participants of the workshop was the need for more exercises like the contact-mechanics challenge described in §3.5. This is not a trivial task, and no general consensus was reached about what could be the best challenge to launch next. However, the need to propose tribology challenges for quantities that can be also experimentally measured in parallel was clearly expressed. In such a way, challenges would not be mainly academic exercises of computing capabilities, but may help set up realistic problems which can have reasonable experimental counterparts. In this context, quantitative comparison with experiments will naturally lead to considering effects not taken into account in the contact-mechanics challenge, such as plasticity, long-range adhesion, large deformations and friction. Those effects could first be assessed separately and then simultaneously with an extensive range of parameters and not just one precise choice. The development of deterministic ways of preparing surfaces (e.g. 3D printing, or micro-milling) opens the way for experimental assessment of the role of various roughness scales on tribological properties, by adding more and more scales in the surface topography.
Considering the contrast between the convergence of interests among the workshop participants and the diversity of cultures and modeling traditions in their respective communities of origin, a need for collaborative platforms for tribologists has emerged. A shared platform, organized via a dedicated website, could include the following sections: (i) open source software provided by research groups, useful also for dissemination purposes; (ii) a collection of contact problem results, reporting, for each case study, the surface topography used as an input for the simulation/experiment, the material parameters and the constitutive model, and a description of the assumptions of the computational model used to obtain the contact response; (iii) a list of simulation and testing facilities of research groups working on contact mechanics, with links to their websites and laboratories, organized according to the major problems of industrial interest. This collaborative platform is envisaged to have an important impact on the community to foster novel round robin campaigns like the challenges mentioned in the previous paragraph, provide material useful for benchmark tests, increase the awareness of companies in the applicability of tribology and contact mechanics research to solve problems of industrial interest and ultimately accelerate tribological research in the interdisciplinary manner necessary to lead to breakthroughs in the field.

Acknowledgements

This review is the result of a Lorentz Center workshop on “Micro/Nanoscale Models for Tribology” held in Leiden between 30 January and 3 February 2017. The workshop was co-organized by M. Ciavarella, A. Fasolino, L. Nicola, J. Scheibert, A.I. Vakis and V.A. Yastrebov and sponsored by the Lorentz Center, the Royal Netherlands Academy of Arts and Sciences (KNAW), the Materials Innovation Institute (M2i), the Groningen University Fund (GUF) and the company Nanovea. Participants came from engineering, materials science, applied physics and chemistry backgrounds with a 9:1 ratio of academia to industry and 85% being (predominantly) modelers and theoreticians versus 15% being (predominantly) experimentalists (for more information, please visit the conference-specific website at https://tribomodels2017.sciencesconf.org/). Interactive discussions were held over a broad range of topics relevant to tribology with emphasis on modeling, facilitated via 7 keynote lectures and 16 round-table discussion sessions. We acknowledge the input of the participants in the writing of this manuscript, both by supplying useful comments and suggestions as well as providing up-to-date and pertinent references.

J.S. acknowledges funding from the People Program (Marie Curie Actions) of the European Union’s Seventh Framework Program (FP7/2007–2013) under Research Executive Agency Grant Agreement PCIG-GA-2011-303871. He also acknowledges support from Institut Carnot Ingénierie@Lyon and from 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).

G.C. and F.B. are supported by H2020 FET Proactive “Neurofibres” grant No. 732344. R.G. is supported by Bonfiglioli Riduttori SpA.

N.M.P. is supported by the European Commission H2020 under the Graphene Flagship Core 1 No. 696656 (WP14 “Polymer composites”) and FET Proactive “Neurofibres” grant No. 732344.

A.C. is supported by the Czech Science Foundation, project 17-24164Y.

P.N. acknowledges the support of the Czech Science Foundation through the project 16-11516Y.

D.D. acknowledges the support of the Engineering and Physical Sciences Research Council (EPSRC) under the Established Career Fellowship grant EP/N025954/1.
References


[335] Hatano T. Scaling properties of granular rheology near the jamming transition. Journal of the Physical Society of Japan 2008;77:123002-.


[450] Coulomb CA. Théorie des machines simples en ayant égard au frottement de leurs parties et à la roideur des cordages.: Bachelier, 1821.


Despite the fact that the task set by the challenge was well-defined in scope (only elastic deformations, no shape but only a single realization of a nominally flat infinite rough surface to consider, modest adhesion, and well-detailed information including some data files to start with), still Martin Müser remarked that reaching the stage where different groups would provide results in the same units, and putting together the amount of information obtained, was not an easy task, involving more than 1,400 email exchanges.