Dissertation zur Erlangung des Doktorgrades der Technischen Fakultät der Albert-Ludwigs-Universität Freiburg im Breisgau

How surface roughness affects adhesion

Antoine Sanner

17.03.2023

Albert-Ludwigs-Universität Freiburg im Breisgau
Technische Fakultät
Institut für Mikrosystemtechnik - IMTEK

Dekan
Prof. Dr. Roland Zengerle

Referenten
Erstgutachter: Prof. Dr. Lars Pastewka
Zweitgutachter: Prof. Dr. Martin H. Müser
Vorsitz: Prof. Dr. Moritz Diehl
Beisitz: Prof. Dr. Patrick Dondl

Datum der mündlichen Prüfung
17.03.2023

Bearbeitungszeit
Juli 2018 bis Dezember 2022

Abstract

At atomic scales, all molecules attract each other, but macroscopic objects usually do not stick. The explanation for this apparent paradox is that most surfaces are rough, so that elastically stiff objects only touch on the top of their asperities. Geckos and insects have compliant fibrillar structures or soft pads at the tip of their feet that conform to surface roughness, sustaining enough adhesion to climb vertical walls. Understanding the role of surface roughness in adhesion is a challenge because surfaces exhibit roughness down to the atomic scale.

In this thesis, my collaborators and I investigate the effect of surface roughness on adhesion in both stiff and compliant contact systems. I model adhesion theoretically, and I help experimentalists analyze surface topography over multiple scales. The combination of my new models and of the comprehensive surface topography characterization by Abhijeet Gujrati (University of Pittsburgh), allows us to unravel the role of surface roughness in adhesion experiments.

Stiff materials do not stick because roughness prevents most of the surfaces to come into the range of molecular attraction. A recent theory quantifies this effect based on an approximate expression for the distribution of interfacial gaps near the contact edge. Joe Monti (Johns Hopkins University) and I benchmark this expression against gap distributions extracted from finely resolved numerical simulations. The theory is valid provided that adhesive stresses are weak and act over a range shorter than a geometrical parameter determined by small-scale roughness.

Elastically soft (jelly-like) objects stick because the elastic penalty to deform into intimate contact is small compared to the gain in surface energy. However, theories based on this simple thermodynamic argument cannot explain the fact that in experiments, the force measured during retraction is often much higher than during indentation. This *adhesion hysteresis* can be caused by material specific irreversibility or elastic instabilities triggered by surface roughness. The role of these instabilities in adhesion hysteresis remains poorly understood because existing numerical and theoretical models cannot account for realistic roughness in soft contacts. I introduce an efficient crack-perturbation model for the contact of rough spheres, enabling large scale simulations with realistic surface roughness. By clarifying the link between adhesion hysteresis and classic pinning problems (for example fracture of heterogeneous materials and wetting angle hysteresis), this model allows me to derive a simple theoretical model linking adhesion hysteresis to surface roughness. In combination with the characterization of surface roughness over multiple scales, my models shed light on the role of elastic instabilities in adhesion experiments.

Surfaces are rough from the macroscopic scale down to the atomic scale, and the lack of comprehensive roughness characterization is the major obstacle towards bringing theory and experiments together. Abhijeet Gujrati and collaborators measured the roughness of four diamond coatings over eight decades of length scales, enabling the application of adhesion

theories on experiments performed with these samples. Besides the experimental challenge of determining roughness down to the atomic scale, an additional obstacle to the documentation of roughness is the technical complexity of established multiscale roughness measures such as the power spectral density. My collaborators and I address this problem by introducing the *scale-dependent roughness parameters* (SDRPs), a new analysis framework that is easy to interpret and to implement. This new analysis, together with several established techniques, is available to use through our web-service *contact.engineering*. We thereby encourage the community to measure, analyze and publish roughness over multiple length scales.

The SDRP analysis computes the fluctuations of slopes and curvatures as a function of the lateral length scale. Slopes and curvatures are important ingredients for rough contact theories, but it remains unclear at which scales they matter. Luke Thimons (University of Pittsburgh) and I show that in macro-scale contacts between ruby spheres and diamond coatings, the roughness that critically affects adhesion is between lateral length scales of 43 nm to 1.8 µm. The large-scale cutoff is related to the finite radius of the spherical indenter, while the unimportance of small scales is due to plastic deformations and the long range of the adhesive interaction (5 nm). To determine the critical range of length scales, as well as the parameters of the adhesive interaction, we analyzed the experimental pull-off forces by combining surface topography characterization and numerical simulations.

Adhesion is critical in applications such as microelectromechanical systems (MEMS), soft robotics and skin adhesives. Our insights provide guidance for practitioners which scales of roughness to control in order to tune adhesion, and our framework for surface topography characterization will allow a better overall understanding of surface topography across the community.

Journal publications

This thesis is partially comprised of the following journal articles and preprints. Full versions of these articles can be found in the appendix.

[**I**] A. Sanner and L. Pastewka, "Why breaking soft contacts is harder than making them", In preparation (2023).

[**II**] M. C. Röttger, A. Sanner, L. A. Thimons, T. Junge, A. Gujrati, J. M. Monti, W. G. Nöhring, T. D. B. Jacobs, and L. Pastewka, "Contact.engineering - Create, analyze and publish digital surface twins from topography measurements across many scales", Surf. Topogr. Metrol. Prop. **10**, 035032 (2022).

[**III**] A. Sanner and L. Pastewka, "Crack-front model for adhesion of soft elastic spheres with chemical heterogeneity", J. Mech. Phys. Solids **160**, 104781 (2022).

[**IV**] A. Sanner, W. G. Nöhring, L. A. Thimons, T. D. Jacobs, and L. Pastewka, "Scale-dependent roughness parameters for topography analysis", Appl. Surf. Sci. Adv. **7**, 100190 (2022).

[**V**] L. A. Thimons, A. Gujrati, A. Sanner, L. Pastewka, and T. D. B. Jacobs, "Hard-material adhesion: which scales of roughness matter?", Exp. Mech. **61**, 1109 (2021).

[**VI**] J. M. Monti, A. Sanner, and L. Pastewka, "Distribution of gaps and adhesive interaction between contacting rough surfaces", Tribol. Lett. **69**, 80 (2021).

[**VII**] A. Gujrati, A. Sanner, S. R. Khanal, N. Moldovan, H. Zeng, L. Pastewka, and T. D. B. Jacobs, "Comprehensive topography characterization of polycrystalline diamond coatings", Surf. Topogr. Metrol. Prop. **9**, 014003 (2021).

Contributed talks to conferences

I presented my research in talks at the following conferences.

Adhesion hysteresis of rough spheres: pinning of an elastic line
European Solid Mechanics Conference 2022 (Galway, Ireland)

Crack-front model for the adhesion hysteresis of rough spheres
Tribology Gordon Research Seminar 2022 (Lewiston, USA)

Crack-front model for the adhesion hysteresis of rough spheres
Contact Mechanics International Symposium 2022 (Lausanne, Switzerland)

Crack-front model for the adhesion hysteresis of rough spheres
Adhesion Society Annual Meeting 2022 (San Diego, USA)

Contact line pinning in soft adhesion
Adhesion Society Annual Meeting 2021 (virtual)

Public software

My collaborators and I provide a free-to-use web-platform for the analysis and publication of measurements of surface topography under the url `https://contact.engineering/`. The code underlying this web-platform, as well as my boundary-element (BEM) and crack-front simulations, is open source and available in the *GitHub* organization `https://github.com/ContactEngineering/`. This computational ecosystem is decomposed in multiple repositories with dedicated functions, *SurfaceTopography*, *ContactMechanics*, *Adhesion*, *CrackFront* and *Topobank*, which I describe individually further below. Our codes also make use of the *Python* packages *muFFT* for the fast Fourier transform, and of *NuMPI* for parallelization with the message-passing interface (MPI) [1] and for minimization algorithms. My main contributions to these projects are the MPI-parallelization of the boundary element method and of the underlying minimization algorithms [2–4], enabling the large reference simulations used in publications [**III**, **VI**]. Till Junge and Lars Pastewka wrote preliminary implementations of the BEM and of the web-application, and Lars Pastewka, Michael C. Röttger and I developed them further. Part of the development of the *ContactEngineering* ecosystem was contributed in the framework of the master's theses of Sindhu Singh [5] and Paul Strauch [6], which I cosupervized with Lars Pastewka. I list the functionalities of these packages and my contributions to the individual repositories in more details below.

muFFT

`https://gitlab.com/muspectre/muspectre`
Main contributors: Lars Pastewka, Till Junge, Antoine Sanner
Hybrid C++/Python, GNU LGPLv3 License
muFFT is a module of the *µSpectre* project [7]. This module contains wrappers to serial and MPI-parallel implementations of the fast Fourier transform, and finite-difference stencils for the computation of derivatives. I implemented a wrapper to the half-complex fast Fourier transform provided by FFTW [8], which Sindhu Singh [5] used to develop a preconditioned formulation of the BEM inspired from Ref. [9].

NuMPI

`https://github.com/IMTEK-Simulation/NuMPI`
Main contributors: Antoine Sanner, Sindhu Singh, Lars Pastewka
Python, MIT License
The *Python* package *NuMPI* is a collection of numerical tools for MPI-parallelized *Python* codes, including minimization algorithms, reduction operations, and parallel file input and output. It uses the *mpi4py* [10] interface to the MPI standard [1]. I am the main maintainer of this repository. I implemented a parallelized version of the L-BFGS algorithm [4]. Sindhu

Singh implemented constraint conjugate gradients algorithms [2, 3, 11, 12] and I parallelized them.

SurfaceTopography

`https://github.com/ContactEngineering/SurfaceTopography`
Main contributors: Lars Pastewka, Michael C. Röttger, Antoine Sanner, Till Junge, Johannes L. Hörmann
Python, MIT License
The *Python* package *SurfaceTopography* provides tools for the processing of raw measurement data, datastructures for the definition of surface topography, multiscale statistical analyses of surface topography, and the generation of synthetic self-affine surfaces. It is the main computational backend behind our web-service *contact.engineering*, and we describe its main functionalities in publication [**II**]. This package implements readers for raw measurement data and implements the automatic detection of tip artifacts introduced in publication [**IV**]. The multiscale analysis of surface roughness includes standard descriptors such as the autocorrealation function (ACF) and the power-spectral density (PSD), as well as our newly established *scale-dependent statistical roughness parameters (SDRPs)* [**IV**]. *SurfaceTopography* implements the stitching and averaging of multiple measurements into multiscale descriptors of the surface topography as established in Refs. [**VII**, 13]. This package also implements the Fourier-filtering algorithm described in Refs. [14, 15] that we used to generate synthetic surfaces. I participated in the general maintenance of the repository. I MPI-parallelized the code, implemented scale-dependent roughness parameters and the detection of tip artifacts [**IV**]. Paul Strauch [6] implemented scale-dependent statistical parameters quantifying deviations from Gaussian distributions.

ContactMechanics

`https://github.com/ContactEngineering/ContactMechanics`
Main contributors: Lars Pastewka, Antoine Sanner, Till Junge, Sindhu Singh
Python, MIT License
The *Python* package *ContactMechanics* implements BEM simulations and analytic models for contact mechanics without adhesion. It relies on *SurfaceTopography* for datastructures defining the contact geometry. Lars Pastewka and I maintained this repository and I MPI-parallelized the code.

Adhesion

`https://github.com/ContactEngineering/Adhesion`
Main contributors: Lars Pastewka, Antoine Sanner, Till Junge, Sindhu Singh
Python, MIT License
The *Python* package *Adhesion* extends *ContactMechanics* with interaction potentials for the BEM simulations of adhesive contacts and implements analytic models of adhesive contacts. Lars Pastewka and I maintained this repository. I MPI-parallelized the code, implemented interaction potentials, and implemented analytic models such as the JKR contact of smooth

spheres [16]. Sindhu Singh implemented preconditioned BEM simulations inspired from Ref. [9].

CrackFront

`https://github.com/ContactEngineering/CrackFront`
Main contributors: Antoine Sanner
Python, MIT License
I implemented my crack-front simulations in the *Python* package *CrackFront*. The code is GPU parallelized using *PyTorch* [17]. It includes the crack-front equations for spheres [**I**, **III**] and for a semi-infinite crack, the mapping from roughness to equivalent work of adhesion heterogeneity, and the crack-propagation algorithm by Rosso and Krauth [18]. I wrote this repository in its entirety.

Topobank

`https://github.com/ContactEngineering/Topobank`
Main contributors: Michael C. Röttger, Lars Pastewka
Python/Web, MIT License
The repository *Topobank* implements the database and the web-interface for our web-service *contact.engineering*. It uses the packages *SurfaceTopography* and *ContactMechanics* to load raw measurement data and to perform the analyses such as SDRPs, PSD and BEM simulations. I participated in the design and in the testing of the web application and I provided support for users via e-mail and a discussion forum.

Contents

1 Introduction

All matter attracts with van der Waals forces [19], and insects and lizards rely on these universal forces to climb vertical walls made of any material [20–22]. Van der Waals interactions are the weakest among intermolecular forces, however they cause strong attractive pressures of order 100 MPa. Acting on a contact area of 1 cm^2, these pressures are strong enough to sustain the weight of a car [21], in contradiction with our daily experience. The main explanation for this "adhesion paradox" is that roughness keeps most of the surface apart from molecular interaction forces [21, 23, 24], which typically act over ranges below a nanometer [19] (Fig. 1.1a,b).

Very soft solids (Fig. 1.1c) are sticky because they are able to conform to the surface roughness [25, 26]. Yet, their adhesion strength is by far not as spectacular as mentioned above, even for perfectly flat interfaces. In this case, the explanation is that interfaces do not separate uniformly. The edge of the contact corresponds to a crack where stresses concentrate due to elastic deformation. As we pull the objects apart, the adhesive bond progressively breaks by

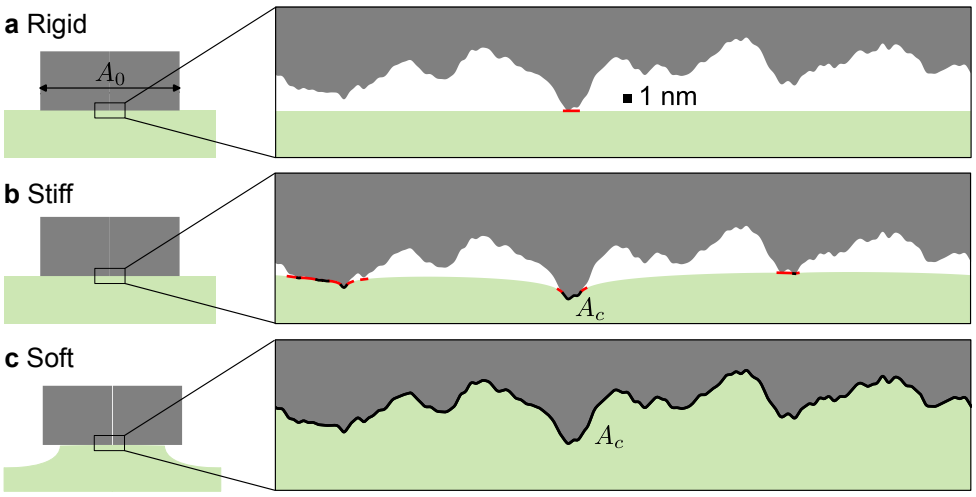


Figure 1.1: Effect of surface roughness on adhesion. **(a)** In the contact of rigid objects, van der Waals forces only act in a vanishingly small area confined to the top of the highest asperity (red line). **(b)** Stiff elastic materials conform to the roughness over a finite area of contact A_c (black line). The actual area of contact A_c is only a small fraction ($\lesssim 10\%$) of the apparent area of contact A_0, so that the majority of the surface does not interact. The adhesive stresses are small compared to the repulsive stresses in the contact area, and the Derjaguin-Muller-Toporov (DMT) approximation neglects the elastic deformations caused by adhesive stresses. **(c)** Soft elastic materials can fully conform to the surface roughness. The large adhesive stresses concentrate at the edge of the contact, forming an adhesive neck, or crack tip. In this Johnson-Kendall-Roberts (JKR) regime, the solids separate by the propagation of this crack.

the zip-like propagation of the crack. The breaking process is governed by Griffith's [27] balance between the elastic strain energy released by propagating the crack and the increase in surface energy associated with separating the surfaces, the work of adhesion.

In soft contacts that behave like cracks, the adhesive stresses cause elastic deformations large compared to the interaction range. This limit of soft materials with strong adhesive stresses confined to small interfacial gaps is named after Johnson Kendall and Roberts (JKR) [16], who described the contact of spheres through fracture mechanics. The opposite limit of stiff materials and weak adhesive stresses is called the DMT limit, because Derjaguin Muller and Toporov [28] neglected any deformation caused by adhesive stresses when computing the force acting between spheres. Bradley [29] computed the force needed to separate rigid spheres (pull-off force). In all these limits, the pull-off force of spheres is proportional to the work of adhesion, so that experiments with spherical probes are commonly used to quantify the strength of adhesive interactions [19].

Surface roughness The major experimental and modeling challenge in contact mechanics is that rough surfaces are fractals: asperities have smaller asperities on top of them [21, 30–33]. The sharp peaks at the small scales are the hardest to conform to and determine the contact area. Other contact properties such as the stiffness depend on the large-scale roughness. Which scales of roughness dominate adhesion depends on the situation and is a matter of current debate.

It is therefore important to statistically characterize surface roughness over a wide range of scales, but measuring roughness down to the atomic scale is a challenge. Early studies relied only on scalar roughness parameters [34–36], which are affected by the resolution of the instrument and are hence not representative of the true surface topography. It is now common to describe roughness via the power spectral density [14, 21], that decomposes a measurement into contributions from different spatial frequencies, and several measurements can be combined in order to construct a comprehensive description of the multiscale roughness [13, 37, 38]. Recent advances in experimental techniques allow the measurement of roughness down to the atomic scale [13, 39].

Role of adhesion in non-sticky contacts Most daily life objects are closer to the stiff, DMT regime, where atomic-scale roughness limits the true area of contact to the highest peaks and thereby suppresses macroscopic adhesion [24, 40]. Adhesive interactions still play an important role because they increase the contact area [24, 40, 41]. The contact area critically affects important properties of the interface like the electric conductivity, heat transfer, fluid leakage, and friction.

Friction commonly obeys Amontons' law [42] (observed earlier by Da Vinci [43]), stating that the lateral force is proportional to the normal force and is independent of the apparent, macroscopic area of contact A_0. A major advance in contact mechanics was to understand that the *actual* area of contact A_c is independent of A_0 and limited to the top of a few asperities [44]. Amontons' law holds because the contact area is proportional to the normal force [45–48], and because the shear stress inside a contact junction is a material constant [49]. While this early work ignored adhesive interactions, Pastewka and Robbins [24] and Müser [40] showed that

for weak adhesive interaction, the contact area increases but remains proportional to the force.

Analytic models Theoretical models link macroscopic contact properties such as the contact area to statistical measures of the surface roughness. The two main categories of analytic models are asperity models and Persson's theory.

Greenwood and Williamson [45] modeled the interface as a set of independent spherical contacts, and showed that the proportionality between area and load emerges from the random distribution of the heights of asperities. Their original work used Hertz's [50] solution of the nonadhesive contact of spheres, but was extended to adhesive contacts using the JKR or DMT theories [34, 51]. While the theory by Bush Gibbson and Thomas [46] accurately predicts the proportionality between force and contact area for small contact area fractions, asperity models break down at large contact area fractions. Furthermore, they fail to predict other properties such as the mean and distribution of interfacial gaps [55], or the stress- and contact-autocorrelation function [0].

Persson's theories [41, 52, 53] accurately predict the contact area from small to full contact area fractions [55, 54, 56]. Borrowing ideas from renormalization group theory [57], Persson describes how the contact area and the distribution of interfacial pressures evolve as roughness at progressively smaller lateral length scales is included. His theory accounts for all length scales of the power spectral density and predicts many contact properties such as the contact area, mean gap and the stored elastic energy. Persson derived two variants for adhesive contacts, one where the contact area is set by Griffith's energy balance (JKR limit) [41] and the other using the DMT approximation [53].

Theoretical models linking adhesion to random surface roughness make simplifying approximations and often focus on limiting cases such as the JKR or DMT limit. A global understanding of the role of surface roughness requires the determination of the accuracy and the range of validity of specific adhesion theories using numerical simulations as a benchmark [24, 40, 55, 56].

Numerical models Numerical simulations describe in detail the contact of a specific realization of the surface topography (e.g. an AFM areal scan or an artificially generated surface) and make only a few basic assumptions. The most popular numerical model is the boundary element method (BEM) [2, 24, 55, 58–60] also called Green's function molecular dynamics (GFMD). It accelerates the computation by using fast Fourier transforms (FFT) and an analytical solution linking surface displacements to pressures, the Green's function. BEM considers only the degrees of freedom at the surface, in contrast to classical finite element or full molecular dynamics simulations, which resolve the displacements in the whole volume. Despite this speed-up, the range of fractal roughness that can be taken into account is limited by the computational resources. Especially the adhesion of soft materials is challenging to describe using these models because the sharp crack tip requires finely discretized meshes [61].

Adhesion hysteresis in soft sticky contacts The role of surface roughness in the adhesion of soft solids is still poorly understood. Dalvi et al. [62] investigated experimentally how the adhesion of soft rubber spheres depends on surface roughness. Using the multiscale

roughness characterization of publication [**VII**], they showed that a theory by Persson and Tosatti [26] quantitatively predicts how surface roughness reduces the adhesive force measured when the contact forms. However, Persson's theories assume that the contact evolves following thermodynamic equilibrium, and thereby fail to explain why adhesion is much stronger during retraction than during approach. This *adhesion hysteresis* [63] can be caused by material specific dissipation mechanisms like viscoelastic dissipation in the bulk of the material or molecular rearrangements at the surface, but it can also emerge in purely elastic systems when surface roughness triggers elastic instabilities. Because they observed that the dissipated energy depends critically on the surface roughness, Dalvi et al. hypothesized that elastic instabilities triggered by surface roughness play the dominant role in their experiment.

Numerical simulations [61, 64–70] and simplified analytical models [36, 71–76] show that the surfaces sometimes enter and leave contact in sudden jumps (elastic instabilities) instead of moving continuously. These elastic instabilities dissipate energy and lead to adhesion hysteresis even in purely elastic systems. This theoretical work indicates that the main parameter governing adhesion hysteresis is the elastic energy required to fully conform to the surface roughness, but there is no theory quantitatively linking adhesion hysteresis to random roughness. Because we lack numerical or theoretical models accounting for realistic surface roughness, it is unclear whether these jumps occur and how much energy they dissipate in real experimental situations.

2 Problem statement and approach

The overarching scientific question of this cumulative dissertation is to understand how surface roughness affects adhesion. This thesis discusses three regimes of adhesion where the solid is either *rigid* (Fig. 1.1a), elastic but *stiff* compared to the adhesive stresses (DMT limit, Fig. 1.1b), or so *soft* that the contact behaves like a crack (JKR limit, Fig. 1.1c). I model adhesion theoretically [**I**, **III**, **VI**], and I help experimentalists to analyze surface topography over multiple scales [**II**, **V**, **VII**]. The combination of my new models [**I**, **III**] and of my collaborators' comprehensive surface topography characterization [**VII**, 77] allows us to unravel the role of surface roughness in adhesion experiments [**V**, 62, 78]. Our scientific insights were enabled by tackling the experimental challenge of statistically characterizing the surface topography from macroscopic to atomic length scales, and the theoretical challenge of developing a model for elastic instabilities that can account for this multiscale roughness. I use large-scale boundary element method simulations as a benchmark for my new model [**I**, **III**] and Joe Monti (Johns Hopkins University) and I use them to test the theory by Pastewka and Robbins [**VI**, 24].

Roughness Experimental datasets where the roughness has been measured and documented over a sufficient range of length scales are still rare. Besides the experimental challenge of determining roughness down to the atomic scale, an additional obstacle to the documentation of roughness is the technical complexity and a lack of consensus on how to statistically describe a surface over multiple scales [14, 79]. Publication [**IV**] and the public database *contact.engineering* described in publication [**II**] aim to simplify the process of analyzing surface roughness over multiple scales.

Surface roughness over all length scales might affect adhesion and other contact properties, but is poorly known down to the atomic length scale. Abhijeet Gujrati (University of Pittsburgh) measured the roughness of four diamond coatings over eight decades of length scales [**VII**, 13, 77], allowing us to test common assumptions on the nature of surface roughness and to apply adhesion theories on experiments performed with these samples [**V**, 62, 80].

Rigid solids Luke Thimons (University of Pittsburgh) measured the adhesion of these diamond coatings against ruby spheres [**V**, 78]. By comparing experimental results with numerical computations on the real surface topography, we investigate which length scale of roughness and which mechanisms critically affect adhesion in these rigid contacts.

Stiff solids The theory by Pastewka and Robbins [24] predicts that roughness at the atomic scale suppresses adhesion, while other theories emphasize the role of large-wavelength rough-

ness [26, 41, 81, 82]. Which lateral length scale is important is still under debate [70], because the range of validity of different adhesion theories is not clear. In publication [**VI**] we use finely discretized boundary element simulations to test Pastewka and Robbins' prediction of the gap distribution in stiff, weakly adhesive contacts. The gap distribution is a key element in their theory for the effect of adhesive interactions on the contact area and for their stickiness criterion.

Soft solids The role of elastic instabilities in adhesion hysteresis between soft materials is poorly understood because no theory or simulation model can describe this effect on realistic surface roughness. In publications [**III**] and [**I**], I use a crack-perturbation approach to develop an efficient coarse-grained model, enabling simulations representative of the experiments by Dalvi et al. [62].

From 3D to 1D in formulating adhesion model The boundary element method uses analytical solutions, the Green's functions, to compute the surface displacements caused by pressures without solving for the three-dimensional equilibrium equations of elasticity. The BEM thereby reduces the dimensionality of the computational problem from 3D to 2D, and the associated speed-up is crucial in the study of the contact between randomly rough surfaces. However, it is not sufficient to describe the adhesion of soft solids on realistic surface roughness.

My crack-front model provides the necessary speed-up by further reducing the dimensionality of the problem to 1D. It describes the position of the contact perimeter, instead of computing the displacements of the whole surface. Here, the crack-perturbation theory by Rice and coworkers [83–86] plays a role similar to the Green's function in the BEM. Rice's equations describe how roughness perturbs the energy balance along the crack-front, and the theory of the contact of smooth spheres by JKR [16] links the adhesion force to the contact area.

In summary, I use efficient computational methods based on the reduction of dimensionality of the 3D elastic problem. These reductions of dimensionality are derived mathematically for simple solids and are exact in the limit of small deformations of the halfspace (BEM) and of the crack front (crack-front model).

3 Modeling assumptions

Most theoretical and numerical models share a common set of basic assumptions. Here I define the main modeling assumptions and the generic contact system used in this thesis.

I describe the quasistatic process of squeezing a rigid indenter into an elastic substrate (indentation), and then pulling it out again (retraction). I consider only displacements normal to the surface and I neglect tangential deformations and stresses. The material is linear elastic and adhesive interactions are fully reversible. This model has hence no dissipation mechanism either in the bulk (such as viscoelasticity) or at the interface (such as friction or time dependent adhesion). I will show in section 7 that energy dissipation and irreversibility in an indentation-retraction cycle can still arise due to elastic instabilities.

The geometry of the rigid indenter is described by the heights $h(x, y)$ (Fig. 3.1), where x, y are Cartesian coordinates in the plane of the undeformed substrate. Here, $h(x, y)$ contains the macroscopic geometry and the surface roughness. The positive direction for heights and surface displacements u point out of the rigid indenter, towards the elastic substrate. The gap between the surfaces is

$$g(x, y) = u(x, y) - (h(x, y) + b), \tag{3.1}$$

where b is the rigid body translation of the indenter.

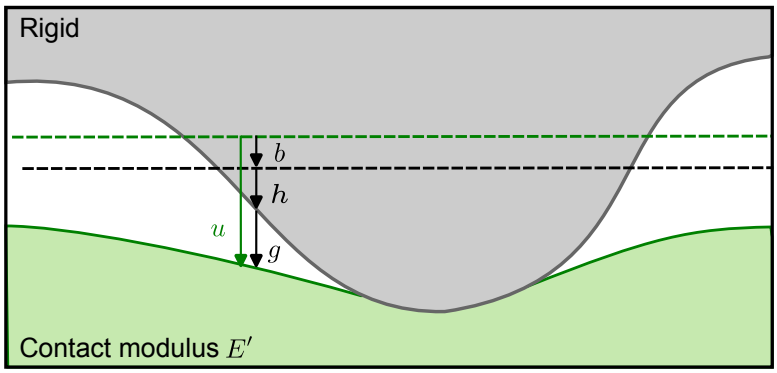


Figure 3.1: Schematic of a contact defining the kinematic relationship between the gaps g, the displacements u, the heights h and the rigid body translation b. The dashed lines represent the surface of the undeformed elastic solid (green) and the mean plane of the surface of the rigid indenter (black).

3.1 Nonadhesive contacts

Elasticity The deformable solid is an isotropic, linear elastic halfspace, for which Young's modulus E and Poisson's ratio ν quantify the stiffness of the material via the contact modulus $E' = E/(1-\nu^2)$ [87]. Linear elastic behavior is valid in the limit of small deformation gradients, but it is a common approximation even for surfaces with larger slopes [26] and describes the behavior of cracks surprisingly well [16]. These assumptions allow describing the contact using only displacements and stresses on the surface of the solid. The surface displacements u caused by the normal pressures σ can be computed by convolution with the Green function G,

$$u(x,y) = \int_A \mathrm{d}x'\,\mathrm{d}y' G(x-x', y-y')\sigma(x',y'). \tag{3.2}$$

Note that I use positive σ for compressive pressures, causing a positive displacement u of the surface towards the bulk. For periodic pressure distributions, this convolution reduces to the simple form [26, 88]

$$\tilde{u}(\vec{q}) = \frac{2}{E'}|\vec{q}|\tilde{\sigma}(\vec{q}), \tag{3.3}$$

where $\vec{q} = (q_x, q_y)$ is the wavevector and the tilde denotes the Fourier transform,

$$\tilde{u}(q_x,q_y) = \int_{-\infty}^{\infty} \mathrm{d}x\,\mathrm{d}y\, e^{-i(q_x x + q_y y)} u(x,y). \tag{3.4}$$

Note that since only the separation between the surfaces matter and the deformations are linear, the contact between two deformable solids with heights h_1 and h_2 can be mapped to the contact with a rigid indenter described here, using $h = h_1 + h_2$. The contact modulus of the elsatic substrate has to be replaced by $1/E' = (1-\nu_1^2)/E_1 + (1-\nu_2^2)/E_2$, where E_1, E_2 and ν_1, ν_2 are the elastic properties of the respective material [87]. In some of the experiments considered in this thesis [62], the most compliant solid is a soft rubber sphere that indents a rough diamond surface. The heights of the rigid indenter are then given by the superposition of the geometry of the sphere and the roughness of the diamond, and the composite modulus approximately corresponds the material properties of the soft sphere.

Equilibrium contact The static equilibrium is the minimum of the mechanical energy Π_{mech} under the constraint that the solids do not interpenetrate. In terms of pressures and displacements, this constraint leads to a linear complementarity problem [89], requiring that the conditions

$$\sigma(x,y) \geq 0,\ g(x,y) \geq 0,\ \sigma(x,y)g(x,y) = 0 \tag{3.5}$$

are satisfied simultaneously in every point (x,y). The region where the gaps are 0 defines the contact area A_c. The inequalities (3.5) require that pressures are compressive inside A_c and vanish outside A_c. The mechanical energy Π_{mech} contains the elastic strain energy

$$U_{\mathrm{el}} = \frac{1}{2}\int \mathrm{d}x\,\mathrm{d}y\,\sigma(x,y)u(x,y) \tag{3.6}$$

and the potential of the applied forces. In this thesis I always prescribe a rigid body penetration b, in which case $\Pi_{\mathrm{mech}} = U_{\mathrm{el}}$.

3.2 Adhesive contacts

Adhesive interactions reduce the energy of the system for small interfacial separations. In our theoretical work, we do not address the physical nature of the surface interaction and assume that these interactions are reversible. We either model adhesion with a generic stress distance relationship (cohesive law) or with a surface energy. Note that while the van der Waals interaction acts over the volume of the contacting bodies, it reduces to a cohesive law within small slope approximations (Derjaguin approximation [90]).

Cohesive law The energy due to adhesive interactions is given by the interaction potential v depending on the local gap between the surfaces,

$$\Pi_{\mathrm{adh}} = \int_{-\infty}^{+\infty} \mathrm{d}x\,\mathrm{d}y\; v(g(x,y)). \tag{3.7}$$

In publications [**V**] and [**VI**], we use the exponential interaction potential,

$$v(g) = -w_{\mathrm{int}} \exp(-g/\rho), \tag{3.8}$$

which is commonly used in numerical simulations [55] for its mathematical convenience and does not represent any specific physical interaction. The length ρ is the range of interaction and w_{int} is the work of adhesion, the energy per surface area gained by bringing the surfaces into contact. The derivative of this potential v' gives the cohesive law determining the pressures outside the contact area

$$\sigma(g) = -v'(g) = -(w_{\mathrm{int}}/\rho) \exp(-g/\rho). \tag{3.9}$$

The adhesive traction is maximal at the contact edge and equals $\sigma_0 = v'(0) = w_{\mathrm{int}}/\rho$ (Fig. 3.2).

The static equilibrium is now given by minimizing the total energy $\Pi = \Pi_{\mathrm{mech}} + \Pi_{\mathrm{adh}}$ under the constraint that $g \geq 0$. Note that now pressures also act outside the contact area and can be tensile in the contact area, however, they cannot exceed $-\sigma_0$. Formally, the non-overlap constraint can be described by an infinitely stiff repulsion in the interaction potential. Many numerical studies with adhesion [24, 53, 61, 65, 70, 91–96] and some theoretical models [82, 97] use a potential with a repulsion of finite stiffness to prevent the solids from overlapping.

Short-ranged limit Adhesive stresses are usually concentrated in a region around the contact perimeter called the cohesive zone. The tensile adhesive stresses stretch the solid at the contact edge (Fig. 3.2), and when these elastic deformations are large compared to the range of interaction, the cohesive zone becomes insignificant so that the perimeter corresponds to a brittle crack. The contact properties no longer depend on the range and shape of the interaction

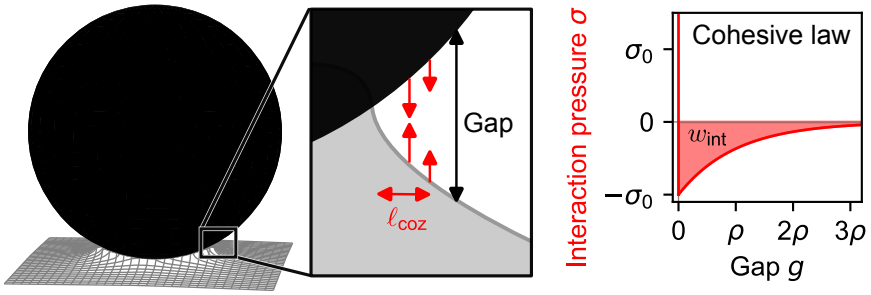


Figure 3.2: The cohesive law gives the relationship between the adhesive interaction pressure σ and the local gap g, and the integral over the cohesive law is the work of adhesion w_{int} (shaded area). For the exponential interaction Eq. (3.9), the maximum adhesive traction $-\sigma_0$ occurs at $g = 0$. When large σ_0 stretch the solid at the edge of the contact and the interaction range ρ is small, adhesive interactions act over a narrow cohesive zone of width ℓ_{coz}, so that the contact corresponds to a brittle crack.

potential so that the adhesive energy only depends on the work of adhesion and the contact area A_c,

$$\Pi_{\mathrm{adh}} = -\int_{A_c} \mathrm{d}x\,\mathrm{d}y\; w_{\mathrm{int}}. \tag{3.10}$$

This description of adhesion corresponds to Griffith's [27] formulation of linear elastic fracture mechanics, where the equilibrium contact area is given by a balance of elastic and surface energies gained per unit contact area. This framework is the basis of my crack-front model [**I**, **III**] and will be discussed in more details in section 7.

Transition between long-ranged and short-ranged adhesion The contact behaves as a brittle crack when the interaction range is short compared to the elastic opening at the edge of the contact. For the contact of spheres this ratio is given by the Tabor parameter [98–101]

$$\mu = \frac{(R w_{\mathrm{int}}^2 / E'^2)^{1/3}}{\rho}, \tag{3.11}$$

stating that soft spheres with large radius R are described by linear elastic fracture mechanics. Note that there exist equivalent criteria based on the width of the cohesive zone or a ratio of stresses called Maugis parameter [102].

For $\mu \gg 1$, the contact is described by the crack-front theory by Johnson Kendal and Roberts (JKR), and for $\mu \ll 1$, adhesive stresses do not cause any elastic deformation, as assumed in the theory of Derjaguin Muller and Toporov (DMT). The concept of Tabor parameter has been generalized to rough surfaces [40, 53], where the radius represents the curvature of an asperity. For any contact geometry, I call the short-ranged limit the JKR limit and the long-ranged limit the DMT limit.

3.3 Boundary element method

The boundary element method (BEM) implements a discretized version of the physical system introduced above with no additional assumptions. It therefore serves as a benchmark for analytical theories [**VI**] and for more coarse-grained simulations models such as my crack front model [**I**, **III**]. The implementation for nonadhesive contacts is available to use in the *contact.engineering* platform and summarized in [**II**]. I accelerate the computation of the elastic response of the material using the fast Fourier transform (FFT) [59, 60, 103]. In order to make use of the fast Fourier transform for nonperiodic contacts, I introduce a padding region that decouples periodic images [94, 104, 105]. The Green's function for nonperiodic contacts is provided in Refs. [87, 106]. I solve for the static equilibrium using constrained minimization algorithms [2–4]. Note that in publications [**III**] and [**I**], I model the JKR limit of vanishing range of interaction, and the cohesive law should be considered as a regularization of the surface energy defined by Eq. (3.10).

4 Roughness

4.1 Surface metrology

Most natural surface are rough over many decades of lateral length scales, and one of the major challenges impeding the prediction of contact properties is the actual measurement and statistical characterization of rough surfaces. Publication [**VII**] characterizes four diamond coatings over an unprecedented range of length scales, from atomic (nm) to mm.

This comprehensive surface topography characterization requires to combine several measurement techniques, because each individual technique can only resolve a limited range of scales (or bandwidth). Abhijeet Gujrati and coworkers measured the large-scale topography using a stylus profilometer (Fig. 4.1a), the medium-sale topography using an atomic force microscope (AFM) (panel b) and the small-scale topography using transmission electron microscopy (TEM) (panel c). Stylus and AFM are based on a mechanical probe scanning the surface. The resolution of these instruments is limited by tip artifacts, because the finite radius of the probe tip smears out the small-scale roughness. In order to resolve surface roughness down to the atomic scale, my experimental collaborators extract profiles from TEM images of side-views or cross-sections of the rough surfaces, allowing to resolve surface roughness down to atomic scales. Details on the measurement procedure are given in Refs. [**VII**, 13, 39].

TEM only provides one-dimensional (1D) profiles of the surface. While AFM and some stylus profilometers allow constructing areal scans of the surface (Fig. 4.2), the data perpendicular to the fast-scan direction has artifacts due to instrument drift. Here I assume that the surfaces are isotropic, so that 1D-profiles yield a complete statistical characterization of the surface topography.

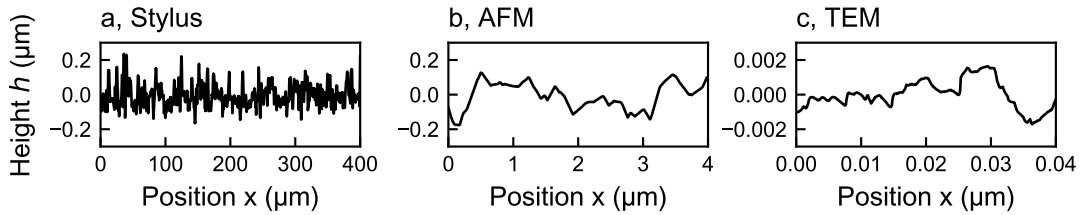


Figure 4.1: Height profiles of nanocrystalline diamond (NCD) at different lateral length scales obtained from three different measurement techniques: (a) stylus profilometry, (b) atomic force microspopy (AFM), and (c) transmission electron microscopy (TEM). The width of the profile decreases by a factor of 100 from instrument to instrument. Note that plots have aspect rations different from unity and that the height axis for TEM is rescaled by a factor of 100 compared to AFM and stylus.

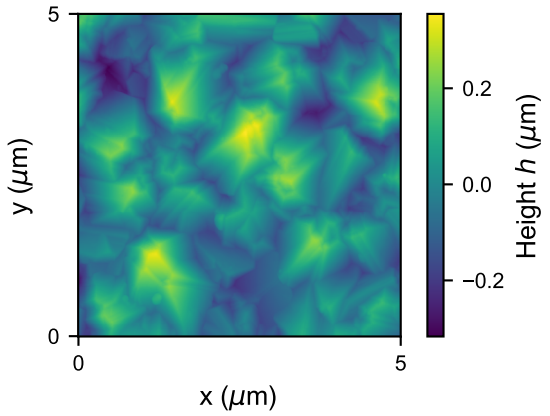


Figure 4.2: Two-dimensional height-map measured with AFM, revealing faceted grains of size $\simeq 1$ μm.

4.2 Multiscale statistical description of surface roughness

Surface topography is commonly described by scalar roughness parameters such as the root-mean-square (rms) fluctuations of heights h_{rms}, slopes h'_{rms} or curvatures h''_{rms}. On NCD, the amplitudes of the heights decreases by two orders of magnitude between the AFM scan of width 4 μm (Fig: 4.1b) and the TEM scan of smaller width 0.04 μm (Fig. 4.1c). Because h_{rms} depends on the width of the measurements, this single scalar value is not representative of the whole surface topography, so that we need multiscale statistical measures for a complete description. In our work, we use the variable bandwidth method (VBM), the autocorrelation function (ACF) and the power spectral density (PSD), which are established multiscale measures, and also introduce an additional framework termed *scale-dependent roughness parameters* (SDRP) [**IV**]. These methods are all available to use on *contact.engineering* and described in publications [**II**] and [**IV**]. I provide a short introduction below.

Reporting h_{rms} as a function of the measurement width L yields the variable bandwidth method (VBM) [**VII**, 13, 31]. The VBM can also be carried out on a single measurement, by subdividing it into smaller sections and computing h_{rms} for each of these sections [107–111].

Similarly to the VBM, the height-difference autocorrelation function A describes how the height fluctuations increase over lateral distances [79]. It is the variance of the difference between heights measured at two points separated by the distance ℓ,

$$A(\ell) = \frac{1}{2}\left\langle \left[h(x+\ell) - h(x)\right]^2 \right\rangle, \tag{4.1}$$

where the angular brackets indicate spacial averaging. The height-difference autocorrelation function is related to the height-height autocorrelation function $\left\langle h(x)h(x+\ell)\right\rangle$ via:

$$A(\ell) = h_{\mathrm{rms}}^2 - \left\langle h(x)h(x+\ell)\right\rangle. \tag{4.2}$$

We mainly use the height-difference autocorrelation function, which we denote autocorrelation function or ACF for brevity.

The power spectral density (PSD) represents the amplitude of the roughness in frequency (Fourier) space [14, 21]. For a 1D-profile of length L, it is defined as

$$C^{\mathrm{1D}}(q_x) = L^{-1}|\tilde{h}(q_x)|^2, \tag{4.3}$$

where $\tilde{h}(q_x)$ is the Fourier transform of the height profile and q_x is the wavevector. The wavelength $\lambda = 2\pi/q_x$ encodes a notion of lateral length scale. The Wiener-Khinchin theorem states that the PSD is the Fourier transform of the height-height autocorrelation function, so that ACF and PSD are mathematically equivalent.

All these measures are straightforward to generalize to areal scans. In this work we assume that the surfaces are isotropic, where only the radial average of these quantities is important. These radially averaged (isotropic) multiscale roughness measures can be obtained from their counterpart computed on 1D profiles. The ACF computed on profiles is identical to the radially averaged 2D-ACF, but the radially averaged 2D-PSD, C^{iso}, and C^{1D} are linked via a convolution [14]. This convolution is difficult to carry out on experimental data because it has a singular kernel. For this reason, we use the approximation

$$C^{\mathrm{iso}}(q) = \frac{\pi}{q} C^{\mathrm{1D}}(q), \tag{4.4}$$

which is strictly valid only for self-affine topographies, where the PSD decays as a power-law (see next section).

Using the PSD to characterize surface topography involves some technical difficulties. Because real measurements are not periodic, they must be multiplied with a window function [14, 112, 113] before taking the Fourier transform. Another problem is that the value C^{1D} has no direct geometrical meaning, and its exact definition differs among researchers [14]. Furthermore, there is no consensus about which definition of the PSD to use, making the comparison of PSDs reported in different publications cumbersome. In our work, we use the conventions established in Ref. [14] to define the Fourier transform and the PSD. Multiscale measures with direct geometrical interpretation such as the ACF and the SDRPs introduced below have the advantage that there is no ambiguity in their definition and that they do not require preliminary windowing of the data [**IV**, 79].

4.3 Self-affine surfaces

The scale-dependence of the rms heights illustrated by the measurements in Fig. 4.1b,c reflects the fractal nature of surface roughness [21, 31–33, 38]. In the special case of *self-affine* surfaces, the rms heights vary with scan size following the power-law $h_{\mathrm{rms}}(L) \propto L^H$, where $H \in (0, 1)$ is the Hurst exponent. A well-known example of self-affine profile is the random walk, for which $H = 0.5$. In the ACF, self-affine scaling implies that $A(\ell) \propto \ell^{2H}$ [114]. For the PSD, $C^{\mathrm{1D}}(q) \propto q^{-1-2H}$ and $C^{\mathrm{iso}} \propto q^{-2-2H}$ [14, 21].

Natural and technical surfaces are often self-affine with $H \simeq 0.8$ below the rolloff wavelength λ_r, above which the profile is random noise [33, 37, 38]. Above λ_r, the rms height is no longer scale-dependent and the PSD and ACF are flat. It is usually assumed that the surface is self-affine down to atomic scales.

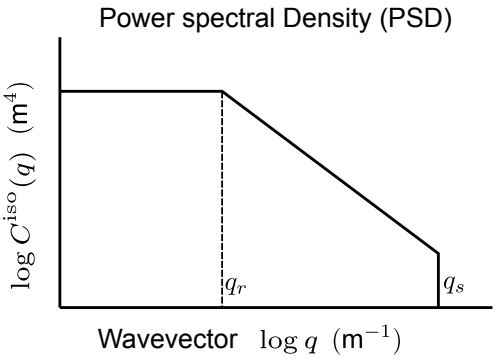


Figure 4.3: Idealized self-affine power spectral density with rolloff wavevector $q_r = 2\pi/\lambda_r$ and short-cutoff wavevector $q_s = 2\pi/\lambda_s$.

Theoretical work, including that of this thesis, idealizes surface roughness as self-affine with Gaussian height distribution. The surface is defined by an isotropic PSD which is flat for wavevectors below $q_r = 2\pi/\lambda_r$, and a power-law for wavevectors between q_r and the short-wavelength cutoff q_s (Fig. 4.3). In numerical work, I generate random realizations of surfaces described by this PSD using the Fourier filtering algorithm described for example in Refs. [14, 15]. This algorithm constructs the height map $h(x, y)$ by a superposition of sine waves with random uncorrelated phases, and amplitudes scaled according to the PSD.

4.4 Scale-dependent roughness parameters

While the value $C^{1\mathrm{D}}$ has no direct physical interpretation, integrals of the PSD yield statistical roughness parameters with clear geometrical meaning. By virtue of Parseval's theorem, the rms height is

$$h_{\mathrm{rms}}^2 = \frac{1}{L}\int_0^L \mathrm{d}x\, h(x)^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}q\, C^{1\mathrm{D}}(q). \tag{4.5}$$

The rms derivatives of order α can be computed similarly via higher moments of the power-spectrum [14, 115] :

$$(h_{\mathrm{rms}}^{(\alpha)})^2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}q\, q^{2\alpha} C^{1\mathrm{D}}(q). \tag{4.6}$$

It is straightforward to generalize Eq. (4.6) to 2D surfaces. For isotropic surfaces, the rms heights and rms curvatures obtained from $C^{1\mathrm{D}}$ or $C^{2\mathrm{D}}$ are identical. Concerning first derivatives, it is important to distinguish between the one-dimensional rms slope $\sqrt{\langle(\partial h/\partial x)^2\rangle}$ and the rms gradients $\sqrt{\langle|\nabla h|^2\rangle} = \sqrt{\langle(\partial h/\partial x)^2 + (\partial h/\partial y)^2\rangle}$, the latter being the quantity used in contact mechanics theories. For isotropic surfaces, the rms gradient is the 1D rms slope multiplied by a factor of $\sqrt{2}$. In publication [**IV**] and in this section, we use the notation h'_{rms} for slopes, but h'_{rms} usually denotes rms gradients.

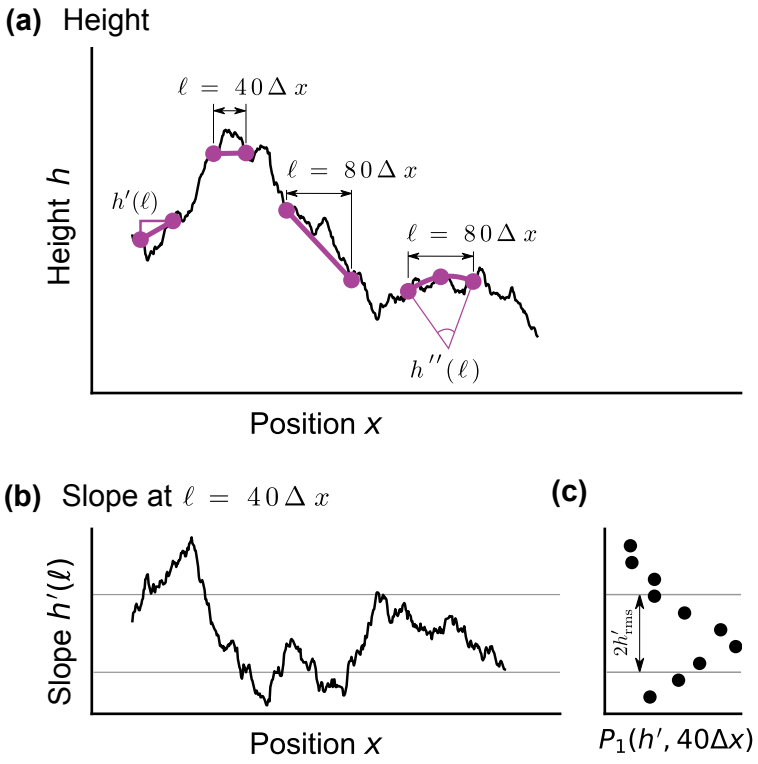


Figure 4.4: Illustration of the basic idea behind the scale-dependent roughness parameters. **(a)** Example line scan showing the computation of slopes $h'(\ell)$ and curvatures $h''(\ell)$ from finite differences. A scale can be attached to this computation by computing these finite differences at different distances ℓ, shown for $\ell = 40\Delta x$ and $\ell = 80\Delta x$ where Δx is the sample spacing. Similarly, the curvature at a finite scale ℓ is given by fitting a quadratic function through three points spaced a distance $\ell/2$. **(b)** Local slope, obtained at a distance scale of $\ell = 40\Delta x$ for the line scan shown in panel (a). The slope is defined for each sample point since we can compute it for overlapping intervals. **(c)** Probability density P_1 of the local slope, obtained from the slope profile shown in panel (b). The rms slope for this length scale is equal to the width of this probability distribution. The figure is reproduced from publication [**IV**] (CC BY 4.0).

Persson [26, 53] introduced a notion of scale-dependence in these rms derivatives by restricting the limits of integration in Eq. (4.6) to a limited range of wavevectors. Persson's theories for contact mechanics are formulated in Fourier space and describe how contact properties evolve as Fourier modes (sine waves) with smaller and smaller wavelength are added to the roughness. The PSD and these scale-dependent roughness parameters play a central role in his contact theories.

4.4.1 Scale-dependent finite differences

In publication [**IV**], we propose a more direct and intuitive way to define scale-dependent roughness parameters. The simplest definition of the derivative uses the finite difference scheme $(h(x+\Delta x) - h(x))/\Delta x$, where Δx is the sample spacing of the measurement. Wang and Müser [79] introduced a notion of scale by applying the scheme on a sample spacing $\eta\Delta x$

(with integer η),

$$h'(x, \eta\Delta x) = \frac{h(x + \eta\Delta x) - h(x)}{\eta\Delta x} \tag{4.7}$$

and Fig. 4.4 illustrates this idea. Equation (4.7) filters out contributions of small-scale roughness to the slope by assuming that the slope is constant over the interval $[x, x + \eta\Delta x]$ and given by the line passing through the two endpoints.

The finite difference scheme for the second derivative fits a second-order polynomial through three sample points and is given by

$$h''(x, \ell) = \frac{h(x + \eta\Delta x) - 2h(x) + h(x - \eta\Delta x)}{(\eta\Delta x)^2}, \tag{4.8}$$

where ℓ is the distance scale. Because Eq. (4.8) implies that the curvature is constant over the stencil length $2\eta\Delta x$, we define the distance scale as $\ell = \alpha\eta\Delta x$, with α the order of the derivative.

The rms fluctuations of these scale-dependent derivatives yield the *scale-dependent roughness parameters* (SDRPs), for example for the slope:

$$h'_{\mathrm{SDRP}}(\ell) = \left\langle \left[\frac{h(x + \ell) - h(x)}{\ell} \right]^2 \right\rangle^{1/2}. \tag{4.9}$$

4.4.2 Relationship with other multiscale measures

The scale-dependent rms derivatives can be computed from the autocorrelation function. For the first derivative, it is apparent from Eqs. (4.1) and (4.7) that [79]

$$h'_{\mathrm{SDRP}}(\ell) = \left[2A(\ell)\right]^{1/2} / \ell. \tag{4.10}$$

Derivatives of higher order can also be computed from the ACF, but require evaluating the ACF in more than one point. Our scale-dependent derivatives are also similar to a generalized form of the variable bandwidth method [116, 117]. While the finite-difference scheme uses only two or three points to fit a polynomial, the generalized VBM makes use of all data points over the length ℓ via a least-squares regression.

Our numerical tests on synthesized self-affine surfaces show that computing scale-dependent derivatives from finite-differences (or the ACF), the VBM or the PSD yields similar results. In comparing the Fourier derivative with other methods, the exact relationship between the distance scale ℓ and the cutoff wavelength λ_c setting the integration bounds in Eq. (4.6) is not evident. Analyzing the finite difference operators in Fourier space shows that the wavelength that contributes the most to $h^{(\alpha)}_{\mathrm{rms}}$ is $2\ell/\alpha$. For the Fourier derivative, the dominating wavelength is the cutoff wavelength λ_c, so that we chose $\lambda_c = 2\ell/\alpha = 2\eta\Delta x$. For the first derivative, this corresponds to Nyquist's sampling theorem stating that the smallest wavelength that can be resolved on a discrete grid is $\lambda_c = 2\Delta x$. In summary, we have shown in publication [**IV**] that scale-dependent derivatives can be obtained from finite difference schemes,

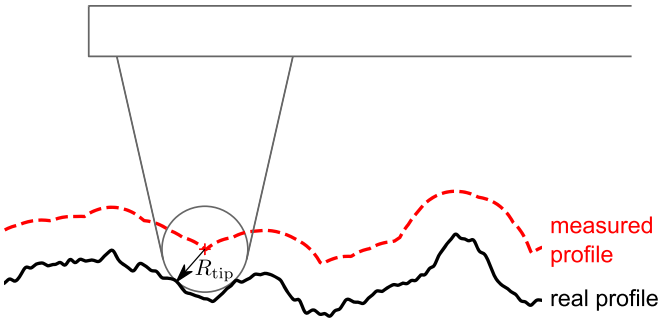


Figure 4.5: Tip-based methods scan the surface with a probe with finite radius R_{tip}. This probe cannot fit into small valleys of the surface and smoothens the curvature of the peaks in the measured profile. The figure is reproduced from publication [**II**] (CC BY 4.0).

the ACF, the generalized VBM or the PSD, and that these four definitions yield equivalent results on self-affine surfaces. However, this relationship is not universal and breaks down in the rolloff region [79], so that future work should compare these methods in other situations such as surfaces with facets (see Fig. 4.2 and section 4.5) or steps.

4.4.3 Scale-dependent distributions of derivatives

The scale-dependent finite differences allow us to extract the full distribution of slopes as a function of length scale (Fig. 4.4c). Established multiscale roughness measures only give information on the variance of heights, slopes and curvatures over scale, and it is commonly assumed that the underlying distribution is Gaussian. We can quantify deviations from Gaussianity for each length scale ℓ by computing higher moments or cumulants (or combinations of them, such as the skewness or kurtosis) of the distribution of the scale-dependent derivative $h'(x, \ell)$ given by Eq. (4.7). In contrast, the PSD, ACF and VBM only contain information about the variance (i.e. the second moment) of this distribution.

4.4.4 Application to tip-artifact detection

An important practical application of the scale-dependent curvature distribution is the detection of tip artifacts. Tip-based methods scan the surface with a probe with finite radius R_{tip} (Fig. 4.5), so that roughness features with curvature larger than $1/R_{\mathrm{tip}}$ cannot be resolved, affecting the measured profile at small lateral length scales. Note that the lateral length scales affected by this artifact depends on the surface topography as well as on R_{tip}. We determine the critical lateral length scale by comparing the maximum of the scale-dependent curvature distribution to the tip radius. Our method is more robust than an alternative method based on the PSD [14], enabling us to automatically remove data with tip artifacts in *contact.engineering* [**II**].

4.5 Roughness of polycrystalline diamond coatings

Diamond coatings are used in various contexts such as microelectromechanical systems [118] (MEMS), biomedical applications [119], and seals and bearings [120, 121]. Surface roughness critically affects their performance, but the nature of surface roughness is rarely known from macroscopic scales down to atomic scales. In publication [**VII**], we characterize the roughness of four different polycrystalline diamond coatings: microcrystalline diamond (MCD), nanocrystalline diamond (NCD), ultrananocrystalline diamond (UNCD), and polished UNCD (pUNCD). These four substrates enable adhesion experiments varying roughness at identical surface chemistry [**V**, 62], and our characterization of topography from millimeters down to atomic distances allows us to quantitatively compare theory and experiment.

Such a comprehensive surface topography description required Abhijeet Gujrati and coworkers to perform multiple measurements using three different techniques, because each instrument only resolves a limited range of length scales, see Sec. 4.1. We build a comprehensive description of the surface topography by stitching together and averaging the PSDs obtained from more than 60 individual measurements for each substrate. Figure 4.6 shows this averaged PSD for UNCD and NCD, as well as the ACF, scale-dependent slopes and scale-dependent curvatures obtained by the same procedure.

The power-spectrum has the characteristic transition between fractal scaling (below $\simeq 1\ \mu\mathrm{m}$) and a rolloff region where the PSD is nearly flat, characteristic of random noise. Similar scaling regions can be seen in the height-difference autocorrelation function. The height fluctuations increase over distance until the rolloff length, where $\sqrt{A}$ becomes independent of scale and gives an intrinsic value for the rms height. Hence, it is important to measure the surface over distances larger than the rolloff length in order to determine the rms height. In contrast, the rms slopes and curvatures increase as we account for progressively smaller length scales by decreasing the spacing of the stencil ℓ. This means that instruments with nanometer resolution such as TEM are required to determine an intrinsic value of slopes and curvatures. Note that while NCD is rougher than UNCD at large length scales, the small-scale roughness of the two surfaces is similar. Since the contact area and the adhesion of stiff elastic materials depend on small-scale roughness, these coatings might have similar tribological performance despite they look so different at large scales.

The PSD below the rolloff cannot be described by a single power-law, so that these surfaces deviate from the common self-affine idealization presented in section 4.3. Self-affinity implies that the PSD scales as q^{-1-2H}, where H is the Hurst exponent and is in the range $[0, 1]$. NCD can be considered self-affine below $\lambda = 37\ \mathrm{nm}$, where a power-law with $H_{\mathrm{smaller}\lambda} = 0.75$ reasonably fits the PSD. In contrast, for $\lambda \in [37\ \mathrm{nm}, 224\ \mathrm{nm}]$, the power spectrum is much steeper, giving an apparent Hurst exponent of $H_{\mathrm{larger}\lambda} = 1.27$. This larger-than-expected exponent implies that NCD is not self-affine in this intermediate range of length scales.

This steep scaling of the PSD is a signature of the faceted grains of NCD that are visible in Fig. 4.2 and introduces kinks in the profiles (Fig. 4.1b). Church and Takacs [122] have mathematically shown that kinks lead to q^{-4} scaling of the PSD, corresponding to an apparent Hurst exponent $H = 1.5$ very similar to the fitted value $H_{\mathrm{larger}\lambda} = 1.27$. Our numerical tests on artificially generated profiles indicate that the wavelength of transition from q^{-4}-scaling to the rolloff corresponds to the mean grain size, and Ref. [123] made a similar observation

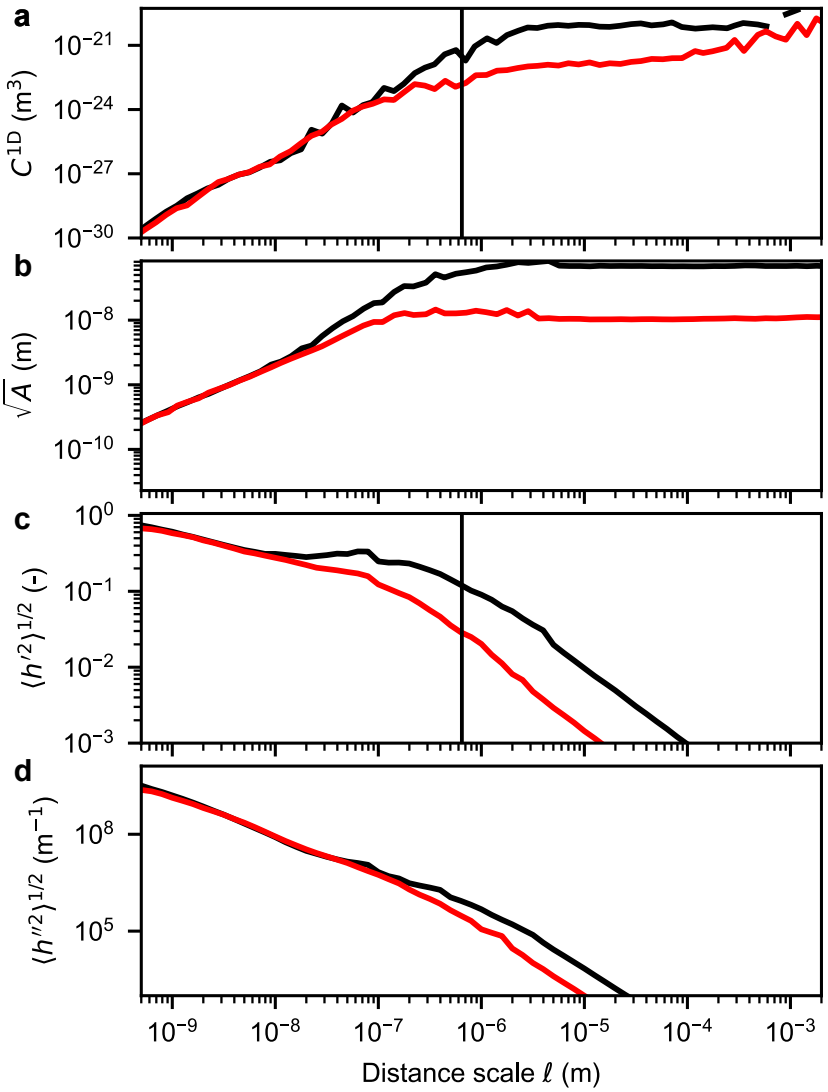


Figure 4.6: Multiscale roughness measures for nanocryscalline diamond (NCD) (black lines) and ultra-nanocrystalline diamond (UNCD) (red lines), as a function of the distance scale ℓ. **(a)** One-dimensional power spectral density C^{1D} (PSD), where the wavelength $\lambda = 2\ell$. The vertical line represents the grain size of NCD. **(b)** Autocorrelation function. **(c)** Scale-dependent rms slopes computed with a finite difference scheme. **(d)** Scale-dependent rms curvatures computed with a finite difference scheme.

on sandstones. The grain size extracted independently on AFM scans (as indicated by the vertical line in Fig. 4.6a) is similar to the rolloff wavelength, confirming our numerical result experimentally.

Faceting has a clear signature in the scale-dependent slope. Crystal facets are flat so that when the spacing between sample points is smaller than the grain size, the measured slopes become independent of ℓ and the curve flattens. The slope rises again for ℓ below 36 nm, because the grain facets are not perfectly flat and have self-affine surface roughness on top of them.

5 Adhesion of rigid spheres

5.1 Introduction

Our goal is to understand how roughness affects the force required to separate two macroscopic solid objects. In publication [**V**], we address this question experimentally for the contact between some of the hardest materials: ruby and diamond. These materials are so stiff that they are effectively rigid and only touch on the top of the tallest asperity. Adhesion studies on hard materials commonly use AFM or colloidal probes and are sensitive only to the nanoscale roughness because of the small radius of the probe [124–126]. By using a spherical probe of 0.5 mm diameter, Luke Thimons' experiments allow us to understand the role of surface roughness up to the macroscopic scale and are representative of technological application.

5.2 Experiment

In these experiments, Luke Thimons brought the ruby sphere into contact by applying a preload of 5 μN and then measured the force needed to detach the sphere. He repeated the experiment more than 2000 times for each of the diamond coatings characterized in publication [**VII**], yielding the distributions of pull-off forces shown in Fig. 5.1 for pUNCD and NCD. The pull-off forces decrease with surface roughness and have large fluctuations approximately described by a log-normal distribution. Similar adhesion distributions were observed in various contexts including particle adhesion [127–130], biological samples and cell adhesion [131, 132], and AFM adhesion [133–135]. Simple analytical models [51, 136] cannot

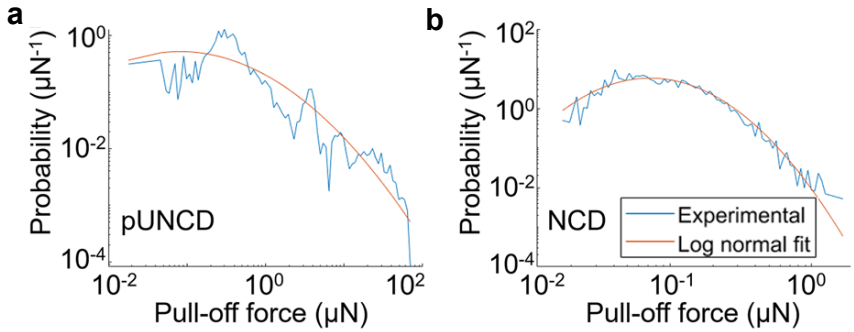


Figure 5.1: Probability distribution of pull-off forces in the contact between a ruby sphere and the diamond coatings pUNCD (a) and NCD (b). Luke Thimons repeated the experiments more than 2000 times for each substrate. The red lines show the fit to a log-normal distribution. The figure is adapted from publication [**V**] (CC BY 4.0).

explain the pull-off force of these materials.

5.3 Numerical model

Luke Thimons and I carried out brute-force numerical calculations on these topographies in order to unravel the role of surface roughness and to identify the main mechanisms. The geometry is given by areal AFM scans of the ruby sphere and of the four diamond substrates. Boundary element simulations of the elastic contact show that adhesive interactions are too weak to deform the material and that at the preload, most of the contact area is plastified. AFM measurements after contact reveal indents that were not present in the same region before contacting, confirming that ruby was deformed plastically. These plastic deformations let the highest peaks of diamond penetrate into the ruby sphere and thereby increase the pull-off force compared to an elastic or purely rigid contacts. We therefore determined the pull-off force by integrating the interaction pressures in a rigid contact, but where the highest peak were initially flattened by plastic deformation. We accounted for plasticity using a simple bearing-area approach where the contact stresses equal the hardness in the plastified area [137], and modeled the adhesive interactions using an exponential cohesive law (Eq. (3.9)).

5.4 Work of adhesion and range of interaction

The cohesive law is parametrized by the work of adhesion w_{int} and the interaction range ρ. We determined these two unknowns by fitting our numerical simulations to the experimental results for the four different rough surfaces. This approach is valid because the four coatings are chemically identical. For $\rho \simeq 5$ nm and $w_{\mathrm{int}} \simeq 46$ mJ/m^2, our numerically predicted average pull-off forces agree with the experiments for all four topographies. The work of adhesion is consistent with findings by Refs. [138–140] and realistic for van der Waals adhesion. However, the interaction range is higher than theoretically expected for atomic interactions such as covalent bonds or van der Waals interactions, which typically act over distances around 0.5 nm [19]. Refs. [140–142] reported similar values for ρ, but the origin for this large range of interaction is still unclear. Possible explanation are electrostatic interactions [143–145], capillary adhesion [146–148] and Casimir forces [141].

5.5 Role of scales

Plastic deformation and the long range of interaction make the pull-off force insensitive to small-scale roughness. We carried out an additional numerical study to identify which scales of roughness affect the normal force. Because an individual AFM scan captures only a limited range of length scales, we extended this surface by superposing synthetic roughness at large and small scales. This synthetic surface roughness was generated using a Fourier filtering algorithm [14, 15] based on the power-spectral density of the specific diamond substrate. The advantage of using such a compound topography instead of a fully synthetic one is that the

AFM scan covers the range of scales where NCD and MCD are faceted. Over these length scales, they differ the most from our synthetic topographies at equal PSD.

Filtering out small-scale roughness with an increasing cutoff wavelength reveals that the features below 43 nm do not affect the pull-off force (Fig. 5.2a). Small scales do not matter because they are flattened by the plastic deformation in the contact area, and because the adhesive interaction is insensitive to gap variations smaller than the interaction range [149]. The role of the interaction range can be understood with the ACF shown in Fig. 4.6b, which characterizes variations in heights (or gaps in a rigid-plastic contact) as a function of lateral distance. For the large interaction range $\rho = 5$ nm that we found, height variations over lateral distances $\ell \lesssim 40$ nm are smaller than the interaction range and do not alter the attractive pressure. In contrast, for the value expected for van der Waals interactions $\rho \simeq 0.5$ nm [19], roughness down to the single digit nanometer scale would matter. Hence, plastic deformation and the larger-than-expected range of interaction make the contact insensitive to small-scale roughness.

By repeating this procedure to filter out large-scale roughness (Fig. 5.2b), we find that roughness above 1.8 μm is unimportant. The unimportance of large scales might be related with the radius of the sphere, but we would need to vary the sphere radius in order to check that hypothesis. A similar numerical study [130] found that the pull-off force is independent of the sphere radius above a certain amplitude of roughness.

5.6 Discussion

The simulations reveal in detail which parts of the surface come into contact and which parts feel attractive interactions. The pull-off force fluctuates a lot because the contact area is limited to the top of one or a few single asperities. For roughness larger than the interaction range, the force is determined by the local geometry of this peak. The adhesion of a spherical asperity depends on the curvature [28, 29], but on our surfaces the contact occurs on the corners of faceted grains and cannot be considered spherical. There is no theory linking pull-off forces to geometrical parameters for our faceted surfaces. In the case of pUNCD, the rms height is small compared to the range of adhesion, so that attractive forces in the regions far from the contact point contribute the most to the total force. The height of the contacted asperity determines the pull-off force by setting the average separation between the surfaces. In both cases, adhesion is now determined by the extreme value statistics of the highest peaks of the surface. Accounting for these finite-size effects in a model is an open theoretical challenge.

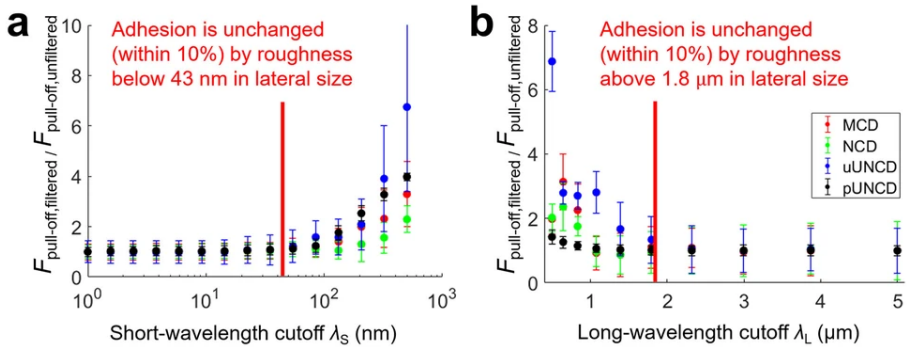


Figure 5.2: The contribution to pull-off force from various length scales can be directly demonstrated by recalculating pull-off force after filtering out small (a) and large (b) scales of roughness. Specifically, the pull-off force calculated from the filtered surfaces is normalized by the pull-off force calculated from the unfiltered surfaces. In panel (a), the x-axis indicates a short-wavelength cutoff, where all roughness below this size scale has been removed. A value near 1 indicates that there is almost no effect on pull-off force of filtering out roughness below that size scale. In panel (b), the x-axis indicates a long-wavelength cutoff, where all roughness above this size is removed. Here, a value of 1 indicates no contribution to pull-off force from roughness above that size scale. The figure is reproduced from publication [**V**] (CC BY 4.0).

6 Theory for the adhesion of stiff elastic contacts

6.1 Introduction

The contact of ruby and diamond coatings is not sticky because only a single asperity holds them together, and the micronewton pull-off force of this asperity is imperceptible. In this section, I consider elastic contacts that comply to the peaks over a finite fraction of the nominal area of contact A_0. This elastic deformation increases the area that falls into the range of molecular attraction, but the adhesive forces now also compete with the repulsive pressures inside the contact area A_{rep}. Pastewka and Robbins [24] showed that most objects are not sticky because the repulsive force increases faster with contact area than the attractive force. In publication [**VI**], Joe Monti and I numerically test the adhesion theory by Pastewka and Robbins [24]. We consider elastic contacts where adhesion is confined to the top of the highest asperities (see Fig. 1.1b), the relative contact area A_{rep}/A_0 is small ($\lesssim 10\%$), and the adhesive tractions are too weak to deform the surface (DMT approximation).

The Pastewka and Robbins model describes the contact between infinite nominally flat surfaces with self-affine roughness. The study of this infinite system allowed the authors to predict the contact area and adhesion of rough spheres via homogenization [94]. When the contact area of the sphere covers a statistically significant number of surface features, the theory of the infinite system gives the relation between mean pressure and contact area inside a representative surface element of the spherical contact. Reported over the nominal area of the system A_0, the pull-off force of a single asperity is vanishingly small, so that the surfaces are not sticky. The surfaces are only sticky when a finite mean pressure is required to separate them, allowing macroscopic spheres to stick with macroscopic forces. At a finite contact area fraction, stickiness depends on the competition between the attractive tractions outside the contact area and the repulsive pressures inside the contact area.

The theory by Pastewka and Robbins predicts how the contact area increases due to adhesive interactions and explains why most objects are not sticky despite the universal van der Waals attractions. However, these molecular attractions still play an important role in nonsticky contacts because they increase the contact area and thereby the friction force [19, 41, 43, 49].

6.2 Pastewka and Robbins model

For nonadhesive contacts with small contact area fractions $A_{\mathrm{rep}}/A_0 \lesssim 10\%$, the normal force F_{rep} is proportional to the contact area A_{rep},

$$F_{\mathrm{rep}}/A_{\mathrm{rep}} = \kappa_{\mathrm{rep}}^{-1} h'_{\mathrm{rms}} E', \tag{6.1}$$

where the proportionality constant $\kappa_{\mathrm{rep}}^{-1} \simeq 1/2$ [46–48]. For an adhesive contact, Pastewka and Robbins split the total force $F = F_{\mathrm{rep}} + F_{\mathrm{att}}$ in a repulsive contribution $F_{\mathrm{rep}} > 0$ and an attractive contribution $F_{\mathrm{att}} < 0$. In the DMT approximation, F_{rep} is still proportional to the area of repulsive contact A_{rep} and given by Eq. (6.1). They argue and verify numerically that F_{att} is proportional to A_{rep} as well, defining the constant $\kappa_{\mathrm{att}}^{-1} = -(F_{\mathrm{att}}/A_{\mathrm{rep}})/(h'_{\mathrm{rms}} E')$. The total force in an adhesive contact is hence proportional to the repulsive contact area and described by the constant

$$\frac{F}{A_{\mathrm{rep}} h'_{\mathrm{rms}} E'} = \kappa^{-1} = \kappa_{\mathrm{rep}}^{-1} - \kappa_{\mathrm{att}}^{-1}. \tag{6.2}$$

When $\kappa_{\mathrm{att}}^{-1}$ exceeds $\kappa_{\mathrm{rep}}^{-1} \simeq 1/2$, the attractive force grows faster with contact area than the repulsive force and the solids spontaneously come into contact. The contact is sticky because a finite mean pressure F/A_0 needs to be overcome to separate it.

The attractive force is the integral of the interaction stress outside the contact area, given by the local gap $g(x)$ via the cohesive law $-v'(g)$ (Eq. (3.9)). In a statistical description, this integral is given by the probability density of gaps $p(g)$,

$$F_{\mathrm{att}} = A_0 \int_{0^+}^{\infty} \mathrm{d}g \, p(g) \, (-v'(g)). \tag{6.3}$$

Using the DMT approximation, $p(g)$ is identical to the distribution of gaps in a nonadhesive contact, shown for a numerical simulation by the blue dots in Fig. 6.1. Before contact, the gap distribution equals the height distribution, which is Gaussian and corresponds to an inverted parabola in the log-log representation in Fig. 6.1. When we squeeze the surfaces together, the repulsive contact stresses flatten the roughness in and close to the contact area and thereby increase the probability of small gaps. The key element in Pastewka and Robbins' theory is an analytical expression approximating the gap distribution near the contact edge in a nonadhesive contact.

When deriving their approximation for the gap distribution, they assumed that the interaction range ρ is short, so that attractive interactions are confined to a narrow rim around the contact perimeter, symbolized in red in Fig. 6.2. They observed in their simulations that the contact perimeter is given by

$$P_{\mathrm{rep}} = \pi A_{\mathrm{rep}}/d_{\mathrm{rep}}, \tag{6.4}$$

with a constant d_{rep}, the typical width of a contacting region [48]. In contrast with compact two-dimensional objects like disks for which $P_{\mathrm{rep}} \propto A_{\mathrm{rep}}^{1/2}$, the contact perimeter is proportional to A_{rep} because the contact patches are fractals with many internal bubbles. This

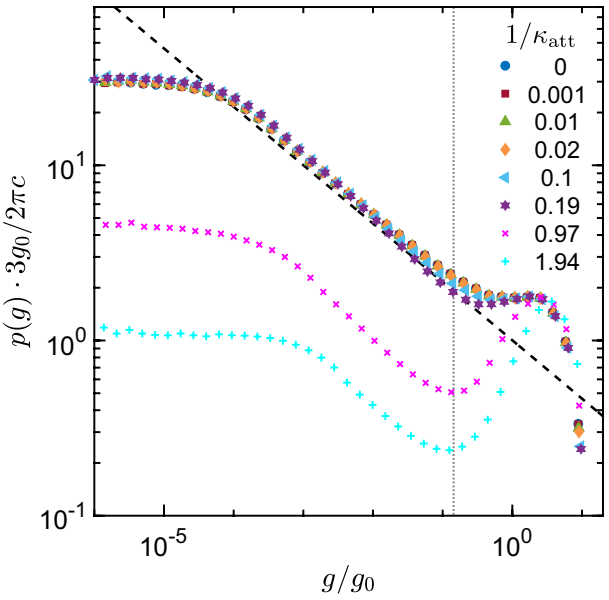


Figure 6.1: Distribution of gaps extracted from boundary element simulations of the contact of a self-affine surface. The strength of the adhesive interaction σ_0 increases with $1/\kappa_{\mathrm{att}}$, and $1/\kappa_{\mathrm{att}} = 0$ (blue circles) corresponds to the nonadhesive case. The power-law $p_n(g)$ from Eq. (6.8) is shown as a dashed black line. Note that we used the prefactor of Eq. (6.8) to normalize the y-axis, with $c = A_{\mathrm{rep}}/A_0$ the relative area of contact. For small gaps, the numerical data deviates from the prediction because of discretization artifacts, which are determined by the ratio between the smallest wavelength in the PSD λ_s and the pixel size Δx. Here, $\lambda_s/\Delta x = 128$, and the simulation was performed on a large grid of 65536 pixels in order to have a large enough region of self-affine scaling in the PSD, $\lambda_r/\lambda_s = 256$. The dotted gray line corresponds to $g/g_0 = \rho/g_0$ with $\rho = 4\Delta x$. The figure is reproduced from publication [**VI**] (CC BY 4.0).

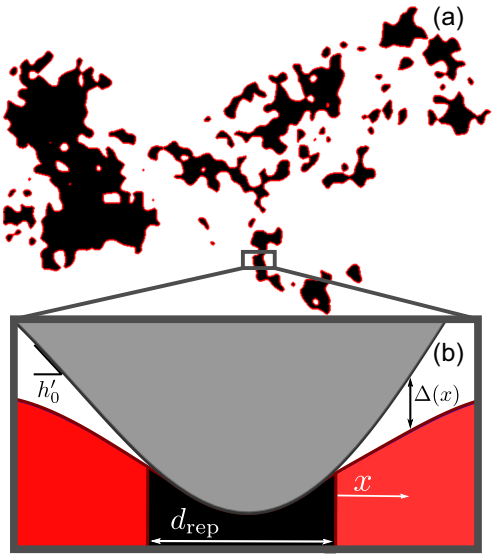


Figure 6.2: (a) Contact map for a self-affine surface with H = 0.8. The area of repulsive contact A_{rep} is the sum over all pixels shown in black, and the contact perimeter P_{rep} is marked in red. **(b)** Schematic cross-section of a contact region created by a contacting asperity, showing the mean contact diameter d_{rep}. The gap Δx between surfaces grows as the lateral distance $x^{3/2}$. The figure is reproduced from publication **[VI]** (CC BY 4.0).

porous, or snake-like contact morphology is the result of roughness at small lateral length scales preventing the surfaces to elastically conform. Pastewka and Robbins argue and verify numerically that the mean width of the contact region is determined by the small-scale geometry of asperities via

$$d_{\mathrm{rep}} \simeq 4h'_{\mathrm{rms}}/h''_{\mathrm{rms}}. \tag{6.5}$$

The width of the contacting regions determines the interfacial separation close to the contact edge. A cross-section through an asperity (Fig. 6.2b) resembles the contact of a cylinder where the gaps separate following the universal power law [87]

$$\Delta(x) = g_0 \left(\frac{x}{d_{\mathrm{rep}}} \right)^{3/2}, \tag{6.6}$$

with

$$g_0 = 4h'_{\mathrm{rms}} d_{\mathrm{rep}}/3 = \frac{16}{3} \frac{[h'_{\mathrm{rms}}]^2}{h''_{\mathrm{rms}}} \tag{6.7}$$

corresponding to how much the asperity penetrates into the substrate. Hence, not only the width of a contacting region is independent of the load, but also the deformed geometry near the perimeter. The area that feels attractive forces is then proportional to P_{rep} and hence to

A_{rep} via Eq. (6.4). More precisely, the probability density at small gap $p_n(g)$ increases linearly with A_{rep} and Eqs. (6.4) to (6.7) yield Pastewka and Robbins' prediction

$$p_n(g) = \frac{A_{\mathrm{rep}}}{A_0} \frac{2\pi}{3g_0} \left(\frac{g_0}{g} \right)^{1/3} . \tag{6.8}$$

Integrating $p_n(g)$ with the cohesive law gives the adhesive force F_{att} and the correction for the load area relationship

$$\kappa_{\mathrm{att}}^{-1} = -\frac{2\pi}{3} \frac{\sigma_0}{h'_{\mathrm{rms}} E'} \left(\frac{\rho}{g_0} \right)^{2/3} . \tag{6.9}$$

The nondimensional term $\sigma_0 / h'_{\mathrm{rms}} E'$ is the ratio between the maximal adhesive stress σ_0 and the repulsive stress in the contact area, and the term ρ / g_0 compares the range of interaction ρ to the local deformation of the substrate by an asperity g_0. Both the mean repulsive stress $h'_{\mathrm{rms}} E'$ and the deformation g_0 are determined by slopes and curvatures, which depend on small lateral length scales. In conclusion, the contact area and the stickiness are determined by the surface roughness at small lateral length scales.

6.3 Numerical validation of the Pastewka and Robbins model

Pastewka and Robbins' model accurately predicts the contact area observed in their weakly adhesive simulations. However, this agreement does not directly validate the individual assumption made on the morphology of contact patches and on the deformed geometry. Pastewka and Robbins verified that their expression for d_{rep} Eq. (6.5) describes the geometry of contact patches in their simulations. Concerning the deformed geometry, numerical results for gap distributions reported in [40, 55, 56, 150] do not resolve small gaps well enough in order to verify Eq. (6.8). In publication [**VI**], Joe Monti and I compare Eq. (6.8) to gap distributions extracted from finely discretized simulations (Fig. 6.1). The numerical results follow the predicted power-law (dashed line) for gaps below $0.1 g_0$. This means that the Pastewka-Robbins model is no longer valid for interaction ranges $\rho \gtrsim 0.1 g_0$ in the case of an exponential cohesive law (Eq. (3.9)).

Our simulations also show that the gap distribution is unaltered by weak adhesive interactions. In the adhesive simulations in Fig. 6.1, the distribution of gaps changes when $\kappa_{\mathrm{att}}^{-1}$ increases from 0.2 to 1, corresponding to where the surfaces become sticky. Müser [40] found that adhesion alters the contact patch morphology and the gap distribution when a generalized Tabor parameter exceeds one. This parameter also depends on the strength of the adhesive interaction and the small-scale roughness, but is not equivalent to $\kappa_{\mathrm{att}}^{-1}$. A more systematic parameter variation is required to conclude when Pastewka and Robbins' assumptions break down due to strong adhesive stresses.

6.4 Discussion

The importance of small-scale features of the surface roughness predicted by Pastewka and Robbins is still actively debated [70, 81, 151, 152], because other models predict that large scales dominate adhesion. The Pastewka-Robbins model is valid in an intermediate regime where the interaction range is shorter than g_0, but at the same time long-ranged enough for the cohesive stresses to be weak (at a fixed work of adhesion w_{int}). In both limits of weak attractive stresses with long range, and strong attractive stresses with short range, large-scale roughness dominates adhesion. For long interaction ranges, the Gaussian peak of the gap distribution at large gaps and the contact stiffness determine the adhesive force, and both quantities are governed by large-scale roughness [53, 81]. For short-ranged adhesion with strong cohesive stresses (JKR regime), the surfaces conform at small scales and the contact area is governed by the elastic energy required to conform to the large-scale roughness [26, 41]. Note that in the case $H < 0.5$ the small-scale roughness remains the hardest to conform to and determines the contact area, but most surfaces have $H > 0.5$ [33, 37, 38]. An additional complication is that the transition between the DMT regime and the JKR regime might occur within the same surface depending on length scale [53, 82].

The numerical studies of Refs. [24, 40] agree with the stickiness criterion $\kappa_{\mathrm{att}}^{-1} > 1/2$. In contrast, Wang and Müser [70] observed that the onset of stickiness is well described by the energetic criterion $w_{\mathrm{int}}/e_{\mathrm{el}} > 0.5$, where e_{el} is the elastic energy for fully conformal contact and usually depends on large length scales [26]. However, their results do not clearly contradict with Pastewka and Robbins either. A discussion of the impact of surface roughness on stickiness unifying different adhesion regimes is still lacking. Further analysis of multiscale contact theories [53, 82] might provide valuable insights.

It is hard to reach conclusions with experimental methods, because roughness is rarely known down to the atomic scale. The comprehensive surface topography characterization in publication [**VII**] might allow such an experiment in the future. Luke Thimons' experiments with ruby spheres [**V**] are not such a test because they transgress two fundamental assumptions of the models. First the ruby deforms plastically and second, the contact is confined to one or a few asperities. In contrast, the Pastewka-Robbins theory considers the limit of an infinite system, where a statistically significant number of surface features come into contact. Since the geometries and materials that Luke Thimons used are technologically relevant, his experiments motivate the development of adhesion theories accounting for finite-size effects and predicting distributions of pull-off forces.

7 Adhesion hysteresis in soft contacts

7.1 Introduction

In the previous section, I focused on stiff materials where the adhesive tractions are too weak to deform the solid significantly, which is called the DMT regime. I now focus on soft materials such as rubber, jelly or the tip of an insect's foot, where the adhesive stresses are strong and confined to a vanishingly small region. These soft contacts are in the JKR regime, where the contact perimeter is a crack described by the balance of elastic and surface energy [16, 27]. Experiments with rubber [36, 62, 153, 154] show that the force needed to break such contacts is higher than the force measured during approach and the origin of this *adhesion hysteresis* is unclear.

I develop a crack-front model for the contact of spheres with heterogeneity in work of adhesion [**III**] or surface roughness [**I**], allowing me to better understand the role of surface roughness in adhesion hysteresis.

7.1.1 Crack-front theory for smooth spheres

Johnson, Kendall and Roberts (JKR) [16] described the contact of smooth spheres using linear elastic fracture mechanics. In their description, the adhesive interaction is represented by a gain of surface energy per unit contact area, the work of adhesion w_{int}, while the non-contact area, or crack face, is pressure free. Their model represents the limit of short interaction ranges and infinite cohesive stress. The contact pressures σ have a tensile singularity as the distance to the edge of the contact area $-\xi$ goes to 0,

$$\sigma(\xi) = -K_{\mathrm{JKR}}(b,a)/\sqrt{2\pi(-\xi)} + \mathcal{O}((-\xi)^{1/2}), \tag{7.1}$$

with the stress-intensity factor

$$K_{\mathrm{JKR}}(b,a) = \left(\frac{a^2}{R} - b\right)\frac{E'}{\sqrt{\pi a}}. \tag{7.2}$$

Here a is the radius of the circular contact area, R is the radius of the sphere and b is the rigid body penetration.

Within my assumptions of small deformations and linear elasticity, stresses and thereby stress-intensity factors superpose linearly as I superpose displacements. JKR obtained K_{JKR} and the normal force F_{JKR} by superposing the solutions for the nonadhesive contact of a sphere [50] and for the circular flat punch under tensile load [100, 155].

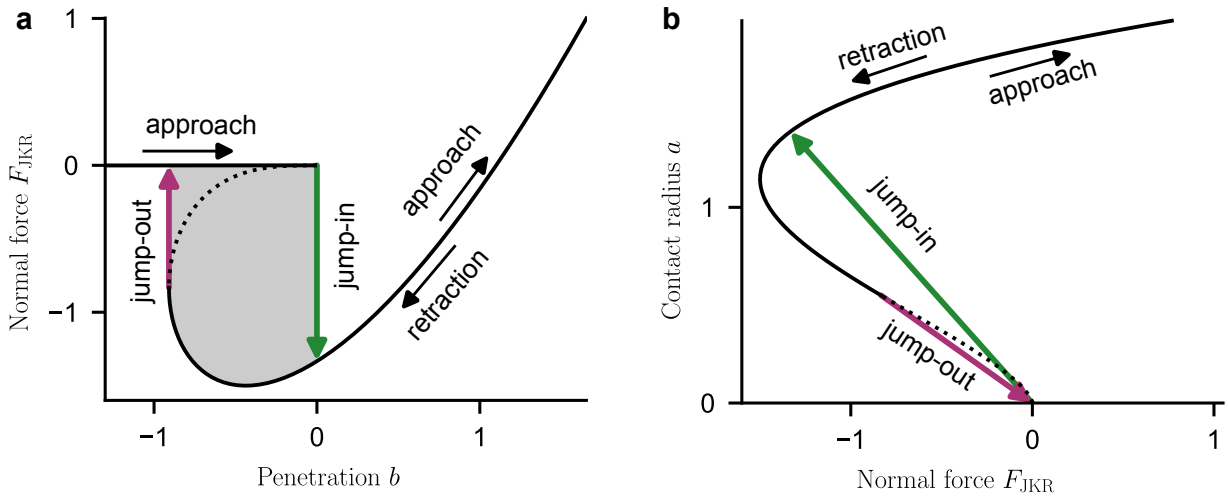


Figure 7.1: Contact of a smooth sphere as predicted by Johnson Kendall and Roberts (JKR) [16]. **(a)** Normal force F_{JKR} versus rigid body penetration b (solid line). For b between -1 and 0, there are two metastable solutions of Eq. (7.4), and the dotted line represents an unstable equilibrium. The shaded area represents the work done by the force during an indentation retraction cycle, corresponding to the energy dissipated during the jump-in and jump-out instabilities (symbolized by the vertical arrows). **(b)** Contact radius as a function of normal force. I have nondimensionalized units following the conventions of Refs. [99, 100] as described in Sec. 7.1.1.

Irwin [156] showed that the amplitude of this stress singularity K_{JKR} determines the cost in elastic energy U_{el} needed to increase the contact area,

$$G_{\mathrm{JKR}}(b,a) = \frac{\partial U_{\mathrm{el}}}{\partial(\pi a^2)} = \frac{K^2_{\mathrm{JKR}}(b,a)}{2E'}. \tag{7.3}$$

G is called the elastic energy release rate and is a generalized force resisting against an increase in contact radius. The equilibrium contact radius is such that this elastic restoring force balances the driving force for new contact area w_{int} [27]:

$$G_{\mathrm{JKR}}(b,a) = w_{\mathrm{int}}. \tag{7.4}$$

Once nondimensionalized using distinct vertical and lateral length units, the JKR contact is parameter free [100, 101, 157], and thus, I present my numerical results in the nondimensional units defined in Refs. [99, 100]. Specifically, lengths along the surface of the half-space (e.g., the contact radius) are normalized by $(3\pi w_{\mathrm{int}} R^2/4E')^{1/3}$, lengths in the vertical direction (e.g., displacements) by $(9\pi^2 w_{\mathrm{int}}^2 R/16E'^2)^{1/3}$ and normal forces by $\pi w_{\mathrm{int}} R$. The equations are in dimensional form but can be nondimensionalized by substituting $R = 1$, $w_{\mathrm{int}} = 1/\pi$ and $E' = 3/4$. Note that this nondimentionalization implies that the pull-off force is proportional to the work of adhesion w_{int}.

Figure 7.1 shows the normal force and contact radius predicted by JKR as I pull the sphere in and out of the contact at prescribed rigid body penetration b. As I approach the sphere that is initially out of contact (panel a), there is no force until they first touch at $b = 0$. However,

during retraction, the surfaces separate at a negative penetration $b \simeq -1$. During a full indentation cycle, the indenter has performed mechanical work corresponding to the shaded area. This energy is dissipated during the jump-in and jump-out instabilities symbolized by the arrows. In the snap-in instability, the contact area is initially out of equilibrium and suddenly increases to a finite value (panel b). This fast, unstable growth of the contact area radiates elastic waves that dissipate the energy, independently of the pulling rate and even for purely elastic materials. In conclusion, energy can be dissipated in purely elastic contacts when the solid jumps between metastable states.

7.1.2 Adiabatic theory of the contact of rough spheres

In the theory by Persson and Tosatti [26] the contact of rough spheres is still described by JKR's equations for a smooth sphere, Eqs. (7.2) to (7.4), but with an apparent work of adhesion w_{PT} containing the effect of surface roughness. This formalism splits the energy balance between macroscopic contributions, represented by G_{JKR}, and interfacial contributions now including the surface roughness in addition to the surface energy w_{int}. Energy conservation implies that w_{PT} is the difference in areal energy between a surface element in contact with the surface roughness and a surface element out of contact. Straining the solid to conform to the surface roughness costs the elastic energy per unit area e_{el} (see Eq. (3.6)), reducing the apparent work of adhesion to

$$w_{\mathrm{PT}} = w_{\mathrm{int}} - e_{\mathrm{el}}. \tag{7.5}$$

w_{PT} is the gain in interfacial energy per unit apparent, or projected, contact area A_{app}, but surface roughness increases the true area of contact A_{true}. Following Persson and Tosatti, this effect increases the surface energy term in Eq. (7.5) by a factor of

$$A_{\mathrm{true}}/A_{\mathrm{app}} \simeq 1 + h'^2_{\mathrm{rms}}/2. \tag{7.6}$$

This increased contact area implies that roughening the surfaces can increase the pull-off force, provided that the material is soft enough for e_{el} to be small. However, Eq. (7.6) is a small-slope approximation that overestimates A_{true} for the diamond coatings [**VII**] considered here, and Persson and Tosatti neglected that stretching the surface from A_{app} to A_{true} also has an energetic cost due to the surface tension of the elastic substrate. Dalvi et al. [62] refined the model to account for these effects when comparing theoretical predictions to their experimental results. In this thesis, I neglect the increase in true area of contact for simplicity and because it does not affect the conclusions in the comparison of my model to the experiments.

7.1.3 Adhesion hysteresis

Dalvi et al. [62] investigated the effect of surface roughness in the adhesion of soft rubber spheres. They varied surface roughness using the four diamond surfaces that Abhijeet Gujrati measured over all length scales [**VII**], allowing them to quantitatively test adhesion theories. Figure 7.2 illustrates a typical force contact radius relationship on the example of NCD. Their refined version of the Persson and Tosatti model captures the effect of surface roughness on

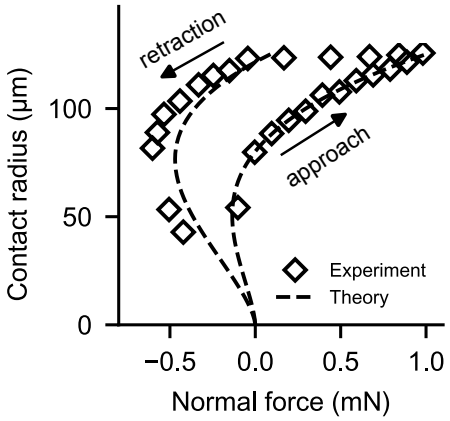


Figure 7.2: Contact radius as a function of the normal force in an adhesion experiment between a rubber sphere and nanocrystalline diamond [62]. The contact area is larger and the force more negative (more adhesive) during retraction than during approach. The sphere has Young's modulus $E = 0.7$ MPa and radius $R = 1.2$ mm. The dashed lines are JKR curves using two distinct apparent work of adhesion for approach w_{appr} and for retraction w_{retr}. The curve during approach is a fit of w_{appr} to the experimental data of Ref. [62], while w_{retr} is a prediction of my theory based on this fit and the power spectral density of the nanocrystalline diamond. This figure is adapted from publication [**I**].

the adhesion measured during approach. The theories presented above assume that the contact follows thermodynamic equilibrium, so that the area should follow the same equilibrium curve during retraction than during approach. However, the force measured during retraction is much higher than during approach (Fig. 7.2), in contrast with this expectation. Such adhesion hysteresis has been reported in several previous experiments [36, 63, 153, 154] and can often be described by two distinct JKR curves with apparent work of adhesion w_{appr} and w_{retr} for approach and retraction, defining the work of adhesion hysteresis $w_{\mathrm{appr}} - w_{\mathrm{retr}}$.

Adhesion hysteresis is most commonly attributed to material specific dissipation mechanisms, such as viscoelastic energy dissipation in the bulk of the material or chemical bonds only formed once the surfaces reached intimate contact [63]. These mechanisms lead to a dependence of the force on the retraction speed and/or the contact time (contact aging). Rate independent adhesion hysteresis can emerge even in purely elastic contacts because of elastic instabilities [63, 64, 72]. Simple analytical models qualitatively explain how surface roughness can trigger elastic instabilities and adhesion hysteresis. These models either focus on the limit of very small contact fractions or on the opposite limit of fully conformal contact.

Asperity models focus on the limit where a small fraction of the area comes into contact [71, 73–76]. Following the ideas of Greenwood and Williamson [45] and Fuller and Tabor [34], surface roughness can be thought of a set of hemispheres with random heights, each of which are independently described by the JKR theory. The force during retraction is higher than during approach because asperities need to be pulled out further out than they entered into contact (each individual asperity behaves like in Fig. 7.1). The energy is dissipated when asperities jump into or pop out of contact. However, modeling the contact area as disconnected patches is not suited for the experiments by Dalvi et al. [62], where the sphere is soft enough for a large fraction of the contact area to come into contact.

For soft materials and small roughness, the contact area is simply connected and described by a single external crack. The surface roughness causes energy barriers that pin the crack front and lead to energy dissipation during depinning instabilities. Guduru [72] and Kesari and Lew [158] demonstrated this effect with a one-dimensional analytical model where roughness is axisymmetric around the sphere apex.

Numerical simulations confirm that two-dimensional random surface roughness can trigger elastic instabilities and thereby cause adhesion hysteresis [61, 64, 65, 68, 70]. However, the crack tip requires fine discretization, preventing boundary element simulations on representative random surface roughness [**III**, 61].

In order to understand the role of elastic instabilities in Dalvi's experiments [62], I need to theoretically predict the adhesion hysteresis caused by elastic instabilities on the diamond coatings characterized in publication [**VII**]. State-of-the-art simulation models are too expensive to account for a realistic surface roughness and analytical theories make oversimplifying assumptions on the geometry of the roughness. The crack-front model that I developed in publications [**I**, **III**] addresses this gap in the limit where the solids can come into fully conformal contact and the contact area is nearly circular. Below I summarize the main ingredients and insights of my model. First I illustrate in more detail how disorder causes adhesion hysteresis on the contact of a sphere with axisymmetric heterogeneity in work of adhesion.

7.2 Pinning of crack fronts by axisymmetric work of adhesion heterogeneity

I now illustrate that adhesion hysteresis can arise in purely elastic contacts on a simple axisymmetric model system similar to Guduru [72] and Kesari [36]. Instead of a wavy sphere, consider a smooth sphere but where the local work of adhesion $W(a)$ depends on the contact radius (Fig. 7.3a). The equilibrium contact radius is now determined by the balance of the elastic force G_{JKR} and the driving force for contact area $W(a)$.

Figure 7.3b shows W as a function of the contact radius a (gray line) along with $G_{\mathrm{JKR}}(b, a)$ at a fixed rigid body penetration $b = 0$. Fluctuations of W lead to several metastable states A, B at a fixed b, so that the contact radius now depends on history. As the solids first touch, the contact radius is initially 0 and suddenly jumps into the metastable state with smallest radius, A. Pushing the sphere further into the substrate shifts the black line representing G_{JKR} to the right. The contact radius increases following the green curve until it looses stability and jumps to the next valley of W. During retraction, the contact radius follows a different path sampling only the regions of highest adhesion and jumping over regions of low adhesion.

In conclusion, the adhesion force and contact radius at a given b (Fig. 7.3c) are higher during retraction than during approach because the fluctuations of W pin the contact line into different metastable states. In the limit of roughness with small wavelength d, the line mainly evolves in jumps between the deepest valleys (approach) and highest crests (retraction) of the work of adhesion heterogeneity (Fig. 7.3d). The force-radius curve then becomes indistinguishable from equilibrium curves predicted by the JKR theory, using two distinct apparent work of adhesion for approach w_{appr} and retraction w_{retr} (Fig. 7.3d). The work of adhesion hysteresis

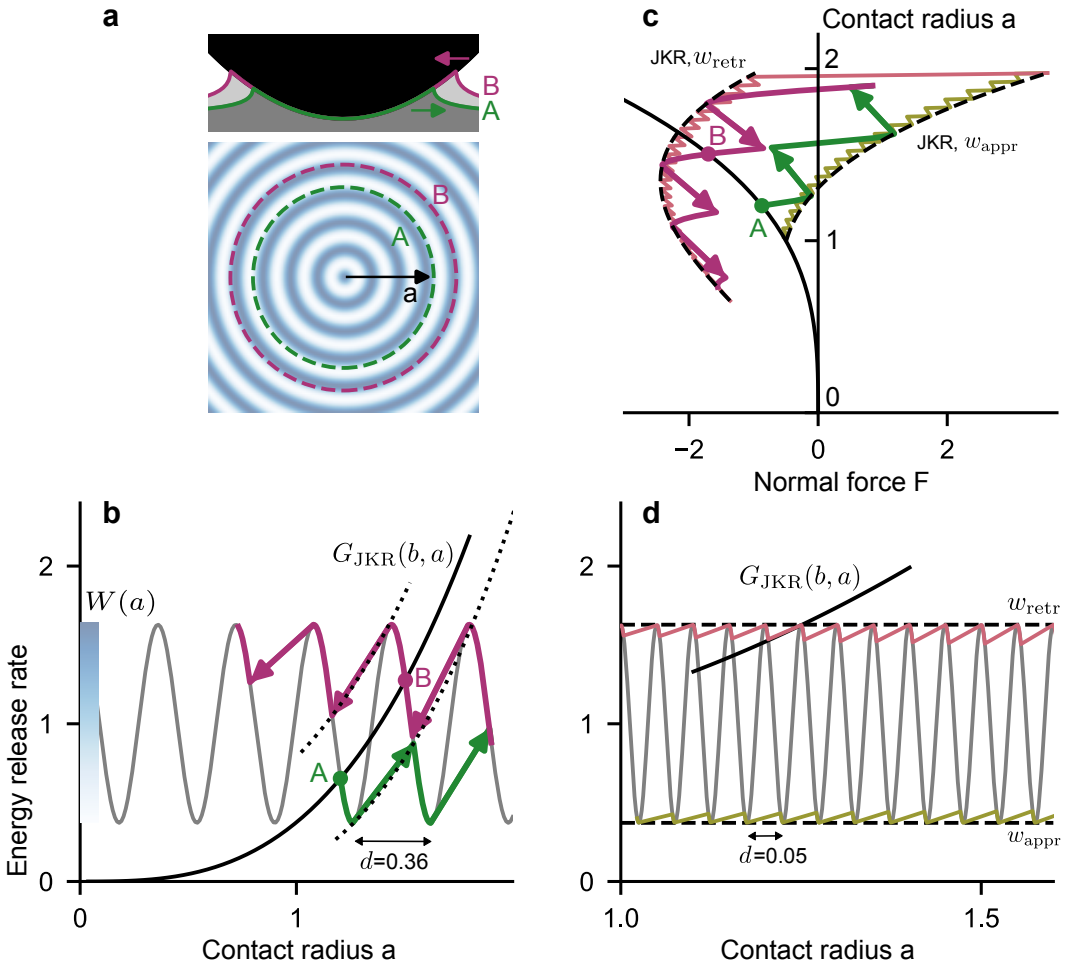


Figure 7.3: Contact of a sphere against axisymmetric work of adhesion heterogeneity $W(a)$ with wavelength d. **(a)** Cross-section of the contact at rigid body penetration $b = 0$ (top) and top view of the axisymmetric work of adhesion heterogeneity $W(a)$ (bottom). The blue color indicates regions of high adhesion. **(b)** Elastic energy release rates in an indentation retraction cycle for a sinusoidal work of adhesion $W(a)$ with wavelength $d = 0.36$ (gray line). The black line shows the elastic energy release rate $G_{\mathrm{JKR}}(b, a)$ as a function of contact radius for fixed rigid body penetration $b = 0$. Fluctuations of $W(a)$ lead to several metastable states A, B at fixed b. During approach, the contact perimeter is pinned in metastable states with low adhesion (green curve), while during retraction the contact perimeter is pinned at higher radii by adhesion peaks (red curve). Arrows indicate elastic instabilities where the contact radius jumps between metastable states. **(c)** The contact radius and the normal force during an indentation retraction cycle for wavelength $d = 0.36$ (darker colors) and $d = 0.05$ (lighter colors). The dashed lines are the prediction by the JKR theory using w_{retr} and w_{appr} for the work of adhesion. The solid black line corresponds to increasing energy release rates at fixed rigid body penetration $b = 0$. **(d)** Energy release rates in an indentation retraction cycle for a work of adhesion heterogeneity with smaller wavelength $d = 0.05$. Note that the slope of G_{JKR} appears to be flatter than in panel (b) because I show a smaller range of contact radii. For short wavelengths, the work of adhesion sampled during approach (light green curve) and retraction (light red curve) stay close to the constant values w_{appr} and w_{retr}. I have nondimensionalized units following the conventions of Refs. [99, 100] as described in Sec. 7.1.1. This figure is adapted from publication [**I**].

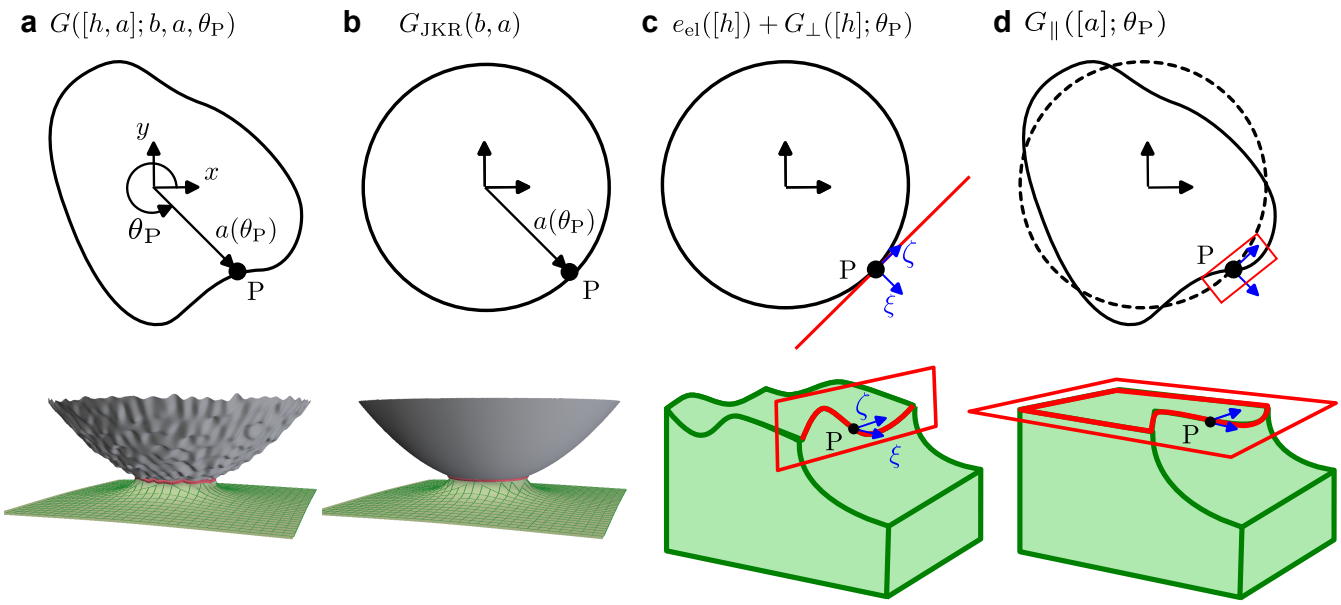


Figure 7.4: Contributions to the energy release in the contact of a sphere with roughness $h(x, y)$ superposed to it. (a) Because of surface roughness, the contact perimeter is no longer circular. I describe it by the contact radius $a(\theta)$, the planar distance between the tip of the sphere and the perimeter of the contact. The energy release rate G at the point P along the crack is decomposed into three contributions, $G = G_{\text{JKR}} + (e_{\text{el}} + G_\perp) + G_\parallel$, illustrated in panels (b) to (d). (b) The energy release rate $G_{\text{JKR}}(a)$ for the smooth contact is given by the theory of Johnson, Kendall and Roberts [16]. (c) Surface roughness leads to out-of-plane displacements of the contact perimeter. This increases the average energy release rate by e_{el} and leads to additional local fluctuations $G_\perp([h], \theta)$. Here, e_{el} is the elastic energy needed to fully conform to surface roughness. (d) The in-plane deflection of the perimeter from circularity leads to the additional contribution $G_\parallel([a], \theta)$. This figure is reproduced from publication [**I**].

$w_{\text{retr}} - w_{\text{appr}}$ is given by the amplitude of W. This toy model shows that fluctuations in the energy balance lead to instabilities and adhesion hysteresis by pinning the crack tip.

7.3 Crack-front model for the contact of rough spheres

By describing the contact of spheres as a circular external crack, the theories by Johnson Kendall and Roberts [16] and Guduru [72] are limited to the one-dimensional, axisymmetric case. Publication [**III**] derives a crack-front model for the contact of spheres against two-dimensional heterogeneity in work of adhesion, where the contact area is no longer circular, and publication [**I**] extends this model to randomly rough surfaces. The crack-front model is a first-order perturbation of the JKR model with respect to the amplitudes of the surface roughness using Rice's weight function theory [83, 84, 159]. It is valid in the limit of small roughness or weak heterogeneity, where the contact area is simply connected and nearly circular. I now summarize the formulation of the crack-front model for rough spheres introduced in more details in the supplementary material of publication [**I**].

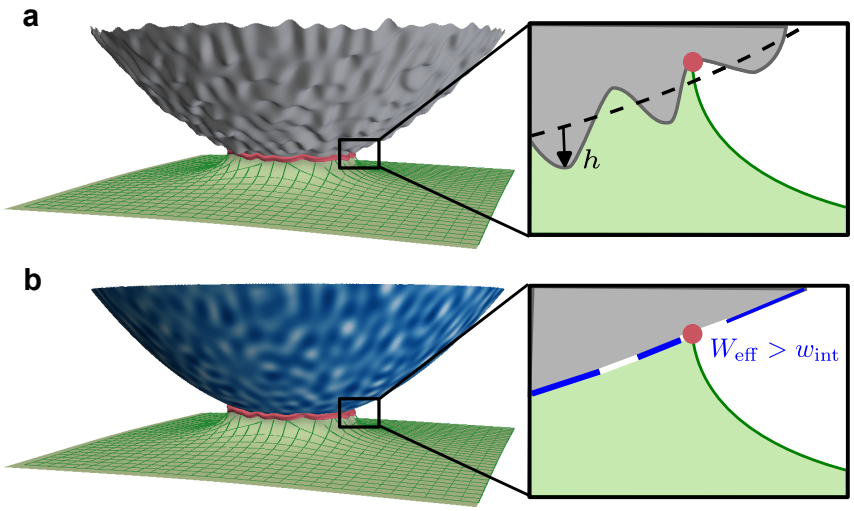


Figure 7.5: The contact of a rough sphere (a) is equivalent to the contact of a sphere with an effective work of adhesion heterogeneity W_{eff} (b). The solid is stretched at the crack tip and surface roughness perturbs this elastic deformation and the elastic energy, which can effectively be described by fluctuations of the work of adhesion. This figure is reproduced from publication [**I**].

7.3.1 Equilibrium equation for the crack front

On rough spheres, Griffith's balance of line forces Eq. (7.4) needs to be evaluated locally for each angle θ along the perimeter:

$$w_{\text{int}} = G([h, a]; b, \theta). \tag{7.7}$$

The energy release rate $G([h, a]; b, \theta)$ is now a functional of the surface roughness $h(x, y)$ and the shape of the contact perimeter $a(\theta)$ (Fig. 7.4a). Note that I explicitly indicate functional dependency by square brackets. I compute G by a perturbation of the circular contact described by JKR (Fig. 7.4b). This perturbation is decomposed in contributions from surface roughness and in-plane perturbation of the crack-shape. The surface roughness h deflects the crack out of the plane and locally perturbs the elastic energy by $G_\perp$ (Fig. 7.4c). In order to satisfy equilibrium with the uniform work of adhesion w_{int}, the perimeter distorts within the plane, perturbing the energy release rate by $G_\parallel$ (Fig. 7.4d). I derive first-order approximations for $G_\perp$ and $G_\parallel$ using crack-perturbation equations developed by Rice and coworkers [83–86].

The fluctuations of the energy release rate caused by the surface roughness $G_\perp$ are given by a simple integral transform of the height profile [86]. Formally, the effect of surface roughness in the energy balance Eq. (7.7) can be remapped into the effective work of adhesion heterogeneity

$$W_{\text{eff}}([h]; a(\theta), \theta) = w_{\text{int}} - e_{\text{el}}([h]) - G_\perp([h]; a(\theta), \theta), \tag{7.8}$$

generalizing Persson and Tosatti's result, Eq. (7.5), to resolve the spatial fluctuations that are critical for hysteresis. This expression establishes that surface roughness is equivalent to work of adhesion heterogeneity, where roughness peaks and valleys correspond to regions of

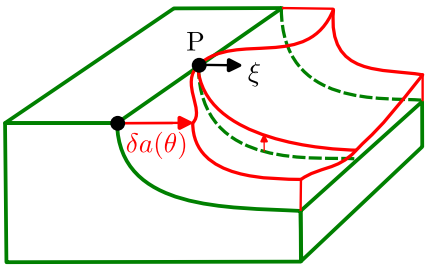


Figure 7.6: Effect of in-plane perturbation of the perimeter on the elastic energy release rate. Moving the crack front within the plane leads to out-of-plane deformations of the crack faces. In a point P that is hold fixed, the energy release rate G is perturbed due to the nonlocal interaction of surface displacements, which is why G is a *functional* of the contact radius $a(\theta)$. This figure is adapted from publication [**I**].

increased and reduced adhesion, respectively. Our mapping from roughness to adhesion is illustrated in Fig. 7.5 and can be intuitively understood as follows. First note that solid is always dilated near the crack tip. In order to conform to a valley, the elastic solid needs to stretch even more, increasing the elastic energy release rate. This additional elastic energy manifests as an effectively decreased *local* work of adhesion. Conversely, conforming to a peak decreases the overall strain near the crack tip and releases elastic energy, leading to an increased effective work of adhesion.

Using this effective work of adhesion, the equilibrium condition is now given by

$$W_{\text{eff}}([h]; a(\theta), \theta) = G_{\text{JKR}}(b, a(\theta)) + G_{\parallel}([a], \theta), \tag{7.9}$$

where $G_{\parallel}([a]; \theta)$ is the elastic restoring force of the line against deviations from circularity within the crack plane. Based on a first-order approximation of the stress intensity factor by Gao and Rice [83, 85],

$$G_{\parallel}([a]; \theta) = c(-\Delta_s)^{1/2} a(\theta), \tag{7.10}$$

with line tension $c = w_{\text{int}}$. The operator $(-\Delta_s)^{1/2}$ is the half-fractional Laplacian with respect to the arclength $\mathrm{d}s = a\,\mathrm{d}\theta$ and can be interpreted as a generalized curvature that scales like a slope.

This curvature describes that the energy release rate is reduced where the perimeter is convex, and Fig. 7.6 illustrates this effect. The green solid represents a small section of the JKR contact with constant radius which I perturb by $\delta a(\theta)$ (red). Advancing the contact area brings the crack-faces closer together even in front of the point P that I hold fixed, because of the nonlocal interaction of the surface displacements. The crack tip is initially stretched so that bringing the crack faces together relieves some of the strain and decreases the elastic energy release rate in point P.

7.3.2 Discussion

There is no unique way to extrapolate the elastic response of the contact line within first-order accuracy in the disorder, and Eq. (7.10) is a slight simplification of the expression I originally

derived in [**III**] for "true" work of adhesion heterogeneity. The current formulation is better suited for surface roughness and is easier to use for theoretical predictions. The crack front possesses an elastic potential [**III**, 160], and I chose my first order approximations so that they respect this property.

I implemented the equilibrium equation (7.9) in an efficient simulation model of the adhesion of rough spheres, allowing me to simulate the indentation retraction process for realistic surface roughness. The predictions of the crack-front model agree with boundary element method simulations on weak adhesion heterogeneity [**III**] and small surface roughness [**I**], but with a computational speed-up of nearly two orders of magnitude. By its efficiency, my new simulation model allows me to predict hysteresis on self-affine synthetic surfaces representative of the experiment in [62] and to conduct a parametric study linking adhesion hysteresis to geometrical measures of the roughness.

The equilibrium condition Eq. (7.9) belongs to the family of equations of the pinning of an elastic line by quenched disorder. These equations of the pinning of elastic lines (or more generally interfaces) describe many phenomena such as sliding friction [161, 162], contact angle hysteresis [163, 164], fracture of heterogeneous materials [165, 166], and pinning by randomness is the subject of many theoretical studies [18, 164, 166–177]. My mapping from surface roughness to an equivalent work of adhesion heterogeneity, Eq. (7.8), makes this analogy explicit and thereby helps to theoretically understand the adhesion hysteresis.

7.4 Theory for adhesion hysteresis by crack pinning

My crack-front simulations and the analogy to the pinning of an elastic line allow me to understand the role of surface roughness in adhesion hysteresis [**I**]. In my simulations on random roughness, adhesion hysteresis is described by two distinct JKR curves with work of adhesion w_{appr} during approach and w_{retr} during retraction (Fig. 7.7). Based on theoretical

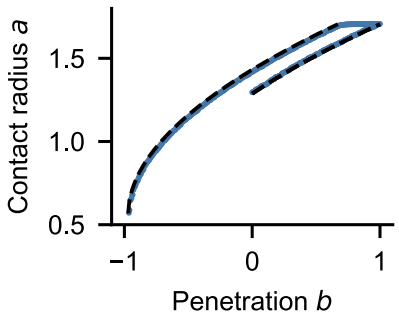


Figure 7.7: Contact radius as a function of the rigid-body penetration in a simulation on self-affine random roughness (nondimensionalized). The dashed lines are JKR curves with work of adhesion w_{appr} and w_{retr} predicted by my theory, Eq. (7.11). The shape of the PSD is shown by the blue circles in the inset of Fig. 7.9 and the elastic energy for fully conformal contact $e_{\mathrm{el}}/w_{\mathrm{int}} = 0.05$. I have nondimensionalized units following the conventions of Refs. [99, 100] as described in Sec. 7.1.1. This figure is adapted from publication [**I**].

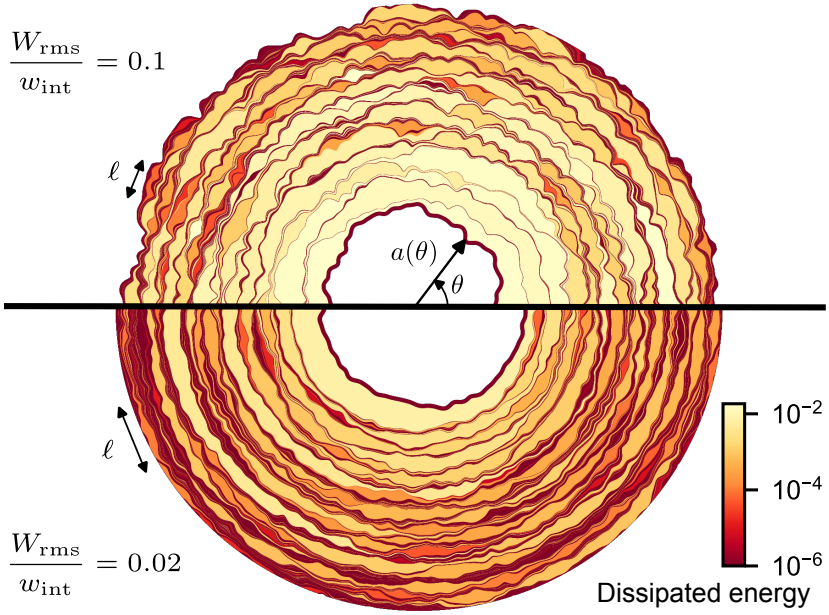


Figure 7.8: Evolution of the contact line during retraction in a crack-front simulation on 2D random roughness. Each colored patch corresponds to an elastic instability during which the perimeter jumps between two pinned configurations (dark lines). The color scale represents the energy dissipated during each instability. The random roughness has a flat power spectral density with nondimensionalized short-wavelength cutoff $\lambda_s = 2d = 0.07$. I have nondimensionalized units following the conventions of Refs. [99, 100] as described in Sec. 7.1.1. This figure is adapted from publication [**I**].

arguments from the pinning literature [164, 166–168],

$$w_{\substack{\mathrm{retr}\\ \mathrm{appr}}} = w_{\mathrm{int}} - e_{\mathrm{el}} \pm k e_{\mathrm{el}}, \tag{7.11}$$

with $k \simeq 3$ determined from crack-front simulations. The reversible term $w_{\mathrm{int}} - e_{\mathrm{el}}$ is the average of the effective work of adhesion field, while the hysteretic contribution comes from the fluctuations of the local effective adhesion $W_{\mathrm{rms}}^2 = 4 e_{\mathrm{el}} w_{\mathrm{int}}$. This scaling of hysteresis with W_{rms}^2 is in contrast with the one-dimensional case presented in section 7.2. Larkin [168] explained that this scaling arises because several asperities *collectively* pin the crack front.

7.4.1 Collective pinning

The key difference with the one-dimensional case is that the contact line is wiggly and evolves in jumps that are spatially localized (Fig. 7.8). The smallest size of these jumps corresponds to the number of asperities required to collectively pin the crack front and is called the Larkin length ℓ [168].

The crucial role of the deformation of the contact line can best be understood on two hypothetical limiting cases. On 2D heterogeneity, the apparent work of adhesion measured is now the average of the work of adhesion over the perimeter of the crack. In the limit of a rigid line

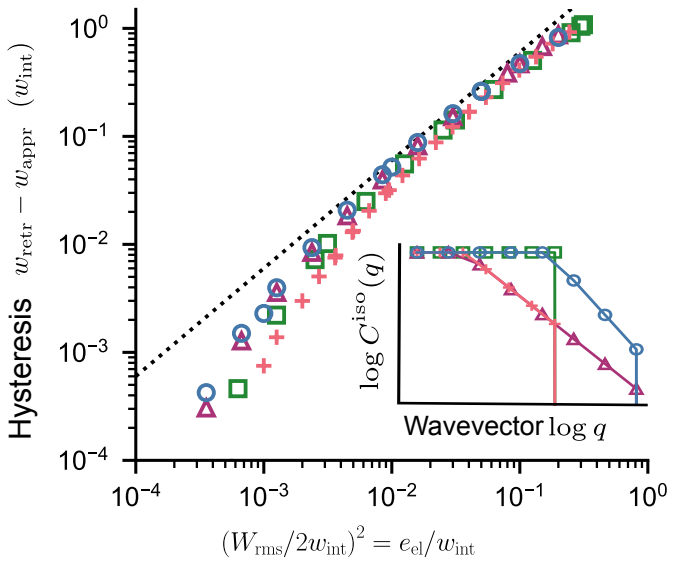


Figure 7.9: Work of adhesion hysteresis as a function of the elastic energy for fully conformal contact e_{el}, or equivalently the variance of the effective work of adhesion W_{rms}^2. The dotted line shows the hysteresis predicted by our theory, Eq. (7.11). The symbols show results of crack-front simulation on randomly rough surfaces with different shapes of the power-spectra represented in the inset. I used a flat PSD with short wavelength cutoff $\lambda_s = 0.005$ (green squares), and three different self-affine PSDs parametrized by the rolloff wavelength λ_r, the short-wavelength cutoff λ_s and the Hurst exponent H. The blue circles correspond to $\lambda_s = 0.000625, \lambda_r = 0.01, H = 0.8$; the purple triangles to $\lambda_s = 0.000625, \lambda_r = 0.1, H = 0.3$; and the pink crosses to $\lambda_s = 0.005, \lambda_r = 0.1, H = 0.3$. I varied e_{el} by scaling the amplitudes of the heights. I have nondimensionalized units following the conventions of Refs. [99, 100] as described in Sec. 7.1.1. This figure is adapted from publication [**I**].

that remains perfectly circular (corresponding to an infinite line tension c), the fluctuations cannot trigger instabilities because they average out over the perimeter. The apparent work of adhesion is identical during approach and retraction and is given by the spacial average $\langle W \rangle$, corresponding to Persson and Tosatti's theory. Without the line tension penalizing deviations from circularity ($c \ll W_{\mathrm{rms}}$), the line can freely distort and meander along valleys during approach and peaks during retraction. Each individual asperity independently triggers instabilities similarly as in the axisymmetric model presented in subsection 7.2 and Fig. 7.3. This hypothetical case would lead to a hysteresis scaling with W_{rms} as in my axisymmetric model.

In my simulations, the line tension $c = w_{\mathrm{int}}$ maintains the line straight at small scales, so that the line is collectively pinned by several asperities over the length ℓ. The hysteresis now additionally depends on the freedom of the line to meander between asperities. Stronger heterogeneities lead to lines with more wiggles that can sample more effectively the deepest valleys and highest peaks (Fig. 7.8), so that the hysteresis increases faster than linearly with W_{rms}. I point to publication [**I**] and to the classic literature [164, 166–168] for the explanation of the exact scaling W_{rms}^2.

Larkin's theory makes strong assumptions on the geometry of the pinning field, and cannot be applied to surface roughness where the work of adhesion field has power-law correlations. Démery et al. [177] predicted that this scaling remains unaffected by the geometry of the

random field by analytically solving a small disorder expansion of the equation of motion of the line. My numerical simulations in [**I**] further confirm that Eq. (7.11) remains valid in the presence of power-law correlations (Fig. 7.9).

7.4.2 Discussion

Equation (7.11) is the first theory quantitatively linking adhesion hysteresis to random surface roughness, and establishes that the elastic energy for fully conformal contact e_{el} is the central parameter governing adhesion hysteresis. In their theory for axisymmetric contacts, Kesari and Lew [158] found that adhesion hysteresis increases linearly with e_{el}, in contrast with my finding. This discrepancy highlights that it is crucial to model how two-dimensional random roughness distorts the contact perimeter. Boundary element simulations on two-dimensional patterns [69] and two-dimensional random roughness [61, 70, 178] confirm the dominant role of e_{el} in adhesion hysteresis, but do not allow conclusions concerning the functional dependence of $w_{\mathrm{retr}} - w_{\mathrm{appr}}$ on this parameter.

My crack-front model allows me to theoretically understand the role of surface roughness in adhesion hysteresis by making use of the equivalence with the classic problem of the pinning of an elastic line by a random field. The broad literature on this topic will allow further insights on adhesion hysteresis. For example, hysteresis disappears for small e_{el} (Fig. 7.9), and the critical e_{el} for hysteresis depends on the size of the heterogeneity. This finite size effect can be understood using a simple argument by Larkin [168] or a more advanced theory by Tanguy and Vettorel [173]. The fluctuations of pinning force and the slight dependence of the adhesion hysteresis on the geometry of the roughness was studied using renormalization group theory and numerical simulations [174, 176]. However, as for the prediction of work of adhesion hysteresis, these theories need to be tested or adapted to self-affine random roughness.

7.5 Role of surface roughness

The dominating parameter for adhesion hysteresis is the elastic energy for fully conformal contact. This elastic energy can be written as

$$e_{\mathrm{el}} = \frac{E'}{4}\left[h_{\mathrm{rms}}^{(1/2)}\right]^2, \tag{7.12}$$

where $h_{\mathrm{rms}}^{(1/2)}$ is the rms half-derivative (or quarter fractional Laplacian) of the heights. This roughness parameter is a generalized measure of the sharpness of peaks sensitive to larger length scales than curvatures and slopes. It is defined by

$$(h_{\mathrm{rms}}^{(\alpha)})^2 = \frac{1}{4\pi}\int \mathrm{d}^2 q\, |q|^{2\alpha} C^{2\mathrm{D}}(\vec{q}), \tag{7.13}$$

which contains the rms heights of the topography, $h_{\mathrm{rms}}^{(0)}$, and the rms gradient of the topography, $h_{\mathrm{rms}}^{(1)}$. Hence, the rms half-derivative is a roughness parameter intermediate between rms height and rms gradient (or slope).

For a self-affine topography, the range of scales determining $h_{\text{rms}}^{(1/2)}$ depends on the Hurst exponent [26]. For $H < 0.5$, it is dominated by small-scale roughness, like the rms slope, while for $H > 0.5$ it depends on the large scales, like the rms height. However, most natural and engineered surfaces have $H > 0.5$ [33, 37, 38], so that my model is consistent with the increase in pull-off force with h_{rms} reported in Refs. [35, 36].

This increase in pull-off force is usually explained by the increase in true area of contact $A_{\text{true}}/A_{\text{app}}$ Eq. (7.6) [26, 62], which increases with the rms gradient of the topography. References [35, 36] only reported h_{rms} so that it is not clear which effect dominated in their experiment. For the experiments by Dalvi et al. [62], Abhijeet Gujrati measured the topography over all scales, allowing me to quantify these two effects. Following Dalvi et al., the increased true area of contact enhances adhesion by a factor of 1.16 for NCD. My theory predicts that crack-front pinning further enhances the work of retraction by a factor of 1.46 in the case of a rubber sphere with Young's modulus $E = 0.7$ MPa. The range of scales that dominates $h_{\text{rms}}^{(1/2)}$ and hence pinning is at the transition between power-law scaling and the flat rolloff at 1 μm, while the surface area increase is determined by roughness below 0.5 μm down to the atomic scale. In conclusion, crack-front pinning is as relevant to the pull-off force as the surface area increase, and soft adhesion depends on roughness on lateral length scales ranging from nanometers up to micrometers.

7.6 Comparison to experiment

For NCD, the hysteresis predicted by my model Eq. (7.11) is very similar to the measurements by Dalvi [62] with a sphere with Young's modulus $E = 0.7$ MPa and radius $R = 1.2$ mm (Fig. 7.2). Since there is no direct way to measure the intrinsic work of adhesion, I extracted $w_{\text{int}} = 65$ mJ/m^2 by fitting w_{appr} to the data, which is in the range expected for van der Waals interaction. The good agreement of the corresponding retraction curve using w_{retr} shows that crack pinning quantitatively reproduces the hysteresis observed in this experiment.

This result confirms that crack pinning plays an important role in adhesion hysteresis, however, my crack-front model does not explain Dalvi's experiments. Figure 7.10 shows how the apparent work of adhesion measured during approach (open symbols) and retraction (full symbols) changes between different substrates. The dashed lines show the predictions of Eq. (7.11) using $w_{\text{int}} = 50$ mJ/m^2 fitted to the data during approach. While my model reasonably describes the behavior during approach, it fails to explain how w_{retr} varies with surface roughness. In particular, the adhesion of UNCD during retraction is significantly higher than predicted by my model.

This discrepancy suggests that an additional mechanism toughens the contact during retraction, such as contact aging [179] or viscoelasticity [154]. Additional experiments varying contact time or velocity are necessary to test this hypothesis. I could not conclude whether my model agrees with experiments from other groups [36, 180] because these authors did not report the surface roughness in enough details.

Since viscoelasticity plays an important role in many experiments [36, 154, 181–183] and in practical application, it would be interesting to extend our crack-front model to account for viscoelasticity [184–187]. This would allow investigating the interplay of random roughness

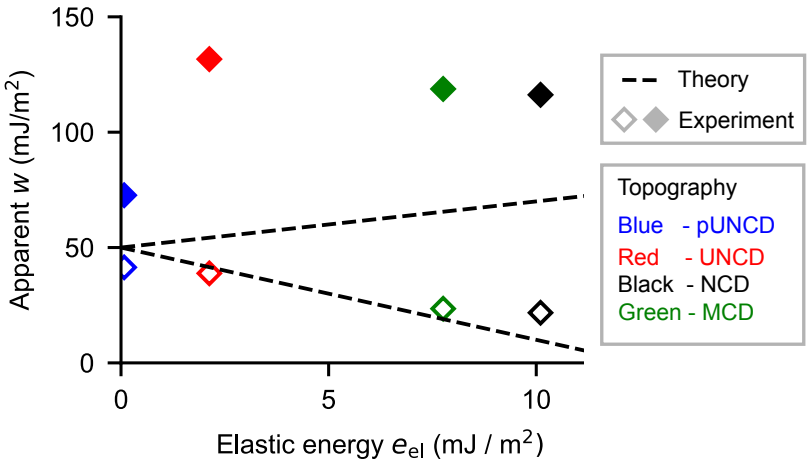


Figure 7.10: Apparent work of adhesion for approach (open symbols) and retraction (closed symbols) reported by Dalvi et al. [62] for adhesion experiments between a rubber sphere and different diamond substrates. These values result from fitting JKR curves to force vs. contact radius measurements similar to Fig. 7.2, and the elastic energies for fully conformal contact e_{el} were computed from the PSDs of the respective topography. The dashed lines represent the predictions from my theory, with $w_{\mathrm{int}} = 50\ \mathrm{mJ/m^2}$ fitted to the data during approach. The sphere has Young's modulus $E = 0.7$ MPa and radius $R = 1.2$ mm.

and viscoelasticity, while viscoelastic formulations of BEM with adhesion were restricted to simple geometries [188–190].

7.7 Limitations of the model

The assumptions I made in my model also need to be checked. My numerical study behind Eq. (7.11) focuses on random Gaussian surfaces, but the surfaces used by Dalvi deviate significantly from this idealization of roughness [**VII**]. NCD is faceted exactly at the scales contributing the most to crack pinning, so that future work should check whether including such features in the simulations affects the predicted hysteresis.

When e_{el} approaches w_{int}, my assumption of fully conformal contact breaks down and my correlation between pull-off force and e_{el} no longer holds. While for large $e_{\mathrm{el}} \gtrsim 2w_{\mathrm{int}}$, the contact area and adhesion nearly vanish [41, 56, 64, 70], it is unclear whether an intermediate regime with large pull-off forces exists.

Higher-order crack perturbation methods [191–193] might provide some insights in this intermediate regime. Bower and Ortiz [192] used such a model to study crack pinning by periodic arrays of tough inclusions, and their results indicate that my theory might underestimate the pull-off force due to crack bridging. Above a critical toughness of the inclusion, the crack circumvents the obstacle rather than crossing it, leaving behind a contact island bridging between the crack faces. These bridges effectively toughen the material because they hold the crack faces together and significantly enhance the force needed to propagate the crack front. However, their work considers flat interfaces, while for rough contacts the interface might be initially weakened because the surfaces do not come into fully conformal contact during

approach.

BEM simulations can account for disconnected contact areas. However, I found very difficult to draw conclusions on adhesion hysteresis because it is computationally challenging to sufficiently discretize the crack tip and at the same time include a representative amount of surface roughness [**III**, 61]. It hence remains unclear how the pull-off force increases and how adhesion hysteresis changes when the contact area is no longer simply connected.

In using my crack-front model and my theoretical predictions, I make the important assumption that the errors due to the first-order approximation are small compared to the predicted hysteresis. This assumption is not guaranteed even in the limit of small roughness, because these errors scale to same order with roughness as the predicted hysteresis [194]. This means that the first-order approximation of the elastic line only allows for conclusions on the scaling of hysteresis. Predictions of the effective toughness of heterogeneous materials [166, 195] and wetting angle hysteresis [164] have the same limitation. Predicting quantitatively the adhesion hysteresis in rough contacts requires to understand the effect of second-order (and possibly higher-order) terms in the equilibrium equation of the line on the adhesion force in rough contacts.

8 Conclusion

My collaborators and I investigated how surface roughness affects adhesion in three different regimes of material compliance: rigid solids that touch on a single, vanishingly small asperity, *stiff* elastic solids that conform over a small fraction of the nominal area of contact, and *soft* solids that fully conform to surface roughness. Abhijeet Gujrati characterized four polycrystalline diamond coatings over an unprecedented range of length scales. Using this data as input for numerical and theoretical models, we unraveled the role of roughness in experiments with rigid ruby spheres and in the contact with soft rubber spheres. Joe Monti and I performed large-scale boundary element method (BEM) simulations that clarify the range of validity of the theory for the adhesion of stiff elastic contacts (DMT limit) by Pastewka and Robbins [24]. For soft spheres that can conform to the surface roughness, I introduced an efficient crack-front simulation model and derived a quantitative theory that links adhesion hysteresis to statistical metrics of roughness representative of real systems.

Theory for adhesion hysteresis I developed a crack-front model for the contact of soft spheres describing how work of adhesion heterogeneity [**III**] or surface roughness [**I**] distorts and pins the perimeter of the contact. Surface roughness is equivalent to work of adhesion heterogeneity, where roughness peaks increase local adhesion and pin the crack front. The crack front is pinned at larger contact radii during retraction than during approach, explaining why breaking contacts is harder than making them even in the absence of material specific energy dissipation. The energy is dissipated during jumps (elastic instabilities) between pinned configurations. Using the power-spectra characterized in publication [**VII**], my model predicts hysteresis similar in magnitude as in experiments performed by Dalvi et al. [62] on the same surface, so that crack pinning plays an important role in adhesion hysteresis.

By establishing the equivalence of rough adhesion with the pinning of an elastic line, my equations pave the way to better understanding adhesion. I used this analogy to derive a quantitative expression showing that the pull-off force increases linearly with the elastic energy for fully conformal contact e_{el}, which is the product of the contact modulus and a simple roughness parameter. For self-affine surfaces with small Hurst exponent, this parameter is determined by roughness at the nanometer scale, but on most surfaces, including the diamond coatings in publication [**VII**], it is dominated by large scale roughness around the rolloff wavelength.

My crack-front simulations allow me to predict adhesion hysteresis on roughness two orders of magnitude finer than the boundary element method. The assumption of nearly circular and simply connected contact area behind this computational speed-up is also the main limitation of my model. When e_{el} exceeds the work of adhesion w_{int}, surfaces no longer conform to each other, and thus the contact area becomes disconnected and can only be described by the

BEM. Understanding hysteresis in this regime of partial contact remains a great computational challenge.

While partial contact might explain some discrepancies between my model and the measurements by Dalvi [62], other mechanisms may also play an important role. Further investigations both in modeling and experiment are required to better understand the role surface roughness in adhesion hysteresis.

Theory for the adhesion of stiff elastic contacts The theory by Pastewka and Robbins [24] describes the adhesion between two nominally flat, infinite rough surfaces in the limit of weak adhesive interactions (DMT limit). Using finely discretized boundary element method simulations, Joe Monti and I show that their equation for the distribution of interfacial gaps is accurate below a critical gap determined by the small-scale geometry of the surfaces. This result means that their prediction of the contact area is valid provided that the interaction range is small enough. Our own work on the contact of ruby and diamond [**V**] and other studies [140–142] show that the range of interaction is larger than usually expected for van der Waals interactions, so that their theory might be applicable to less contact systems than assumed before.

We also verified that the gap distribution remains unaffected by weak adhesive interactions. However, the transition between DMT-like contacts and contacts where adhesive stresses significantly affect the deformation remains poorly understood.

Contact of rigid spheres Luke Thimons conducted thousands of adhesion tests between a macroscopic ruby sphere and the diamond coatings. By comparing these experimental results to brute-force numerical computation on the rough surfaces measured in publication [**VII**], we determine the parameters of the adhesive interaction and show that the pull-off force is mostly affected by roughness with wavelength between 43 nm and 1.8 μm. Roughness below 43 nm does not matter because small-scale features deform the ruby plastically and because the range of interaction $\rho = 5$ nm is large compared to height fluctuations over small lateral distances. The range of interaction is larger than expected for van der Waals interaction ($\simeq 0.5$ nm), and its physical origin remains unclear.

The pull-off forces have large fluctuations, approximately described by a log-normal distribution. These fluctuations occur because the contact is confined to one or a few of the highest asperities. Predicting the magnitude and distribution of these pull-off forces theoretically is an open challenge.

Nature of surface roughness Surface roughness is fractal, and often idealized as self-affine in theoretical work. The comprehensive surface topography characterization in publication [**VII**] reveals that polycrystalline diamond coatings deviate from this idealization, with different regions of scaling of the power spectral density (PSD). At small scales, the roughness is self-affine and is similar between all unpolished surfaces. However, at large scales, the surfaces with larger grains have higher roughness. These grains are faceted, and the surface is not self-affine at scales similar to the grain size. Faceting leads to a characteristic q^{-4}-scaling in the PSD and a constant scale-dependent slope.

The numerical study in publication [**V**] accounts for faceting, but we did not check whether faceting has an effect on adhesion hysteresis. We also need to test whether scale-dependent roughness parameters defined via finite differences and Fourier filtering are still equivalent on such surfaces.

Besides self-affinity, theoretical work usually assumes that the distributions of heights, slopes and curvatures are Gaussian. The scale-dependent derivatives introduced in publication [**IV**] provide a tool allowing us to test this assumption over different length scales in future work.

Statistical characterization of surface roughnesss Our work illustrates that it is important to analyze surface topography over multiple scales and to document surface roughness in publications. Publication [**IV**] and the web service *contact.engineering* described in publication [**II**] make the process of analyzing and publishing surface roughness easier than before.

The statistics of slopes and curvatures are used in simple contact theories [24, 26], and thus it is important to understand how these quantities depend on the scale. Publication [**IV**] demonstrates that scale-dependent derivatives can be computed using a new approach that is easy to interpret and to implement. This approach is based on finite-difference schemes, and my collaborators and I show that for self-affine surfaces, it is equivalent to the common approach which uses Fourier filtering. By demonstrating four equivalent ways to compute scale-dependent roughness parameters, we clarified the relationship between scale-dependent derivatives, the autocorrelation function (ACF), the variable bandwidth method (VBM) and the PSD.

The web-platform *contact.engineering* described in publication [**II**] is a public database where the community can publish and analyze surface topography in a standardized way. It automatizes the workflow used in publication [**VII**] to combine several measurements into a comprehensive statistical representation of the surface roughness. If in the future, researchers publish their measurements on our public database, it will become easy to compare surface roughness among different publications and for theoreticians to apply their models to real experiments.

contact.engineering also allows the user to compute force-area relationships by performing nonadhesive BEM simulations. However, a BEM simulation is restricted to an individual areal scan (such as an AFM measurement) and the result is often not representative of the whole surface. An important enhancement would be to implement the multiscale contact theories introduced by Persson [41, 52, 53, 82], that account for all scales of the PSD or ACF [79].

Conclusion Which scale of roughness matters in adhesion? The answer is not unique, as the force needed to break a contact might either be reduced or enhanced by increasing roughness amplitudes. For infinite contacts between stiff materials, it is expected by theory that roughness close to the atomic scale makes most of the interfaces unsticky. However, the pull-off force in Luke Thimons' experiments with ruby spheres is not affected by roughness below 43 nm, and is still affected by roughness up to 1.8 μm. Here, plastic deformations, a long range of attraction and the finite size of the contact play an important role. For stiff contacts, the rougher surfaces are the least adhesive. On the other hand for soft spheres, the

pull-off force increases with surface roughness, because roughness increases the true area of contact and because it pins the crack front. The true area of contact increases with surface slopes, set by small-scale roughness, while the strength of pinning is increased by roughness at large scales, around the rolloff wavelength. Hence, the adhesion of soft contacts might be controlled either by tuning small-scale roughness or large-scale roughness.

Our results provide guidelines for practitioners on how to vary roughness to control the tribological performance of an interface. Being able to precisely tune the adhesion of soft solids is important for applications such as pick and place systems [196–198] and skin adhesives [199–202]. Adhesion between hard solids can lead to catastrophic failure of microelectromechanical devices [203], and this problem might be suppressed by roughening the surfaces at the appropriate length scales.

Acknowledgements

First of all, I am deeply thankful to Lars Pastewka for his support and enthusiasm, and for all I learned during my thesis. Interacting with him deepened my understanding of physics, and taught me the craft of research as well as the craft of writing scientific software. I highly appreciate his generosity and his benevolent leadership.

I thank the members of the jury for taking the time to evaluate my thesis.

A great part of this thesis is a close collaboration with the experimental group of Tevis D. B. Jacobs (University of Pittsburgh), grounding my theoretical work to reality. It was always a pleasure to attend our regular "roughness teleconferences", and I thank Luke, Abhijeet, Manyu, Arushi, Nate and Tevis for the interesting discussions and the nice atmosphere in these teleconferences across the Atlantic. I also thank my coauthors, especially Luke, Abhijeet, Wolfram, Joe and Tevis for the constructive and enjoyable collaboration.

Regular discussions with my lab mates Beck Andrews and Lucas Frérot deepened my understanding of fracture mechanics, contact mechanics and numerical optimization. My understanding of crack pinning benefitted from discussions with Patrick Dondl of the department of mathematics, and with Mathias Lebihain from École des Ponts PariTech. I also acknowledge useful discussions with Nityanshu Kumar and Ali Dhinojwala (University of Akron) concerning the experiments in Ref. [62], and discussions with Martin H. Müser (Universität Saarland), which in particular initiated the Master's thesis of Sindhu Singh on preconditioned boundary element simulations.

I am indebted to Mark O. Robbins. The few hours of discussion I had the chance to share with him were enlightening and gave an important change of course to my thesis.

I thank all my lab mates and other colleagues for the nice atmosphere, for sharing problems, their advice, and for helping out. I am grateful to Lucas, Maria, Beck, Anna, Jan M and Jan G for their crucial advice in the final phase of my thesis. I especially thank Lucas for helping me improve the presentation for my defense and Maria for her comments on my manuscript. I thank my friends, among which I count many colleagues, for helping me to escape from my work during bike tours, jam sessions, beers, and other excursions.

I am indebted to my parents, Adrien and Martina Sanner, for their limitless support.

Funding was provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2193/1 – 390951807 and by the European Research Council (StG-757343). I performed numerical simulations on bwForCluster NEMO (University of Freiburg, DFG grant INST 39/963-1 FUGG). I stored data on bwSFS (University of Freiburg, DFG grant INST 39/1099-1 FUGG) using the data management software *dtool* [204, 205], and I thank my lab mate Johannes Hörmann for helping me to use this tool. I thank Till Junge for implementing the original BEM code upon which I built my work, Sindhu Singh for implementing the constrained conjugate-gradients algorithm, and Laura Mahoney for writing assistance.

References

[1] Message Passing Interface Forum, *MPI: A Message-Passing Interface Standard Version 2.2* (Sept. 2009).

[2] I. A. Polonsky and L. M. Keer, "A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques", Wear **231**, 206 (1999).

[3] R. Bugnicourt, P. Sainsot, D. Dureisseix, C. Gauthier, and A. A. Lubrecht, "FFT-based methods for solving a rough adhesive contact: description and convergence study", Tribol. Lett. **66**, 29 (2018).

[4] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization", SIAM J. Sci. Comput. **16**, 1190 (1995).

[5] S. Singh, "Optimization algorithms for contact mechanics", MA thesis (University of Freiburg, 2021).

[6] P. Strauch, "Classification of rough surface topographies", MA thesis (University of Freiburg, 2022).

[7] `https://gitlab.com/muspectre/muspectre`.

[8] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3", Proc. IEEE **93**, 216 (2005).

[9] Y. Zhou, M. Moseler, and M. H. Müser, "Solution of boundary-element problems using the fast-inertial-relaxation-engine method", Phys. Rev. B **99**, 144103 (2019).

[10] L. Dalcin and Y.-L. L. Fang, "mpi4py: Status update after 12 years of development", Comput. Sci. Eng. **23**, 47 (2021).

[11] E. A. H. Vollebregt, "The bound-constrained conjugate gradient method for non-negative matrices", J. Optim. Theory Appl. **162**, 931 (2013).

[12] V. Rey, G. Anciaux, and J.-F. Molinari, "Normal adhesive contact on rough surfaces: Efficient algorithm for fft-based bem resolution", Comput. Mech. **60**, 69 (2017).

[13] A. Gujrati, S. R. Khanal, L. Pastewka, and T. D. B. Jacobs, "Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales", ACS Appl. Mater. Interf. **10**, 29169 (2018).

[14] T. D. B. Jacobs, T. Junge, and L. Pastewka, "Quantitative characterization of surface topography using spectral analysis", Surf. Topogr. Metrol. Prop. **5**, 013001 (2017).

[15] S. B. Ramisetti, C. Campañá, G. Anciaux, J.-F. Molinari, M. H. Müser, and M. O. Robbins, "The autocorrelation function for island areas on self-affine surfaces", J. Phys. Condens. Matter **23**, 215004 (2011).

[16] K. L. Johnson, K. Kendall, and A. D. Roberts, "Surface energy and the contact of elastic solids", Proc. R. Soc. London, Ser. A **324**, 301 (1971).

[17] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library", in Advances in Neural Information Processing Systems, Vol. 32 (2019).

[18] A. Rosso and W. Krauth, "Roughness at the depinning threshold for a long-range elastic string", Phys. Rev. E **65**, 025101 (2002).

[19] J. N. Israelachvili, *Intermolecular and Surface Forces* (Academic Press, London, 1991).

[20] K. Autumn, Y. A. Liang, S. T. Hsieh, W Zesch, W. P. Chan, T. W. Kenny, R Fearing, and R. J. Full, "Adhesive force of a single gecko foot-hair", Nature **405**, 681 (2000).

[21] B. N. J. Persson, O. Albohr, U. Tartaglino, A. I. Volokitin, and E. Tosatti, "On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion", J. Phys.: Condens. Matter **17**, R1 (2004).

[22] Y. Zhou, A. Robinson, U. Steiner, and W. Federle, "Insect adhesion on rough surfaces: Analysis of adhesive contact of smooth and hairy pads on transparent microstructured substrates", J. R. Soc. Interface **11**, 20140499 (2014).

[23] K. Kendall, *Molecular Adhesion and Its Applications: The Sticky Universe* (Springer, Berlin; New York, 2001).

[24] L. Pastewka and M. O. Robbins, "Contact between rough surfaces and a criterion for macroscopic adhesion", Proc. Natl. Acad. Sci. U.S.A. **111**, 3298 (2014).

[25] C. Creton and L. Leibler, "How does tack depend on time of contact and contact pressure?", J. Polym. Sci. B Polym. Phys. **34**, 545 (1996).

[26] B. N. J. Persson and E. Tosatti, "The effect of surface roughness on the adhesion of elastic solids", J. Chem. Phys. **115**, 5597 (2001).

[27] A. A. Griffith and G. I. Taylor, "VI. The phenomena of rupture and flow in solids", Philos. Trans. R. Soc. London, Ser. A **221**, 163 (1921).

[28] B. V. Derjaguin, V. M. Müller, and Y. P. Toporov, "Effect of contact deformations on the adhesion of particles", J. Colloid Interface Sci. **53**, 314 (1975).

[29] R. S. Bradley, "The cohesive force between solid surfaces and the surface energy of solids", Lond. Edinb. Dubl. Phil. Mag. **13**, 853 (1932).

[30] J. F. Archard, "Elastic deformation and the laws of friction", Proc. R. Soc. A **243**, 190 (1957).

[31] R. S. Sayles and T. R. Thomas, "Surface topography as a nonstationary random process", Nature **271**, 431 (1978).

[32] B. B. Mandelbrot and B. B. Mandelbrot, *The Fractal Geometry of Nature*, Vol. 1 (WH freeman New York, 1982).

[33] B. B. Mandelbrot, D. E. Passoja, and A. J. Paullay, “Fractal character of fracture surfaces of metals”, Nature **308**, 721 (1984).

[34] K. N. G. Fuller and D. Tabor, “The effect of surface roughness on the adhesion of elastic solids”, Proc. R. Soc. London, Ser. A **345**, 327 (1975).

[35] G. A. D. Briggs and B. J. Briscoe, “The effect of surface topography on the adhesion of elastic solids”, J. Phys. D: Appl. Phys. **10**, 2453 (1977).

[36] H. Kesari, J. C. Doll, B. L. Pruitt, W. Cai, and A. J. Lew, “Role of surface roughness in hysteresis during adhesive elastic contact”, Philos. Mag. Lett. **90**, 891 (2010).

[37] T. Candela, F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, and E. E. Brodsky, “Roughness of fault surfaces over nine decades of length scales”, J. Geophys. Res. – Sol. Ea. **117**, B08409 (2012).

[38] B. N. J. Persson, “On the fractal dimension of rough surfaces”, Tribol. Lett. **54**, 99 (2014).

[39] S. R. Khanal, A. Gujrati, S. B. Vishnubhotla, P. Nowakowski, C. S. Bonifacio, L. Pastewka, and T. D. B. Jacobs, “Characterization of small-scale surface topography using transmission electron microscopy”, Surf. Topogr.: Metrol. Prop. **6**, 045004 (2018).

[40] M. H. Müser, “A dimensionless measure for adhesion and effects of the range of adhesion in contacts of nominally flat surfaces”, Tribol. Int. **100**, 41 (2016).

[41] B. N. J. Persson, “Adhesion between an elastic body and a randomly rough hard surface”, Eur. Phys. J. E **8**, 385 (2002).

[42] G. Amontons, “De la résistance causée dans les machines”, Mem. Acad. R. Med. Belg. **A**, 275 (1699).

[43] B. N. J. Persson, *Sliding Friction* (Springer Berlin Heidelberg, 2000).

[44] F. P. Bowden and D. Tabor, “The area of contact between stationary and between moving surfaces”, Proc. R. Soc. London, Ser. A **169**, 391 (1939).

[45] J. A. Greenwood and J. B. P. Williamson, “Contact of nominally flat surfaces”, Proc. R. Soc. London, Ser. A **295**, 300 (1966).

[46] A. W. Bush, R. D. Gibson, and T. R. Thomas, “The elastic contact of a rough surface”, Wear **35**, 87 (1975).

[47] B. N. J. Persson, “Elastoplastic contact between randomly rough surfaces”, Phys. Rev. Lett. **87**, 116101 (2001).

[48] S. Hyun, L Pei, J.-F. Molinari, and M. O. Robbins, “Finite-element analysis of contact between elastic self-affine surfaces”, Phys. Rev. E **70**, 026117 (2004).

[49] F. P. Bowden and D. Tabor, *The Friction and Lubrication of Solids* (Oxford University Press, 1950).

[50] H. Hertz, “Über die Berührung fester elastischer Körper”, J. Reine Angew. Math. **92**, 156 (1881).

[51] D. Maugis, “On the contact and adhesion of rough surfaces”, J. Adhes. Sci. Technol. **10**, 161 (1996).

[55] M. H. Müser, W. B. Dapp, R. Bugnicourt, P. Sainsot, N. Lesaffre, T. A. Lubrecht, B. N. J. Persson, K. Harris, A. Bennett, K. Schulze, S. Rohde, P. Ifju, W Gregory Sawyer, T. Angelini, H. A. Esfahani, M. Kadkhodaei, S. Akbarzadeh, J.-J. Wu, G. Vorlaufer, A. Vernes, S. Solhjoo, A. I. Vakis, R. L. Jackson, Y. Xu, J. Streator, A. Rostami, D. Dini, S. Medina, G. Carbone, F. Bottiglione, L. Afferrante, J. Monti, L. Pastewka, M. O. Robbins, and J. A. Greenwood, “Meeting the contact-mechanics challenge”, Tribol. Lett. **65**, 118 (2017).

[0] C. Campañá, M. H. Müser, and M. O. Robbins, “Elastic contact between self-affine surfaces: comparison of numerical stress and contact correlation functions with analytic predictions”, J. Phys. Condens. Matter **20**, 354013 (2008).

[52] B. N. J. Persson, “Theory of rubber friction and contact mechanics”, J. Chem. Phys. **115**, 3840 (2001).

[53] B. N. J. Persson and M Scaraggi, “Theory of adhesion: role of surface roughness”, J. Chem. Phys. **141**, 124701 (2014).

[54] N. Prodanov, W. B. Dapp, and M. H. Müser, “On the contact area and mean gap of rough, elastic contacts: Dimensional analysis, numerical corrections, and reference data”, Tribol. Lett. **53**, 433 (2013).

[56] A. Wang and M. H. Müser, “Gauging Persson theory on adhesion”, Tribol. Lett. **65**, 103 (2017).

[57] K. G. Wilson, “Problems in physics with many scales of length”, Sci. Am. **241**, 158 (1979).

[58] J. J. Kalker and Y. V. Randen, “A minimum principle for frictionless elastic contact with application to non-Hertzian half-space contact problems”, J. Eng. Math. **6**, 193 (1972).

[59] H. M. Stanley and T Kato, “An FFT-based method for rough surface contact”, J. Tribol. **119**, 481 (1997).

[60] C. Campañá and M. H. Müser, “Practical Green’s function approach to the simulation of elastic semi-infinite solids”, Phys. Rev. B **74**, 075420 (2006).

[61] A. Wang, Y. Zhou, and M. H. Müser, “Modeling adhesive hysteresis”, Lubricants **9**, 17 (2021).

[62] S. Dalvi, A. Gujrati, S. R. Khanal, L. Pastewka, A. Dhinojwala, and T. D. B. Jacobs, “Linking energy loss in soft adhesion to surface roughness”, Proc. Natl. Acad. Sci. U.S.A. **116**, 25484 (2019).

[63] Y. L. Chen, C. A. Helm, and J. N. Israelachvili, “Molecular mechanisms associated with adhesion and contact angle hysteresis of monolayer surfaces”, J. Phys. Chem. **95**, 10736 (1991).

[64] N. Mulakaluri and B. N. J. Persson, "Adhesion between elastic solids with randomly rough surfaces: Comparison of analytical theory with molecular-dynamics simulations", Europhys. Lett. **96**, 66003 (2011).

[65] S. Medina and D. Dini, "A numerical model for the deterministic analysis of adhesive rough contacts down to the nano-scale", Int. J. Solids Struct. **51**, 2620 (2014).

[66] W. B. Dapp and M. H. Müser, "Contact mechanics of and Reynolds flow through saddle points: On the coalescence of contact patches and the leakage rate through near-critical constrictions", Europhys. Lett. **109**, 44001 (2015).

[67] G. Carbone, E. Pierro, and G. Recchia, "Loading-unloading hysteresis loop of randomly rough adhesive contacts", Phys. Rev. E **92**, 062404 (2015).

[68] W. Deng and H. Kesari, "Molecular statics study of depth-dependent hysteresis in nano-scale adhesive elastic contacts", Modell. Simul. Mater. Sci. Eng. **25**, 055002 (2017).

[69] V. L. Popov, Q. Li, I. A. Lyashenko, and R. Pohrt, "Adhesion and friction in hard and soft contacts: theory and experiment", Friction (2021).

[70] A. Wang and M. H. Müser, "Is there more than one stickiness criterion?", Friction (2022).

[71] B. Zappone, K. J. Rosenberg, and J. Israelachvili, "Role of nanometer roughness on the adhesion and friction of a rough polymer surface and a molecularly smooth mica surface", Tribol. Lett. **26**, 191 (2007).

[72] P. R. Guduru, "Detachment of a rigid solid from an elastic wavy surface: Theory", J. Mech. Phys. Solids **55**, 445 (2007).

[73] Z. Wei, M.-F. He, and Y.-P. Zhao, "The effects of roughness on adhesion hysteresis", J. Adhes. Sci. Technol. **24**, 1045 (2010).

[74] J. A. Greenwood, "Reflections on and extensions of the Fuller and Tabor theory of rough surface adhesion", Tribol. Lett. **65**, 159 (2017).

[75] W. Deng and H. Kesari, "Depth-dependent hysteresis in adhesive elastic contacts at large surface roughness", Sci. Rep. **9**, 1639 (2019).

[76] G. Violano and L. Afferrante, "Roughness-induced adhesive hysteresis in self-affine fractal surfaces", Lubricants **9**, 7 (2021).

[77] A. Gujrati, "Understanding nanoscale surface roughness and its effect on macroscale adhesion", PhD thesis (University of Pittsburgh, 2020).

[78] L. A. Thimons, "Understanding the effects of multi-scale surface roughness on the contact properties of hard-material interfaces", PhD thesis (University of Pittsburgh, 2021).

[79] A. Wang and M. H. Müser, "On the usefulness of the height-difference-autocorrelation function for contact mechanics", Tribol. Int. **123**, 224 (2018).

[80] D. Maksuta, S. Dalvi, A. Gujrati, L. Pastewka, T. D. B. Jacobs, and A. Dhinojwala, "Dependence of adhesive friction on surface roughness and elastic modulus", Soft Matter **18**, 5843 (2022).

[81] M. Ciavarella, "A very simple estimate of adhesion of hard solids with rough surfaces based on a bearing area model", Meccanica **53**, 241 (2018).

[82] J. Joe, M. Thouless, and J. Barber, "Effect of roughness on the adhesive tractions between contacting bodies", J. Mech. Phys. Solids **118**, 365 (2018).

[83] J. R. Rice, "First-order variation in elastic fields due to variation in location of a planar crack front", J. Appl. Mech. **52**, 571 (1985).

[84] J. R. Rice, in *Fracture mechanics: perspectives and directions (twentieth symposium)*, edited by R. Wei and R. Gangloff (American Society for Testing and Materials, Philadelphia, USA, 1989), pp. 29–57.

[85] H. Gao and J. R. Rice, "Nearly circular connections of elastic half spaces", J. Appl. Mech. **54**, 627 (1987).

[86] P. M. Anderson and J. R. Rice, "The stress field and energy of a three-dimensional dislocation loop at a crack tip", J. Mech. Phys. Solids **35**, 743 (1987).

[87] K. L. Johnson, *Contact Mechanics* (Cambridge University Press, 1985).

[88] H. M. Westergaard, "Bearing pressures and cracks: Bearing pressures through a slightly waved surface or through a nearly flat part of a cylinder, and related problems of cracks", J. Appl. Mech. **6**, A49 (1939).

[89] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd, Springer Series in Operations Research (Springer, New York, 2006).

[90] B. V. Derjaguin, "Untersuchungen über die Reibung und Adhäsion, IV", Kolloid-Zeitschrift **69**, 155 (1934).

[91] J. A. Greenwood, "Adhesion of elastic spheres", Proc. R. Soc. London, Ser. A **453**, 1277 (1997).

[92] J. Q. Feng, "Contact behavior of spherical elastic particles: A computational study of particle adhesion and deformations", Colloids Surf., A **172**, 175 (2000).

[93] J.-J. Wu, "The jump-to-contact distance in atomic force microscopy measurement", J. Adhes. **86**, 1071 (2010).

[94] L. Pastewka and M. O. Robbins, "Contact area of rough spheres: Large scale simulations and simple scaling laws", Appl. Phys. Lett. **108**, 221601 (2016).

[95] J. Monti, P. M. McGuiggan, and M. O. Robbins, "Effect of roughness and elasticity on interactions between charged colloidal spheres", Langmuir **35**, 15948 (2019).

[96] A. Ghanbarzadeh, M. Faraji, and A. Neville, "Deterministic normal contact of rough surfaces with adhesion using a surface integral method", Proc. R. Soc. London, Ser. A **476**, 20200281 (2020).

[97] M. H. Müser, "Rigorous Field-Theoretical Approach to the Contact Mechanics of Rough Elastic Solids", Phys. Rev. Lett. **100**, 055504 (2008).

[98] D. Tabor, "Surface forces and surface interactions", J. Colloid Interface Sci. **58**, 2 (1977).

[99] E. Barthel, "Adhesive elastic contacts: JKR and more", J. Phys. D: Appl. Phys. **41**, 163001 (2008).

[100] D. Maugis, *Contact, Adhesion and Rupture of Elastic Solids* (Springer, Berlin; New York, 2010).

[101] M. H. Müser, "Single-asperity contact mechanics with positive and negative work of adhesion: Influence of finite-range interactions and a continuum description for the squeeze-out of wetting fluids", Beilstein J. Nanotechnol. **5**, 419 (2014).

[102] D. Maugis, "Adhesion of spheres: The JKR-DMT transition using a Dugdale model", J. Colloid Interface Sci. **150**, 243 (1992).

[103] L. Pastewka, T. A. Sharp, and M. O. Robbins, "Seamless elastic boundaries for atomistic calculations", Phys. Rev. B **86**, 075459 (2012).

[104] R. W. Hockney, in *Methods in computational physics, Vol. 9*, edited by B. A. Alder, S Fernbach, and M Rotenberg (Academic Press, New York, 1970), pp. 135–211.

[105] S. Liu, Q. Wang, and G. Liu, "A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses", Wear **243**, 101 (2000).

[106] A. E. H. Love, "IX. The stress produced in a semi-infinite solid by pressure on part of the boundary", Philos. Trans. R. Soc. Lond. A **228**, 377 (1929).

[107] M. J. Cannon, D. B. Percival, D. C. Caccia, G. M. Raymond, and J. B. Bassingthwaighte, "Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series", Physica A **241**, 606 (1997).

[108] G. Moreira, J. da Silva, and S. Kamphorst, "On the fractal dimension of self-affine profiles", J. Phys. A **27**, 80794089 (1994).

[109] J. Schmittbuhl, J.-P. Vilotte, and S. Roux, "Reliability of self-affine measurements", Phys. Rev. E **51**, 131 (1995).

[110] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series", Chaos **5**, 82 (1995).

[111] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of DNA nucleotides", Phys. Rev. E **49**, 1685 (1994).

[112] K. M. M. Prabhu, *Window Functions and Their Applications in Signal Processing* (Boca Raton, FL: CRC Press, 2013).

[113] J. M. Elson and J. M. Bennett, "Calculation of the power spectral density from surface profile data", Appl. Opt. **34**, 201 (1995).

[114] B. B. Mandelbrot and J. W. V. Ness, “Fractional brownian motions, fractional noises and applications”, SIAM Rev. **10**, 422 (1968).

[115] P. R. Nayak, “Random process model of rough surfaces”, J. Lubr. Technol. **93**, 398 (1971).

[116] P. Podsiadlo, M. Wolski, and G. W. Stachowiak, “Novel directional blanket covering method for surface curvature analysis at different scales and directions”, Tribol. Lett. **65**, 2 (2017).

[117] P. Podsiadlo, M. Wolski, and G. W. Stachowiak, “Directional signatures of surface texture”, Tribol. Lett. **67**, 109 (2019).

[118] O. Auciello, S. Pacheco, A. V. Sumant, C. Gudeman, S. Sampath, A. Datta, R. W. Carpick, V. P. Adiga, P. Zurcher, Z. Ma, H.-c. Yuan, J. A. Carlisle, B. Kabius, J. Hiller, and S. Srinivasan, “Are diamonds a MEMS’ best friend?”, IEEE Microw. Mag. **8**, 61 (2007).

[119] P. A. Nistor and P. W. May, “Diamond thin films: giving biomedical applications a new shine”, J. R. Soc. Interface **14**, 20170382 (2017).

[120] J. Otschik, A. Schrüfer, J. Thelke, M. Kirchhof, and S. Schmaderer, “Diamond seal faces – the benefits”, World Pumps **2010**, 18 (2010).

[121] H. Zeng, R. Jarvik, G. Catausan, N. Moldovan, and J. Carlisle, “Diamond coated artificial cardiovascular devices”, Surf. Coat. Technol. **302**, 420 (2016).

[122] E. L. Church and P. Z. Takacs, “Effects of the nonvanishing tip size in mechanical profile measurements”, in Proc. SPIE 1332, Optical Testing and Metrology III: Recent Advances in Industrial Optical Inspection (1991), pp. 504–514.

[123] J. M. Boffa, C. Allain, and J. P. Hulin, “Experimental analysis of fracture rugosity in granular and compact rocks”, Eur. Phys. J. Appl. Phys. **2**, 281 (1998).

[124] S. N. Ramakrishna, L. Y. Clasohm, A. Rao, and N. D. Spencer, “Controlling adhesion force by means of nanoscale surface roughness”, Langmuir **27**, 9972 (2011).

[125] O. Laitinen, K. Bauer, J. Niinimäki, and U. A. Peuker, “Validity of the Rumpf and the Rabinovich adhesion force models for alumina substrates with nanoscale roughness”, Powder Technol. **246**, 545 (2013).

[126] A. Kumar, T. Staedler, and X. Jiang, “Role of relative size of asperities and adhering particles on the adhesion force”, J. Colloid Interface Sci. **409**, 211 (2013).

[127] G. Salazar-Banda, M. Felicetti, J. Gonçalves, J. Coury, and M. Aguiar, “Determination of the adhesion force between particles and a flat surface, using the centrifuge technique”, Powder Technol. **173**, 107 (2007).

[128] K. K. Lam and J. M. Newton, “Investigation of applied compression on the adhesion of powders to a substrate surface”, Powder Technol. **65**, 167 (1991).

[129] B. N. J. Persson, “Influence of humidity on the binding of stone fragments via capillary bridges”, Europhys. Lett. **137**, 46001 (2022).

[130] B. N. J. Persson and J. Biele, "On the stability of spinning asteroids", Tribol. Lett. **70**, 34 (2022).

[131] Y. Chen, H. J. Busscher, H. C. van der Mei, and W. Norde, "Statistical analysis of long- and short-range forces involved in bacterial adhesion to substratum surfaces as measured using atomic force microscopy", Appl. Environ. Microbiol. **77**, 5065 (2011).

[132] M. Sztilkovics, T. Gerecsei, B. Peter, A. Saftics, S. Kurunczi, I. Szekacs, B. Szabo, and R. Horvath, "Single-cell adhesion force kinetics of cell populations from combined label-free optical biosensor and robotic fluidic force microscopy", Sci. Rep. **10**, 61 (2020).

[133] P. Tolias, G. Riva, M. D. Angeli, S. Ratynskaia, G. Daminelli, C. Lungu, and C. Porosnicu, "Adhesive force distributions for tungsten dust deposited on bulk tungsten and beryllium-coated tungsten surfaces", Nucl. Mater. Energy **15**, 55 (2018).

[134] G. W. Tormoen, J. Drelich, and J. Nalaskowski, "A distribution of AFM pull-off forces for glass microspheres on a symmetrically structured rough surface", J. Adhes. Sci. Technol. **19**, 215 (2005).

[135] M. Götzinger and W. Peukert, "Particle adhesion force distributions on rough surfaces", Langmuir **20**, 5298 (2004).

[136] Y. I. Rabinovich, J. J. Adler, A Ata, R. K. Singh, and B. M. Moudgil, "Adhesion between nanoscale rough surfaces: I. Role of asperity geometry", J. Colloid Interface Sci. **232**, 10 (2000).

[137] F. P. Bowden and D. Tabor, "Mechanism of metallic friction", Nature **150**, 197 (1942).

[138] C. Mastrangelo and C. Hsu, "A simple experimental technique for the measurement of the work of adhesion of microstructures", in Technical Digest IEEE Solid-State Sensor and Actuator Workshop (1992), pp. 208–212.

[139] A. V. Sumant, D. S. Grierson, J Gerbi, J Carlisle, O Auciello, and R. W. Carpick, "Surface chemistry and bonding configuration of ultrananocrystalline diamond surfaces and their effects on nanotribological properties", Phys. Rev. B **76**, 235429 (2007).

[140] Y. Jiang and K. T. Turner, "Measurement of the strength and range of adhesion using atomic force microscopy", Extreme Mech. Lett. **9**, 119 (2016).

[141] F. W. Delrio, M. P. de Boer, J. A. Knapp, D. E. Reedy Jr., P. J. Clews, and M. L. Dunn, "The role of van der Waals forces in adhesion of micromachined surfaces", Nat. Mater. **4**, 629 (2005).

[142] D. S. Grierson, J. Liu, R. W. Carpick, and K. T. Turner, "Adhesion of nanoscale asperities with power-law profiles", J. Mech. Phys. Solids **61**, 597 (2013).

[143] A. E. Kovalev and S. N. Gorb, "Charge contribution to the adhesion performance of polymeric microstructures", Tribol. Lett. **48**, 103 (2012).

[144] T. Tang, C.-Y. Hui, and A. Jagota, "Adhesive contact driven by electrostatic forces", J. Appl. Phys. **99**, 054906 (2006).

[145] M. Feshanjerdi and A. Malekan, "Contact electrification between randomly rough surfaces with identical materials", J. Appl. Phys. **125**, 165302 (2019).

[146] M. P. de Boer and P. C. T. de Boer, "Thermodynamics of capillary adhesion between rough surfaces", J. Colloid Interface Sci. **311**, 171 (2007).

[147] F. W. DelRio, M. L. Dunn, and M. P. de Boer, "Capillary adhesion model for contacting micromachined surfaces", Scr. Mater. **59**, 916 (2008).

[148] M. Sedighi, V. B. Svetovoy, and G. Palasantzas, "Capillary-force measurement on SiC surfaces", Phys. Rev. E **93**, 062803.

[149] J. Joe, M. Scaraggi, and J. R. Barber, "Effect of fine-scale roughness on the tractions between contacting bodies", Tribol. Int. **111**, 52 (2017).

[150] A. Almqvist, C. Campañá, N. Prodanov, and B. N. J. Persson, "Interfacial separation between elastic solids with randomly rough surfaces: Comparison between theory and numerical techniques", J. Mech. Phys. Solids **59**, 2355 (2011).

[151] G. Violano, L. Afferrante, A. Papangelo, and M. Ciavarella, "On stickiness of multiscale randomly rough surfaces", J. Adhes. **97**, 509 (2019).

[152] A. Tiwari, J. Wang, and B. N. J. Persson, "Adhesion paradox: why adhesion is usually not observed for macroscopic solids", Phys. Rev. E **102**, 042803 (2020).

[153] M. K. Chaudhury and G. M. Whitesides, "Direct measurement of interfacial interactions between semispherical lenses and flat sheets of poly(dimethylsiloxane) and their chemical derivatives", Langmuir **7**, 1013 (1991).

[154] A. Tiwari, L. Dorogin, A. I. Bennett, K. D. Schulze, W. G. Sawyer, M. Tahir, G. Heinrich, and B. N. J. Persson, "The effect of surface roughness and viscoelasticity on rubber adhesion", Soft Matter **13**, 3602 (2017).

[155] I. N. Sneddon, "Boussinesq's problem for a flat-ended cylinder", Math. Proc. Cambridge Philos. Soc. **42**, 29 (1946).

[156] G. R. Irwin, "Analysis of stresses and strains near the end of a crack transversing a plate", J. Appl. Mech. **24**, 361 (1957).

[157] V. M. Muller, V. S. Yushchenko, and B. V. Derjaguin, "On the influence of molecular forces on the deformation of an elastic sphere and its sticking to a rigid plane", J. Colloid Interface Sci. **77**, 91 (1980).

[158] H. Kesari and A. J. Lew, "Effective macroscopic adhesive contact behavior induced by small surface roughness", J. Mech. Phys. Solids **59**, 2488 (2011).

[159] V. Lazarus, "Perturbation approaches of a planar crack in linear elastic fracture mechanics: A review", J. Mech. Phys. Solids **59**, 121 (2011).

[160] A. Salvadori and F. Fantoni, "Weight function theory and variational formulations for three-dimensional plane elastic cracks advancing", Int. J. Solids Struct. **51**, 1030 (2014).

[161] C. Caroli and P. Nozières, "Hysteresis and elastic interactions of microasperities in dry friction", Eur. Phys. J. B **4**, 233 (1998).

[162] B. N. J. Persson and E. Tosatti, “Theory of friction: Elastic coherence length and earthquake dynamics”, Solid State Commun. **109**, 739 (1999).

[163] J. F. Joanny and P. G. de Gennes, “A model for contact angle hysteresis”, J. Chem. Phys. **81**, 552 (1984).

[164] M. O. Robbins and J. F. Joanny, “Contact angle hysteresis on random surfaces”, Europhys. Lett. **3**, 729 (1987).

[165] H. Gao and J. R. Rice, “A first-order perturbation analysis of crack trapping by arrays of obstacles”, J. Appl. Mech. **56**, 828 (1989).

[166] V. Démery, A. Rosso, and L. Ponson, “From microstructural features to effective toughness in disordered brittle solids”, Europhys. Lett. **105**, 34003 (2014).

[167] Y. Imry and S.-k. Ma, “Random-field instability of the ordered state of continuous symmetry”, Phys. Rev. Lett. **35**, 1399 (1975).

[168] A. I. Larkin and Y. N. Ovchinnikov, “Pinning in type II superconductors”, J. Low. Temp. Phys. **34**, 409 (1979).

[169] D. S. Fisher, “Threshold behavior of charge-density waves pinned by impurities”, Phys. Rev. Lett. **50**, 1486 (1983).

[170] A. A. Middleton, “Asymptotic uniqueness of the sliding state for charge-density waves”, Phys. Rev. Lett. **68**, 670 (1992).

[171] L. A. N. Amaral, A.-L. Barabasi, H. A. Makse, and H. E. Stanley, “Scaling properties of driven interfaces in disordered media”, Phys. Rev. E **52**, 4087 (1995).

[172] Y. Zhou, “Critical dynamics of contact lines”, PhD thesis (Johns Hopkins University, Baltimore, Maryland, USA, 2000).

[173] A. Tanguy and T. Vettorel, “From weak to strong pinning I: a finite size study”, Eur. Phys. J. B **38**, 71 (2004).

[174] A. Rosso, P. Le Doussal, and K. J. Wiese, “Numerical calculation of the functional renormalization group fixed-point functions at the depinning transition”, Phys. Rev. B **75**, 220201 (2007).

[175] P. Dondl and K. Bhattacharya, “Effective behavior of an interface propagating through a periodic elastic medium”, Interfaces Free Boundaries **18**, 91 (2016).

[176] A. B. Kolton, S. Bustingorry, E. E. Ferrero, and A. Rosso, “Uniqueness of the thermodynamic limit for driven disordered elastic interfaces”, J. Stat. Mech., P12004 (2013).

[177] V. Démery, V. Lecomte, and A. Rosso, “The effect of disorder geometry on the critical force in disordered elastic systems”, J. Stat. Mech., P03009 (2014).

[178] Q. Li, R. Pohrt, and V. L. Popov, “Adhesive strength of contacts of rough spheres”, Front. Mech. Eng. **5**, 7 (2019).

[179] I. A. Lyashenko and V. L. Popov, “The effect of contact duration and indentation depth on adhesion strength: experiment and numerical simulation”, **65**, 13 (2020).

[180] P. R. Guduru and C Bull, "Detachment of a rigid solid from an elastic wavy surface: Experiments", J. Mech. Phys. Solids **55**, 473 (2007).

[181] J. A. Greenwood and K. L. Johnson, "The mechanics of adhesion of viscoelastic solids", Philos. Mag. A **43**, 697 (1981).

[182] M. K. Chaudhury, T. Weaver, C. Y. Hui, and E. J. Kramer, "Adhesive contact of cylindrical lens and a flat sheet", J. Appl. Phys. **80**, 30 (1996).

[183] B. Lorenz, B. A. Krick, N. Mulakaluri, M. Smolyakova, S. Dieluweit, W. G. Sawyer, and B. N. J. Persson, "Adhesion: role of bulk viscoelasticity and surface roughness", J. Phys.: Condens. Matter **25**, 225004 (2013).

[184] J. A. Greenwood, "The theory of viscoelastic crack propagation and healing", J. Phys. D: Appl. Phys. **37**, 2557 (2004).

[185] B. N. J. Persson and E. A. Brener, "Crack propagation in viscoelastic solids", Physical Review E **71**, 036123 (2005).

[186] J. Chopin, A. Bhaskar, A. Jog, and L. Ponson, "Depinning dynamics of crack fronts", Phys. Rev. Lett. **121**, 235501 (2018).

[187] G. Albertini, M. Lebihain, F. Hild, L. Ponson, and D. S. Kammer, "Effective toughness of heterogeneous materials with rate-dependent fracture energy", Phys. Rev. Lett. **127**, 035501 (2021).

[188] M. H. Müser and B. N. J. Persson, "Crack and pull-off dynamics of adhesive, viscoelastic solids", Europhys. Lett. **137**, 36004 (2022).

[189] F. Pérez-Ràfols, J. S. V. Dokkum, and L. Nicola, "On the interplay between roughness and viscoelasticity in adhesive hysteresis", J. Mech. Phys. Solids **170**, 105079 (2023).

[190] C. Müller, M. Samri, R. Hensel, E. Arzt, and M. H. Müser, "Revealing the co-action of viscous and multistability hysteresis in an adhesive, nominally flat punch: a combined numerical and experimental study", Preprint at `https://arxiv.org/abs/2211.12963`, (2022).

[191] A. F. Bower and M. Ortiz, "Solution of three-dimensional crack problems by a finite perturbation method", J. Mech. Phys. Solids **38**, 443 (1990).

[192] A. F. Bower and M. Ortiz, "A three-dimensional analysis of crack trapping and bridging by tough particles", J. Mech. Phys. Solids **39**, 815 (1991).

[193] J.-B. Leblond, S. Patinet, J. Frelat, and V. Lazarus, "Second-order coplanar perturbation of a semi-infinite crack in an infinite body", Eng. Fract. Mech. **90**, 129 (2012).

[194] M. Lebihain, "Large-scale crack propagation in heterogeneous materials: an insight into the homogenization of brittle fracture properties", PhD thesis (Sorbonne Université, 2019).

[195] M. Lebihain, L. Ponson, D. Kondo, and J.-B. Leblond, "Effective toughness of disordered brittle solids: A homogenization framework", J. Mech. Phys. Solids **153**, 104463 (2021).

[196] A. Carlson, A. M. Bowen, Y. Huang, R. G. Nuzzo, and J. A. Rogers, "Transfer printing techniques for materials assembly and micro/nanodevice fabrication", Adv. Mater. **24**, 5284 (2012).

[197] R. Hensel, K. Moh, and E. Arzt, "Engineering micropatterned dry adhesives: from contact theory to handling applications", Adv. Funct. Mater. **28**, 1800865 (2018).

[198] N. Deneke, A. L. Chau, and C. S. Davis, "Pressure tunable adhesion of rough elastomers", Soft Matter **17**, 863 (2021).

[199] J. M. Karp and R Langer, "Dry solution to a sticky problem", Nature **477**, 42 (2011).

[200] M. K. Kwak, H.-E. Jeong, and K. Y. Suh, "Rational design and enhanced biocompatibility of a dry adhesive medical skin patch", Adv. Mater. **23**, 3949 (2011).

[201] Z. Yu and H. Cheng, "Tunable adhesion for bio-integrated devices", Micromachines **9**, 529 (2018).

[202] D.-G. Hwang, K. Trent, and M. D. Bartlett, "Kirigami-inspired structures for smart adhesion", ACS Appl. Mater. Interfaces **10**, 6747 (2018).

[203] R. Maboudian, "Adhesion and friction issues associated with reliable operation of MEMS", MRS Bull. **23**, 47 (1998).

[204] T. S. Olsson and M. Hartley, "Lightweight data management with dtool", PeerJ **7**, e6562 (2019).

[205] J. L. Hörmann and L. Pastewka, "Lightweight research data management with dtool : a use case", Universität Ulm, 29 (2022).

Zusammenfassung

Obwohl sich an der atomaren Skala alle Moleküle anziehen, haften die meisten makroskopischen Objekte nicht. Die Erklärung für dieses scheinbare Paradox ist, dass die meisten Oberflächen rau sind, so dass sich Objekte mit steifen elastischen Eigenschaften nur auf ihren höchsten Rauheitsspitzen berühren. Die Fähigkeit von Geckos und Insekten, senkrechte Wände hochzuklettern beruht darauf, dass ihre Fußspitzen nachgiebige fibrilläre Strukturen oder weiche Polster besitzen, die sich an die Rauheit anschmiegen können. Die genaue Rolle von Oberflächenrauheit in Haftung zu verstehen ist eine Herausforderung, weil Oberflächen bis zur atomaren Skala rau sind.

Meine Kollaboratoren und ich untersuchen den Einfluss von Oberflächenrauheit auf die Haftung sowohl steifer als auch weicher Materialien. Ich entwickle theoretische Modelle und helfe Experimentatoren die Topographie skalenübergreifend zu analysieren. Indem wir meine neuen Modelle und Abhijeet Gujrati's skalenübergreifende Charakterisierung von Oberflächentopographie kombinieren, gelingt es uns, die Rolle von Oberflächenrauheit in Adhäsionsexperimenten zu entziffern.

Steife Materialien haften nicht, weil die Rauheit den größten Teil der Oberfläche außerhalb der Reichweite molekularer Attraktionskräfte hält. Eine letztlich entwickelte Theorie quantifiziert diesen Effekt mithilfe einer Näherungsgleichung für die Wahrscheinlichkeitsverteilung von Spalthöhen nahe der Kontaktkante. Die Theorie ist akkurat unter der Voraussetzung, dass adhäsive Spannungen schwach sind, und sie über eine geringere Reichweite wirken als ein geometrischer Parameter, der von der Rauheit auf kleinen Skalen abhängt.

Weiche Objekte (wie Pudding) haften, weil die elastische Verformungsenergie, die benötigt wird, um sich der Rauheit anzuschmiegen, klein ist im Vergleich zur Oberflächenenergie, die durch intimen Kontakt gewonnen wird. Auf diesem simplen thermodynamischen Argument beruhende Theorien können jedoch nicht erklären, warum die Kraft, die benötigt wird um die Objekte zu trennen, in vielen Experimenten deutlich größer ist als die Kraft, die während der Entstehung des Kontaktes gemessen wird. Diese *Adhäsionshysterese* wird oft durch materialspezifische Irreversibilität erklärt, sie kann aber auch aufgrund von elastischen Instabilitäten entstehen, die von der Rauheit ausgelöst werden. Die Rolle dieser elastischen Instabilitäten in Adhäsionshysterese ist stets unklar weil existierende numerische und analytische Modelle keine realistische Rauheit berücksichtigen können. Ich führe ein effizientes Riss-Perturbationsmodell ein, das große Simulationen mit realistischer Oberflächenrauheit ermöglicht. Dieses Modell etabliert die Analogie zwischen Haftung von rauen Oberflächen und anderen, schon besser verstandenen Phänomenen, wie zum Beispiel die Propagation von Rissen in heterogenen Materialien oder die Kontaktwinkelhysterese von Tropfen. Ich verwende Erkenntnisse über diese verwandten Phänomene um eine simple Theorie herzuleiten, die die Hysterese als Funktion der statistischen Eigenschaften der Oberflächentopographie vorhersagt. In Kombination mit der skalenübergreifenden Oberflächencharakterisierung klärt mein

Modell die Rolle elastischer Instabilitäten in Adhäsionsexperimenten auf.

Weil Oberflächen von der makroskopischen Skala bis zur atomaren Skala rau sind, ist der Mangel an skalenübergreifender Oberflächencharakterisierung das größte Hindernis, Experimente und Theorie zu vereinbaren. Abhijeet Gujrati und Kollaboratoren haben die Topographie von vier Diamantbeschichtungen über acht Dekaden von Längenskalen gemessen. Damit haben sie es ermöglicht, Adhäsionstheorien mit Messungen der Haftung dieser Beschichtungen zu vergleichen. Nicht nur Rauheit auf der atomaren Skala zu messen ist eine Herausforderung, auch die Komplexität etablierter multiskaliger Rauheitsmetriken erschwert die Dokumentation von Oberflächentopographie. Wir gehen dieses Problem an, indem wir einen neuen Ansatz zur Analyse der Topographie einführen, der einfach zu interpretieren und zu implementieren ist. Wir nennen diese Analysen "*scale-dependent roughness parameters*" (SDRPs), ins Deutsche übersetzt *skalenabhängige Rauheitsparameter*. Die SDRPs, zusammen mit weiteren, etablierten Rauheitsmetriken, können auch auf unserer freien web-Plattform *contact.engineering* berechnet werden. *contact.engineering* ist eine Datenbank, die es Wissenschaftlern ermöglicht, Messungen der Oberflächentopographie zu publizieren und in einem vereinheitlichten Verfahren zu analysieren. Wir regen die wissenschaftliche Gemeinschaft dazu an, Oberflächentopographie über mehrere Skalen zu messen, zu analysieren und zu publizieren, indem wir diesen Prozess vereinfachen. Dadurch fördern wir den Ausbau des Wissens über die Rauheit von Oberflächen.

Die SDRP-Analyse berechnet die Fluktuationen der Steigung und der Krümmung der Oberfläche auf unterschiedlichen Längenskalen. Steigungen und Krümmungen sind wichtige Parameter in Theorien zum Kontakt von rauen Oberflächen, jedoch ist stets unklar, auf welcher Längenskala sie wichtig sind. Luke Thimons hat die Ablösekraft zwischen makroskopischen Rubinkugeln und Diamantbeschichtungen gemessen, und wir zeigen, dass die Rauheit auf lateralen Skalen zwischen 43 nm und 1.8 μm diese Haftung kritisch beeinflusst. Kleine Längenskalen sind unwichtig, weil die Rubinkugel sich plastisch verformt, und weil die Reichweite der Attraktionskräfte lang ist ($\simeq$5 nm). Die Oberflächentopographie auf Skalen länger als 1.8 μm spielt aufgrund des endlichen Radius der Kugel keine Rolle. Wir bestimmen den kritischen Längenbereich der Oberflächenrauheit sowie die Parameter der adhäsiven Interaktionskräfte, indem wir die experimentell gemessenen Ablösekräfte mit Vorhersagen aus numerischen Simulationen vergleichen. Die numerischen Simulationen sind mithilfe der skalenübergreifenden Topographiecharakterisierung von Abhijeet Gujrati parametrisiert.

Die Forschung über Haftung findet Anwendung in mikroelektromechanischen Systemen (MEMS), Softrobotik und Hautklebebändern. Die Erkenntnisse aus dieser Dissertation geben Leitlinien für Ingenieure, auf welcher Längenskala sie die Rauheit ändern müssen, um die Haftung zu kontrollieren. Unsere Werkzeuge zur Analyse von Topographie fördern ein besseres Verständnis von Oberflächenrauheit in der wissenschaftlichen Gemeinschaft.

Publication I

Why breaking soft contacts is harder than making them

Antoine Sanner[1,2] and Lars Pastewka[1,2,*]

[1] *Department of Microsystems Engineering (IMTEK), University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany*

[2] *Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany*

Insects, pick-and-place systems, engineered adhesives, and soft robots employ soft materials to stick to surfaces even in the presence of roughness. Experiments ubiquitously show that the force required for making contact is lower than for releasing it, a phenomenon known as the adhesion hysteresis.[1,2] The common explanation for this hysteresis is either contact aging or viscoelasticity, ignoring the influence of surface roughness.[3,4] Here, we use an efficient numerical crack-perturbation model[5–7] to simulate the contact of soft solids on rough surfaces across orders of magnitude in length scale, allowing us to study realistic rough geometries.[8,9] The simulations show that adhesion hysteresis emerges even for perfectly elastic contacts and in the absence of contact aging and viscoelasticity. The adhesive contact line is pinned by surface heterogeneity,[10–12] leading to a hysteresis between advancing and receding motion. Our model quantitatively reproduces the hysteresis observed in experiments and allows us to derive analytical predictions for its magnitude, requiring as input only simple statistical measures of the rough geometry.[13] Our results explain why adhesion hysteresis is ubiquitous and show that soft pads in nature and engineering are efficient in adhering even to surfaces with significant roughness.

Two solids stick to each other because of attractive van-der-Waals or capillary interactions at small scales.[3] The strength of these interactions is commonly described by the intrinsic work of adhesion w_{int}, the energy that is gained by these interactions per surface area of intimate contact. While for hard substrates roughness limits this area to the highest protrusions,[14,15] soft solids are sticky because they can deform to come into contact over a large portion of the rough topography. The overall strength of the adhesive joint is then determined by the balance of the energy gained by making contact and the elastic energy spent for conforming to the surface. In the presence of roughness, this energy balance defines an effective work of adhesion,[16]

$$W = w_{\mathrm{int}} - e_{\mathrm{el}}, \tag{1}$$

where e_{el} is the elastic energy per unit contact area required to conform to the roughness. This effective work of adhesion is most commonly measured from the pull-off force $F_{\mathrm{pulloff}} = -3\pi WR/2$ of a soft spherical probe (see Fig. 1a) with radius R.[17] As shown in Fig. 1b, experiments typically follow different paths during approach and retraction, leading to different effective works of adhesion for making and breaking contact.

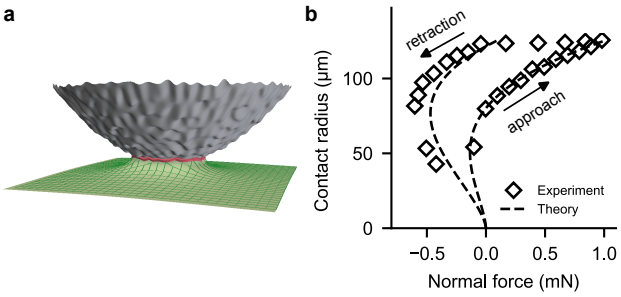


FIG. 1. **Phenomenology of adhesive contact. (a)** Many contacts can be described as spheres making contact with a flat surface. For soft materials, microscopic interactions are strong enough that they deform significantly near the contact edge. **(b)** Dependency of contact radius versus normal force. The pull-off force is the most negative force on these curves. The diamonds are experimental results from Dalvi et al.,[2] that show hysteresis between approach and retraction. The dashed lines are obtained from the JKR model, with two distinct apparent works of adhesion for approach w_{appr} and for retraction w_{retr}. The values for w_{appr} and w_{retr} are a prediction of our theory, Eq. (6), with only a single adjustable parameter, the intrinsic work of adhesion w_{int}. The rough surface is nanocrystalline diamond (NCD), and the sphere has Young's modulus $E = 0.7$ MPa and radius $R = 1.2$ mm.

In this letter, we present an analytic theory that allows us to predict these effective works of adhesion and thereby the adhesive hysteresis. For soft spherical probes, we can describe the circular contact perimeter as a crack (see red line in Fig. 1a). The crack front is in equilibrium when Griffith's criterion is fulfilled:[18] The energy per unit area required for opening the crack, W, is equal to the energy released from the elastic deformation, $G\delta A = W\delta A$, where δA is the contact area swept by the crack front. A more common way of writing this equation is in terms of a balance of forces rather than energies,

$$G = W, \tag{2}$$

where both the energy release rate G and W should be interpreted as forces per unit crack length. Johnson, Kendall and Roberts (JKR)[17] derived the expression for

* lars.pastewka@imtek.uni-freiburg.de

the energy release rate G for a smooth spherical indenter, $G = G_{\rm JKR}(b, a)$. Equation (2) then allows to evaluate not just the pull-off force, but all functional dependencies between rigid body displacement b, contact radius a and normal force F during contact. In the presence of roughness, we will show below that W becomes a local quantity that describes the effective work of adhesion at a specific position on the surface. Equation (2) must then hold independently for each point on the contact perimeter.

In order to illustrate why hysteresis emerges from spatial fluctuations in W, we first discuss a simple toy model. We assume that $W(a)$, rather than being random, varies in concentric rings of wavelength d as a function of distance a from the apex of the contacting sphere (Fig. 2a). Figure 2b also shows $W(a)$ alongside $G_{\rm JKR}(b, a)$ for a fixed b. Because of the spatial variations of W, there are multiple solutions to Eq. (2) indicated by the labels A and B. Moving into contact from the solution denoted by A leads to an instability where the solution A disappears, at which the contact radius jumps to the next ring of $W(a)$. This samples the lower values of W shown by the green line in Fig. 2b. Conversely, moving out of contact progresses along a different path that samples the higher values of $W(a)$, shown by the red line. The combination of fluctuations in W and the elastic restoring force $G_{\rm JKR}$ acts like a ratchet that makes increasing the contact area easier than decreasing it.

In the limit of roughness with small wavelength, $d \to 0$, $G_{\rm JKR}$ does not decrease significantly before the contact line arrests at the next peak (see Fig. 2c). In this limit, the contact line samples the minimum values $w_{\rm appr}$ of W during approach and the maximum values $w_{\rm retr}$ during retraction. The functional relationship between b, a and F then becomes identical to the JKR solution for smooth bodies, but with a work of adhesion that differs between approach ($w_{\rm appr}$) and retraction ($w_{\rm retr}$, see Fig. 2d). In the limit of small wavelength, the hysteresis $w_{\rm retr} - w_{\rm appr}$ becomes equal to the peak-to-peak amplitude of $W(a)$.[21]

For non-axisymmetric work of adhesion fields, the contact line is no longer perfectly circular (see Fig. 3a). The energy release rate G then becomes a functional of the contact shape $a(s)$, here parameterized by the length s of the corresponding path along the contact circle. Based on the crack-perturbation theory by Gao and Rice,[5,7,22] we recently derived the approximate expression $G([a]; s) = G_\circ(a(s)) + G_\parallel([a]; s)$ with[5,6]

$$G_\parallel([a]; s) = c(-\Delta_s)^{1/2} a(s), \tag{3}$$

where we used the square brackets to indicate a functional dependency. Equation (3) can be interpreted as the restoring force of an elastic line that penalizes excursions from circularity through the fractional Laplacian $(-\Delta_s)^{1/2}$. The fractional Laplacian represents the effect of the non-local interaction of surface displacements on the energy release rate, and Fig. 3b illustrates this effect. Supplementary Section S-I derives this expression and shows that near equilibrium the stiffness of the line is given by $c = w_{\rm int}$.

Numerical solution of Eq. (2) (see Supplementary Section S-II) on a random field $W(x, y)$ with lateral correlation of length d yields force-area curves similar to those of our toy model. The key difference is that the contact line now advances and recedes in jumps (Fig. 3a) that are localized over a spatial length ℓ, the Larkin length.[10–12,23,24] Between these jumps, the contact line is pinned. At the same rigid body penetration, pinning occurs at lower contact radii in approach than during retraction, leading to a hysteresis in apparent adhesion described by two JKR curves with constant apparent work of adhesion $w_{\rm appr}$ and $w_{\rm retr}$ (Fig. 3c), similar to the curves obtained from our 1D toy model (Fig. 2d). Our numerical data shows that the magnitude of hysteresis, $w_{\rm retr} - w_{\rm appr} \propto W_{\rm rms}^2$ (Fig. 3d).

We now summarize why this scaling emerges in our adhesive problem, starting with the discussion of two limiting cases. In the limit $c \to 0$ the line is floppy and deviations from circularity are not penalized. In this individual-pinning limit,[12,25,26] each angle θ along the contact perimeter independently yields our toy model and we obtain $w_{\rm retr} - w_{\rm appr} \propto W_{\rm rms}$. In the opposite limit, $c \to \infty$ the line is stiff and the contact remains circular. There is no hysteresis, $w_{\rm retr} - w_{\rm appr} = 0$, and the contact radius is obtained from the JKR expression evaluated for the mean work of adhesion, $\langle W \rangle$.

Our simulations are in an intermediate regime, characterized by local jumps over length ℓ or $N = \ell/d$ pinning sites. The line is stiff over ℓ and hence samples a coarse-grained work of adhesion field $W^{(\ell)}$ with $W_{\rm rms}^{(\ell)} = W_{\rm rms}/\sqrt{N}$. From force-balance Eq. (3) we obtain that an excursion of the contact line by distance δa over this length leads to a restoring force $\delta G \propto c\delta a/\ell$, which must balance $W_{\rm rms}^{(\ell)}$. We note that $\delta a \approx d$, which is the distance to the closest local stable configuration.[10,11] The equilibrium condition $\delta G = W_{\rm rms}^{(\ell)}$ then yields

$$N \propto \left(c/W_{\rm rms}\right)^2 \tag{4}$$

or $\ell = Nd$ for the Larkin length. This means, the magnitude of the hysteresis must scale as

$$w_{\rm retr} - w_{\rm appr} \propto W_{\rm rms}^{(\ell)} \propto W_{\rm rms}^2/c, \tag{5}$$

exactly as observed in our simulations. Identical results were obtained previously for cracks in heterogeneous media.[12,27]

We now turn to discussing topographic roughness, rather than heterogeneity in the work of adhesion. For this we need to consider excursions of the contact line normal to the surface in additional to the lateral excursions that are described by the contact radius $a(\theta)$ (see Fig. 4)a. First note that the solid is always dilated near the crack tip. In order to conform to a valley, the elastic solid needs to stretch even more, requiring elastic energy. Using the same arguments that lead to Eq. (1), this additional elastic energy manifests as an effective decreased

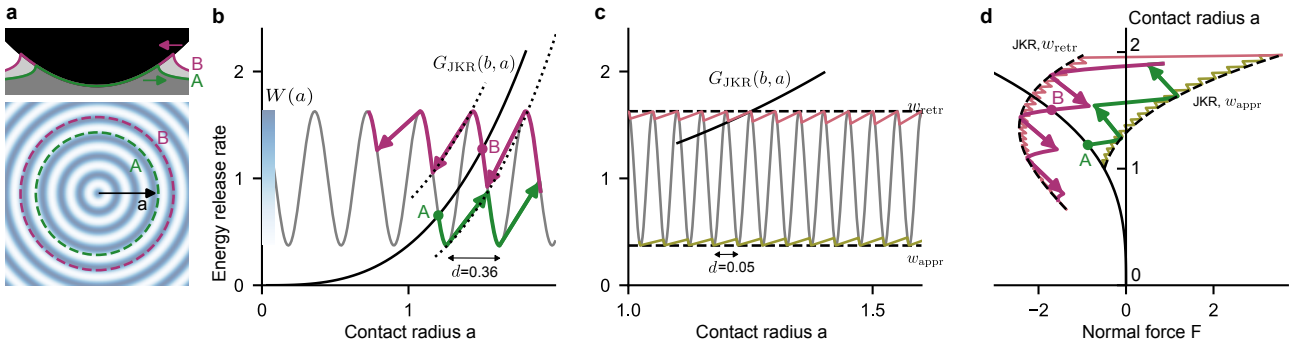


FIG. 2. **Contact of a sphere against an axisymmetric work of adhesion heterogeneity $W(a)$ with wavelength d. (a)** Cross-section of the contact at rigid body penetration $b = 0$ (top) and top view of the axisymmetric work of adhesion heterogeneity $W(a)$ (bottom). The blue color indicates regions of high adhesion. **(b)** Elastic energy release rates in an indentation retraction cycle for a sinusoidal work of adhesion $W(a)$ with wavelength $d = 0.36$ (gray line). The black line shows the elastic energy release rate $G_{\mathrm{JKR}}(b, a)$ as a function of contact radius for fixed rigid body penetration $b = 0$. Fluctuations of $W(a)$ lead to several metastable states A, B at fixed b. During approach, the contact perimeter is pinned in metastable states with low adhesion (green curve), while during retraction the contact perimeter is pinned at higher radii by adhesion peaks (red curve). Arrows indicate elastic instabilities where the contact radius jumps between metastable states. **(c)** Energy release rates in an indentation retraction cycle for a work of adhesion heterogeneity with smaller wavelength $d = 0.05$. Note that the slope of G_{JKR} appears to be flatter than in panel (b) because we show a smaller range of contact radii. For short wavelengths, the work of adhesion sampled during approach (light green curve) and retraction (light red curve) stay close to the constant values w_{appr} and w_{retr}. **(d)** The contact radius and the normal force during an indentation retraction cycle for wavelength $d = 0.36$ (darker colors) and $d = 0.05$ (lighter colors). The dashed lines are the prediction by the JKR theory using w_{retr} and w_{appr} for the work of adhesion. The solid black line corresponds to increasing energy release rates at fixed rigid body penetration $b = 0$. Energy release rates are displayed in units of w_{int} and lengths and forces have been nondimensionalized following the conventions of Refs.[19,20] as described in the Supplemental Material.

local work of adhesion. Conversely, conforming to a peak decreases the overall strain near the crack tip and releases elastic energy, leading to an increased effective work of adhesion. Supplementary Section S-I derives an integral transformation for transforming a topographic field $h(x, y)$ into an effective work of adhesion field $W_{\mathrm{eff}}(x, y)$. Supplementary Section S-III also shows that a crack-front simulation on $W_{\mathrm{eff}}(x, y)$ yields results virtually indistinguishable from an exact boundary-element calculation.

The effective work of adhesion field has the intuitive property that its mean corresponds to the Persson-Tosatti expression, Eq. (1). Furthermore, it has local fluctuations with amplitude $W_{\mathrm{rms}} = \sqrt{2 w_{\mathrm{int}} e_{\mathrm{el}}}$ that determine the adhesion hysteresis. These arguments on simple roughness suggest that the main parameter determining the hysteresis is e_{el}. We carried out crack-front simulations on self-affine randomly rough topographies (Fig. 3d) with varying parameters to confirm that the work of adhesion during approach and retraction is indeed given by

$$w_{\substack{\mathrm{ret}\\\mathrm{appr}}} = w_{\mathrm{int}} - e_{\mathrm{el}} \pm k e_{\mathrm{el}}, \tag{6}$$

with a numerical factor of $k \approx 3$.

Our theoretical prediction reproduces the hysteresis measured by Dalvi et al.[2] when contacting hard rough diamond substrates with soft rubber spheres. The roughness of the diamond substrates used in Dalvi's experiments were characterized from atomic to macroscopic scales,[8,9] allowing us to quantitatively test our theory. Since there is no direct way to measure the intrinsic work of adhesion, we extracted $w_{\mathrm{int}} = 65$ mJ m^{-2} by fitting our w_{appr} to the data, which is in the range expected for van der Waals interaction. The final prediction of our model is shown by the dashed line in Fig. 1b. The good agreement of the corresponding retraction curve, w_{retr}, shows that crack pinning quantitatively reproduces the hysteresis observed in this experiment.

The central parameter governing hysteresis is the elastic energy for fully conformal contact. This energy can written as

$$e_{\mathrm{el}} = \frac{E'}{4} \left[h_{\mathrm{rms}}^{(1/2)} \right]^2, \tag{7}$$

where E' is the elastic contact modulus[28] and $h_{\mathrm{rms}}^{(1/2)}$ is a geometric descriptor of the rough topography. In terms of the power-spectral density of the rough topography[13] C^{iso}, we define

$$(h_{\mathrm{rms}}^{(\alpha)})^2 = \frac{1}{4\pi^2} \int \mathrm{d}^2 q \, |\vec{q}|^{2\alpha} C^{\mathrm{iso}}(|\vec{q}|), \tag{8}$$

where $\vec{q}$ is the wavevector. This expression contains the rms amplitude of the topography, $h_{\mathrm{rms}}^{(0)}$, the rms gradient of the topography, $h_{\mathrm{rms}}^{(1)}$, as well as arbitrary derivatives of order α. The elastic energy is given by a roughness parameter intermediate between rms heights and rms gra-

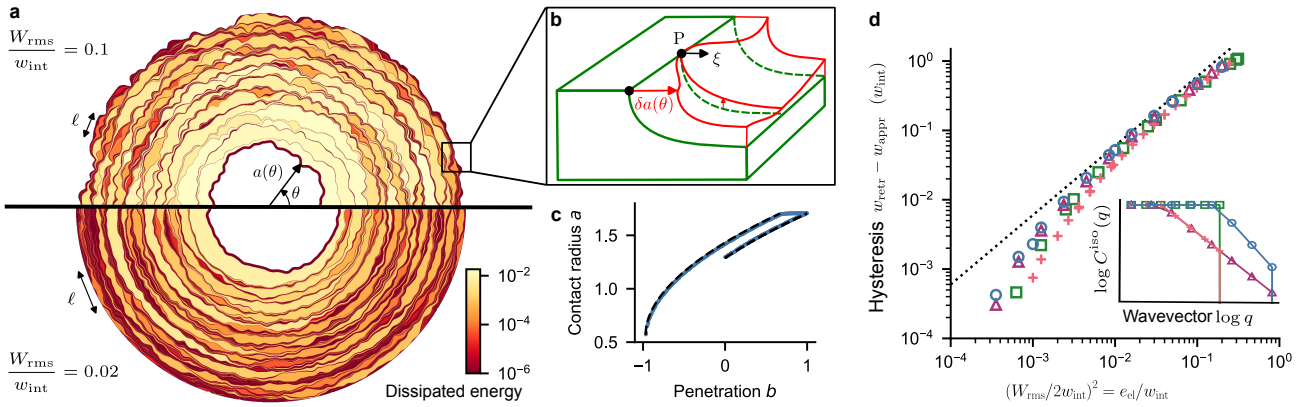


FIG. 3. **Crack-front pinning by two-dimensional random roughness. (a)** Evolution of the contact line during retraction in a crack-front simulation on 2D random roughness. Each colored patch corresponds to an elastic instability during which the perimeter umps between two pinned configurations (dark lines). The color scale represents the energy dissipated during each instability. The random roughness has a flat power spectral density with short-wavelength cutoff $\lambda_s = 0.07$. **(b)** Effect of in-plane perturbation of the perimeter on the elastic energy release rate. The green solid represents a small section of the JKR contact with constant radius that we perturb by $\delta a(\theta)$ (red). Moving the crack front within the plane leads to out-of-plane deformations of the crack faces. In a point P that we hold fixed, the energy release rate G is perturbed due to the non-local interaction of surface displacements, which is why G is a *functional* of the contact radius $a(\theta)$. **(c)** Contact radius as a function of the rigid body penetration in a simulation on self-affine random roughness. The shape of the PSD is shown by the blue circles in the inset of panel (d) and the elastic energy for fully conformal contact $e_{\rm el}/w_{\rm int} = 0.05$. The dashed lines are JKR curves with work of adhesion $w_{\rm appr}$ and $w_{\rm retr}$ predicted by our theory Eq. (6). **(d)** Work of adhesion hysteresis as a function of the elastic energy for fully conformal contact $e_{\rm el}$, or equivalently the variance of the effective work of adhesion $W^2_{\rm rms}$. The dashed line shows the hysteresis predicted by our theory, Eq. (6). The symbols show results of crack-front simulation on randomly rough surfaces with different shapes of the power-spectra represented in the inset. We used a flat PSD with short-wavelength cutoff $\lambda_s = 0.005$ (green squares), and three different self-affine PSDs parameterized by the rolloff wavelength λ_r, the short-wavelength cutoff λ_s and the Hurst exponent H. The blue circles correspond to $\lambda_s = 0.000625, \lambda_r = 0.01, H = 0.8$; the purple triangles to $\lambda_s = 0.000625, \lambda_r = 0.1, H = 0.3$; and the pink crosses to $\lambda_s = 0.005, \lambda_r = 0.1, H = 0.3$. We varied $e_{\rm el}$ by scaling the amplitudes of the heights. Units have been nondimensionalized following the conventions of Refs.[19,20] as described in the Supplemental Material.

dients. For a self-affine topography, this parameters differ for Hurst exponents H above and below $1/2$.[16] For $H < 0.5$, it is dominated by small-scale roughness, like the rms slope, while for $H > 0.5$ it depends on the large scales, like the rms height. However, most natural and engineered surfaces have $H > 0.5$.[29–31] Our model is then consistent with the increase in pull-off force with $h_{\rm rms}$ reported in Refs.[21,32] We note that most measurements report insufficient details on surface roughness to allow definite conclusions on the applicability of a certain contact model.

The only experiment for which we know the topography across all scales is the one shown in Fig. 1 (and reported in Refs.[2,8,9]). For nanocrystalline diamond (NCD), the range of scales that dominates $h^{(1/2)}_{\rm rms}$ and hence contact-line pinning is at the transition between power-law scaling and the flat rolloff at 1 μm, a length scale that is accessible with an atomic-force microscope. We illustrate the respective scales that contribute to $h^{(\alpha)}_{\rm rms}$ in Fig. 5.

It is important to emphasize that our model is valid for soft materials. Increasing $h^{(1/2)}_{\rm rms}$ increases adhesion only as long as the energy needed to fully conform the surface roughness $e_{\rm el}$ is lower than the gain in surface energy $w_{\rm int}$.[15,34,35] When $e_{\rm el}$ approaches $w_{\rm int}$, our assumption of fully conformal contact breaks down. When the objects are able to only conform partially, our correlation between pull-off force and scalar descriptors of roughness no longer holds. Experiments then often observe that pull-off force decreases with $h_{\rm rms}$, as for example reported in the classic adhesion experiment by Fuller and Tabor.[36] We note unlike the theory presented here for soft solids and our understanding of nonadhesive contact,[15] there is presently no unifying theory that quantitatively describes contact behavior in this intermediate region. Large scale simulations with boundary-element methods are needed to better understand this intermediate regime.[14,35,37,38]

In summary, the pinning of the contact line explains why breaking soft contacts is always harder than making them, even in the absence of material specific dissipation. Roughness peaks increase local adhesion, which pins the crack front and increases the pull-off force. By describing rough adhesion as the pinning of an elastic line, we were able to derive parameter-free, quantitative expressions for the hysteresis in terms of a simple statistical roughness parameter. This analogy paves the way to better understanding the role of surface roughness in adhesion,

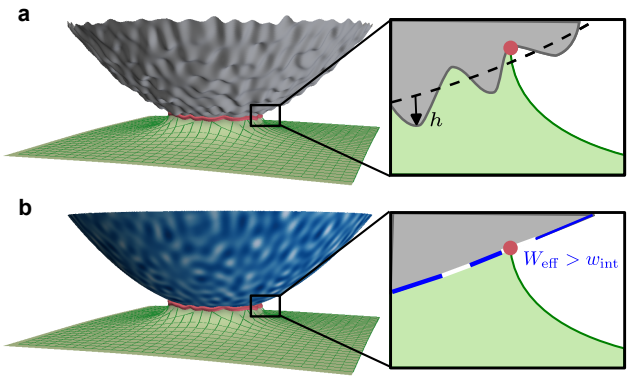


FIG. 4. **Mapping topographic roughness to effective work of adhesion.** The contact of a rough sphere (a) is equivalent to the contact of a sphere with an effective work of adhesion heterogeneity $W_{\rm eff}$ (b). The solid is stretched at the crack tip and surface roughness perturbs this elastic deformation. The associated perturbation of the elastic energy can effectively be described by fluctuations of the work of adhesion.

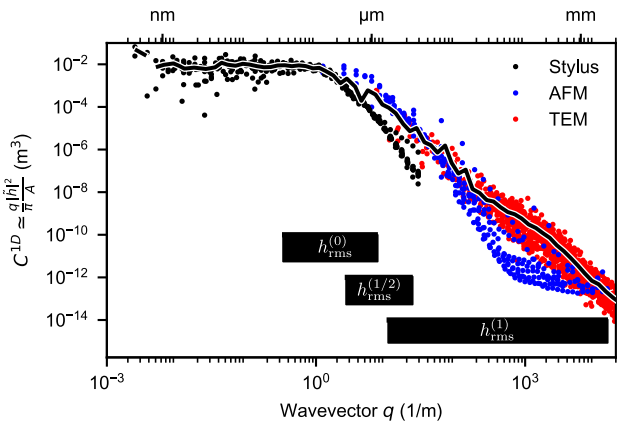


FIG. 5. **Roughness parameters.** Power spectral density (PSD) extracted from more than 60 measurements of Ref.,[9] combining stylus profilometry, atomic force microscopy (AFM), transmission electron microscopy (AFM) and transmission electron microscopy (TEM). Black bars indicate the range of scales that dominates $h^{(\alpha)}_{\rm rms}$ (Eq. (8)). Specifically, short wavelengths and long wavelengths beyond this range respectively contribute to only 10% of the value of $(h^{(\alpha)}_{\rm rms})^2$. Evaluating Eq. (8) requires the 2D or isotropic power-spectral density of the surface topography, while only the 1D PSD is known. Following Refs.,[2,13] we converted the 1D PSD $C^{\rm 1D}$ to the isotropic 2D PSD using the approximation $C^{\rm iso}(q) \simeq \frac{\pi}{q} C^{\rm 1D}(q)$. The data used in this figure is available online in Ref.[33]

and provides guidance for practitioners which scales of roughness to control in order to tune adhesion.

ACKNOWLEDGEMENTS

We thank Tevis D. B. Jacobs, Ali Dhinojwala, W. Beck Andrews, Patrick Dondl, Lucas Frérot, Mathias Lebihain and Mark Robbins for insightful discussions. This work was funded by the Deutsche Forschungsgemeinschaft (grant EXC-2193/1 – 390951807) and the European Research Council (StG 757343).

[1] Chaudhury, M. K. & Whitesides, G. M. Direct measurement of interfacial interactions between semispherical lenses and flat sheets of poly(dimethylsiloxane) and their chemical derivatives. *Langmuir* **7**, 1013–1025 (1991).

[2] Dalvi, S. *et al.* Linking energy loss in soft adhesion to surface roughness. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25484–25490 (2019).

[3] Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic Press, London, 1991).

[4] Chen, Y. L., Helm, C. A. & Israelachvili, J. N. Molecular mechanisms associated with adhesion and contact angle hysteresis of monolayer surfaces. *J. Phys. Chem.* **95**, 10736–10747 (1991).

[5] Gao, H. & Rice, J. R. Nearly circular connections of elastic half spaces. *J. Appl. Mech.* **54**, 627–634 (1987).

[6] Sanner, A. & Pastewka, L. Crack-front model for adhesion of soft elastic spheres with chemical heterogeneity. *J. Mech. Phys. Solids* **160**, 104781 (2022).

[7] Rice, J. R. Weight function theory for three-dimensional elastic crack analysis. In Wei, R. & Gangloff, R. (eds.) *Fracture Mechanics: Perspectives and Directions (Twentieth Symposium)*, 29–57 (American Society for Testing and Materials, Philadelphia, USA, 1989).

[8] Gujrati, A., Khanal, S. R., Pastewka, L. & Jacobs, T. D. B. Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales. *ACS Appl. Mater. Interf.* **10**, 29169–29178 (2018).

[9] Gujrati, A. *et al.* Comprehensive topography characterization of polycrystalline diamond coatings. *Surf. Topogr. Metrol. Prop.* **9**, 014003 (2021).

[10] Larkin, A. I. & Ovchinnikov, Y. N. Pinning in type II superconductors. *J. Low. Temp. Phys.* **34**, 409–428 (1979).

[11] Robbins, M. O. & Joanny, J. F. Contact angle hysteresis on random surfaces. *Europhys. Lett.* **3**, 729–735 (1987).

[12] Démery, V., Rosso, A. & Ponson, L. From microstructural features to effective toughness in disordered brittle solids. *Europhys. Lett.* **105**, 34003 (2014).

[13] Jacobs, T. D. B., Junge, T. & Pastewka, L. Quantitative characterization of surface topography using spectral analysis. *Surf. Topogr. Metrol. Prop.* **5**, 013001 (2017).

[14] Pastewka, L. & Robbins, M. O. Contact between rough surfaces and a criterion for macroscopic adhesion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3298–3303 (2014).

[15] Persson, B. N. J., Albohr, O., Tartaglino, U., Volokitin, A. I. & Tosatti, E. On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. *J. Phys.: Condens. Matter* **17**, R1–R62 (2004).

[16] Persson, B. N. J. & Tosatti, E. The effect of surface roughness on the adhesion of elastic solids. *J. Chem. Phys.* **115**, 5597–5610 (2001).

[17] Johnson, K. L., Kendall, K. & Roberts, A. D. Surface energy and the contact of elastic solids. *Proc. R. Soc. London, Ser. A* **324**, 301–313 (1971).

[18] Griffith, A. A. & Taylor, G. I. VI. The phenomena of rupture and flow in solids. *Philos. Trans. R. Soc. London, Ser. A* **221**, 163–198 (1921).

[19] Barthel, E. Adhesive elastic contacts: JKR and more. *J. Phys. D: Appl. Phys.* **41**, 163001 (2008).

[20] Maugis, D. *Contact, Adhesion and Rupture of Elastic Solids* (Springer, Berlin; New York, 2010).

[21] Kesari, H., Doll, J. C., Pruitt, B. L., Cai, W. & Lew, A. J. Role of surface roughness in hysteresis during adhesive elastic contact. *Philos. Mag. Lett.* **90**, 891–902 (2010).

[22] Rice, J. R. First-order variation in elastic fields due to variation in location of a planar crack front. *J. Appl. Mech.* **52**, 571–579 (1985).

[23] Imry, Y. & Ma, S.-k. Random-field instability of the ordered state of continuous symmetry. *Phys. Rev. Lett.* **35**, 1399–1401 (1975).

[24] Monti, J. M. & Robbins, M. O. Sliding friction of amorphous asperities on crystalline substrates: Scaling with contact radius and substrate thickness. *ACS Nano* **14**, 16997–17003 (2020).

[25] Joanny, J. F. & de Gennes, P. G. A model for contact angle hysteresis. *J. Chem. Phys.* **81**, 552–562 (1984).

[26] Patinet, S., Vandembroucq, D. & Roux, S. Quantitative prediction of effective toughness at random heterogeneous interfaces. *Phys. Rev. Lett.* **110**, 165507 (2013).

[27] Démery, V., Lecomte, V. & Rosso, A. The effect of disorder geometry on the critical force in disordered elastic systems. *J. Stat. Mech.* P03009 (2014).

[28] Johnson, K. L. *Contact Mechanics* (Cambridge University Press, 1985).

[29] Mandelbrot, B. B., Passoja, D. E. & Paullay, A. J. Fractal character of fracture surfaces of metals. *Nature* **308**, 721–722 (1984).

[30] Candela, T. *et al.* Roughness of fault surfaces over nine decades of length scales. *J. Geophys. Res. – Sol. Ea.* **117**, B08409 (2012).

[31] Persson, B. N. J. On the fractal dimension of rough surfaces. *Tribol. Lett.* **54**, 99–106 (2014).

[32] Briggs, G. A. D. & Briscoe, B. J. Effect of surface roughness on rolling friction and adhesion between elastic solids. *Nature* **260**, 313–315 (1976).

[33] Gujrati, A., Khanal, S. R., Pastewka, L. & Jacobs, T. D. B. Nanocrystalline diamond (version 1). https://doi.org/10.57703/ce-cjy6s (2020). NCD surface topography measured using TEM, AFM, and Stylus Profilometry.

[34] Mulakaluri, N. & Persson, B. N. J. Adhesion between elastic solids with randomly rough surfaces: Comparison of analytical theory with molecular-dynamics simulations. *Europhys. Lett.* **96**, 66003 (2011).

[35] Wang, A. & Müser, M. H. Is there more than one stickiness criterion? *Friction* (2022).

[36] Fuller, K. N. G. & Tabor, D. The effect of surface roughness on the adhesion of elastic solids. *Proc. R. Soc. London, Ser. A* **345**, 327–342 (1975).

[37] Medina, S. & Dini, D. A numerical model for the deterministic analysis of adhesive rough contacts down to the nano-scale. *Int. J. Solids Struct.* **51**, 2620–2632 (2014).

[38] Popov, V. L., Pohrt, R. & Li, Q. Strength of adhesive contacts: Influence of contact geometry and material gradients. *Friction* **5**, 308–325 (2017).

Supplementary Material for "Why breaking soft contacts is harder than making them"

Antoine Sanner[1,2]**, and Lars Pastewka**[1,2]

[1] Department of Microsystems Engineering (IMTEK), University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany

[2] Cluster of Excellence *liv*MatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany

S-I. CRACK-FRONT MODEL

Our goal is to model the contact of a rough sphere on a deformable elastic flat (Fig. S-1a). The contact perimeter of the adhesive contact between a smooth sphere and flat can be regarded as a circular crack (Fig. S-1b). This is the basis of the Johnson, Kendall and Roberts (JKR) model for adhesion [1]. JKR derived an expression for the elastic energy release rate G_{JKR} for this spherical geometry, and balanced it with the intrinsic work of adhesion, $G_{\text{JKR}} = w_{\text{int}}$. Here, we extend this result to rough spheres, where the crack shape deviates from circularity. Surface roughness perturbs the shape of the crack in the surface normal direction. This perturbs the local balance of energy, leading to additional deviation of the crack shape in direction parallel to the surface.

Figure S-1 illustrates this decomposition in terms of the energy release rate G. The surface roughness h locally perturbs the elastic energy by $G_\perp$ (Fig. S-1c) [2] and the perimeter distorts within the plane to satisfy equilibrium with the uniform work of adhesion w_{int} (Fig. S-1d). As we show in detail in this supplementary material, we describe the effect of surface roughness by an effective work of adhesion field

$$W_{\text{eff}}([h]; a(\theta), \theta) = w_{\text{int}} - e_{\text{el}}([h]) - G_\perp([h]; a(\theta), \theta), \tag{S-1}$$

where e_{el} is the elastic energy required to fully conform to the surface roughness and the square brackets indicate a functional dependency.

The effect of the in-plane deflection on the elastic energy $G_\parallel$ was derived by Gao and Rice [3] and later extended by us to spheres [4]. In our simulations, the equilibrium condition

$$W([h]; a(\theta), \theta) = G_{\text{JKR}}(b, a(\theta)) + G_\parallel([a], \theta) \tag{S-2}$$

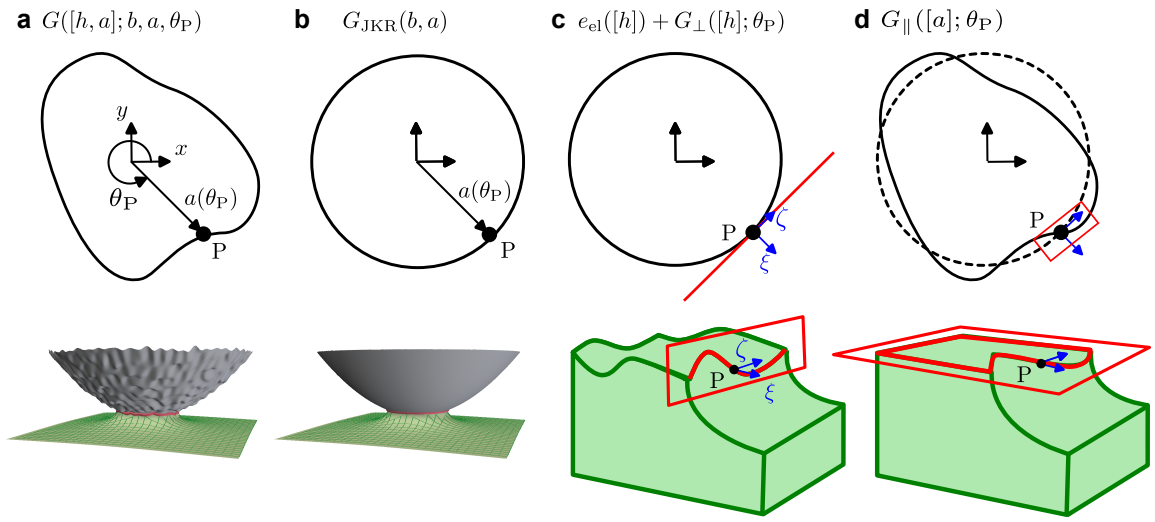


FIG. S-1. We consider the contact of a sphere of radius R with roughness $h(x,y)$ superposed to it. **(a)** Because of surface roughness, the contact perimeter is no longer circular. We describe it by the contact radius $a(\theta)$, the planar distance between the tip of the sphere and the perimeter of the contact. The energy release rate G at the point P along the crack is decomposed into three contributions, $G = G_{\rm JKR} + (e_{\rm el} + G_\perp) + G_\parallel$, illustrated in panels (b) to (d). **(b)** The energy release rate $G_{\rm JKR}(b,a)$ for the smooth contact is given by the theory of Johnson, Kendall and Roberts [1]. **(c)** Surface roughness leads to out-of-plane displacements of the contact perimeter. This increases the average energy release rate by $e_{\rm el}$ and leads to additional local fluctuations $G_\perp([h],\theta)$. Here, $e_{\rm el}$ is the elastic energy needed to fully conform to surface roughness. **(d)** The in-plane deflection of the perimeter from circularity leads to the additional contribution $G_\parallel([a],\theta)$.

determines the contact radius with $\mathcal{O}(h^2)$ errors in the strength of the disorder. The left hand side represents the driving force to increase the contact radius that fluctuates according to the surface roughness, while the right hand side represents the elastic response of the line that only depends on the spherical geometry and the material properties. The numerical implementation follows Refs. [4, 5] and is summarized in supplementary material S-II. We validate our equations by comparing crack-front simulations to boundary element method simulations in supplementary material S-III. Equations (S-1) and (S-2) establish the equivalence between the adhesion of rough spheres and the classic problem of the pinning of an elastic-line by quenched disorder [5–10].

A. Axisymmetric contact: The JKR model

We consider the contact of a sphere (to be exact, a paraboloid) adhering an elastic half-space at a fixed rigid body penetration b (Fig. S-1a). This case can be mapped to the contact of two spheres with the same composite radius R and contact modulus E' [11]. When only one half-space deforms, $E' = E/(1-\nu^2)$, where E is Young's modulus and ν is Poisson's ratio. Fracture mechanics typically considers the contact of two elastic half-spaces where $E' = E/2(1-\nu^2)$. We assume the contact is frictionless and consider only vertical displacements of the half space.

The equilibrium radius and force for a perfect sphere against the axisymmetric work of adhesion heterogeneity $W(a)$ is given by the JKR theory [1, 12, 13]. JKR described the adhesion of a paraboloid with radius R by superposing the displacements and the stress fields of the nonadhesive Hertzian contact [14] and the circular flat punch under tensile load [15].

The contact pressures p have a tensile singularity as the distance to the edge of the contact $-\xi$ goes to 0,

$$p(\xi) = -K_{\mathrm{JKR}}/\sqrt{2\pi(-\xi)} + \mathcal{O}((-\xi)^{1/2}), \tag{S-3}$$

with the stress intensity factor

$$K_{\mathrm{JKR}} = \left(\frac{a^2}{R} - b\right)\frac{E'}{\sqrt{\pi a}}. \tag{S-4}$$

Here and below we use the subscript JKR to indicate the circular contact to a smooth sphere. The energy release rate depends solely on the amplitude of this singularity [16]

$$G_{\mathrm{JKR}} = K_{\mathrm{JKR}}^2/(2E') \tag{S-5}$$

and the equilibrium condition $G_{\mathrm{JKR}}(b,a) = W(a)$ yields the contact radius. The normal force is given by

$$F_{\mathrm{JKR}}(a,b) = \frac{4E'}{3R}a^3 + 2aE'\left(b - \frac{a^2}{R}\right). \tag{S-6}$$

Once nondimensionalized using distinct vertical and lateral length units, the JKR contact is parameter free [13, 17, 18], and we present our numerical results in the nondimensional units defined in Refs. [12, 13]. Specifically, lengths along the surface of the half-space (e.g., the contact radius) are normalized by $(3\pi w_{\mathrm{int}}R^2/4E')^{1/3}$, lengths in vertical direction (e.g., displacements) by $(9\pi^2 w_{\mathrm{int}}^2 R/16E'^2)^{1/3}$ and normal forces by $\pi w_{\mathrm{int}}R$. The equations are in dimensional form but can be nondimensionalized by substituting $R = 1$, $w_{\mathrm{int}} = 1/\pi$ and $E' = 3/4$.

B. Circular contact with surface roughness: Out-of-plane perturbation of the elastic energy release rate

We now determine the energy release rate at the perimeter of the contact with a *rough* sphere but where the contact perimeter remains circular (Fig.S-1c). We denote the respective energy release by $G_\circ([h]; b, a, \theta)$, where the brackets indicate a functional dependency on the height field $h(x, y)$ that describes the roughness. Out-of-plane deflections of the surface of the solid make the elastic energy release rate $G_\circ([h]; b, a, \theta)$ fluctuate along the contact perimeter, and θ parameterizes the angle along the perimeter of the circular crack front. In the main text and in our simulations, we formally describe this perturbation of the energy release rate by the effective work of adhesion W_{eff}. In order to justify this mapping, we first discuss the *true* elastic energy release rate $G_\circ$ and show that the effects of the spherical geometry, surface roughness and in-plane distortion of the crack-front are decoupled.

The JKR contact is the superposition of the (adhesive) flat punch [15] and the Hertz solution [14]. For the rough sphere, we now additionally superpose the stresses and displacements needed to conform to the surface roughness. We do not need to determine the whole distribution of contact stresses because the energy release rate only depends on the stress intensity factor at the contact edge via Irwin's relation [16]

$$G_\circ([h]; b, a, \theta) = \left\{K_{\text{JKR}}(b, a) + K_\perp([h]; a, \theta)\right\}^2 /(2E'), \tag{S-7}$$

where $K_\perp$ captures the effect of roughness. $K_\perp$ can be thought of as the stress intensity factor in the conforming contact of a flat punch with roughness h at zero external load. Note that the stress intensity factors of the JKR solution and the influence of roughness can be superposed linearly, because in linear elasticity we can simply superpose stresses originating from different geometric contributions.

1. Stress intensity factor caused by roughness at the tip of a semi-infinite crack

We compute $K_\perp$ approximately by treating the contact as a semi-infinite crack (Fig. S-2a,b), i.e. the roughness features are small compared to the contact radius. We describe the semi-infinite crack in the coordinate system ξ, ζ, where ξ points in the normal to the crack front with $\xi < 0$ in the contacting area. ζ points parallel to it. This is essentially a locally rotated coordinated system

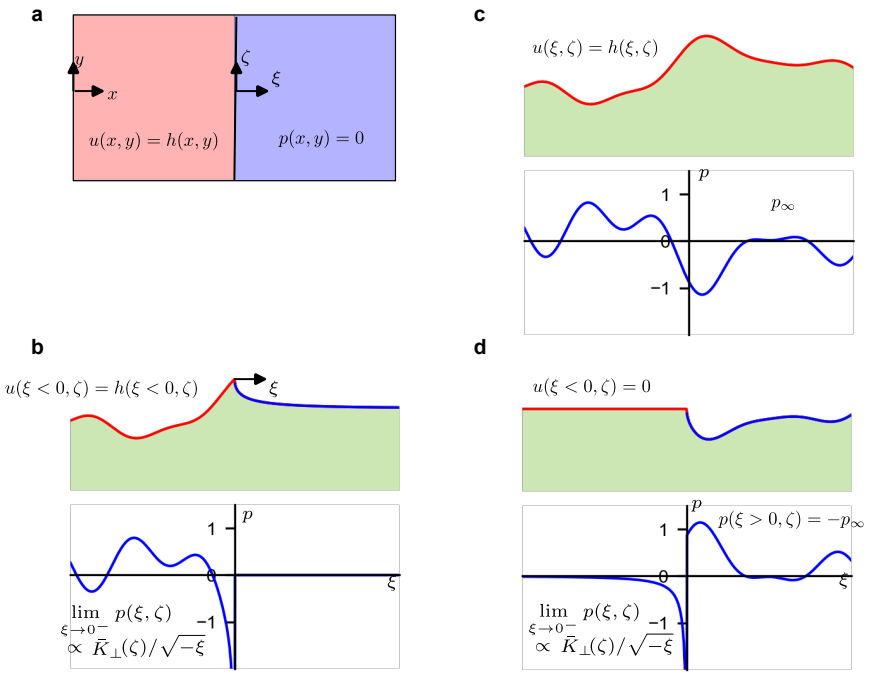


FIG. S-2. We compute the stress intensity factor caused by surface roughness for a straight crack $\bar{K}_\perp$ using a superposition. **(a)** We consider a semi-infinite crack located at x, for which the positive x and ξ directions point towards the cracked area. The local coordinate system ξ, ζ is centered on the crack tip, so that $\xi < 0$ corresponds to the contact area. **(b)** For $\xi > 0$, the surface is free to move vertically and the pressure $p = 0$. For $\xi < 0$, the solid fully conforms to the surface roughness so that the displacements u are prescribed to be equal to the heights h. Note that the positive direction for displacements and heights, corresponding to roughness peaks, points into the elastic halfspace (downwards). In the contact area, surface roughness causes fluctuating contact pressures $p(x, y)$ with stress intensity factor $\bar{K}_\perp(x, y)$. We compute $\bar{K}_\perp(x, y)$ by superposing the solutions of two elastic problems (c) and (d). **(c)** Displacements and pressures in an uncracked contact with the roughness h. The displacements $u(x, y) = h(x, y)$ cause the pressure distribution $p_\infty(x, y)$ **(d)** Semi-infinite crack with pressures applied on his crack faces. We apply the pressures $-p_\infty$ so that the pressures cancel out on the crack faces when superposing to (b). The displacements are 0 in the contact area so that the contact condition $u = h$ remains satisfied for $\xi < 0$ after superposition. Loading the crack faces at fixed displacements in the contact area causes the stress intensity factor. This stress intensity factor corresponds to $\bar{K}_\perp$ because there is no stress singularity in solution (b).

at the angle θ on the crack, as shown in Fig. S-1c. The semi-infinite crack hence represents a small subsection of the circular perimeter centered at $\xi = 0$ and $\zeta = 0$.

We compute the stress intensity factor by a classic superposition [19, 2.6.4 Full Stress Field for Mode-I Crack in an Infinite Plate], where we first compute the pressures needed to conform the surface roughness in the absence of a crack (Fig. S-2c) and subsequently cancel out these pressures on the crack faces ($\xi > 0$) (Fig. S-2d). Loading the crack-faces while keeping the displacements fixed in the contact area ($\xi < 0$) leads to the stress intensity factor $K_\perp(\zeta)$.

The pressures needed to conform to the surface roughness in the infinite contact are [20, 21]:

$$\tilde{p}_\infty(\vec{q}) = \frac{E'}{2}|\vec{q}|\tilde{h}(\vec{q}), \tag{S-8}$$

where $\vec{q} = (q_x, q_y)$ is the wavevector and the tilde denotes the Fourier transform,

$$\tilde{h}(q_x, q_y) = \int_{-\infty}^{\infty} \mathrm{d}x\, \mathrm{d}y\, e^{-i(q_x x + q_y y)} h(x, y). \tag{S-9}$$

The stress intensity factor at the edge of the contact results from the crack-face loading needed to cancel p_∞ outside the contact area,

$$\overline{K}_\perp([h]; x, y) = \int_x^\infty \mathrm{d}x_\mathrm{P} \int_{-\infty}^{\infty} \mathrm{d}y_\mathrm{P}\, k(x_\mathrm{P} - x, y_\mathrm{P} - y) \left\{-p_\infty([h]; x_\mathrm{P}, y_\mathrm{P})\right\}. \tag{S-10}$$

The quantity $\bar{K}_\perp$ is the stress intensity factor at position y along the tip of a crack advanced to position x (Fig. S-2a). The bar over $\bar{K}_\perp$ indicates that the result is valid for a straight crack. The crack-face weight function [22]

$$k(\xi, \zeta) = \frac{\sqrt{2/\pi^3}\sqrt{\xi}}{\xi^2 + \zeta^2}, \tag{S-11}$$

is the stress intensity factor at the origin of a semi-infinite crack caused by a unit point force at (ξ, ζ). Evaluating the convolution Eq. (S-10) for each position of the crack x yields a two-dimensional field of stress intensity factors, which can be most easily represented in terms of its Fourier modes,

$$\tilde{\bar{K}}_\perp(q_x, q_y) = -\frac{E'}{\sqrt{2}}\sqrt{iq_x + |q_y|}\tilde{h}(q_x, q_y) \tag{S-12}$$

with

$$\bar{K}_\perp([h]; x, y) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \mathrm{d}q_x\, \mathrm{d}q_y\, \tilde{\bar{K}}_\perp(q_x, q_y) e^{i(q_x x + q_y y)}. \tag{S-13}$$

Note that $\bar{K}_\perp$ has zero average (because of symmetry of the elastic surface response) and that the solids overlap where the stress intensity factor is negative. Our final result has no overlap

provided that $|\bar{K}_\perp| < K_{\mathrm{JKR}}$. Anderson and Rice [2] derived an equation equivalent to Eq. (S-12) to understand the interaction of crack tips with dislocations.

We now detail the steps leading from Eq. (S-10) to Eq. (S-12). The pressures needed to conform to the surface roughness in the infinite contact p_∞ are easier to express in Fourier space, see Eq. (S-8). Using the Heaviside step function $\Theta(\xi)$, we now define the weight function on the whole plane as

$$f(\xi, \zeta) = \Theta(\xi) k(\xi, \zeta). \tag{S-14}$$

This allows us to extend the integration bound on Eq. (S-10) to infinity. The convolution theorem then yields the simple expression

$$\tilde{\bar{K}}_\perp(q_x, q_y) = \tilde{f}^*(q_x, q_y)(-\tilde{p}_\infty(q_x, q_y)), \tag{S-15}$$

where the star is the complex conjugate. We now compute the Fourier transform of this generalized weight-function

$$\tilde{f}(q_x, q_y) = \int_{-\infty}^{\infty} \mathrm{d}\xi \, \mathrm{d}\zeta \, \Theta(\xi) k(\xi, \zeta) e^{-iq_x\xi} e^{-iq_y\zeta}. \tag{S-16}$$

Using that

$$\int_{-\infty}^{\infty} \mathrm{d}y \, e^{-iq_y y} k(x, y) = \int_{-\infty}^{\infty} \mathrm{d}y \, e^{-iq_y y} \frac{\sqrt{2/\pi^3}\sqrt{x}}{x^2 + y^2} = \sqrt{\frac{2}{\pi}} \frac{e^{-x|q_y|}}{\sqrt{x}} \tag{S-17}$$

and evaluating the step function, we get a classic Laplace transform [23, Eq. 29.3.4]

$$\tilde{f}^*(q_x, q_y) = \sqrt{2} \int_0^{\infty} \mathrm{d}x \, \frac{1}{\sqrt{\pi x}} e^{-(|q_y| - iq_x)x} = \frac{\sqrt{2}}{\sqrt{|q_y| - iq_x}}. \tag{S-18}$$

Inserting Eq. (S-8) into Eq. (S-15) and using that $|q| = \sqrt{|q_y| - iq_x}\sqrt{|q_y| + iq_x}$ yields Eq. (S-12).

C. Circular contact with surface roughness: Effective work of adhesion

We now switch from the straight crack back to the contact of a sphere. The first step is to approximate the stress intensity factor $K_\perp$ by the result for the straight crack, $\bar{K}_\perp$, obtained above. This approximation requires us to rotate the straight crack to be tangential to the contact circle, i.e. to rotate it by the angle θ that gives the circumferential position (Fig. S-1c). Note that for isotropic random fields, this rotation becomes inconsequential and we do not carry it out for the results shown in the main text. We do carry out this rotation when comparing the crack-front results to the boundary element method, shown in Sec. S-III below.

We now rewrite Eq. (S-7) as

$$G_\circ([h];b,a,\theta) = G_{\text{JKR}}(b,a) + K_{\text{JKR}}(b,a)K_\perp([h];a,\theta)/E' + K_\perp^2([h];a,\theta)/(2E'). \tag{S-19}$$

The term $K_\perp$ is stochastic, as it describes the influence of surface roughness, which is typically a random field. Since $K_\perp$ is linear in h, its spatial average $\langle K_\perp \rangle_{a,\theta}$ vanishes. For a random field with a short correlation length, even partial averages over just the angle θ must vanish. This means the middle summand in Eq. (S-19) does not contribute to the average energy release rate. However, the variance $\langle K_\perp^2 \rangle_{a,\theta}$ must be positive and nonzero. Parseval's theorem tells us that,

$$\left\langle K_\perp^2/2E' \right\rangle_{a,\theta} = \frac{E'}{16\pi^2} \int \mathrm{d}q_x \, \mathrm{d}q_y \, |q| C^{\text{2D}}(q_x,q_y) = e_{\text{el}}, \tag{S-20}$$

where $C^{\text{2D}}(q_x,q_y) = (L_x L_y)^{-1} |\tilde{h}(q_x,q_y)|^2$ is the power spectral density of the heights [24] and L_x, L_y are the period of the system in the respective direction. Note that while we consider the limit of an infinite system size $L_x, L_y \to \infty$, C^{2D} remains finite. The variance gives the elastic energy e_{el} for fully conformal contact. The average of Eq. (S-19) then becomes

$$\langle G_\circ \rangle_{a,\theta} = G_{\text{JKR}} + e_{\text{el}}. \tag{S-21}$$

This equation is equivalent to a classic results by Persson and Tosatti [21]. They approximated equilibrium by $\langle G_\circ \rangle_{a,\theta} = w_{\text{int}}$. Formally, this can be described by the equilibrium of a smooth sphere, where $G_\circ = G_{\text{JKR}}$, with the effective work of adhesion $W = w_{\text{int}} - e_{\text{el}}$. This approximation only works in the adiabatic limit. Fluctuations become crucial when they are able to pin the crack front.

We now show how to generalize Persson and Tosatti's result to describe local fluctuations. This means we need to consider the effect of the second term in Eq. (S-19),

$$G_\perp([h];a,\theta) = K_{\text{JKR}}(b,a)K_\perp([h];a,\theta)/E', \tag{S-22}$$

that disappears in the average but represents the leading-order effect of roughness on the fluctuations of $G_\circ$. $G_\perp$ depends on the geometry and position of the indenter via K_{JKR}. This coupling between macroscopic boundary conditions and the microscopic disorder is a second-order effect of the roughness, which we can neglect because our final equilibrium equation determines the crack shape with first-order accuracy only. To first order in h, we approximate $K_{\text{JKR}} \approx \sqrt{2w_{\text{int}}E'}$, yielding

$$G_\perp([h];a,\theta) \approx \sqrt{2w_{\text{int}}/E'} K_\perp([h];a,\theta). \tag{S-23}$$

This first-order approximation allows us to describe the effect of surface roughness by the equivalent quenched disorder in work of adhesion

$$W_{\text{eff}}([h]; a, \theta) = w_{\text{int}} - e_{\text{el}}([h7]) - G_{\perp}([h]; a, \theta). \tag{S-24}$$

The effective work of adhesion Eq. (S-24) contains only the essential leading-order contributions of the roughness and is independent of the macroscopic geometry, so that our results generalizes to other adhesion setups where our approximations are valid.

Our mapping to an effective work of adhesion establishes a link to the pinning of elastic lines by quenched disorder. Theoretical work on the pinning of elastic lines [7–10] allow us to link the hysteresis in apparent adhesion to the root-mean-square (rms) fluctuations of W. Inserting Eq. (S-12) into Eq. (S-23) and (S-24) yields

$$W_{\text{rms}} = \sqrt{\left\langle W_{\text{eff}}^2 - \langle W_{\text{eff}} \rangle_{a,\theta}^2 \right\rangle_{a,\theta}} = 2\sqrt{e_{\text{el}} w_{\text{int}}} = h_{\text{rms}}^{(1/2)} \sqrt{E' w_{\text{int}}}. \tag{S-25}$$

The quantity $h_{\text{rms}}^{(1/2)}$ is the rms half-derivative (or quarter fractional Laplacian) given by

$$\left[h_{\text{rms}}^{(\alpha)}\right]^2 = \frac{1}{4\pi^2} \int \mathrm{d}^2 q \, |\vec{q}|^{2\alpha} C^{\text{2D}}(\vec{q}) \tag{S-26}$$

with $\alpha = 1/2$.

The fluctuations of W_{eff} are linear in the roughness amplitudes. Since our assumption of fully conformal contact requires that $e_{\text{el}} < w_{\text{int}}$, W_{rms} is larger than the (second-order) shift in average adhesion. This strong *linear* perturbation of the local energy arises because the solid is stretched by a distance $u(\xi) \propto \sqrt{\xi} K/E'$ with $K = \sqrt{2 w_{\text{int}} E'}$ at an equilibrium crack tip. In valleys, the solid needs to stretch even more, increasing the elastic energy and decreasing the effective adhesion, while on roughness peaks, the effective adhesion *increases* because the solid needs to stretch less than for a perfect sphere (see also Fig. 4 of the main text). The amplitude of these energy fluctuations are given by $h_{\text{rms}}^{(1/2)}$, a generalized measure of the sharpness of peaks sensitive to larger length scales than curvatures and slopes. For self-affine roughness, this parameter is dominated either by large scales like the rms height, or by small scales like slope and curvatures, depending on the Hurst exponent [21].

D. Non-circular contact: In-plane perturbation of the elastic energy release rate

Above, we discussed the effect of out-of-plane perturbation on a perfectly circular contact. In reality, the contact shape will deviate from circularity. We now compute the energy release rate at

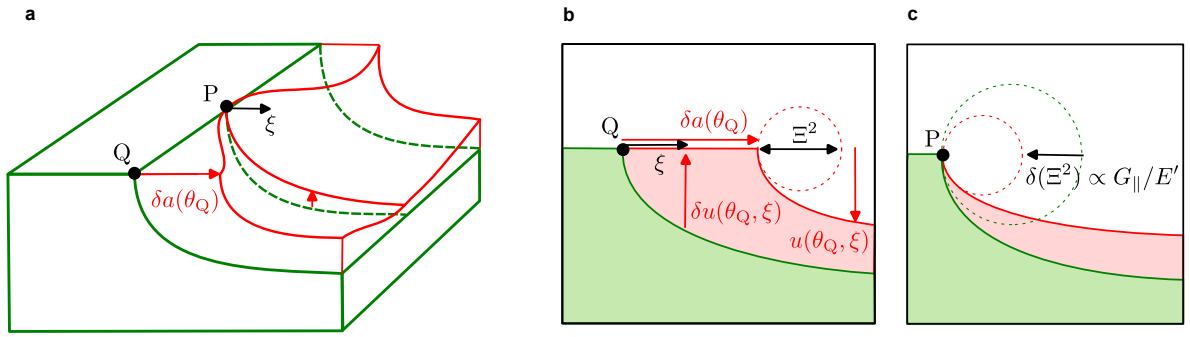


FIG. S-3. Effect of in-plane perturbations of the crack front on the energy release rate, $G_{\parallel}$. (a) The green solid represent a small section of the circular reference configuration with constant radius $a(\theta_P)$ that we perturb by $\delta a(\theta_Q, \theta_P) = a(\theta_Q) - a(\theta_P)$ (red). Advancing the contact area brings the crack faces closer together even in front of the point P that we hold fixed, because of the nonlocal interaction of the surface displacements. (b) At the crack tip, the displacements $u(\xi) \simeq \Xi\sqrt{\xi}$ with displacement intensity factor Ξ, so that closing the crack faces requires displacements $\delta u(\theta_Q, \xi) = \Xi(\theta_Q)\sqrt{\xi}$. The length Ξ^2 is the out-of-plane diameter at the crack tip (red circle) and corresponds to the elastic energy release via $G \propto E'\Xi^2$. At equilibrium, this diameter is proportional to the elastoadhesive length $\ell_a = w_{\text{int}}/E'$. (c) The diameter of the crack tip decreases by $\delta(\Xi^2) = 2\Xi\Xi_{\parallel}$ as the crack faces come together at the point P.

point P on a nearly circular outer contact to a rough sphere,

$$G([h, a]; b, \theta_{\text{P}}) = G_{\circ}([h]; b, a(\theta_{\text{P}}), \theta_{\text{P}}) + G_{\parallel}([a]; \theta_{\text{P}}), \tag{S-27}$$

with first-order accuracy in the deviation from circularity $\delta a(\theta, \theta_{\text{P}}) = a(\theta) - a(\theta_{\text{P}})$, see Fig. S-3. Our approximation is based on Gao and Rice's [3] first-order perturbation of the stress intensity factor at the perimeter of an initially circular external crack. Their result applies to arbitrary indenter geometries, where the stress intensity-factor can vary along the perimeter [4, 25, 26], as is the case here due to surface roughness. We show that the first-order effect of the in-plane perturbation $G_{\parallel}$ is independent of the out-of-plane geometry, so that the contact of a rough sphere is equivalent to the contact of a smooth sphere discussed in Ref. [4] using the effective work of adhesion heterogeneity Eq. (S-24). In Ref. [4], we only considered the case where $G_{\circ}$ is uniform over the contact perimeter, such as when $G_{\circ} = G_{\text{JKR}}$. Here, we highlight the key changes required when $G_{\circ}$ is a function of θ.

The line-elasticity emerges from the elastic coupling of the surface displacements caused by

moving an initially stretched crack tip. The more the solid is stretched, the larger the elastic energy required to distort the contact line. In this section, we discuss perturbation of stress intensity factor in terms of displacements rather than pressures and introduce a displacement intensity factor [27, 28],

$$\Xi = \sqrt{8/\pi} K/E', \tag{S-28}$$

in order to shorten the notation. Close to the crack tip, the geometry of the solid is described by

$$u(\xi, \theta) = \Xi(\theta)\sqrt{\xi} + \mathcal{O}(\xi^{3/2}), \tag{S-29}$$

so that Ξ^2 corresponds to the diameter at the crack tip, see Fig. S-3b. This length gives a geometric interpretation of the energy release rate via Eq. (S-5),

$$G = (\pi/16) E' \Xi^2. \tag{S-30}$$

At equilibrium, $G = w_{\text{int}}$, so that the diameter of the crack tip $\Xi_{\text{eq}}^2 = (16/\pi)\ell_{\text{a}}$, with the elastoadhesive length $\ell_{\text{a}} = w_{\text{int}}/E'$ [29].

Rice explained how distorting an initially circular contact perimeter affects the energy release rate in a point P that we hold fixed, see Fig. S-3a. Making use of symmetries of the elastic potential, Rice showed that the crack-face weight-function describes how the surface of the solid moves as we distort the crack front within the plane. For the energy release rate, only the perturbation of crack-face displacements close to the crack tip matter. They are described by [3]

$$\begin{aligned}\Xi_{\parallel}([h,a]; b, \theta_P) =& \Xi([h,a]; b, \theta_{\text{P}}) - \Xi_{\circ}([h]; b, a(\theta_{\text{P}}), \theta_{\text{P}}) \\ =& -\operatorname{PV} \int_0^{2\pi} \mathrm{d}\theta_{\text{Q}}\, a(\theta_{\text{P}}) \frac{\delta a(\theta_{\text{Q}}) \Xi_{\circ}([h]; b, a(\theta_{\text{Q}}), \theta_{\text{Q}})}{||\vec{r_{\text{P}}} - \vec{r_{\text{Q}}}||^2} + \mathcal{O}([a - a(\theta_{\text{P}})]^2),\end{aligned} \tag{S-31}$$

where $\Xi_{\circ}$ is the displacement intensity factor for the perfectly circular contact including roughness. The kernel of the integral was obtained from the $\xi \to 0$ limit of the crack-face weight-function of a circular external crack by Gao and Rice [3, 30, 31]. Equations (S-30) and (S-31) combined with the results from section S-I B yield the energy release rate for the nearly circular contact to a rough sphere.

Equation (S-31) captures the dominating effect of the long-ranged elastic coupling of the surface displacements on the energy release rate. We illustrate this effect in Fig. S-3, where we hold $a(\theta_P)$ fixed and advance the contact area in the neighborhood, corresponding to a locally convex

perturbation of the contact perimeter (Fig. S-3a). Closing the adhesive neck over the surface element $\mathrm{d}\theta\, a(\theta_\mathrm{Q})\delta a(\theta_\mathrm{Q})$ requires the displacement $\delta u(\theta_\mathrm{Q},\xi) = \Xi(\theta_\mathrm{Q})\sqrt{\xi}$ (Fig. S-3b), which perturbs the whole surface of the solid with amplitudes decaying with distance as $||\vec{r}_P - \vec{r}_Q||^{-2}$. We hold the crack front locally pinned in θ_P, yet this nonlocal interaction along the crack front brings the crack faces together (Fig. S-3c) and thereby reduces the energy release rate $G(\theta_P)$.

We now show that within our first-order approximation, the reduction of G for convex a discussed in the previous paragraph is independent of the indenter geometry and corresponds to a generalized curvature, the half-fractional Laplacian $(-\Delta_s)^{1/2}a(s)$, where $s = a\,\theta$ is an arclength along the contact perimeter. The principal value integral in Eq. (S-31) depends on roughness, indenter geometry and indenter position via $\Xi_\circ$, but this coupling of the in-plane elastic response to the out-of-plane geometry is only $\mathcal{O}(\delta a^2)$. Near equilibrium, where

$$\Xi_\circ([h]; b, a(\theta), \theta) + \mathcal{O}([a_\mathrm{eq} - a_\mathrm{eq}(\theta)]) = \Xi_\mathrm{eq} = \sqrt{\frac{16 w_\mathrm{int}}{\pi E'}}, \tag{S-32}$$

deviations of $\Xi_\circ$ from the material property Ξ_eq manifest in Eq. (S-31) through the second-order term $\Xi_\circ \delta a$. Approximating $\Xi_\circ$ by the constant value Ξ_eq simplifies the principal value integral to

$$\Xi_\parallel([a];\theta_\mathrm{P}) \approx -\Xi_\mathrm{eq}\,\mathrm{PV}\int_0^{2\pi} \mathrm{d}\theta_\mathrm{Q}\, a(\theta_\mathrm{P})\frac{\delta a(\theta_\mathrm{Q})}{||\vec{r_\mathrm{P}} - \vec{r_\mathrm{Q}}||^2} = \Xi_\mathrm{eq}\,(-\Delta_s)^{1/2}\,a(\theta_\mathrm{P}). \tag{S-33}$$

Here, the half-fractional Laplacian of the contact radius with respect to the arclength $\mathrm{d}s = a(\theta_\mathrm{P})\,\mathrm{d}\theta_\mathrm{P}$ is defined by

$$(-\Delta_s)^{1/2}\,a(\theta_\mathrm{P}) = \frac{1}{a(\theta_\mathrm{P})}\,(-\Delta_\theta)^{1/2}\,a(\theta_\mathrm{P}) = \frac{1}{a(\theta_\mathrm{P})}\sum_{n\in\mathbb{Z}\setminus\{0\}} |n|\;\tilde{a}_n\, e^{in\theta_\mathrm{P}}, \tag{S-34}$$

where $\tilde{a}_n$ are the coefficients of the Fourier series

$$a(\theta) = \sum_\mathrm{Z} \tilde{a}_n\, e^{in\theta}. \tag{S-35}$$

The wavelength of a Fourier mode is $\ell_n = 2\pi a(\theta)|n|$. The Fourier amplitude of $(-\Delta_s)^{1/2}\,a(\theta_\mathrm{P})$, $\tilde{a}_n/\ell_n$, is the slope of the Fourier mode, but unlike slopes, the maxima and minima of the fractional Laplacian are in phase with maxima and minima of a. Hence, $(-\Delta_s)^{1/2}\,a$ can be interpreted as a generalized curvature scaling like a slope.

Equations (S-32) and (S-33) yield the first-order perturbation of the energy release rate

$$G_\parallel([a],\theta_\mathrm{P}) \approx \frac{\pi}{8}\Xi_\mathrm{eq}\Xi_\parallel([a],\theta_\mathrm{P}) = w_\mathrm{int}\,(-\Delta_s)^{1/2}\,a(\theta_\mathrm{P}), \tag{S-36}$$

describing that the line penalizes deviations from circularity with a strength proportional to the equilibrium energy release rate w_{int} and a generalized curvature. This means that for a fixed jump-depth $\delta a = d$, it is easier to deflect the line over a wider lateral section ℓ, $\delta G = w_{\text{int}} d/\ell$, explaining why a row of several asperities can *collectively* pin the crack front while an individual asperity cannot [6].

S-II. NUMERICAL IMPLEMENTATION OF THE CRACK-FRONT MODEL

Our numerical simulations use the algorithm by Rosso and Krauth [5] to solve for the equilibrium configurations (metastable states) visited by the crack front as we pull the sphere in and out of the contact. We discretize the crack front in N collocation points at equally spaced angles θ following Ref. [4].

The surface roughness $h(x, y)$ is a Gaussian random field, where the height spectrum $\tilde{h}(q_x, q_y)$ has uncorrelated phases and random amplitudes scaling according to the PSD, and defines the effective work of adhesion field via Eqs. (S-12) and (S-24). Equation (S-12) describes the stress intensity factor for a straight crack that is rotated to be tangential to the contact circle. Note that the prefactor in Eq. (S-12) is a complex number that introduces a minor phase-shift between W_{eff} and h in the direction normal to the front. While this phase-shift, and thereby the orientation of the crack, are important when comparing deterministically the crack-front model to the BEM, they have no effect on the power-spectrum of W_{eff} and on the work of adhesion hysteresis. When the correlation length is much smaller than the contact radius, the heights decorrelate before the orientation of the crack changes significantly along the perimeter. For this reason, and because the rotation becomes computationally intractable on large grids, we generate the effective work of adhesion fields used in the main text using a constant orientation of the crack.

S-III. VALIDATION AGAINST THE BOUNDARY ELEMENT METHOD

We compare the crack-front model to a boundary element method (BEM) simulation to validate our mapping from surface roughness to an effective work of adhesion heterogeneity. The implementation of the BEM and the parameters of the simulation are similar to Ref. [4], where we validated the crack-front model for spheres with heterogeneous work of adhesion. In the BEM simulation we perform here, the sphere is rough the work of adhesion is uniform. The surfaces

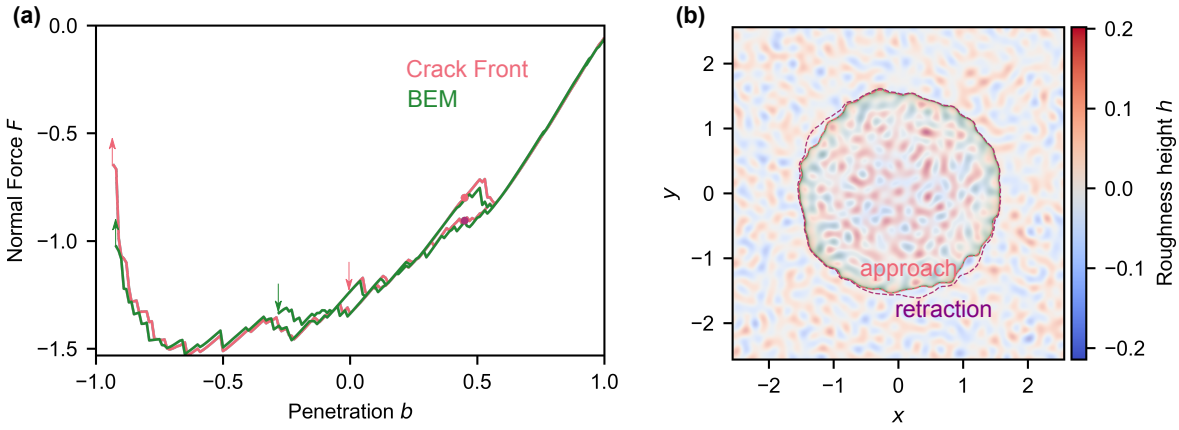


FIG. S-4. **(a)** Force-penetration curves from a boundary element method (BEM) and a crack-front simulation on the random roughness shown in panel (b). The arrows indicate the jump into contact and the jump out of contact instabilities. **(b)** Contact areas at the penetration $b = 0.45$, indicated by the dots in the force-penetration curve, on top of the surface topography. The tensile pressures of the contact mechanics simulation during approach are shown in green, so that the perimeter of the contact is indicated by the darkest green pixels. The dashed lines represent the contact perimeter calculated with the crack-front model during approach (pink) and retraction (purple). The BEM simulation was discretized on a 1024×1024 grid with pixel size $\ell_{\mathrm{pix}} = 0.005$. The roughness is a random Gaussian field with a flat power spectrum at wavelengths above the cutoff wavelength $\lambda_{\mathrm{r}} = 0.2$ and 0 below. The interaction is a cubic polynomial with a cutoff distance $g_c = 0.24$, corresponding to a cohesive zone size $\ell_{\mathrm{coz}} = (\pi/36) g_c^2/\ell_a \simeq 0.012$. In both simulations, we increased the penetration b in steps of 0.01 until the maximum penetration $b_{\max} = 1$ was reached and then decreased it until pull off. The results are nondimensionalized following the conventions of Refs. [12, 13].

interact with a cubic cohesive law with a hard-wall repulsion. Our implementation of the BEM is described in detail in Ref. [4] and is based on [32–36].

Figure S-4 shows a BEM and a crack front simulation on random roughness with $e_{\mathrm{el}}/w_{\mathrm{int}} \simeq 0.03$ and a power spectrum that is flat for wavelengths above the correlation length $\lambda_{\mathrm{r}} = 0.2$ and 0 below. The force-penetration curves computed with the BEM and the crack-front model nearly overlap and contact perimeters agree well, confirming that the contact of rough spheres is equivalent to the pinning of a crack by the effective work of adhesion heterogeneity W_{eff}. Note that

in the BEM, the jump into contact instability occurs too early because of the finite interaction range [4, 37–39]. This particular event converges slowly with interaction range, while the remainder of the force-penetration curve, including depinning instabilities, is well converged. Other discrepancies in the force-penetration curves are due to the linearization in the crack-front model.

---

[1] K. L. Johnson, K. Kendall, and A. D. Roberts, Surface energy and the contact of elastic solids, Proc. R. Soc. London, Ser. A **324**, 301 (1971).

[2] P. M. Anderson and J. R. Rice, The stress field and energy of a three-dimensional dislocation loop at a crack tip, J. Mech. Phys. Solids **35**, 743 (1987).

[3] H. Gao and J. R. Rice, Nearly circular connections of elastic half spaces, J. Appl. Mech. **54**, 627 (1987).

[4] A. Sanner and L. Pastewka, Crack-front model for adhesion of soft elastic spheres with chemical heterogeneity, J. Mech. Phys. Solids **160**, 104781 (2022).

[5] A. Rosso and W. Krauth, Roughness at the depinning threshold for a long-range elastic string, Phys. Rev. E **65**, 025101 (2002).

[6] Y. Imry and S.-k. Ma, Random-field instability of the ordered state of continuous symmetry, Phys. Rev. Lett. **35**, 1399 (1975).

[7] A. I. Larkin and Y. N. Ovchinnikov, Pinning in type II superconductors, J. Low. Temp. Phys. **34**, 409 (1979).

[8] M. O. Robbins and J. F. Joanny, Contact angle hysteresis on random surfaces, Europhys. Lett. **3**, 729 (1987).

[9] V. Démery, A. Rosso, and L. Ponson, From microstructural features to effective toughness in disordered brittle solids, Europhys. Lett. **105**, 34003 (2014).

[10] V. Démery, V. Lecomte, and A. Rosso, The effect of disorder geometry on the critical force in disordered elastic systems, J. Stat. Mech. , P03009 (2014).

[11] K. L. Johnson, *Contact Mechanics* (Cambridge University Press, 1985).

[12] E. Barthel, Adhesive elastic contacts: JKR and more, J. Phys. D: Appl. Phys. **41**, 163001 (2008).

[13] D. Maugis, *Contact, Adhesion and Rupture of Elastic Solids* (Springer, Berlin; New York, 2010).

[14] H. Hertz, über die Berührung fester elastischer Körper, J. Reine Angew. Math. **92**, 156 (1881).

[15] I. N. Sneddon, Boussinesq's problem for a flat-ended cylinder, Math. Proc. Cambridge Philos. Soc.

42, 29 (1946).

[16] G. R. Irwin, Analysis of stresses and strains near the end of a crack transversing a plate, J. Appl. Mech. **24**, 361 (1957).

[17] V. M. Muller, V. S. Yushchenko, and B. V. Derjaguin, On the influence of molecular forces on the deformation of an elastic sphere and its sticking to a rigid plane, J. Colloid Interface Sci. **77**, 91 (1980).

[18] M. H. Müser, Single-asperity contact mechanics with positive and negative work of adhesion: Influence of finite-range interactions and a continuum description for the squeeze-out of wetting fluids, Beilstein J. Nanotechnol. **5**, 419 (2014).

[19] A. T. Zehnder, *Fracture Mechanics*, Lecture Notes in Applied and Computational Mechanics No. 62 (Springer Science+Business Media, London ; New York, 2012).

[20] H. M. Westergaard, General solution of the problem of elastostatics of an n-dimensional homogeneous isotropic solid in an n-dimensional space, Bull. Am. Math. Soc. **41**, 695 (1935).

[21] B. N. J. Persson and E. Tosatti, The effect of surface roughness on the adhesion of elastic solids, J. Chem. Phys. **115**, 5597 (2001).

[22] H. Tada, P. C. Paris, and G. R. Irwin, *The Stress Analysis Of Cracks Handbook*, 3rd ed. (ASME Press, New York, 2000).

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Vol. 55 (US Government printing office, 1964).

[24] T. D. B. Jacobs, T. Junge, and L. Pastewka, Quantitative characterization of surface topography using spectral analysis, Surf. Topogr. Metrol. Prop. **5**, 013001 (2017).

[25] J. R. Rice, Three-dimensional elastic crack tip interactions with transformation strains and dislocations, Int. J. Solids Struct. Topics in Continuum Mechanics, **21**, 781 (1985).

[26] N. M. Borodachev, Contact problem for an elastic half-space with a near-circular contact area, Soviet Applied Mechanics **27**, 118 (1991).

[27] M. L. Williams, On the stress distribution at the base of a stationary crack, J. Appl. Mech. **24**, 109 (1957).

[28] R. Hartranft and G. Sih, The use of eigenfunction expansions in the general solution of three-dimensional crack problems, Indiana Univ. Math. J. **19**, 123 (1969).

[29] C. Creton and M. Ciccotti, Fracture and adhesion of soft materials: A review, Rep. Prog. Phys. **79**, 046601 (2016).

[30] J. R. Rice, Weight function theory for three-dimensional elastic crack analysis, in *Fracture Mechanics: Perspectives and Directions (Twentieth Symposium)*, edited by R. Wei and R. Gangloff (American Society for Testing and Materials, Philadelphia, USA, 1989) pp. 29–57.

[31] G. M. L. Gladwell, ed., *Contact Problems*, Solid Mechanics and Its Applications, Vol. 155 (Springer Netherlands, Dordrecht, 2008).

[32] R. W. Hockney, The potential calculation and some applications, in *Methods in Computational Physics, Vol. 9*, edited by B. A. Alder, S. Fernbach, and M. Rotenberg (Academic Press, New York, 1970) pp. 135–211.

[33] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A limited memory algorithm for bound constrained optimization, SIAM J. Sci. Comput. **16**, 1190 (1995).

[34] H. M. Stanley and T. Kato, An FFT-based method for rough surface contact, J. Tribol. **119**, 481 (1997).

[35] C. Campañá and M. H. Müser, Practical Green's function approach to the simulation of elastic semi-infinite solids, Phys. Rev. B **74**, 075420 (2006).

[36] L. Pastewka and M. O. Robbins, Contact area of rough spheres: Large scale simulations and simple scaling laws, Appl. Phys. Lett. **108**, 221601 (2016).

[37] J.-J. Wu, The jump-to-contact distance in atomic force microscopy measurement, J. Adhes. **86**, 1071 (2010).

[38] M. Ciavarella, J. A. Greenwood, and J. R. Barber, Effect of Tabor parameter on hysteresis losses during adhesive contact, J. Mech. Phys. Solids **98**, 236 (2017).

[39] A. Wang, Y. Zhou, and M. H. Müser, Modeling adhesive hysteresis, Lubricants **9**, 17 (2021).

Publication II

Surface Topography: Metrology and Properties

PAPER

Contact.engineering—Create, analyze and publish digital surface twins from topography measurements across many scales

Michael C Röttger[1], **Antoine Sanner**[1,5], **Luke A Thimons**[3], **Till Junge**[2], **Abhijeet Gujrati**[3], **Joseph M Monti**[4], **Wolfram G Nöhring**[1], **Tevis D B Jacobs**[3,*] and **Lars Pastewka**[1,5,*]

[1] Department of Microsystems Engineering, University of Freiburg, 79110 Freiburg, Germany
[2] Laboratory for Multiscale Mechanics Modeling, EPFL, CH-1015 Lausanne, Switzerland
[3] Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA 15261, United States of America
[4] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD 21218, United States of America
[5] Cluster of Excellence *liv*MatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, 79110 Freiburg, Germany
[*] Authors to whom any correspondence should be addressed.

E-mail: tjacobs@pitt.edu **and** lars.pastewka@imtek.uni-freiburg.de

Keywords: roughness, friction, adhesion, contact mechanics, spectral analysis, digital twin, FAIR data

OPEN ACCESS

RECEIVED
28 March 2022

REVISED
4 July 2022

ACCEPTED FOR PUBLICATION
2 August 2022

PUBLISHED
14 September 2022

Abstract

The optimization of surface finish to improve performance, such as adhesion, friction, wear, fatigue life, or interfacial transport, occurs largely through trial and error, despite significant advancements in the relevant science. There are three central challenges that account for this disconnect: (1) the challenge of integration of many different types of measurement for the same surface to capture the multi-scale nature of roughness; (2) the technical complexity of implementing spectral analysis methods, and of applying mechanical or numerical models to describe surface performance; (3) a lack of consistency between researchers and industries in how surfaces are measured, quantified, and communicated. Here we present a freely-available internet-based application (available at https://contact.engineering) which attempts to overcome all three challenges. First, the application enables the user to upload many different topography measurements taken from a single surface, including using different techniques, and then integrates all of them together to create a digital surface twin. Second, the application calculates many of the commonly used topography metrics, such as root-mean-square parameters, power spectral density (PSD), and autocorrelation function (ACF), as well as implementing analytical and numerical calculations, such as boundary element modeling (BEM) for elastic and plastic deformation. Third, the application serves as a repository for users to securely store surfaces, and if they choose, to share these with collaborators or even publish them (with a digital object identifier) for all to access. The primary goal of this application is to enable researchers and manufacturers to quickly and easily apply cutting-edge tools for the characterization and properties-modeling of real-world surfaces. An additional goal is to advance the use of open-science principles in surface engineering by providing a FAIR database where researchers can choose to publish surface measurements for all to use.

1. Introduction: The challenges impeding a scientific and practical understanding of roughness-dependent surface properties

Surface topography controls the function of material interfaces [1]. More than 100 years ago, Binder [2] showed that the electrical resistance between two flat surfaces in contact was far higher than would be expected given the material properties and overall geometry of the contact. In the following years, it became clear that, because of surface roughness, the area of true contact between materials is typically far lower than the apparent area of their macroscopic contact [3, 4]. Since then it has been shown that virtually all engineering materials, even the most highly polished ones, have surface roughness over some range of length scales [5]. For emerging manufacturing approaches, such as additive manufacturing

(3D printing) the effect of surface finish can be even more dramatic, with severe consequences e.g. for fatigue [6]. However, whether mature or advanced manufacturing techniques are used, this surface roughness significantly alters surface properties—including adhesion, friction, lubrication, elastic and plastic deformation under load, electrical transport, and thermal transport [7–10].

Over the last sixty years, a wide variety of analytical and numerical models have been proposed to describe roughness-dependent properties [3, 11–16], as described in Sect. 5. Especially in the last two decades, significant advances have been made in accounting for the multi-scale nature of roughness. However, there remains no consensus in the scientific community about which models best describe real surfaces under which conditions. Even the most effective models have only limited application in industrial contexts to improve product performance. To date, the modification of surface finish is mostly empirical, and it remains impossible to rationally design the optimal surface topography to precisely tune a given surface property [17–19].

The disconnect between scientific advancement of roughness theories and experimental validation and use of these theories arises because of three central challenges, as described below.

Central challenge 1: The difficulty of integrating different surface measurements to capture the multi-scale nature of roughness. Surface roughness is known to exist at many length scales and therefore cannot be fully described from any single measurement, such as a line scan from a stylus profilometer or a topographic map from an atomic force microscope (AFM). Instead, a comprehensive description requires many surface measurements across a wide range of magnifications, including using different instruments and techniques. However, it is challenging to fuse these disparate measurements into one comprehensive description of a surface, because of their different size scales, resolutions, collection/analysis software, and file formats.

Central challenge 2: *The technical complexity of implementing spectral analysis methods, and of applying mechanical or numerical models to describe surface performance*: Spectral analysis methods can be complex, often relying on signal processing tools, while technical implementation choices affect the precise value of the outcome. (See, for example, the wide array of calculations and units that are commonly used for the power spectral density [20].) Furthermore, analytical and numerical models of contact tend to be even more challenging to implement, often requiring complex calculations or even numerical solvers that are not accessible for typical experimentalists in either research or manufacturing contexts. This technical complexity and the ambiguity surrounding specific implementations present a significant barrier to the validation and use of roughness models to real-world surfaces.

Central challenge 3: A lack of consistency between researchers and industries in how surfaces are measured, quantified, and communicated: The characterization and specification of surface roughness varies enormously. Manufacturing contexts favor simple scalar parameters that are easy to measure and that are specified clearly using reference standards (e.g., ISO 4287 and ASME B46.1), yet there is not even agreement about which of the dozens of scalar parameters should be measured and specified (R_a, R_q, R_z, $R_{\Delta q}$ etc.). In the context of scientific research, more comprehensive descriptions are preferred such as the power spectral density and the autocorrelation function, but there is a lack of agreement about which is most meaningful. Furthermore, the very act of sharing surface data (with collaborators, or publishing it with a scientific paper) is made complicated by proprietary software applications with varying file formats, and a lack of agreement about how to save and report raw data.

These three central challenges cause a gap in the current understanding of topography, which must be filled in order to predict and control roughness-dependent properties. The remainder of the paper describes a software platform that attempts to close this gap.

2. Contact.engineering: overview and guiding principles

In order to address all three of the central challenges discussed in the previous section, we have created a freely-available internet-based application for rough-surface analysis that can be accessed at https://contact.engineering/. The source code behind this application is freely available[6] such that power-users can directly use the computational engine for surface topography analysis in their workflows. The present paper describes the choices made for the web application that is accessible at *contact.engineering*. These choices are implemented as defaults in the computational engine and serve to set a standard for topography analysis. Users of the back-end Python code can deviate from these choices by overriding these defaults. The application and back-end code are designed around functionalities and guiding principles that address each of the central challenges laid out in the previous section (see figure 1).

We note that a variety of software tools for handling topography data already exist. A popular tool is the open-source software *Gwyddion* [21] that is used by many researchers around the world. Commercial offerings with related functionality include *OmniSurf*

[6] See https://github.com/ContactEngineering/.

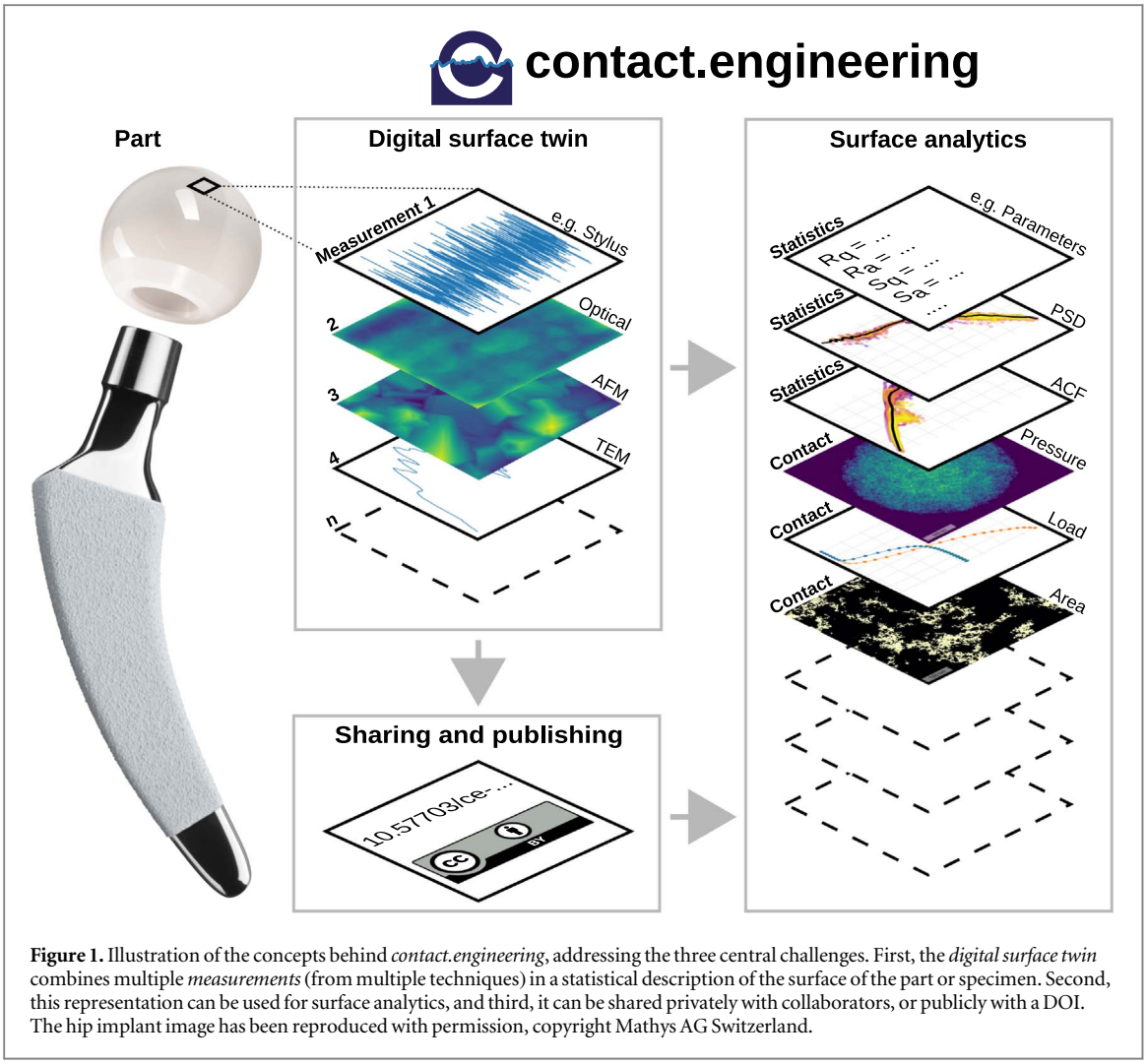


Figure 1. Illustration of the concepts behind *contact.engineering*, addressing the three central challenges. First, the *digital surface twin* combines multiple *measurements* (from multiple techniques) in a statistical description of the surface of the part or specimen. Second, this representation can be used for surface analytics, and third, it can be shared privately with collaborators, or publicly with a DOI. The hip implant image has been reproduced with permission, copyright Mathys AG Switzerland.

[22] and *MountainsMap* [23] or software developed by metrology manufacturers and shipped with instruments (e.g. *Talysurf*). While each of these excellent tools serves an important purpose in characterizing topography, there are remaining gaps that these are not designed to address. For instance, while some of the existing tools can do very sophisticated correlative microscopy where the same region of interest is characterized with multiple techniques (e.g. optical and scanning electron microscopy), this is a different functionality than combining many different types of measurements in many locations for the statistical characterization of a material's surface (Challenge 1). Additionally, without the integration of information from multiple different scales and sources, it is difficult or impossible to accurately predict functional properties of surfaces (Challenge 2). Finally, it is common for the source code of commercial software, and thus the implementation of mathematical algorithms, to be hidden from the user. Therefore, these tools cannot be used for fulfilling reproducibility requirements of scientific research [24]. The aim of *contact.engineering* is to address the remaining gaps left by current offerings.

2.1. Addressing central challenge 1: The creation of *digital surface twins*

The application allows the user to stitch together many individual measurements of the same surface, such that many limited-size-scale measurements can be combined into a complete statistical representation of the whole surface, called a *digital surface twin*.

Many instruments exist for measuring surface topography. The most popular instruments use either optical methods (white light interferometry, confocal microscopy, and others) or contact-based techniques (atomic force microscopy, stylus profilometry, and others). Any individual measurement of a surface is inherently incomplete, limited in scan size and resolution, and subject to any artifacts of that method. This raw topography data is uploaded to the application as a *measurement*.

Many different *measurements* are collected in a *digital surface twin*, which is an abstract representation of the surface of a physical object. We have implemented analysis tools (described below) that can compute properties of the digital surface twin by combining the individual properties of the *measurements*. While a single topographic measurement can never be representative of the full surface, a collection

of many measurements can comprise a comprehensive statistical description, especially if the same surface is measured many times with many different instruments at various locations, scan sizes, and resolutions. The main idea behind *contact.engineering* is to provide a natural way of combining many measurements for the comprehensive description of the surface topography of a specimen, component, or system.

We call this collection of measurements a digital surface twin to reflect the intent, which is to create a virtual model designed to accurately reflect a physical object, i.e. the surface. The virtual version is intended to be a construct of a wide variety of independent measurements, and to enable calculations and simulations to be performed on the construct which could not be accurately performed from the independent measurements in isolation. In the more general context of manufacturing, this digital surface twin would be a part of the full digital twin of a component or system. In the context of scientific research, the digital surface twin is useful to analyze the functional properties of a surface and their dependence on the multi-scale topography.

2.2. Addressing central challenge 2: Open-source digital workflows for statistical analysis of surface topography and the prediction of properties

The data uploaded to *contact.engineering* are analyzed through a set of digital analysis workflows. There is a continuously growing number of available analysis workflows, for example for statistical analysis or contact mechanics. These workflows can operate on *measurements*, *digital surface twins*, or both, and produce results that are displayed to the user through the web application. The currently available workflows are described in this paper.

Briefly, the application enables statistical analysis to describe surfaces (described in Sect. 4) for the calculation of mathematically consistent scale-dependent statistical descriptors, such as the power spectral density [20], the height-difference autocorrelation function [25], the variable bandwidth method [26–30], and the newly established scale-dependent roughness parameters (SDRP) analyses [31]. Additionally, the application performs mechanics calculations to compute contact properties (described in Sect. 5), such as the contact area, contact stress, and contact stiffness as a function of load, through the application of numerical models of continuum mechanics. All computational workflows are composed of open-source software.

All of the available analysis workflows are automatically run on each uploaded topography. Some of these have parameters that can be changed by the user, which results in the re-running of the workflow with the new parameters. The database stores *all* workflow results such that requesting a calculation with identical parameters will immediately yield the result already stored in the database. The idea is that results are available to the user at the click of a button without incurring long delays, even for those workflows that require significant computational effort.

2.3. Addressing central challenge 3: Enabling (optional) data sharing, fulfilling open-data and FAIR principles

Digital surface twins are securely stored on *contact.engineering*, but enable the user to choose to *share* the surfaces and analyses with colleagues and collaborators, or to *publish* them for the general public, for example, as a companion to a journal article. This publishing of digital surface twins generates a digital object identifier (DOI) [32] that points back to the full collection of measurements that comprises the digital surface twin. Attached to these published digital surface twins are the workflow results described above, such that not only the underlying raw data is available but also a set of derived properties. Once logged in, a user can individualize the parameters of the workflows of these published surfaces and hence directly reuse this data for future scientific work.

Users can log into the system using an ORCID [33] identifier (which anyone can create), or can access a limited functionality without logging in. As with scientific publications, the ORCID identifier of the owner of a digital surface twin is attached to the published versions, and that owner has responsibility for the correctness of the data. Co-authors can be specified upon publication of the data set.

Anyone can search the database of published digital surface twins and download the underlying raw data from the repository. The download contains the original (as-uploaded) dataset as well as a normalized variant that uses the binary, self-describing *NetCDF* format [34] that represents the measured data after potential detrending and filling of missing values (see Sect. 3.2). *NetCDF* is directly readable with common mathematical software tools such as *MATLAB* or *Python/numpy*. This conversion of vendor-specific data formats to *NetCDF* hence enables subsequent reuse of the data in custom analysis scripts. In this way, the website aids the FAIR principles of data handling [35]—making data findable, accessible, interoperable, and reusable. In particular in tribology, solutions for FAIR data are presently lacking [36]. To date no commercial or open-source software solution offers the possibility to share and publish topography data according to these principles; *contact.engineering* provides this needed capability.

3. Data handling: The creation, sharing, and publishing of digital surface twins

3.1. Data formats

A digital surface twin can be created by uploading a collection of measurements from a single physical

specimen (see figure 1). Individual measurements can take various forms, and derive from a wide variety of different measurement approaches. Many topography instruments, including optical profilometry and scanning probe microscopy, measure the height h over a square area. These are referred to as *area scans.* These measurements are therefore described by a function $h(x, y)$, where x and y describe the position in the plane of the surface. The lateral ordinates x and y run from 0 to L_x or L_y, respectively, the size of the topographic map. Topographies are exclusively stored on a *regular rectilinear grid* such that the function $h(x, y)$ is discretized into $N_x \times N_y$ equally spaced points, $h_{m,n} = h(m\Delta x, n\Delta y)$, with grid spacing Δx and Δy, $L_x = N_x \Delta x$ and $L_y = N_y \Delta y$ where $m \in [0, N_x - 1]$ and $n \in [0, N_y - 1]$.

Other topography measurements, such as many stylus profilometers, yield the height of the sample along a line profile, and are called *line scans*. Most commonly, line scans can be stored on regular grids such that $h_m = h(m\Delta x)$. However, an alternative way of producing line scans is to take an optical- or electron-microscope image of a cross-section or side-view of a material, and then to use edge-finding or other image analysis to extract its surface contour [37, 38]. This typically yields line scans on a nonuniform grid, $h_m = h(x_m)$, but with irregularly spaced x_m.

A variety of data formats exists for area scans, and *contact.engineering* supports reading several native formats (e.g., with file suffixes .di, .spm, .opd, .opdx, .zon, and many others). When reading height data from a native format, we extract additional metadata from these files such as the date of the measurement or the name of the instrument, so that this information does not need to be provided manually by the user. As the application continues to develop, we will further expand the number of supported data formats and the amount of metadata that is automatically extracted. For any unsupported format, or unconventional approach (such as cross-sectioning and contour digitization), topography data can be uploaded using generic formats. As generic data formats for height data, we support plain text, *MATLAB*, *numpy*, *NetCDF* and *HDF5* files. Plain text files are parsed for metadata (header) information and we support standardized metadata as for example provided by the text export function in *Gwyddion* [21]. Line scans on irregular grids can only be uploaded in text format where the text file contains rows of (x, y)-coordinates separated by a space.

A specific feature of irregular grids is that reentrant surfaces can be represented, i.e. those for which there are cliffs or overhangs leading to more than one value of height for a single lateral position. Some methods, in particular cross-sectioning and contour digitization, lead to a discretization of the surface in terms of a set of points (x_k, h_k) connected by straight lines (linear interpolation) that describe the bounding surface. A description of a surface in terms of a discretized curve can therefore be incompatible with a description in terms of a function $h(x)$ where a single position x is assigned a single height value. The web application supports line scans on nonuniform grids but has only limited support for reentrant surfaces at present.

3.2. Data preprocessing

Experimental measurements often need to be purified from instrument artifacts before performing analyses. *contact.engineering* implements some basic preprocessing filters to remove such artifacts, so that preprocessing parameters are documented and editable and the original data is stored.

Sample tilt and artificial curvature from the tool can be removed by subtracting the first- or second-order polynomial that minimizes the RMS height of the detrended profile (least-squares fit). Otherwise the mean value of the heights is set to 0.

When using optical profilometers, it is common that individual patches of the surface are left undefined. While contact calculations can still be performed leaving these values undefined, computing the power-spectrum or autocorrelation function requires the interpolation of the missing values. Our method to choose these values is to minimize the RMS slope of the resulting topography, which implies filling each island of missing values (patch) using a harmonic function. The harmonic function is constructed by solving the Laplace equation with the values on the edge of the patch as Dirichlet boundary conditions. This construction leads to exact filling for linear fields and has the property that the mean of the interpolated field is equal to the mean of the height over the edge (by virtue of the mean value property of harmonic functions). For line scans, this is simply a linear interpolation of missing data. For area scans, this leads to well-behaved interpolations without jumps even if large patches are missing.

3.3. Data sharing

Once created, the digital surface twin, along with all corresponding analysis (see Sect. 4 and 5), is securely stored on the database. The digital surface twins and their analysis can also be readily shared with colleagues and collaborators within the same institution or across the world. The owner(s) of the topography control access rights (such as view-only, editing, etc.). Additionally, the digital surface twin can be published, such that anyone can access the data for viewing, downloading, verifying, reusing, or re-running *contact.engineering* workflows. Publication of a digital surface twin triggers the creation of a DOI, which links to the twin. All analyses of the twin and individual measurements can be reached from the DOI landing page and retriggered with user-defined parameters. These DOIs can be referenced in publications, facilitating the goal of open data in the scientific community, and can be used for satisfying data availability requirements of

funding agencies. This publication function could even be used by a manufacturer to demonstrate a high-performance surface finish of a novel coating or manufacturing process to the scientific community or a customer.

4. Statistical analysis: Scalar parameters and correlation functions

The web application displays the uploaded topography and also computes scalar roughness parameters, scale-dependent roughness parameters [31], the power spectral density [20], the height-difference autocorrelation function [25], and performs a variable bandwidth analysis [26–30]. These methods are described briefly below, including the specific algorithms used and recommendations for their use in the context of surface metrology.

4.1. Spectral analysis using a Fourier representation

To generate synthetic topographies [20, 39], to measure Hurst exponents [29, 40], or to formulate contact theories [12, 13], it is often convenient to work with a spectral (Fourier) representation $\tilde{h}(q, k)$ of the topography rather than the topography $h(x,y)$ itself. The Fourier representation is required to compute the PSD as described in more detail below. In computing the Fourier transform, we follow the conventions laid out in [20]. The discrete Fourier transform (DFT) of the discrete topography map $h_{m,n}$ is given by

$$\tilde{h}_{o,p} = \Delta x \Delta y \sum_{m,n} h_{m,n} e^{-i(q_o x_m + k_p y_n)}, \tag{1}$$

with pixel size $\Delta x = L_x/N_x$ and $\Delta y = L_y/N_y$. The wavevector $\vec{q}_{o,p} = (q_o, k_p)$, with $q_o = 2\pi o/L_x$ and $k_p = 2\pi p/L_y$. The inverse DFT is then given by

$$h_{m,n} = \frac{1}{L_x L_y} \sum_{o,p} \tilde{h}_{o,p} e^{i(q_o x_m + k_p y_n)} \tag{2}$$

Note that it is important to lay out conventions for the DFT, as different authors use different prefactors in equations (1) and (2). We use a Fast-Fourier-Transform (FFT) algorithm to numerically compute the DFT.

4.2. Computing derivatives

Statistical analysis often involves the computation of a local derivative of the heights $h(x, y)$, e.g. $\partial h/\partial x$. For a discrete set of data $h_{x,y}$, there are multiple ways of computing a derivative that lead to different numerical answers. The simplest example is the forward-differences expression for the derivative,

$$\left(\frac{\partial h}{\partial x}\right)_{x_m, y_n} \approx (D_x h)_{m,n} \equiv \frac{h_{m+1,n} - h_{m,n}}{\Delta x} \tag{3}$$

and the central-differences expression for the second derivative

$$\left(\frac{\partial^2 h}{\partial x^2}\right)_{x_m, y_n} \approx (D_{xx} h)_{m,n} \equiv \frac{h_{m+1,n} - 2h_{m,n} + h_{m-1,n}}{\Delta x^2} \tag{4}$$

Higher-order schemes can be systematically derived, for example from finite-difference approximations. Equation (3) is used in the computation of the root-mean-square slope parameter $S_{\Delta q}$ described in Sect. 4.4. More details on such scalar roughness measures are given in the next section.

Within an analysis in terms of the Fourier representation of the topography, the derivative is usually computed by taking the analytical derivative of the Fourier-interpolated function. The result of this construction differs from equation (3) and has the additional problem that it may introduce Gibbs ringing into the solution [41]. However, we can also represent discrete derivatives (of the form given by equations (3) and (4)) using their respective Fourier representations. Inserting equation (2) into (3) yields

$$\begin{aligned}(D_x h)_{m,n} &\equiv \frac{1}{L_x L_y} \sum_{o,p} \frac{e^{iq_o \Delta x} - 1}{\Delta x} \tilde{h}_{o,p} e^{i(q_o x_m + k_p y_n)} \\ &= \frac{1}{L_x L_y} \sum_{o,p} \tilde{D}_x(q_o, k_p) \tilde{h}_{o,p} e^{i(q_o x_m + k_p y_n)}\end{aligned} \tag{5}$$

with

$$\tilde{D}_x(q, k) = \frac{e^{iq\Delta x} - 1}{\Delta x} \tag{6}$$

Here $\tilde{D}_x(q_x, q_y)$ is the Fourier-space representation of the discrete derivative operator D_x given by equation (3). (D_x is an operator while $\tilde{D}_x$ is a complex number.) We can turn any discrete derivative defined in terms of a local 'stencil' into a scalar function such as equation (6). Note that the Fourier derivative (in x-direction) is given by $\tilde{D}_x^{\mathscr{F}}(q, k) = iq$ and that $\tilde{D}_x(q, k) \to D_x^{\mathscr{F}}(q, k)$ for $q\Delta x \ll 1$. The discrete representation of the derivative in not unique but this limiting behavior must be true for any representation. Evaluating derivatives in real-space (via equation (3) or similar) or in Fourier-space (via equation (5)) is numerically *identical*. Within *contact.engineering*, we use a unique representation of the derivative for real-space and Fourier-space analysis. The derivative operators used by *contact.engineering* are summarized in table 1.

4.3. Nonperiodic surfaces and windowing

The web application supports both nonperiodic and periodic measurements. Truly periodic surfaces are a rare case that typically only occur in roughness models [20, 39] or computer simulations [42, 43]. Treatment of periodic surfaces is straightforward, as the DFT inherently assumes periodicity. When uploading a topography measurement, the user can declare whether it should be treated as periodic or nonperiodic.

Fourier analysis of nonperiodic surfaces is slightly more complicated. Nonperiodicity introduces ringing

Table 1. Discrete derivative operators used for surface topography analysis.

Operator	Real-space representation	Fourier-space representation
Derivative in x-direction	$(D_x h)_{m,n} \equiv \frac{h_{m+1,n} - h_{m,n}}{\Delta x}$	$\tilde{D}_x = \frac{e^{iq\Delta x} - 1}{\Delta x}$
Derivative in y-direction	$(D_y h)_{m,n} \equiv \frac{h_{m,n+1} - h_{m,n}}{\Delta y}$	$\widetilde{D}_y = \frac{e^{ik\Delta y} - 1}{\Delta y}$
Second derivative in x-direction	$(D_{xx} h)_{m,n} = \frac{h_{m+1,n} - 2h_{m,n} + h_{m-1,n}}{\Delta x^2}$	$\tilde{D}_{xx} = \frac{e^{iq\Delta x} - 2 + e^{-iq\Delta x}}{\Delta x^2}$
Second derivative in y-direction	$(D_{yy} h)_{m,n} = \frac{h_{m,n+1} - 2h_{m,n} + h_{m,n-1}}{\Delta y^2}$	$\widetilde{D}_{yy} = \frac{e^{ik\Delta y} - 2 + e^{-ik\Delta y}}{\Delta y^2}$

Table 2. Scalar parameters and their continuous and discrete expressions.

Name	Symbol	Continuous	Discrete
		Area properties	
RMS height	$S_q \approx h_{rms}$	$h_{rms}^2 = \frac{1}{L_x L_y} \int h^2(x, y) \mathrm{d}x\mathrm{d}y$	$S_q^2 = \frac{1}{N_x N_y} \sum_{m,n} h_{m,n}^2$
RMS gradient	$S_{\Delta q} \approx h'_{rms}$	$h'^2_{rms} = \frac{1}{L_x L_y} \int \lvert\nabla h\rvert^2 \mathrm{d}x\mathrm{d}y$	$S_{\Delta q}^2 = \frac{1}{N_x N_y} \sum_{m,n} [(D_x h)_{m,n}^2 + (D_y h)_{m,n}^2]$
RMS curvature	$S_{\Delta^2 q} \approx h''_{rms}$	$h''^2_{rms} = \frac{1}{4L_x L_y} \int (\nabla^2 h)^2 \quad \mathrm{d}x\mathrm{d}y$	$S_{\Delta^2 q}^2 = \frac{1}{4N_x N_y} \sum_{m,n} [(D_{xx} h)_{m,n} + (D_{yy} h)_{m,n}]^2$
		Profile properties	
RMS heights	$R_q \approx h_{rms}$	$h_{rms}^2 = \frac{1}{L_x} \int h^2(x) \mathrm{d}x$	$R_q^2 = \frac{1}{N_x} \sum_m h_m^2$
RMS slope	$R_{\Delta q} \approx h'_{rms}$	$h'^2_{rms} = \frac{1}{L_x} \int \left(\frac{\partial h}{\partial x}\right)^2 \mathrm{d}x$	$R_{\Delta q}^2 = \frac{1}{N_x} \sum_m (D_x h)_m^2$
RMS curvature	$R_{\Delta^2 q} \approx h''_{rms}$	$h''^2_{rms} = \frac{1}{L_x} \int \left(\frac{\partial^2 h}{\partial x^2}\right)^2 \mathrm{d}x$	$R_{\Delta^2 q}^2 = \frac{1}{N_x} \sum_m (D_{xx} h)_m^2$

artifacts into the Fourier transform that has a signature in the power spectral density $\propto q^2$ and that hides meaningful trends in the data. As is well described in the signal-processing literature, this can be overcome by *windowing*. We define the windowed topography by

$$h_{\text{windowed}}(x, y) = h(x, y) w(x, y) \tag{7}$$

where $w(x, y)$ is the windowing function. We use a Hann window, but care needs to be taken that the window is spherically symmetric for topography maps and that it is normalized according to $\int w^2(x, y) \mathrm{d}x\mathrm{d}y = 1$. (See [20] for a detailed discussion.) For periodic surfaces, $w(x, y) = 1$. Some analysis outlined below is carried out on the windowed topography $h_{\text{windowed}}(x, y)$, but for simplicity we will refer to windowed and bare topographies as $h(x, y)$.

4.4. Scalar roughness parameters

The web application also displays the distribution of heights, slopes, and curvatures (computed via the respective operators in table 1) in the uploaded measurement. It also computes the root-mean-square values of these quantities, with the appropriate discrete representations for the first and second derivative. Table 2 summarizes these parameters and the symbols we use. One-dimensional (1D, profile) and two-dimensional (2D, area) versions of these numbers are computed and reported. For area scans, we report profile properties in x- and y-directions. These properties are averages over the respective consecutive line scans in the area scan. Note that for isotropic area scans, the RMS slope the 2D value is larger than the 1D value by a factor of $\sqrt{2}$, i.e., $S_{\Delta q} = \sqrt{2} R_{\Delta q}$. We therefore explicitly call the 2D value $S_{\Delta q}$ *RMS gradient* and not RMS slope to highlight this difference.

Our RMS parameters are similar to R- and S-parameters of common standards (SEMI MF1811 or ISO 4287) but we would like to point out a couple of differences. First, ISO 16610 recommends the application of two Gaussian filters: one small-wavelength filter to remove instrument noise; and one large-wavelength filter to distinguish 'roughness' from 'waviness'. Here we do not apply such filters; The properties reported by *contact.engineering* are all computed on the *unfiltered* data that is only corrected using detrending and filling of missing values as described in section 3.2. For the detection of instrument noise and other artifacts, we apply a reliability cut-off based on tip-radius or instrument resolution. As for the separation of roughness and waviness, there is not a clear justification in the scientific literature for a sharp cutoff of size scales —therefore it is preferable to consider all of the multi-scale topography data in determining which scales matter, rather than making *a priori* assumptions. Second, there are distinctions in the way that slopes are

calculated. For instance, SEMI MF1811 is inconsistent in proposing the forward differences (also shown in table 1) for a real-space calculation of $R_{\Delta q}$ but the Fourier derivative for the reciprocal space calculation of this quantity. ISO 4287 on the other hand proposes a sixth-order finite-difference scheme that will introduce significant smoothing at small scales. By contrast, the present web app computes these quantities in a way that avoids these inconsistencies, as shown in table 1.

Overall, while scalar parameters provide simple, clear descriptors, they must be interpreted with care. The RMS parameters of the *measurement* are not necessarily representative of the RMS parameters of the *physical surface* as represented by the digital surface twin. As shown in [37, 44, 45], the RMS parameters for a given measurement will vary by orders of magnitude depending on the scale at which they are measured. It can be useful to make relative comparisons of RMS parameters between topographies, as long as the measurements were performed in the same manner. However, the technique and size-range should always be explicitly specified along with the value. If it is instead desired to describe the parameters of the full *physical surface* rather than a *measurement*, then a stitched-together PSD may be used, with Parseval's law, to compute RMS parameters (as described in [20]). However, even these descriptors cannot reflect the size-dependent variation. In order to describe the height, slope, and curvature of a surface *as a function of size*, the scale-dependent roughness parameter (SDRP) analysis can be used, as described in the next section.

4.5. Scale-dependent roughness parameters (SDRP)

The purpose of the scale-dependent roughness parameters analysis is to explicitly show how the height, slope, and curvature vary as a function of the size scale that is being considered. We have described this approach in detail in [31], so only a brief description is included here.

The RMS parameters described in the previous section are all computed at the smallest relevant scale. For example, to compute the local slope, the discrete expression

$$(D_x h)_{m,n} \equiv \frac{h_{m+1,n} - h_{m,n}}{\Delta x} \tag{8}$$

considers the height difference between two adjacent pixels. We can generalize this derivative to

$$(D_x^{(\eta)} h)_{m,n} \equiv \frac{h_{m+\eta,n} - h_{m,n}}{\eta \Delta x} \tag{9}$$

which now considers the height difference between two pixels at distance $\ell = \eta \Delta x$. Note that the scale factor η in equation (9) is an integer, but a similar expression can be defined for fractional scale factors, which however requires interpolation of the height values. We use linear interpolation for evaluating equation (9) at arbitrary ℓ.

Using the scaled derivative to compute the RMS values summarized in table 2 yields the scale-dependent roughness parameters (SDRPs). Of particular interest is the scale-dependent RMS slope $h'_{\mathrm{rms}}(\ell)$ and the scale-dependent RMS curvature $h''_{\mathrm{rms}}(\ell)$. The true value of the underlying physical surface is the short distance limit $\ell \to 0$ of these properties, but care must be taken in interpreting this limit as it is most strongly affected by instrumental noise.

4.6. Power spectral density (PSD)

As discussed in detail in [20], there are many different mathematically valid forms of the PSD. The web application makes use of the conventions recommended in [20], and computes two different types of PSDs: the radially-averaged 2D-PSD C^{iso}; and the one-dimensional PSD C^{1D} (referred to in [20] as $C^{\mathrm{1D+}}$; however, the '+' is omitted here for simplicity). The 1D-PSD of a single line scan is given by $C_o^{1D} = L_x^{-1} |\tilde{h}_o|^2$, and the 2D-PSD of an area scan is given by $C_{o,p}^{2D} = (L_x L_y)^{-1} |\tilde{h}_{o,p}|^2$. We compute $C^{\mathrm{iso}}(q)$ by (radially) averaging $C_{o,p}^{2D}$ over bins logarithmically spaced in the magnitude of the wavevector $q = |\vec{q}_{o,p}|$.

The 'correct' PSD to use for analysis of 2D measurements will depend on the surface, the measurement technique and parameters, and the desired application:

- The radially-averaged 2D-PSD $C^{\mathrm{iso\,iso}}$ (units of m^4) is commonly used because it contains a full statistical description of the surface for an isotropic measurement. However, for a measurement that contains anisotropy due to instrumentation artifacts (as many AFM scans do), C^{iso} will deviate from a 'true' description of the surface.
- The 1D-PSD C^{1D} (units of m^3) is computed as the average of the individual PSDs computed from each line in a single direction. This is computed along both the x- and y-directions and the deviation between them serves as another measure of anisotropy (real or artifacted). For measurements that introduce artificial anisotropy (such as AFM), the C^{1D} that is computed along the fast-scan direction may represent the most accurate possible description of the surface.

For comparisons between multiple surfaces, any of the above PSDs can be used and researchers should choose the most appropriate one based on anisotropy and ease of comparison. However, for quantitative calculations, such as the implementation of theoretical roughness models, a particular PSD must be used which is determined by the derivation of the model. For example, in the models developed by Persson (for example, [1, 46]), a modified version of C^{iso} is required [20], $C^{\mathrm{Persson}} = C^{\mathrm{iso}}/4\pi^2$, to account for the differences in normalization between those models and the present calculations.

Radially-averaged 2D-PSDs can be converted into an equivalent one-dimensional (profile) representation via

$$\overline{C}^{1D}(q) \approx \frac{q}{\pi} C^{iso}(q) \tag{10}$$

Note that this conversion is valid only where the PSD is self-affine; more information is given in [20]. To facilitate comparison between area scans and line scans, the web application displays both the pure 1D-PSDs C^{1D} and the 2D-PSD C^{iso} in the equivalent 1D representation (in units of m^3) as obtained from equation (10).

4.7. Height-difference autocorrelation function (ACF)

The height-difference autocorrelation function (ACF), also sometimes called the structure function [47], is the real-space equivalent of the PSD. It is formally defined as

$$A(\ell) = \frac{1}{2L} \int (h(x) - h(x+\ell))^2 \mathrm{d}x \tag{11}$$

i.e., as the correlation of the difference of heights between two points at distance ℓ on the surface. The ACF and PSD are connected by the Wiener-Khinchin theorem; one is the Fourier transform of the other. More information can be found in [25].

The ACF can be used to directly estimate some roughness parameters. For self-affine surfaces with Hurst exponent H, the ACF is a power-law $A(\ell) \propto \ell^{2H}$. The height-difference ACF furthermore saturates at the RMS height for distances where heights become uncorrelated, which we can formally write as $h_{rms}^2 = \lim_{\ell \to \infty} A(\ell)$. Dividing the square-root of the ACF by the distance yields the scale-dependent RMS slope of the surface, $h'_{rms}(\ell) = \sqrt{2A(\ell)}/\ell$. This is because equation (11) contains the first-order forward differences expression for the derivative (for more information see [31]). It also means that for small distances, the slope of the ACF corresponds to the RMS slope, $h'_{rms} \equiv h'_{rms}(0) = \lim_{\ell \to 0} \sqrt{2A(\ell)}/\ell$.

It is important to note that the web application displays the ACF on a double-logarithmic scale. The slope in this double-logarithmic scale is then proportional to the Hurst exponent H, not the RMS slope. The ACF needs to be plotted on a linear scale to estimate the slope. Alternatively, the slope can be estimated from the limiting value at small ℓ of the scale-dependent RMS slope described in the previous section. Like the PSD, the ACF can be reported for line scans, i.e., along certain directions, or as the radial average over all directions:

$$A(\ell) = \frac{1}{2\pi} \int \frac{1}{L_x L_y} \iint \frac{1}{2} (h(\vec{r}) - h(\vec{r} + \Delta\vec{r}(\ell, \phi)))^2 \mathrm{d}^2 r \mathrm{d}\phi \tag{12}$$

with $\Delta\vec{r}(\ell, \phi) = (\ell \cos\phi, \ell \sin\phi)$. Unlike the PSD, the units for $A(\ell)$ are independent of whether we compute it for line scans, along certain directions, or as the radial average.

4.8. Variable bandwidth method (VBM)

While both ACF and PSD are commonly used to analyze the statistical properties of rough surfaces, *contact.engineering* also implements a slightly less common method for statistical analysis, the variable bandwidth method (VBM) [26–30]. The VBM directly encodes that, for a self-affine topography, the roughness parameters are intrinsically dependent on the lateral size of the measurement. The VBM displays the RMS height h_{rms} as a function of the lateral size of the measurement, which we call the *bandwidth*. In the simplest incarnation, this is the size of the physical measurement, as for example demonstrated in [37, 44, 45].

The VBM approach can also be carried out on a single measurement. The measurement is subdivided into a checkerboard pattern (see inset to figure 2) of squares of equal lateral length $\ell = L/\zeta$ where we call ζ the *magnification*. We then compute h_{rms} within each of these squares (after tilt correction in the individual square) and report the value $\langle h_{rms} \rangle$ that is obtained as the average over all squares. This procedure is repeated for magnifications that are an integer power of 2, i.e. $\zeta = 1,2,4,8,\ldots$. An example of this procedure is shown in figure 2. The VBM can be thought of as a specific example of the more general SDRP analysis described above, though the implementation is slightly different (see [31] for more details.)

At the largest bandwidth ($\zeta = 1$), this simply yields the RMS height of the full measurement. For self-affine topographies, $\langle h_{rms}(\ell) \rangle \propto \ell^H$. The VBM has a simple intuitive interpretation in that it reports the width of the height distribution at different scales. While the one-dimensional PSD and ACF depend on the direction (or is reported as a radial average over different directions), the VBM has no directional dependence.

4.9. Reliability analysis

The SDRP, PSD, ACF, and VBM constitute four different methods that can be used to look at statistical properties of a measurement as a function of *scale* (denoted above by length ℓ or wavevector q). The measurement is accurate only over a portion of the full scale of the measurement. In particular, the smallest scales can be artifacted, because of instrumental noise, tip-radius artifacts, the diffraction limit, or other physical mechanisms that limit resolution. *Contact.engineering* has the capability to remove the inaccurate, artifacted portions of SDRP, PSD, ACF and VBM. This is of particular importance when stitching together individual measurements (see next section).

For optical profilometry, the user is asked to provide the resolution limit of the instrument. All data on

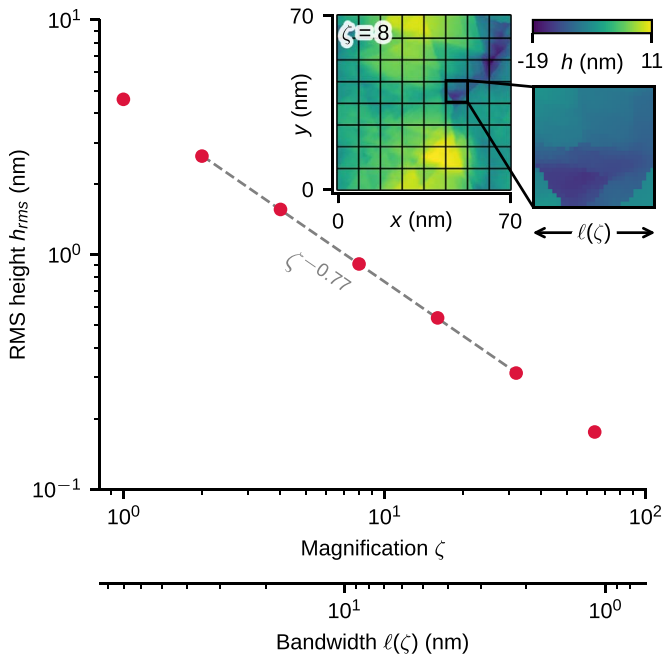


Figure 2. Variable-bandwidth analysis of the surface of a high entropy alloy from a molecular dynamics simulation [42].

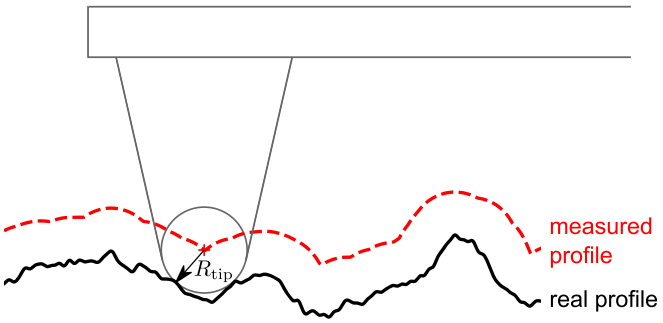


Figure 3. Scanning the topography of a rough surface with a stylus can only capture features with a radius of curvature larger than the radius of the stylus tip. Peaks are smoothened to the tip radius and valleys turn into cusps.

length scales below this resolution limit is eliminated from the respective scale-dependent analysis.

For tip-based profilometry, the user is asked to provide the tip radius R_{tip} which is used to automatically detect tip artifacts. We use a tip radius analysis technique that originated in the work of Church & Takacs [48, 49] and was recently extended by the present authors [31]. Briefly, if the probe-tip radius R_{tip} is larger than the radius of curvature of a valley, the scanning probe cannot fit into this valley on the rough topography. In this case, the peaks on the measured height data are rounded with a radius of curvature that cannot be smaller than the tip curvature. The valleys of the topography then develop cusps (figure 3). Church & Takacs pointed out [48] that the cusps lead to a (one-dimensional) PSD that scales as $C^{1D}(q) \propto q^{-4}$.

We have recently extended this idea [31], but we look at the rounded peaks instead of the cusps. We compute the local curvature at a certain scale ℓ via the scale-dependent second derivative, i.e., equation (9) or rather its equivalent for the second derivative. We then find the length ℓ below which the maximum of the (negative) curvature rises above the user-specified tip curvature,

$$\max_{x_m}\left[-\frac{D^2_{(\ell)}}{D_{(\ell)}x^2}h(x_m)\right] > \frac{c}{R_{tip}}. \tag{13}$$

Any ℓ for which this condition is fulfilled must be unreliable, as a scanning probe with tip radius R_{tip} is unable to scan regions with this curvature. The factor c must be of order unity; we use $c = 1/2$ based on numerical experiments (see [31]). Criterion equation (13) hence adjusts the reliable region of the scale-dependent surface roughness data, with rougher surfaces generally leading to a larger unreliable region. Note that equation (13) expects peaks on the rough topography to have positive height values and care must be taken not to accidentally upload height data that has been flipped upside down.

4.10. The stitching together of surface descriptors: SDRP, PSD, ACF, and VBM

Almost all geometric properties of rough surfaces depend on scale, yet the existing engineering practice is to report only single values, such as R_q or S_q. *contact.engineering* therefore promotes the use of SDRP, PSD, ACF, or VBM to report properties as a function of *scale*. The utility of reporting these properties as a function of scale is that measurements taken at different resolution, and even with different instruments, can be easily stitched together. The web application therefore displays the individual measurements of a surface in a single plot, exposing the underlying statistical structure of the physical surface.

The collection of all measurements of the same physical specimen constitute our digital surface twin. To obtain a single representation for SDRP, PSD, ACF, and VBM, we compute the averaged curves. Averages are taken over individual measurements after reliability analysis, i.e., portions of the measurement detected to be unreliable are not included in the average. The average is computed as the arithmetic mean over the datasets at specific logarithmically spaced collocation points. Note that we currently assign identical weights to each dataset, even though an individual line scan contains fewer points than an area scan. The underlying rationale is that consecutive lines on an area scan are correlated and it is presently unclear how to account for this when selecting weights for the averaging procedure.

For SDRPs, we compute the individual curves for all measurements at exactly the same collocation points. This requires interpolation of the line or area scan for the computation of the discrete derivative (see equation (9)) at fractional η. For PSD, ACF, and VBM we collect the measurements into logarithmically spaced bins over which we carry out the average. All averages are carried out over 1D (profile) representations of the respective analysis technique, since 2D representations are not available for line scans that may be part of the digital surface twin. In section 6 we will show examples of how this feature can be used to obtain a comprehensive description of the topography of a surface that goes beyond the individual measurement.

5. Contact calculations: Relationship between topography and surface properties, such as contact area, load, and normal displacement

5.1. Overview of common models to describe surface performance

A wide variety of models have been proposed to link the (elastic or plastic) mechanical deformation of rough topography to contact area or contact stiffness as intermediate properties from which other functional properties can be derived (e.g. [50]). These have been extensively reviewed elsewhere [1, 51–53]. In the broadest sense, they can be categorized into bearing-area models, independent-asperity models, renormalization group theory, and 'brute-force' numerical models. Bearing-area models assume that the surfaces in contact are rigid and treat contact as interpenetration of the rough topography; the contact area is then simply a slit-island analysis [54] of the rough topography. The resulting behavior of contact area versus penetration is often called the Abott-Firestone curve [55]. Independent-asperity contact models (such as the famous Greenwood and Williamson model [3] and its extensions, such as [56]) approximate roughness as a series of non-interacting hemispherical asperities. Persson's scaling theory [12, 13] and the approach by Joe, Thouless, & Barber [57] solve the contact problem at a specified scale given that the solution at larger scales is known, leading to a renormalization of contact properties. Bearing-area, independent-asperity and renormalization models require as input statistical descriptions of the topography in the form of RMS parameters, the power spectral density, or others. By contrast, 'brute-force' numerical calculation solve for the elastic deformation using a *specific* realization of topography [16, 53, 58–78]. While these calculations require fewer simplifications than the other categories and so may result in more accurate predictions, they are resource-intensive to run and do not typically yield simple equations that are generalizable to other surfaces with similar statistics.

All four categories of rough-surface models have been proven successful in certain contexts, but all are based on simplifying assumptions, and it remains difficult to experimentally validate their predictions and to directly apply them to real-world surfaces to optimize roughness for a desired surface property. At present, the web application focuses primarily on the fourth class of models: the numerical implementation of continuum contact mechanics. However, additional computational workflows are constantly being added.

5.2. Assumptions behind the boundary element method (BEM)

Contact.engineering computes exactly the elastic deformation of two frictionless contacting topographies for a linear-elastic, isotropic solid, and is able to consider plasticity using approximate models. The geometry considered is that of a rigid rough surface on an elastic half-space, and we only consider the displacements of the surface in the normal direction. For frictionless contact, the only error introduced by the latter approximation is that asperities can be displaced within the plane of contact, leading to slightly different contact geometries. This latter approximation is commonly employed in state-of-the-art contact calculations [16, 53, 58–78] and allows to map the contact of any two elastic solids with arbitrary geometry onto

that of a rigid rough surface on an elastic flat surface. This approach is exact for the contact of two elastic solids with Poisson ratio $\nu = 1/2$. We note that consideration of the relative motion of the two solids during contact, introduced by differences in the elastic moduli, would require frictional constitutive laws, significantly complicating the analysis.

All results with units of pressure (such as normal pressure or normal force) are expressed in terms of the effective contact modulus E^*, with $1/E^* = (1 - \nu_1^2)/E_1 + (1 - \nu_2^2)/E_2$, where E_1 and E_2 are the Young's moduli of the two contacting bodies and ν_1 and ν_2 their Poisson ratios [79]. Conversion to real pressure units is therefore the responsibility of the user. For plastic calculations, as discussed below, there is then only a single material parameter p_Y/E^*, the ratio of hardness p_Y to the modulus E^*. For the contact of two rough surfaces, the user is at present required to take a measurement of each surface, $h_1(x, y)$ and $h_2(x, y)$, and then to create a compound topography $h_1(x, y)+h_2(x, y)$ for upload. We note that the analysis is carried out for individual measurements or compound topographies and cannot be averaged to be representative for a full digital surface twin. Contact properties also depend on the scale of the measurement and care has to be taken in their interpretation. In particular, contact area depends on small-scale properties (in particular h'_{rms}) [11, 14, 80] while contact stiffness depends on the largest scale (in particular h_{rms}) [68, 81, 82].

5.3. Solution of the elastic problem

The elastic deformation of the substrate is computed using a boundary element method that considers just the degrees of freedom of the surface and treats the bulk as an elastic half-space. Such boundary element methods have been extensively described in the literature. Examples include direct summation of the (regularized) Boussinesq-Cerutti solution [79] for point load or multilevel summation techniques [60]. We solve the elastic problem in reciprocal space using a fast Fourier transform (FFT) technique, that is similar to the following: the approach of Stanley & Kato [59]; the DC-FFT technique [83]; and related techniques for atomic lattices [84–86].

The FFT can be used to accelerate the computation of a convolution in real-space. In linear elasticity, the surface displacements $u(x, y)$ are related to the surface pressure $p(x, y)$ through a linear operator, the elastic surface Green's function $G(x, y)$:

$$\begin{aligned} u(x, y) &= \int_{A_G} G(x - x', y - y')p(x', y')dx'dy' \\ &= G(x, y) * p(x, y) \end{aligned} \tag{14}$$

The Fourier transform of equation (14) turns the convolution $G * p$ into a simple product

$$\tilde{u}(q, k) = \tilde{G}(q, k)\tilde{p}(q, k), \tag{15}$$

where a tilde $\tilde{f}$ indicates the Fourier transform of a function f:

$$\tilde{f}(q, k) = \int_{A_G} f(x, y)e^{-iqx-iky}\mathrm{d}x\mathrm{d}y \tag{16}$$

Note that the integral is carried out over the area of support of the Green's function, A_G. For nonperiodic calculations, this differs from the nominal area of the measurement A_0 by a factor of 4, since the computation requires a padding region as described below.

The Green's function for the elastic problem can be analytically or semi-analytically derived for periodic [84–88] and non-periodic [58, 89, 90] systems in the continuum limit [87, 89, 90], for atomic lattices [84–86] and for substrates of finite thickness [91–94]. The web service uses the continuum expressions and can handle periodic and non-periodic systems.

The linear-elastic Green's function for periodic pressure distributions takes the particularly simple form in reciprocal (Fourier) space [88]

$$\tilde{G}(\vec{q}) = 2/(E^*|\vec{q}|). \tag{17}$$

Surface topographies are typically provided as measurements of rectangular sections of size $L_x \times L_y$, measured on a homogeneous grid of $N_x \times N_y$ pixels. The nominal measurement resolution is then $\Delta x = L_x/N_x$ and $\Delta y = L_y/N_y$. (The true resolution can of course be lower because of instrumental artifacts, see discussion above.) Data and calculations are therefore provided on a discrete grid. The discrete version of equation (14) replaces the integral with a sum. Equation (16) then becomes the discrete Fourier transform, which we compute with a standard FFT algorithm on the same grid as the uploaded surface. This discretization limits the range of wavevectors that enter the periodic Green's function, equation (17), to $\tilde{G}_{mn} = \tilde{G}(2\pi m/L_x, 2\pi n/L_y)$ with $m \in [0,1,\ldots, \lfloor(N_x - 1)/2\rfloor, -\lfloor N_x/2\rfloor,\ldots, -1]$ and $n \in [0,1,\ldots, \lfloor(N_y - 1)/2\rfloor, -\lfloor N_y/2\rfloor,\ldots,-1]$.

The situation for nonperiodic contacts is more complicated. First, we can no longer straightforwardly use equation (17) for the Green's function because this would require a numerical grid of infinite size to decouple images. Carrying out the inverse Fourier transform of equation (17) analytically for this case yields the Boussinesq-Cerutti functions that describe the response of the elastic solid to a concentrated normal force [79],

$$G_{\mathrm{BC}}(\vec{r}) = 1/(\pi E^*|\vec{r}|). \tag{18}$$

This function diverges as $r \to 0$ and must be regularized. The typical approach is to distribute the contacting force over the area associated with the discrete surface element (of size $\Delta x \times \Delta y$) and assume constant pressure on that element [58]. Our elements are rectangles and we use the expression for constant pressure on a rectangular surface area [79, 95]. We note that more sophisticated schemes such as linear or quadratic interpolation of pressure between discretization points have been described in the literature

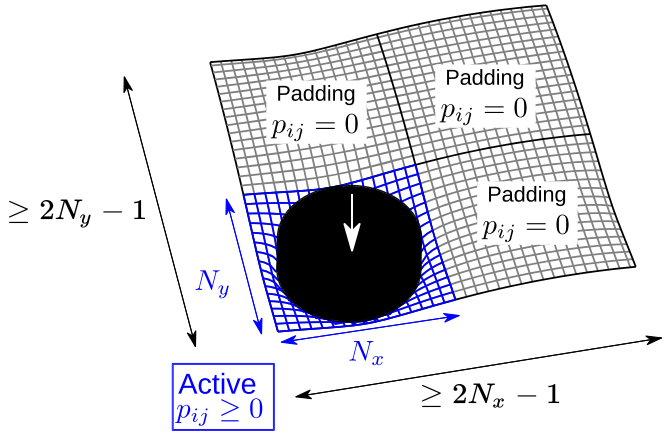


Figure 4. Illustration of the algorithm used for decoupling of periodic images into nonperiodic domains in the FFT-accelerated boundary-element method. The method uses the nonperiodic Green's functions and the FFT to compute the convolution of the Green's function with the pressure field p_{ij}. Since the Green's function extends by N_x grid points in both positive and negative x-direction (N_y in y-direction), we employ a full computational domain of size ($2N_x$-1) × ($2N_y$-1) to decouple the images. The padding region must carry zero pressure and displacement values obtained there are discarded.

[58, 89, 96], but practical differences only occur at small scales where other effects (e.g. measurement artifacts) dominate.

The final Green's function for nonperiodic systems that *contact.engineering* uses is [79]

$$\begin{aligned}
&\pi E^* G(x, y) \\
&= (x + a) \ln \left[\frac{(y + b) + \{(x + a)^2 + (y + b)^2\}^{1/2}}{(y - b) + \{(x + a)^2 + (y - b)^2\}^{1/2}} \right] \\
&+ (y + b) \ln \left[\frac{(x + a) + \{(x + a)^2 + (y + b)^2\}^{1/2}}{(x - a) + \{(x - a)^2 + (y + b)^2\}^{1/2}} \right] \\
&+ (x - a) \ln \left[\frac{(y - b) + \{(x - a)^2 + (y - b)^2\}^{1/2}}{(y + b) + \{(x - a)^2 + (y + b)^2\}^{1/2}} \right] \\
&+ (y - b) \ln \left[\frac{(x - a) + \{(x - a)^2 + (y - b)^2\}^{1/2}}{(x + a) + \{(x + a)^2 + (y - b)^2\}^{1/2}} \right]
\end{aligned} \tag{19}$$

where $a = \Delta x/2$ and $b = \Delta y/2$ are half the grid spacing in x- and y-directions.

The use of the nonperiodic real-space Green's function, equation (19), with the FFT convolution technique, equation (15), requires an additional trick to decouple periodic images. This trick was described by Hockney in the context of the solution of the Poisson equation for electrostatic problems [97] and later employed, e.g., to decouple images in plane-wave density functional calculations [98]. In contact mechanics, this method was introduced by multiple authors to decouple periodic images [73, 83]. In brief (see also figure 4), we solve the equation (14) on a grid of at least $(2N_x - 1) \times (2N_y - 1)$ grid points (and area A_G) and require that all pressures $p_{ij} \equiv 0$ for $i \geqslant N_x$ or $j \geqslant N_y$. The reciprocal space Green's function is then obtained by discretizing equation (19) on a regular grid, $G_{mn} = G(\Delta xm, \Delta yn)$ with $m \in [0,1,\ldots,N_x - 1, -(N_x - 1), \ldots,-1]$ and $n \in [0,1,\ldots,N_y - 1, -(N_y-1),\ldots,-1]$, and then computing the discrete Fourier transform of G_{mn}. Note that this is the minimum grid size required to decouple the periodic images. Large grid sizes are possible and sometimes beneficial for optimal FFT performance.

The regions with $p_{ij} \equiv 0$ are padding regions that decouple the images (see figure 4). This works because the Green's function G_{kl} has a maximum range of $N_x - 1$ in x- and $N_y - 1$ in y-direction. Because of the padding region, any two points in the 'active' region (see figure 4) are at most $(N_x - 1, N_y - 1)$ apart and any distance vector crossing the padding region is longer than $(N_x - 1, N_y - 1)$. Note that displacements u_{kl} within the padding region have no physical meaning: They come from a superposition of repeating images and must be ignored. The web app returns just pressure and displacements in the active region with area A_0; the inactive padding region is hidden from the user.

5.4. Solution of the contact problem

The previous section discussed the calculation of the relationship between local pressure and elastic displacement of the substrate's surface. We use an FFT-based algorithm, but this FFT algorithm can in principle be replaced by any other (equivalent) solution of this problem described in the literature, such as direct summation [58] or multilevel summation [60]. In our experience, the FFT-based formulation yields the best numerical performance across application scenarios.

The second part to any solution of a contact problem is the contact condition itself. The continuum mechanics picture, which is also the one employed at present in *contact.engineering,* is that of two impenetrable solids. One of those is flat and deformable. When not contacted, it is located at $u_{ij} \equiv 0$. The counterbody is described by the discrete function h_{ij}. In mathematical terms, the contact condition then

becomes a linear complementarity problem [58]:

$$p_{ij} \geqslant 0,\ u_{ij} - h_{ij} \geqslant 0,\ p_{ij}(u_{ij} - h_{ij}) = 0. \qquad (20)$$

We use the constrained conjugate gradient algorithm of Polonsky & Keer [60] to solve equation (20). The variable in Polonsky & Keer's algorithm is the pressure p_{ij}. The conjugate gradient scheme optimizes p_{ij} so that the overlap $h_{ij}-u_{ij}$ is close to zero within the contacting regions $I_c = \{(i, j)|p_{ij} > 0\}$. After each step, the algorithm constrains $p_{ij} \geqslant 0$ and whenever a point outside the contact region (defined by $p_{ij} = 0$) enters into contact ($u_{ij} \leqslant h_{ij}$), the conjugate gradient iteration is reset. More information on this algorithm can be found in [60]. Our implementation of this algorithm also allows to specify an upper limit to the pressure on the surface, corresponding to an indentation hardness p_y [62, 99].

We note that the algorithm is fast at low contact area but requires many iterations to converge at large contact area. By default, we stop the algorithm after 100 iterations. These non-converging calculations are reported to the user but the dataset is shown with translucent markers to make the user aware that the data may be unreliable. The user can change the number of iterations up to a hard upper limit of 1000.

6. Examples of use

In the following section, we present two examples of how to use *contact.engineering*. We note that all figures presented here are reproduced as shown by the web service to illustrate that publication-ready figures can be directly obtained from *contact.engineering*. The first example illustrates the use of statistical analysis techniques to stitch together many measurements, while the second example illustrates contact calculations on individual topographies.

6.1. Ultrananocrystalline diamond

In order to illustrate that many measurements can be stitched together to arrive at a statistical representation of a rough surface, we analyze the data of more than 100 measurements of an ultrananocrystalline diamond film reported in [37, 45]. The dataset contains topography data obtained from stylus profilometry, atomic force microscopy (AFM), and also using a less-common microscopy approach: side-views and cross-sections in a transmission electron microscope (TEM). The TEM and stylus profilometry approaches yield line scans while the AFM yields area scans. This illustrates that it is possible to combine and analyze structurally distinct data within a single digital surface twin. We note that we are repeating the analysis of [31, 37, 45]; this shows that anyone can repeat this sophisticated analysis without the need for complex mathematical calculations or specialized knowledge.

The whole dataset was uploaded to create a digital surface twin, and then analyzed using the PSD, ACF, VBM, and scale-dependent curvature, as shown in figure 5. All data has been tilt-corrected before analysis. Data from the scanning probe techniques (AFM and stylus) has been filtered to remove unreliable portions (see discussion on tip artifacts in section 4.9); the remaining dataset contains only data deemed reliable and the averages (thick black curves) are only computed over this reliable dataset. The plot illustrates that the individual measurements of all four methods line up and can be used to obtain a statistical representation of the surface.

All analysis methods show a region of power-law scaling, which occurs at large wavevectors for the PSD (figure 5(a)) and at short distances or bandwidths for ACF, VBM, and SDRP (figures 5(b)–(d)). The VBM (figure 5(c)) saturates at the RMS height of the surface of roughly 10 nm. The ACF saturates at the square of this value. The SDRP shows that the curvature is a function of scale that appears to grow unbounded as resolution increases (distance decreases). Note that the curvature shown in figure 5(d) cannot be straightforwardly extracted from the other plots, showing that the SDRP is useful to obtain parameters that are straightforward to interpret geometrically but difficult to obtain by other means. The ACF (figure 5(b)) shows larger deviation from the average, e.g., at distances in the range 10-100 nm. Those are artifacts introduced by tilt-correcting individual measurements: Tilt correction forces the ACF to go to zero at the size of the measurement, resulting in a downtick of the ACF curve. None of the other techniques is sensitive to this artifact.

To illustrate the advancement of FAIR data practices, we note that this digital surface twin of the ultrananocrystalline diamond has been published and is DOI-accessible at [100]. This means that anyone can access the surface, including all of the underlying raw topography data, for examination or re-use. Furthermore, all of the above analysis can be repeated, and any other desired analysis can be run, by anyone who accesses the site.

6.2. Smooth- and rough-sphere contacts

As an example of the use of elastic contact-mechanics simulations, we simulate the contact of a rough sphere. This digital surface twin is not based on real-world measurements but on computer-generated topography, defined on a 1024-by-1024 grid with a pixel size of 1 nm by 1 nm. Elasticity was computed using free boundaries as described in section 5. The sphere was generated by superposing a nominally flat self-affine random field generated by Fourier-synthesis [20, 39] and a paraboloid.

The relationship between contact area and load (mean pressure) is shown in figure 6(a), as displayed by *contact.engineering*. For reference, we also show

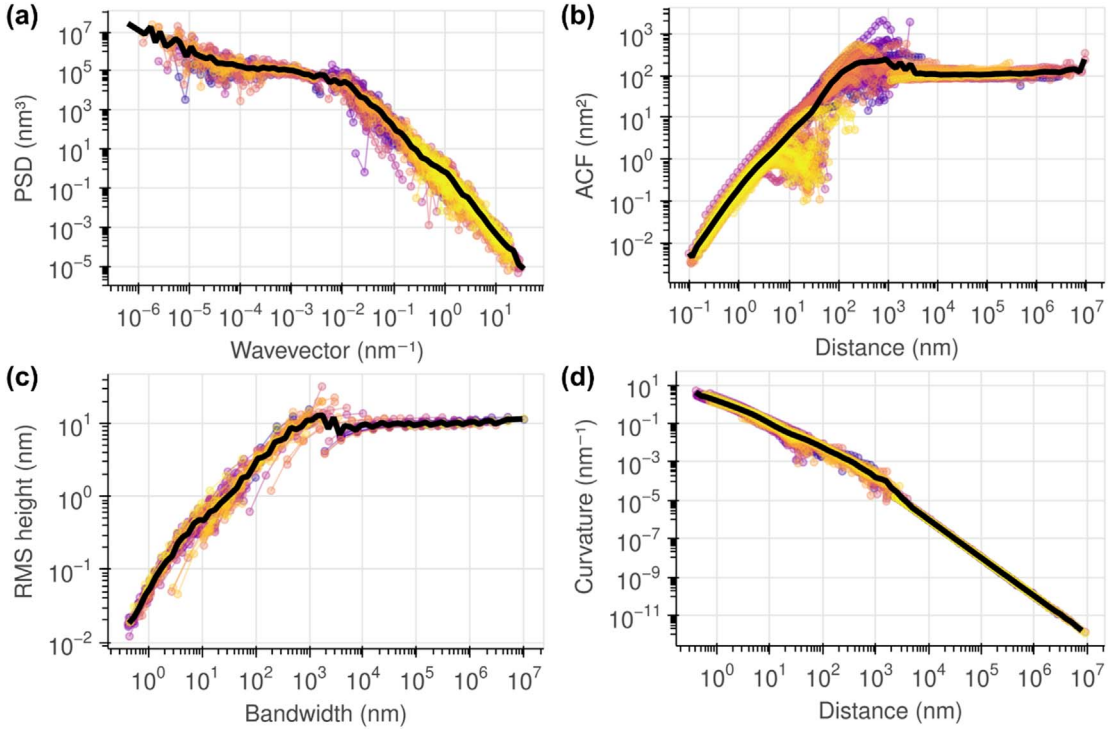


Figure 5. Comprehensive statistical topography analysis of an ultrananocrystalline diamond film. The dataset contains transmission electron microscopy, atomic force microscopy and stylus profilometry data, bridging more than 7 orders of magnitude in length. The data from this digital surface twin is stitched together into a statistical representation of the underlying physical surface using (a) the one-dimensional power-spectral density (PSD, C^{1D}), (b) the height-difference autocorrelation function (ACF), (c) the variable-bandwidth method (VBM) and (d) the scale-dependent curvature (from the SDRP analysis). Colored lines are individual measurement results and the thick black line is the average of those measurements. The individual panels (a), (b), (c), and (d) are reproduced here as obtained from the web application. This dataset is available at [100].

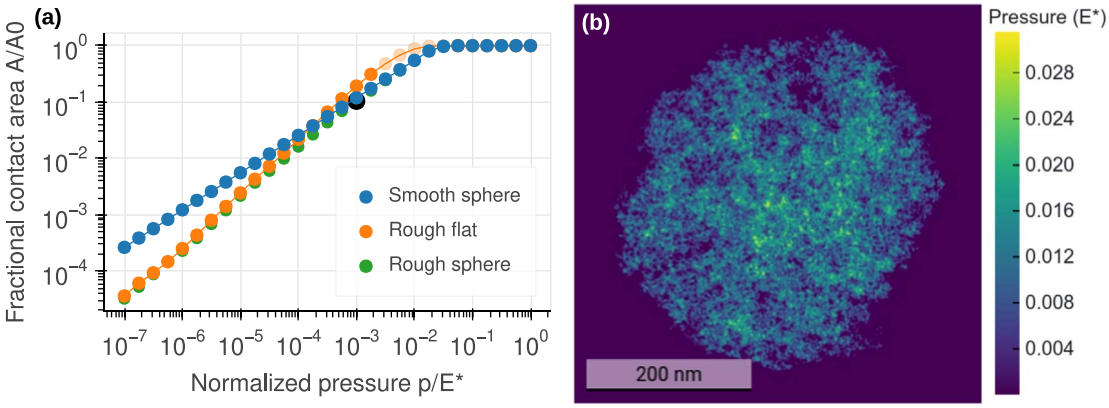


Figure 6. Contact of a rough sphere on an elastic flat. (a) Fractional contact area A/A_0 as a function of normalized pressure p/E^*, where the pressure p is computed over the reference area A_0. For comparison, we also show the results of a smooth sphere and a rough surface that is nominally flat. (b) Spatial map of contact pressures for the rough sphere at a load indicated by the black circle in (a). The individual panels (a) and (b) are reproduced here as obtained from the web application. This dataset is available at [101].

contact areas for the smooth paraboloid and for the roughness alone, where we used periodic boundary conditions for the nominally flat roughness. It can be observed that the contact-area-versus-pressure curve corresponds to the rough surface at low loads but crosses over to Hertzian for high pressures, as described in [73]. For the periodic nominally flat roughness at high contact ratios, some simulations did not reach convergence in the number of allowed iterations (1000). These simulations are indicated by a translucent datapoint.

In addition to the plot of contact area versus load, *contact.engineering* allows the visualization of a map of the contact pressure for each simulation (figure 6(b)), as well as the gap distribution and other quantities. The contact pressure reflects the rough topography of the sphere, but appears highest in the center and lowest as the edges as would be predicted by Hertz theory.

The specific pressure map of figure 6(b) is obtained right where the rough sphere crosses over from flat-on-flat to Hertz behavior and is indicated by the black circle in figure 6(a). The pressure is more homogeneous below this crossover (lower loads) and becomes more Hertzian-like at higher loads. More details can be found in [73].

7. Software implementation and infrastructure

We here briefly comment on the software implementation of the methods described here. All software is published under the MIT license and available at https://github.com/ContactEngineering and we encourage engagement from the scientific community. We follow established practices in development of scientific software [102], such as continuous integration with an appropriate automated testing framework to ensure software quality.

We distinguish between back-end and front-end modules. Back-end modules handle the topography data and carry out numerical analyses. The package *SurfaceTopography* implements filters for reading topography measurement data and performing statistical analysis (PSD, ACF, etc.) on it. It also has provisions for computing averages over multiple measurements. The package *ContactMechanics* implements boundary-element calculations for obtaining contact area, pressure, and other associated quantities. These back-end codes use *numpy* [103] for numerical calculations and implement all the scientific, algorithmic functionality described in this paper. The back-end is parallelized using the message passing interface (MPI) and can be used independently of the web front, for example on high-performance computing systems.

The web front end *TopoBank* is based on *Django* that provides bookkeeping of uploaded data via interaction with an underlying *PostgreSQL* database and a storage system, currently a *NetApp StorageGRID* instance running georedundantly in Freiburg and Tübingen, Germany. *Django* also renders HTML pages and handles user interaction. *TopoBank* is split into a manager that handles upload and visualization of measurements and digital surface data (including their metadata), and an analysis module that orchestrates running pre-defined analyses and visualizing them. Analyses themselves are short functions that call the respective analysis methods in *SurfaceTopography* or *ContactMechanics* and place results in the storage system. Analysis functions are distributed among compute nodes and prioritized via the *Celery* task queue. All analyses are computed asynchronously, i.e., they are placed in the *Celery* task queue and become available for visualization once they have run. This means that once they have run, the results are instantly available to the user for visualization. Some analyses allow specification of parameters. Changing a parameter only triggers a new calculation if no calculation with the same parameter set has already run. Calculations with a default parameter set run immediately after upload for all analyses.

8. Summary and conclusions

We have described a web-based application that is designed to standardize the analysis of topography data and the calculation of surface performance. The *contact.engineering* application addresses the three central challenges of surface analysis, which are described in the Introduction section. First, the application puts a focus on collecting many measurements for the surface of a real-world specimen or device, at various length scales and using different instruments, and integrating them into a digital surface twin of the real-world incarnation. The measurements are stitched together to describe all measured scales of the physical surface. The more measurements that are uploaded and the wider the variety of length scales, the more accurate the representation of the digital surface twin, with the goal of a *comprehensive* description of multi-scale surface topography. Second, the web application implements advanced analysis tools, including statistical metrics to describe the comprehensive surface topography, as well as mechanical models to predict the surface performance. Third, the application allows users to securely store, share, and optionally publish the digital surface twins and all associated analysis, in a manner that implements a FAIR data policy [35]. The published datasets receive digital object identifiers (DOIs) and can be cited in publications, in reports to funding agencies, or even in advertising a new material or technique. We hope that the service will become a central repository of surface topography data for scientific research and industry applications.

This document describes the current state of the service; we are constantly evolving it towards better usability and more analysis features. Right now, the *digital surface twin* only contains topography data, but in the future will incorporate an even broader set of surface characterizations, such as chemical composition, as well as computational models for the evolution of roughness [42, 43]. We will also work towards incorporating standardized ontologies that are currently being developed for tribological problems [36]. The overarching goal is to get ever closer to a complete digital surface twin that accurately reflects surface performance.

To aid in this process, we would like to encourage suggestions from the scientific and industry communities about ways to further improve data standardization, data analysis, and the advancement of understanding and application of surface finish as a means to improve surface performance. A starting point is the discussion forum on the

GitHub repository at https://github.com/ContactEngineering/TopoBank/discussions or by email to the authors of this article. The web service itself is free of charge to the scientific community and accessible at https://contact.engineering/.

Acknowledgments

This article is dedicated, and the authors are deeply indebted, to the late Mark O. Robbins, who was involved in developing the basic ideas behind *contact.engineering* and starting the project. We thank Tobias Amann, Roland Bennewitz, Claas Bierwisch, Sitangshu Chatterjee, Michele Ciavarella, Qinglin Deng, Martin Dienwiebel, Robert Jackson, Tobias Kling, Mena Klittich, Richard Leute, Zhuohan Li, Gianpietro Moras, Michael Moseler, Nathaniel Orndorf, Antonio Papangelo, Chris Thom and Yang Xu for patience in testing early versions of *contact.engineering*, reporting bugs and suggesting changes to interface and functionality. We are indebted to Johannes Hörmann, Jan Leendertse and Dirk von Suchodoletz for enlightening discussions on open data and research data management, Kolja Glogowski and Marcel Tschöpe for help with the storage infrastructure, and Saher Seeman and Jannis Seyfried for enabling DOI registration. We thank Mathys AG Bettlach (Switzerland) for providing the image of their hip implant and Michal Rössler for designing the contact.engineering logo, both shown in figure 1. The work presented in this paper is the joint outcome of multiple projects, funded by the European Research Council (StG 757343), the Deutsche Forschungsgemeinschaft (grants PA 2023/2, EXC 2193/1—390951807 and INST 39/1099-1 FUGG) and the U.S. National Science Foundation (CMMI-1727378 and CMMI-1844739).

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.57703/ce-5cz7a, https://doi.org/10.57703/ce-8ppx5.

ORCID iDs

Michael C Röttger https://orcid.org/0000-0002-7111-8570
Antoine Sanner https://orcid.org/0000-0002-7019-2103
Luke A Thimons https://orcid.org/0000-0003-4511-1807
Till Junge https://orcid.org/0000-0001-8188-9363
Wolfram G Nöhring https://orcid.org/0000-0003-4203-755X
Tevis D B Jacobs https://orcid.org/0000-0001-8576-914X
Lars Pastewka https://orcid.org/0000-0001-8351-7336

References

[1] Persson B N J, Albohr O, Tartaglino U, Volokitin A I and Tosatti E 2004 On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion *J. Phys. Condens. Matter* **17** R1
[2] Binder L 1912 Der Widerstand von Kontakten *Elektrotechnik und Maschinenbau* **30** 781
[3] Greenwood J A and Williamson J B P 1966 Contact of nominally flat surfaces *Proc. R. Soc. A* **295** 300
[4] Holm R 1967 *Electric Contacts* 4th edn (Heidelberg, Berlin: Springer-Verlag)
[5] Gui C, Elwenspoek M, Tas N and Gardeniers J G E 1999 The effect of surface roughness on direct wafer bonding *J. Appl. Phys.* **85** 7448
[6] Yadollahi A and Shamsaei N 2017 Additive manufacturing of fatigue resistant Materials: Challenges and Opportunities *Int. J. Fatigue* **98** 14
[7] Bowden F P and Tabor D 1950 *The Friction and Lubrication of Solids* (Oxford: Oxford University Press)
[8] Fuller K N G and Tabor D 1975 The effect of surface roughness on the adhesion of elastic solids *Proc. R. Soc. Lond. A Math. Phys. Sci.* **345** 327
[9] Rabinowicz E 1995 *Friction and Wear of Materials* (New York: Wiley)
[10] Gotsmann B and Lantz M A 2013 Quantized thermal transport across contacts of rough surfaces *Nat. Mater.* **12** 59
[11] Bush A W, Gibson R D and Thomas T R 1975 The elastic contact of a rough surface *Wear* **35** 87
[12] Persson B N J 2001 Theory of rubber friction and contact mechanics *J. Chem. Phys.* **115** 3840
[13] Persson B N J 2001 Elastoplastic contact between randomly rough surfaces *Phys. Rev. Lett.* **87** 116101
[14] Hyun S, Pei L, Molinari J-F and Robbins M O 2004 Finite-element analysis of contact between elastic self-affine surfaces *Phys. Rev. E* **70** 026117
[15] Pei L, Hyun S, Molinari J-F and Robbins M O 2005 Finite element modeling of elasto-plastic contact between rough surfaces *J. Mech. Phys. Solids* **53** 2385
[16] Pastewka L and Robbins M O 2014 Contact between rough surfaces and a criterion for macroscopic adhesion *Proc. Natl. Acad. Sci. U. S. A.* **111** 3298
[17] Etsion I 2005 State of the art in laser surface texturing *J. Tribol.* **127** 248
[18] Gachot C, Rosenkranz A, Hsu S M and Costa H L 2017 A critical assessment of surface texturing for friction and wear improvement *Wear* **372–373** 21
[19] Lu P and Wood R J K 2020 Tribological performance of surface texturing in mechanical applications—a review *Surf. Topogr.* **8** 043001
[20] Jacobs T D B, Junge T and Pastewka L 2017 Quantitative characterization of surface topography using spectral analysis *Surf. Topogr.: Metrol. Prop.* **5** 013001
[21] Nečas D and Klapetek P 2012 Gwyddion: an open-source software for spm data analysis *Open Physics* **10** 181
[22] OmniSurf Surface Profile Analysis Software *Digital Metrology* (https://digitalmetrology.com/solution/omnisurf/)
[23] Mountains Surface Imaging *Analysis and Metrology Software, Digital Surf* (https://digitalsurf.com/)
[24] Barba L A 2019 Praxis of reproducible computational science *Comput. Sci. Eng.* **21** 73
[25] Wang A and Müser M H 2018 On the usefulness of the height-difference-autocorrelation function for contact mechanics *Tribol. Int.* **123** 224
[26] Moreira G, da Silva J K L and Kamphorst S O 1994 On the fractal dimension of self-affine profiles *J. Phys. A Math. Gen.* **27** 80794089
[27] Peng C-K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 Mosaic organization of DNA nucleotides *Phys. Rev. E* **49** 1685
[28] Peng C K, Havlin S, Stanley H E and Goldberger A L 1995 Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series *Chaos* **5** 82

[29] Schmittbuhl J, Vilotte J-P and Roux S 1995 Reliability of self-affine measurements *Phys. Rev. E* **51** 131
[30] Cannon M J, Percival D B, Caccia D C, Raymond G M and Bassingthwaighte J B 1997 Evaluating scaled windowed variance methods for estimating the hurst coefficient of time series *Physica A* **241** 606
[31] Sanner A, Nöhring W G, Thimons L A, Jacobs T D B and Pastewka L 2022 Scale-dependent roughness parameters for topography analysis *Appl. Surf. Sci. Adv* **7** 100190
[32] Paskin N 2009 Digital object identifier (DOI®) system *Encyclopedia of Library and Information Sciences* ed M J Bates and M N Maack 3rd ed. (Boca Raton, FL: CRC Press) 1586–92
[33] Haak L L, Fenner M, Paglione L, Pentz E and Ratner H 2012 ORCID: a system to uniquely identify researchers *Learn. Publ.* **25** 259
[34] Rew R and Davis G 1990 NetCDF: an interface for scientific data access *IEEE Comput. Graph. Appl.* **10** 76
[35] Wilkinson M D *et al* 2016 The FAIR guiding principles for scientific data management and stewardship *Sci. Data.* **3** 160018
[36] Garabedian N T *et al* 2022 Generating FAIR research data in experimental tribology *Sci. Data* **9** 315
[37] Gujrati A, Khanal S R, Pastewka L and Jacobs T D B 2018 Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales *ACS Appl. Mater. Interfaces* **10** 29169
[38] Khanal S R, Gujrati A, Vishnubhotla S B, Nowakowski P, Bonifacio C S, Pastewka L and Jacobs T D B 2018 Characterization of small-scale surface topography using transmission electron microscopy *Surf. Topogr.: Metrol. Prop.* **6** 045004
[39] Ramisetti S B, Campañá C, Anciaux G, Molinari J-F, Müser M H and Robbins M O 2011 The autocorrelation function for island areas on self-affine surfaces *J. Phys. Condens. Matter* **23** 215004
[40] Mandelbrot B B, Passoja D E and Paullay A J 1984 Fractal character of fracture surfaces of metals *Nature* **308** 721
[41] Gottlieb D and Shu C-W 1997 On the gibbs phenomenon and its resolution *SIAM Rev.* **39** 644
[42] Hinkle A R, Nöhring W G, Leute R, Junge T and Pastewka L 2020 The emergence of small-scale self-affine surface roughness from deformation *Sci. Adv.* **6** eaax0847
[43] Nöhring W G, Hinkle A R and Pastewka L 2022 Nonequilibrium plastic roughening of metallic glasses yields self-affine topographies with strain-rate and temperature-dependent scaling exponents *Phys. Rev. Materials* **6** 075603
[44] Sayles R S and Thomas T R 1978 Surface topography as a nonstationary random process *Nature* **271** 431
[45] Gujrati A, Sanner A, Khanal S R, Moldovan N, Zeng H, Pastewka L and Jacobs T D B 2021 Comprehensive topography characterization of polycrystalline diamond coatings *Surf. Topogr.: Metrol. Prop.* **9** 014003
[46] Persson B N J and Tosatti E 2001 The effect of surface roughness on the adhesion of elastic solids *J. Chem. Phys.* **115** 5597
[47] Thomas T R 1998 *Rough Surfaces* 2nd edn (London: Imperial College Press)
[48] Church E L and Takacs P Z 1989 Instrumental effects in surface finish measurement Proc. SPIE, Surface Measurement and Characterization **1009** 46–55
[49] Church E L and Takacs P Z 1991 Effects of the nonvanishing tip size in mechanical profile measurements *Proc. SPIE, Optical Testing and Metrology III: Recent Advances in Industrial Optical Inspection* **1332** 504–14
[50] Barber J R 2003 Bounds on the electrical resistance between contacting elastic rough bodies *Proc. R. Soc. A* **459** 53
[51] Ciavarella M, Murolo C and Demelio G 2006 On the elastic contact of rough surfaces: numerical experiments and comparisons with recent theories *Wear* **261** 1102
[52] Jackson R L and Green I 2011 On the modeling of elastic contact between rough surfaces *Tribol. Trans.* **54** 300
[53] Müser M H *et al* 2017 Meeting the contact-mechanics challenge *Tribol. Lett.* **65** 118
[54] Mandelbrot B B 1982 *The Fractal Geometry of Nature* (San Francisco, CA: Freeman)
[55] Abott E J and Firestone F A 1933 Specifying surface quality: a method based on accurate measurement and comparison *Mech. Eng.* **55** 569
[56] Ciavarella M, Delfine V and Demelio G 2006 A 'Re-Vitalized' greenwood and williamson model of elastic contact between fractal surfaces *J. Mech. Phys. Solids* **54** 2569
[57] Joe J, Thouless M D and Barber J R 2018 Effect of roughness on the adhesive tractions between contacting bodies *J. Mech. Phys. Solids* **118** 365
[58] Kalker J J and Randen Y 1972 A minimum principle for frictionless elastic contact with application to non-hertzian half-space contact problems *J. Eng. Math.* **6** 193
[59] Stanley H M and Kato T 1997 An FFT-based method for rough surface contact *J. Tribol.* **119** 481
[60] Polonsky I A and Keer L M 1999 A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques *Wear* **231** 206
[61] Wu J-J 2006 Numerical analyses on elliptical adhesive contact *J. Phys. D Appl. Phys.* **39** 1899
[62] Almqvist A, Sahlin F, Larsson R and Glavatskih S 2007 On the dry elasto-plastic contact of nominally flat surfaces *Tribol. Int.* **40** 574
[63] Mulakaluri N and Persson B N J 2011 Adhesion between elastic solids with randomly rough surfaces: comparison of analytical theory with molecular-dynamics simulations *EPL* **96** 66003
[64] Pohrt R and Popov V L 2012 Normal contact stiffness of elastic solids with fractal rough surfaces *Phys. Rev. Lett.* **108** 104301
[65] Pohrt R, Popov V L and Filippov A E 2012 Normal contact stiffness of elastic solids with fractal rough surfaces for one- and three-dimensional systems *Physical Review E* **86** 026710
[66] Putignano C, Afferrante L, Carbone G and Demelio G 2012 The influence of the statistical properties of self-affine surfaces in elastic contacts: a numerical investigation *J. Mech. Phys. Solids* **60** 973
[67] Prodanov N, Dapp W B and Müser M H 2013 On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections, and reference data *Tribol. Lett.* **53** 433
[68] Pastewka L, Prodanov N, Lorenz B, Müser M H, Robbins M O and Persson B N J 2013 Finite-size scaling in the interfacial stiffness of rough elastic contacts *Phys. Rev. E* **87** 062809
[69] Medina S and Dini D 2014 A numerical model for the deterministic analysis of adhesive rough contacts down to the nano-scale *Int. J. Solids Struct.* **51** 2620
[70] Vollebregt E A H 2014 The bound-constrained conjugate gradient method for non-negative matrices *J. Optim. Theory Appl.* **162** 931
[71] Bemporad A and Paggi M 2015 Optimization algorithms for the solution of the frictionless normal contact between rough surfaces *Int. J. Solids Struct.* **69–70** 94
[72] Yastrebov V A, Anciaux G and Molinari J-F 2015 From infinitesimal to full contact between rough surfaces: evolution of the contact area *Int. J. Solids Struct.* **52** 83
[73] Pastewka L and Robbins M O 2016 Contact area of rough spheres: large scale simulations and simple scaling laws *Appl. Phys. Lett.* **108** 221601
[74] Hulikal S, Bhattacharya K and Lapusta N 2017 The relation between a microscopic threshold-force model and macroscopic models of adhesion *Acta Mech. Sin.* **33** 508
[75] Bazrafshan M, de Rooij M B, Valefi M and Schipper D J 2017 Numerical method for the adhesive normal contact analysis based on a dugdale approximation *Tribol. Int.* **112** 117
[76] Bugnicourt R, Sainsot P, Dureisseix D, Gauthier C and Lubrecht A A 2018 FFT-based methods for solving a rough adhesive contact: description and convergence study *Tribol. Lett.* **66** 29

[77] Frérot L, Bonnet M, Molinari J-F and Anciaux G 2019 A fourier-accelerated volume integral method for elastoplastic contact *Comput. Methods Appl. Mech. Eng.* **351** 951
[78] Monti J M, Pastewka L and Robbins M O 2022 Fractal geometry of contacting patches in rough elastic contacts *J. Mech. Phys. Solids* 104797
[79] Johnson K L 1985 *Contact Mechanics* (Cambridge: Cambridge University Press)
[80] Campañá C and Müser M H 2007 Contact mechanics of real vs. randomly rough surfaces: a green's function molecular dynamics study *EPL* **77** 38005
[81] Akarapu S, Sharp T and Robbins M O 2011 Stiffness of contacts between rough surfaces *Phys. Rev. Lett.* **106** 204301
[82] Campañá C, Persson B N J and Müser M H 2011 Transverse and normal interfacial stiffness of solids with randomly rough surfaces *J. Phys. Condens. Matter* **23** 085001
[83] Liu S, Wang Q and Liu G 2000 A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses *Wear* **243** 101
[84] Campañá C and Müser M H 2006 Practical Green's function approach to the simulation of elastic semi-infinite solids *Phys. Rev. B* **74** 075420
[85] Pastewka L, Sharp T A and Robbins M O 2012 Seamless elastic boundaries for atomistic calculations *Phys. Rev. B* **86** 075459
[86] Monti J M, Pastewka L and Robbins M O 2021 Green's function method for dynamic contact calculations *Phys. Rev. E* **103** 053305
[87] Amba-Rao C L 1969 Fourier transform methods in elasticity problems and an application *J. Franklin Inst.* **287** 241
[88] Müser M H 2008 Rigorous field-theoretical approach to the contact mechanics of rough elastic solids *Phys. Rev. Lett.* **100** 055504
[89] Li J and Berger E J 2001 A boussinesq-cerruti solution set for constant and linear distribution of normal and tangential load over a triangular area *J. Elast.* **63** 137
[90] Pohrt R and Li Q 2014 Complete boundary element formulation for normal and tangential contact problems *Phys. Mesomech.* **17** 334
[91] Carbone G, Lorenz B, Persson B N J and Wohlers A 2009 Contact mechanics and rubber friction for randomly rough surfaces with anisotropic statistical properties *Eur. Phys. J. E* **29** 275
[92] Venugopalan S P, Nicola L and Müser M H 2017 Green's function molecular dynamics: including finite heights, shear, and body fields *Modell. Simul. Mater. Sci. Eng.* **25** 034001
[93] Sebastiano M 2017 Greensche Funktionen für elastische Substrate endlicher Dicke *B.Sc. thesis* Karlsruhe Institute of Technology
[94] Müser M H 2021 Elastic contacts of randomly rough indenters with thin sheets, membranes under tension, half spaces, and beyond *Tribol. Lett.* **69** 25
[95] Love A E H 1892 *A Treatise on the Mathematical Theory of Elasticity* (Cambridge: Cambridge University Press)
[96] Li J and Berger E J 2003 A semi-analytical approach to three-dimensional normal contact problems with friction *Comput. Mech.* **30** 310
[97] Hockney R W 1970 *The potential calculation and some applications Methods in Computational Physics* ed B A Alder, S Fernbach and M Rotenberg 9 (New York: Academic) 135–211
[98] Barnett R N and Landman U 1993 Born–oppenheimer molecular-dynamics simulations of finite systems: Structure and dynamics of $(H_2O)_2$ *Phys. Rev. B* **48** 2081
[99] Weber B, Suhina T, Junge T, Pastewka L, Brouwer A M and Bonn D 2018 Molecular probes reveal deviations from amontons' law in multi-asperity frictional contacts *Nat. Commun.* **9** 888
[100] Gujrati A, Khanal S R, Pastewka L and Jacobs T D B 2020 *Ultrananocrystalline Diamond (UNCD) (Version 2)* contact.engineering (https://doi.org/10.57703/ce-5cz7a)
[101] Sanner A and Pastewka L 2022 *Rough Sphere (Version 1)* contact.engineering (https://doi.org/10.57703/ce-8ppx5)
[102] Wilson G *et al* 2014 Best practices for scientific computing *PLoS Biol.* **12** e1001745
[103] Harris C R *et al* 2020 Array programming with numpy *Nature* **585** 357

Publication III

J. Mech. Phys. Solids 160 (2022) 104781

Contents lists available at ScienceDirect

Journal of the Mechanics and Physics of Solids

journal homepage: www.elsevier.com/locate/jmps

Crack-front model for adhesion of soft elastic spheres with chemical heterogeneity

Antoine Sanner, Lars Pastewka *

Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany
Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany

ARTICLE INFO

Keywords:
Adhesion and adhesives
Hysteresis
Elastic material
Contact mechanics
Crack propagation and arrest

ABSTRACT

Adhesion hysteresis can be caused by elastic instabilities that are triggered by surface roughness or chemical heterogeneity. However, the role of these instabilities in adhesion hysteresis remains poorly understood because we lack theoretical and numerical models accounting for realistic roughness. Our work focuses on the adhesion of soft elastic spheres with low roughness or weak heterogeneity, where the indentation process can be described as a Griffith-like propagation of a nearly circular external crack. We discuss how to describe the contact of spheres with chemical heterogeneity that leads to fluctuations in the local work of adhesion. We introduce a variational first-order crack-perturbation model and validate our approach using boundary-element simulations. The crack-perturbation model faithfully predicts contact shapes and hysteretic force-penetration curves, provided that the contact perimeter remains close to a circle and the contact area is simply connected. Computationally, the crack-perturbation model is orders of magnitude more efficient than the corresponding boundary element formulation, allowing for realistic heterogeneity fields.

1. Introduction

At atomic scales, all molecules attract each other, but macroscopic objects usually do not stick. The explanation for this apparent paradox is that most surfaces are rough, so that the objects only touch on the top of their asperities (Kendall, 2001; Persson et al., 2004; Pastewka and Robbins, 2014). Geckos and insects have compliant fibrilar structures or soft pads at the tip of their feet that conform to surface roughness, sustaining enough adhesion to climb vertical walls (Autumn et al., 2000; Persson et al., 2004; Zhou et al., 2014). Similarly, a soft layer makes tapes adhere to diverse surfaces (Creton and Leibler, 1996). Being able to tune or switch off adhesion is important for applications such as pick and place systems (Carlson et al., 2012; Hensel et al., 2018; Deneke et al., 2021) or skin adhesives (Karp and Langer, 2011; Kwak et al., 2011; Yu and Cheng, 2018; Hwang et al., 2018).

Soft materials stick to rough surfaces if the elastic energy needed to conform to surface roughness is small compared to the surface energy gained by contact (Fuller and Tabor, 1975; Briggs and Briscoe, 1977; Persson and Tosatti, 2001). In an indentation experiment with such materials, the force needed to break the contact is often higher than the force measured during indentation (Chen et al., 1991). This observation contradicts expectations from classical theories on smooth surfaces (Johnson et al., 1971) and theories for rough contacts that assume the contact follows thermodynamic equilibrium (Persson and Tosatti, 2001; Persson, 2002a,b). This *adhesion hysteresis* can be caused by several mechanisms like viscoelasticity, molecular rearrangements and elastic instabilities. Recent experiments with spherical indenters indicate that in some cases, elastic instabilities triggered by surface roughness play

* Corresponding author at: Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany.
E-mail addresses: antoine.sanner@imtek.uni-freiburg.de (A. Sanner), lars.pastewka@imtek.uni-freiburg.de (L. Pastewka).

https://doi.org/10.1016/j.jmps.2022.104781
Received 15 August 2021; Received in revised form 21 December 2021; Accepted 1 January 2022
Available online 19 January 2022

the dominant role (Kesari et al., 2010; Dalvi et al., 2019). Details of how roughness gives rise to elastic instabilities remain poorly understood, mainly because we lack theoretical models accounting for realistic surface roughness.

In this paper, we discuss how to describe the contact of elastic spheres with chemical heterogeneity, local fluctuations in the work of adhesion, using a first-order crack-perturbation model. Understanding the hysteresis caused by quenched disorder in the work of adhesion is a first step towards a general model for geometrically rough surfaces. Our model will permit efficient numerical simulations of realistic system sizes and clarify the link between adhesive contact mechanics and classic theories on pinning of elastic lines.

Depending on the compliance of the elastic material, hysteresis arises from disconnected patches snapping in and out of contact or depinning instabilities in the motion of the contact perimeter, the crack front. Brute-force numerical methods like the boundary-element method (BEM) can capture these different regimes of instabilities (Medina and Dini, 2014; Dapp and Müser, 2015; Carbone et al., 2015; Deng and Kesari, 2017; Wang et al., 2021). However, for soft materials it is challenging to sufficiently discretize the adhesive neck (or crack tip) and at the same time include a representative amount of surface roughness (Wang et al., 2021). Insufficient discretization in boundary element models can additionally lead to "lattice trapping", that causes artificial pinning of the crack front alike physical lattice trapping in atomic crystal (Thomson et al., 1971).

More coarse-grained models either concentrate on the two opposite limits of very low contact fraction or full contact. Asperity models describe surface roughness as a set of spherical peaks with random heights (Greenwood and Williamson, 1966; Fuller and Tabor, 1975), that dissipate energy during snap in and out of contact instabilities (Zappone et al., 2007; Wei et al., 2010; Greenwood, 2017; Deng and Kesari, 2019; Violano and Afferrante, 2021). This type of approximation breaks down when asperities coalesce and a larger fraction of the surface comes into contact. In the limit of small roughness and low stiffness, the contact area is simply connected. In that limit, the surface roughness causes energy barriers that pin the crack front and the crack front dissipates energy during depinning instabilities. This phenomenon has to date only been studied on one-dimensional roughness (Guduru, 2007; Guduru and Bull, 2007; Kesari et al., 2010; Kesari and Lew, 2011; Carbone et al., 2015).

The pinning of a crack front by quenched disorder in the work of adhesion (or equivalently, the fracture toughness) is better understood (Gao and Rice, 1989; Schmittbuhl et al., 1995; Xia et al., 2012; Démery and Ponson, 2014; Xia et al., 2015; Chopin et al., 2015; Ponson, 2016; Lebihain et al., 2021). An essential step towards this understanding was Rice's description of a semi-infinite crack as an elastic line with long-range elasticity (Rice, 1985a). This equation belongs to the class of elastic interfaces pinned by a random field and the current understanding on fracture of heterogeneous media benefited from works in other fields (Larkin and Ovchinnikov, 1979; Fisher, 1983; Robbins and Joanny, 1987; Middleton, 1992; Amaral et al., 1995; Zhou, 2000; Rosso and Krauth, 2002; Tanguy and Vettorel, 2004; Rosso et al., 2007).

Rice's first-order perturbation was applied to several geometries with finite sizes (Lazarus, 2011; Patinet et al., 2013). To our knowledge, this perturbation approach was never applied to the contact of spheres or similar indenters, except recently by Argatov (2021) to investigate the effect of indenter ellipticity on the pull-off force. Applying crack perturbation to spheres is of particular interest because most of the experiments reporting adhesion hysteresis were performed using spherical indenters (Chaudhury and Whitesides, 1991; Chen et al., 1991; Kesari et al., 2010; Dalvi et al., 2019). Crack perturbation is a promising approach to understand how surface roughness affects adhesion hysteresis in indentation experiments where the contact area is simply connected. The first step towards this end is to understand how the adhesion hysteresis depends on the geometry and strength of quenched disorder in the work of adhesion.

In this paper, we discuss how to describe the contact of spheres with fluctuating work of adhesion as a perturbation from the homogeneous, perfectly circular contact. In the circular contact, the stress intensity factor is provided by the Johnson–Kendall–Roberts (JKR) theory (Johnson et al., 1971; Barthel, 2008), and its first functional derivative by Gao and Rice (1987a). We use these equations to write first-order approximations of the distortion of the contact area caused by work of adhesion heterogeneity. The extrapolation of the contact line's elastic response can be done in several ways. In the past , either the stress intensity factor (Gao and Rice, 1989; Fares, 1989) or the energy release rate (Bonamy et al., 2008; Ponson and Bonamy, 2010; Patinet et al., 2013; Chopin et al., 2015; Ponson, 2016) was linearly extrapolated. However, the latter approaches violate the symmetry of the second derivatives of the elastic energy. We here propose a third equation that satisfies these symmetries and therefore constitutes a variational approach to crack perturbation. In order to validate our model and to discriminate between different variants of the perturbation, we compare the crack-perturbation models to finely discretized BEM simulations. Fig. 1 illustrates schematically the two types of models used in this paper.

2. Problem definition

We consider the contact of a sphere (to be exact, a paraboloid) adhering to an elastic half-space at a fixed rigid body penetration Δ. This case can be mapped to the contact of two spheres with the same composite radius R and contact modulus E' (Johnson, 1985). When only one half-space deforms, $E' = E/(1-\nu^2)$, where E is Young's modulus and ν is Poisson's ratio. Fracture mechanics typically considers the contact of two elastic half-spaces where $E' = E/2(1-\nu^2)$. We assume the contact is frictionless and consider only vertical displacements.

The equilibrium contact area minimizes the total energy $\Pi = \Pi_{\text{mech}} + \Pi_{\text{surf}}$. The mechanical energy Π_{mech} contains the elastic strain energy U_{el} and the potential of external forces. The surface energy Π_{surf} results from adhesive interactions. We consider the limit of infinitesimally short interaction range (JKR-limit), where

$$\Pi_{\text{surf}} = \int_{A_c} \mathrm{d}x\, \mathrm{d}y\, w(x,y). \tag{1}$$

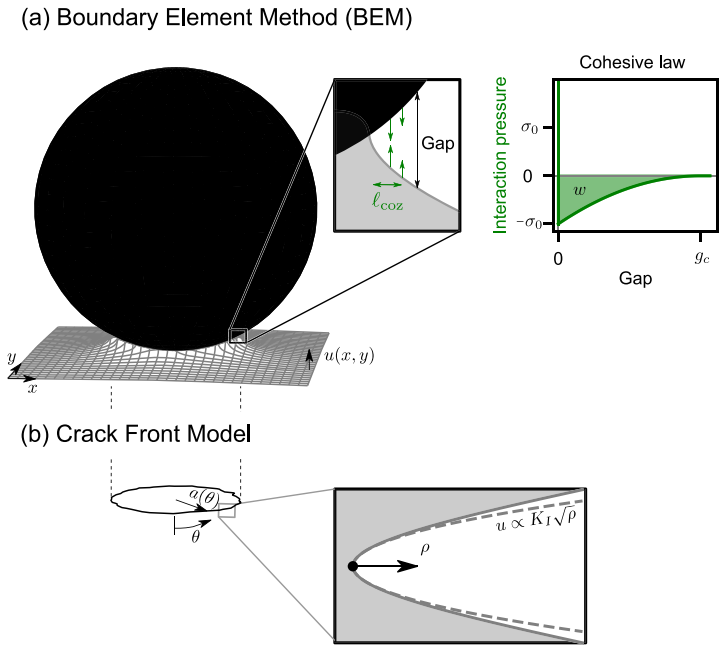


Fig. 1. Schematic illustration of the two types of models used in this work to describe the contact of a sphere. (a) In the boundary-element method (BEM), the surface displacements $u(x, y)$ are computed using the Green's function of the elastic half-space and adhesive attractions are modeled using a cohesive law that relates the interaction pressure to the gap between the two surfaces. The integral over the cohesive law gives the work of adhesion w. Spatial heterogeneity is encoded into the cohesive law. (b) In the crack-front model, the state of the nearly circular contact is described by the contact radius as a function of angle $a(\theta)$. The heterogeneous work of adhesion distorts the shape of the crack front. The elastic response of the contact line is approximated using the first-order variation of the stress intensity factor with respect to the contact shape developed by Gao and Rice (1987a).

A_c is the contact area and the work of adhesion $w(x, y)$ depends on the position (x, y). Note that the work of adhesion is the same for a receding crack (indentation) as for an opening crack (retraction), i.e. there is no *intrinsic* hysteresis in the work of adhesion.

Once nondimensionalized using distinct vertical and lateral length units, the JKR contact is parameter free (Muller et al., 1980; Maugis, 2010; Müser, 2014). We present our numerical results in nondimensional units (Maugis, 2010) and indicate by an asterisk* when quantities are normalized. Specifically, lengths along the surface of the half-space (e.g., the contact radius) are normalized by $(3\pi w_m R^2/4E')^{1/3}$, lengths in vertical direction (e.g., displacements) by $(9\pi^2 w_m^2 R^2/16E'^2)^{1/3}$ and normal forces by $\pi w_m R$. w_m is the median work of adhesion. The equations are in dimensional form but can be nondimensionalized by substituting $R = 1$, $w_m = 1/\pi$ and $E' = 3/4$.

3. Reference model

We used the BEM to validate the crack-front models. BEM computes the surface displacements $u(x, y)$ (Fig. 1a) minimizing the total energy Π at prescribed rigid body penetration Δ. This method makes no assumptions on the contact morphology, allowing for holes and disconnected contact islands, but it is computationally too expensive to properly discretize contacts with small heterogeneities.

The mechanical energy $\Pi_{\text{mech}} = U_{\text{el}}$ and its gradients were computed using the Green's function for pressures that are constant on each pixel (Love, 1929; Johnson, 1985). While the Green's function is nonperiodic, we accelerated calculations with a fast Fourier transform (Stanley and Kato, 1997; Campañá and Müser, 2006; Pastewka et al., 2012) by introducing a padding region that decouples periodic images (Hockney, 1970; Liu et al., 2000; Pastewka and Robbins, 2016).

The surface energy needs to be regularized using an interaction potential ϕ between the surfaces (cohesive law),

$$\Pi_{\text{surf}} = \int_{-\infty}^{+\infty} \mathrm{d}x\ \mathrm{d}y\ \phi\left(w(x, y), g(x, y)\right), \tag{2}$$

where $w(x, y)$ is the (spatially varying) work of adhesion and $g(x, y)$ is the gap between the surfaces. The results converge to the JKR limit for small pixel size and interaction range, i.e. high Tabor parameter μ_{T} (Muller et al., 1980; Maugis, 1992; Greenwood, 1997; Müser, 2014), but the point of instability where the two interfaces jump into contact converges particularly slowly with the interaction range (Wu, 2010; Ciavarella et al., 2017). Wang et al. (2021) could slightly retard the premature jump into contact by using a potential with a cutoff. Our choice of cohesive law,

$$\phi(w, g) = \begin{cases} -w\left(1 - \frac{g}{g_c}\right)^3, & g < g_c \\ 0, & \text{else,} \end{cases} \tag{3}$$

is based on the same observation. Eq. (3) describes a cubic potential where w is the work of adhesion, g is the gap and g_c the cutoff distance. The derivative of ϕ (interaction pressure) is illustrated in Fig. 1a. The maximum attractive stress $\sigma_0 = 3w/g_c$ occurs at the perimeter of the contact area, where $g = 0$. To model (chemical) heterogeneity, the interaction parameters are different in each pixel. We keep g_c constant, so that σ_0 and w vary proportionally from pixel to pixel.

The surfaces repel each other with a hard-wall potential, which is implemented as inequality constraints ($g(x, y) \geq 0$) in the minimization algorithms. We used the constrained quasi-newton algorithm L-BFGS-B (Byrd et al., 1995) and a constrained conjugate gradient algorithm (Polonsky and Keer, 1999; Vollebregt, 2014; Bugnicourt et al., 2018) to minimize the energy. The latter algorithm is parallelized with the message passing interface and was used to produce precise reference solutions for the contact shape, but could not be used in the presence of instabilities. Note that in general, these minimization algorithms have better convergence properties when the potential has a continuous second derivative, justifying our choice of a third-order polynomial. BEM with hard-wall repulsion and finite-ranged attraction (Müser, 2014, 2016; Müser et al., 2017; Wang and Müser, 2017; Bazrafshan et al., 2017; Rey et al., 2017; Bugnicourt et al., 2018; Monti et al., 2021) or soft (Lennard-Jones type) repulsion (Greenwood, 1997; Feng, 2000; Wu, 2010; Medina and Dini, 2014; Pastewka and Robbins, 2014, 2016; Persson and Scaraggi, 2014; Monti et al., 2019; Ghanbarzadeh et al., 2020; Wang et al., 2021) have been used in the past to study the adhesion of spheres and rough surfaces. During retraction, these models are similar to fiber-bundle models of quasi-brittle fracture (Batrouni et al., 2002; Schmittbuhl et al., 2003; Stormo et al., 2012; Gjerden et al., 2013, 2014) and threshold-force models (Pohrt and Popov, 2015; Hulikal et al., 2017; Li et al., 2019).

The finite interaction range introduces a cohesive zone around the contact perimeter in which the surfaces attract each other. This cohesive zone needs to be small in comparison to the scale of the heterogeneity (Chen et al., 2008) in order to mimic a JKR-like contact. We estimate the width of the cohesive zone using a Dugdale model and assume it is small in comparison to the contact radius (Maugis, 2010):

$$\ell_{\mathrm{coz}} = \frac{\pi E' w}{4\sigma_0^2} \tag{4}$$

For uniform work of adhesion, Müser (2014) showed that the (nondimensionalized) contact radius difference with the JKR limit is asymptotically $\sim \ell^*_{\mathrm{coz}} \sim \mu_{\mathrm{T}}^{-2}$, where $\ell^*_{\mathrm{coz}} = \ell_{\mathrm{coz}}/(3\pi w_m R^2/4E')^{1/3}$ is the nondimensionalized cohesive zone width, see Section 2. We use ℓ^*_{coz} as a proxy for interaction range and Tabor parameter μ_{T} (Tabor, 1977; Muller et al., 1980; Maugis, 1992). We chose the pixel size ℓ_{pix} small enough so that further grid refinement affects the contact radius and the force less than decreasing the interaction range.

4. Crack-front model

4.1. Axisymmetric contact: The JKR model

We briefly review the JKR model of the adhesion of a sphere against an elastic half-space. JKR described the contact by the balance of elastic and surface energy of a circular crack (Griffith and Taylor, 1921). At equilibrium, the contact radius a is such that the increment of mechanical energy equals the increment of surface energy:

$$\frac{\partial \Pi_{\mathrm{mech}}(a, \Delta)}{\partial a} = 2\pi a w(a) \tag{5}$$

Π_{mech} is the mechanical potential and w is the work of adhesion at the perimeter of the contact. In this paper, we consider only boundary conditions with fixed rigid body penetration Δ, where Π_{mech} is the elastic energy U_{el}.

The elastic energy release rate,

$$G_{\mathrm{JKR}}(a, \Delta) = \frac{1}{2\pi a}\frac{\partial \Pi_{\mathrm{mech}}(a, \Delta)}{\partial a}, \tag{6}$$

can be related to the intensity K of the singularity of the stress field near the crack tip (Irwin, 1957):

$$G = \frac{K^2}{2E'}. \tag{7}$$

JKR obtained the stress intensity factor K_{JKR} for sphere on flat contact by superposing contact pressures of the Hertzian sphere (Hertz, 1881) and of the circular flat punch (Sneddon, 1946; Maugis, 2010):

$$K_{\mathrm{JKR}}(a, \Delta) = \left(\frac{a^2}{R} - \Delta\right)\frac{E'}{\sqrt{\pi a}}. \tag{8}$$

Here and below we use the subscript JKR to indicate that the contact area is perfectly circular. Inserting Eqs. (6) to (8) into Eq. (5) yields an equilibrium equation for the contact radius as a function of the rigid body penetration. The normal force follows from the same superposition (Johnson et al., 1971; Maugis, 2010):

$$F_{\mathrm{JKR}}(a, \Delta) = \frac{4E'}{3R}a^3 + 2aE'\left(\Delta - \frac{a^2}{R}\right). \tag{9}$$

The increment in surface energy is the integral of the work of adhesion within the area swept by the contact line. The JKR theory applies to any axisymmetric work of adhesion, where w depends on the radius in Eq. (5).

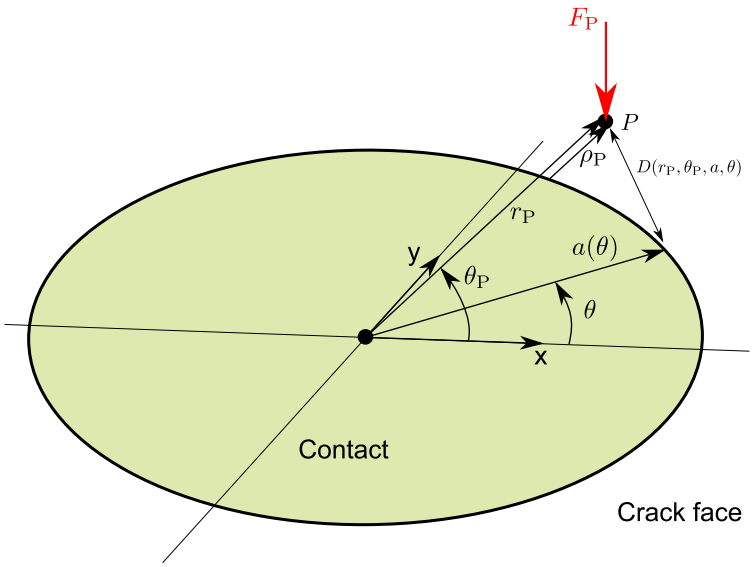


Fig. 2. The crack-face weight function k is central in the derivation of the first-order variation of the stress intensity factor. $k_\mathrm{JKR}(a, r_\mathrm{P}, \theta_\mathrm{P}, \theta)$ corresponds to the stress intensity factor at angle θ when a unit point force F_P is applied on the face of a circular external crack at $r_\mathrm{P}, \theta_\mathrm{P}$. The origin of the coordinate system is aligned with the tip of the spherical indenter.

4.2. Equilibrium condition for non-circular contacts

We now consider the perimeter of a nearly circular contact. The function $a(\theta)$ represents the planar distance between the point at angle θ on the contact perimeter (or crack front) and the tip of the sphere. The mechanical potential $\Pi_\mathrm{mech}([a])$ is now a *functional* of the contact radius a. (We use square brackets $[\cdot]$ to indicate a functional dependence.) The functional derivative $\delta\Pi_\mathrm{mech}/\delta a(\theta)$ is defined by

$$\delta\Pi_\mathrm{mech}([\delta a]) = \int_0^{2\pi} \mathrm{d}\theta \delta a(\theta) \frac{\delta\Pi_\mathrm{mech}}{\delta a(\theta)}, \tag{10}$$

where $\delta\Pi_\mathrm{mech}([\delta a])$ is the first variation of the mechanical potential due to a perturbation of the crack front δa (Giaquinta and Hildebrandt, 2004; Finnis, 2010). We omit the penetration Δ and the contact shape a from the arguments for brevity.

$\delta\Pi_\mathrm{mech}([\delta a])$ depends on the energy release rate $G(\theta)$ along the crack front,

$$\delta\Pi_\mathrm{mech}([\delta a]) = \int_0^{2\pi} \mathrm{d}\theta a(\theta) \delta a(\theta) G(\theta), \tag{11}$$

so that

$$\frac{\delta\Pi_\mathrm{mech}}{\delta a(\theta)} = a(\theta) G(\theta). \tag{12}$$

The mechanical energy release rate $G(\theta)$ is the local work per surface area needed to close the crack faces and is linked to the stress intensity factors $K(\theta)$ by Irwin's relation, Eq. (7).

The stationarity of the total potential, i.e. the Griffith criterion, requires that the mechanical energy release rate at point θ on the contact perimeter equals the local work of adhesion at the point $(a(\theta), \theta)$ on the heterogeneous plane:

$$G(\theta) = w(a(\theta), \theta) \tag{13}$$

G is available in closed form only for simple crack and contact shapes. Rice's weight function theory allows us to compute efficient approximations of G when perturbed from a reference configuration.

4.3. First-order variation of the stress intensity factor

Rice showed that the sensitivity of the stress intensity factor to the contact shape $\delta K/\delta a$ is linked to the crack-face weight function. The crack-face weight function $k([a]; P, \theta)$ is the stress intensity factor caused by a unit force at point P outside the contact area and is known for simple geometries (Tada et al., 2000). While Rice's first variation of K was almost exclusively applied to flat-on-flat contact geometries (planar crack problems), $\delta K/\delta a$ is unaffected by the presence of the Hertzian displacements in the contact area (Rice, 1985b; Borodachev, 1991). To make explicit that the first-order variation of K for circular connections (Gao and Rice, 1987a) applies also to the contact of spheres, we review the main steps leading to $\delta K/\delta a$.

We apply a force F_P on the elastic half-space at point P outside the contact area (on the crack face). This point force causes the displacement u_P at P. At fixed rigid body penetration Δ and fixed force F_P, the mechanical potential is the Legendre transform of the elastic strain energy U_el,

$$\Pi_\mathrm{mech}(F_\mathrm{P}) = \min_{u_\mathrm{P}} \left\{ U_\mathrm{el}(u_\mathrm{P}) - u_\mathrm{P} F_\mathrm{P} \right\}, \tag{14}$$

with derivatives

$$\left.\frac{\partial \Pi_\mathrm{mech}}{\partial F_\mathrm{P}}\right|_a = -u_\mathrm{P} \qquad \text{and} \qquad \left.\frac{\delta \Pi_\mathrm{mech}}{\delta a(\theta)}\right|_{F_\mathrm{P}} = \left.\frac{\delta U_\mathrm{el}}{\delta a(\theta)}\right|_{u_\mathrm{P}} = G(\theta) a(\theta). \tag{15}$$

U_el is the strain energy and $-u_\mathrm{P} F_\mathrm{P}$ is the potential of the applied force F_P. Because the second derivatives of Π_mech are continuous, the order of differentiation does not matter (Schwartz's theorem):

$$\frac{\delta u_\mathrm{P}}{\delta a(\theta)} = -\frac{\partial G(\theta)}{\partial F_\mathrm{P}} a(\theta). \tag{16}$$

From Eq. (7) and $\partial K / \partial F_\mathrm{P} = k([a]; P, \theta)$ we obtain

$$\frac{\delta u_\mathrm{P}}{\delta a(\theta)} = -\frac{1}{E'} K(\theta) k([a]; P, \theta) a(\theta). \tag{17}$$

Setting $F_\mathrm{P} = 0$ (for arbitrary P) we see that the crack-face weight function k describes how the crack faces deform when the crack front moves.

Knowing that the displacements near the crack tip correspond to the stress intensity factor, we can now deduce how the stress intensity factor is affected by a perturbation of the contact radius δa. At the perpendicular distance ρ_P away from the point θ_P on the crack tip,

$$u(\rho_\mathrm{P}, \theta_\mathrm{P}) = \frac{4}{E'} \sqrt{\frac{\rho_\mathrm{P}}{2\pi}} K(\theta_\mathrm{P}) + \mathcal{O}(\rho_\mathrm{P}^{3/2}). \tag{18}$$

From Eq. (18), the perturbation of the stress intensity factor becomes

$$\begin{aligned} \delta K(\theta_\mathrm{P}) &= \lim_{\rho_\mathrm{P} \to 0} \frac{E'}{4} \sqrt{\frac{2\pi}{\rho_\mathrm{P}}} \int_0^{2\pi} \mathrm{d}\theta \; \delta a(\theta) \frac{\delta u(\rho_\mathrm{P}, \theta_\mathrm{P})}{\delta a(\theta)} \\ &= -\lim_{\rho_\mathrm{P} \to 0} \sqrt{\frac{\pi}{8\rho_\mathrm{P}}} \int_0^{2\pi} \mathrm{d}\theta a(\theta) \; \delta a(\theta) K(\theta) k([a]; \rho_\mathrm{P}, \theta_\mathrm{P}, \theta), \end{aligned} \tag{19}$$

for δa restricted to $\delta a(\theta_\mathrm{P}) = 0$. Rice introduced this restriction to ensure that $\delta \rho_\mathrm{P} = 0$. As a consequence, Eq. (19) is not a first-order variation because δa is not arbitrary. This restriction will be relaxed later by expanding the circular reference configuration by $\delta a(\theta_\mathrm{P})$.

When the contact perimeter is a circle ($a(\theta) = \text{const.}$), we know K from Johnson et al. (1971) and k from Galin (Galin, 2008), enabling us to evaluate Eq. (19).

The crack-face weight function for the JKR contact is the same as for a circular connection. For a circular connection with radius a and fixed displacements at infinity (Gao and Rice, 1987a; Rice, 1989; Galin, 2008),

$$k_\mathrm{JKR}(a, r_\mathrm{P}, \theta_\mathrm{P}, \theta) = \frac{\sqrt{(r_\mathrm{P}^2 - a^2)/(a\pi^3)}}{D^2(r_\mathrm{P}, \theta_\mathrm{P}, a, \theta)}, \tag{20}$$

where $D^2(r_\mathrm{P}, \theta_\mathrm{P}, a, \theta) = a^2 + r_\mathrm{P}^2 - 2 a r_\mathrm{P} \cos(\theta_\mathrm{P} - \theta)$ is the square of the distance between the point of application of the force on the crack face, $P = (r_\mathrm{P}, \theta_\mathrm{P})$, and the point on the crack-front where the stress intensity factor is calculated, θ. Note that k_JKR is not a functional of the whole radius, since in JKR the underlying contact area is a circle with radius a.

First applying a uniform perturbation of magnitude $\delta a(\theta_\mathrm{P})$ to the circular reference contact (radius $\tilde{a}_0$) and then using $K = K_\mathrm{JKR}$ and inserting Eq. (20) into Eq. (19) yields the first-order variation of $K(\theta_\mathrm{P})$ at $a = \tilde{a}_0 = \text{const.}$:

$$\delta K(\theta_\mathrm{P}) = \frac{\partial K_\mathrm{JKR}(\tilde{a}_0)}{\partial \tilde{a}_0} \delta a(\theta_\mathrm{P}) - \frac{K_\mathrm{JKR}(\tilde{a}_0)}{8\pi} \operatorname{PV} \int_0^{2\pi} \mathrm{d}\theta \; \frac{\tilde{a}_0 (\delta a(\theta) - \delta a(\theta_\mathrm{P}))}{D^2(\tilde{a}_0, \theta_\mathrm{P}, \tilde{a}_0, \theta)} \tag{21}$$

We point to Gao and Rice (1987a,b) for details on how the limit $\rho_\mathrm{P} \to 0$ leads to the Cauchy principal value integral $\operatorname{PV} \int$. Alternatively, this limit can be taken without applying the preliminary uniform perturbation, leading to a finite part of Hadamard instead of a Cauchy principal value integral (Salvadori and Fantoni, 2014).

Note that Eqs. (8), (20) and (21) are exact only for an infinite elastic halfspace. Furthermore, we assume that the stress field at the crack tip is only mode-I, which requires either that the contact is frictionless or that the lateral displacements of the two contacting solids match. Lateral displacements match when the Poisson ratio $\nu = 0.5$ or when the two contacting materials have the same elastic properties. Piccolroaz et al. (2007) derived δK for a straight crack between two halfspaces with mismatching elastic properties; Legrand et al. (2011) and Xia et al. (2015) discussed the perturbation of a straight mode-I crack in thin plates.

We now recast Eq. (21) using the Fourier series of a. Inserting

$$a(\theta) = \sum_{n \in \mathbb{Z}} \tilde{a}_n e^{in\theta} \tag{22}$$

into Eq. (21) yields

$$\delta K(\theta_\mathrm{P}) = \frac{\partial K_\mathrm{JKR}(\tilde{a}_0)}{\partial \tilde{a}_0}\delta a(\theta_\mathrm{P}) + \frac{K_\mathrm{JKR}(\tilde{a}_0)}{\tilde{a}_0}\sum_{n\in\mathbb{Z}\setminus\{0\}}\frac{|n|}{2}\,\delta\tilde{a}_n\,e^{in\theta_\mathrm{P}}. \tag{23}$$

From the inverse of Eq. (22),

$$\tilde{a}_n = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\; e^{-in\theta}a(\theta), \tag{24}$$

we obtain $\delta\tilde{a}_n/\delta a(\theta) = \exp(-in\theta)/2\pi$. The functional derivative of K then becomes

$$\frac{\delta K(\theta_\mathrm{P})}{\delta a(\theta)} = \frac{\partial K_\mathrm{JKR}(\tilde{a}_0)}{\partial \tilde{a}_0}\delta(\theta-\theta_\mathrm{P}) + \frac{K_\mathrm{JKR}(\tilde{a}_0)}{2\pi\tilde{a}_0}\sum_{n\in\mathbb{N}}|n|\cos n(\theta_\mathrm{P}-\theta), \tag{25}$$

which depends only on $|\theta-\theta_\mathrm{P}|$. The translational invariance reflects the axial symmetry of the unperturbed state. As will be discussed in a few paragraphs, θ and θ_P commute because $\delta K/\delta a(\theta)$ is related to the second derivative of the energy.

The key step leading to $\delta K/\delta a(\theta)$ was recognizing the symmetry of the derivatives of the elastic energy, Eq. (16), that follows from the smoothness of the elastic energy. Hence, Eq. (25) not only applies to flat indenters (Rice, 1985a; Gao and Rice, 1987a,b; Rice, 1989), but also to contacts against spheres or any smooth indenters. All the relevant information of the axisymmetric contact geometry is in K_JKR and its derivatives. Knowing the first-order variation of K when the contact is perfectly circular allows us to approximate the equilibrium Eq. (13) to first order in the perturbation from circularity.

4.4. Quadratic approximation of the deformation energy

The straightforward approach to construct a first-order model is to linearly extrapolate either G or K. However, these approximations do not conserve the variational property of aG, i.e. that it is the gradient of a potential, Eq. (12). We now guess a quadratic approximation for the elastic energy and verify a posteriori that it possesses the exact first and second derivatives at $a(\theta) = \tilde{a}_0$. Our guess for the elastic energy is

$$U_\mathrm{el} = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\; U_\mathrm{JKR}(a(\theta)) + \pi G_\mathrm{JKR}(\tilde{a}_0)\sum_{n\in\mathbf{Z}}|n||\tilde{a}_n|^2 \tag{26}$$

with first derivative

$$\frac{\delta U_\mathrm{el}}{\delta a(\theta)} = \frac{1}{2\pi}\frac{\partial U_\mathrm{JKR}(a(\theta))}{\partial a(\theta)} + G_\mathrm{JKR}(\tilde{a}_0)\sum_{n\in\mathbf{Z}}|n|\tilde{a}_n e^{in\theta} + \frac{1}{2}\frac{\partial G_\mathrm{JKR}(\tilde{a}_0)}{\partial\tilde{a}_0}\sum_{n\in\mathbf{Z}}|n||\tilde{a}_n|^2 \tag{27}$$

and second derivative

$$\begin{aligned}\frac{\delta^2 U_\mathrm{el}}{\delta a(\theta)\delta a(\theta_\mathrm{P})} =& \frac{1}{2\pi}\frac{\partial^2 U_\mathrm{JKR}(a(\theta))}{\partial a^2(\theta)}\delta(\theta-\theta_\mathrm{P}) + \frac{1}{2\pi}G_\mathrm{JKR}(\tilde{a}_0)\sum_{n\in\mathbf{Z}}|n|e^{in(\theta-\theta_\mathrm{P})} \\ &+ \frac{1}{2\pi}\frac{\partial G_\mathrm{JKR}(\tilde{a}_0)}{\partial\tilde{a}_0}\sum_{n\in\mathbf{Z}}|n|\tilde{a}_n(e^{in\theta}+e^{in\theta_\mathrm{P}}) + \frac{1}{4\pi}\frac{\partial^2 G_\mathrm{JKR}(\tilde{a}_0)}{\partial\tilde{a}_0^2}\sum_{n\in\mathbf{Z}}|n||\tilde{a}_n|^2.\end{aligned} \tag{28}$$

U_JKR is the elastic energy in the perfectly circular contact (Johnson et al., 1971) and $\partial U_\mathrm{JKR}/\partial a = 2\pi a G_\mathrm{JKR}(a)$. It is straightforward to show that Eq. (28) for $a(\theta) = \tilde{a}_0$ gives Eq. (25). We determine the contact shape by solving (see also Eq. (12))

$$\frac{\delta U_\mathrm{el}}{\delta a(\theta)} - w(\theta, a(\theta))a(\theta) = 0 \tag{29}$$

using a minimization algorithm (see Section 4.7). We will call this model CF-E when presenting our results.

The normal force applied on the indenter follows directly from our approximation of the energy:

$$F([a];\Delta) = \frac{\partial U_\mathrm{el}([a];\Delta)}{\partial\Delta} \tag{30}$$

Using Eq. (26),

$$F([a];\Delta) = \frac{1}{2\pi}\int_0^{2\pi}\mathrm{d}\theta\; F_\mathrm{JKR}(a(\theta),\Delta) + \pi\frac{\partial G_\mathrm{JKR}(\tilde{a}_0,\Delta)}{\partial\Delta}\sum_{n\in\mathbf{Z}}|n||\tilde{a}_n|^2. \tag{31}$$

4.5. Other first-order approximations

As a benchmark we also consider two common first-order approximations of the crack front. The first approximation linearly extrapolates the stress intensity factor (Gao and Rice, 1989; Fares, 1989),

$$K(\theta) = K_\mathrm{JKR}(\tilde{a}_0) + \frac{\partial K_\mathrm{JKR}(\tilde{a}_0)}{\partial\tilde{a}_0}\left(a(\theta)-\tilde{a}_0\right) + \frac{K_\mathrm{JKR}(\tilde{a}_0)}{\tilde{a}_0}\sum_{n\in\mathbf{Z}}\frac{|n|}{2}\tilde{a}_n e^{in\theta}, \tag{32}$$

and solves for

$$K(\theta) = K_\mathrm{c}(\theta, a(\theta)) = \sqrt{2E' w(\theta, a(\theta))}. \tag{33}$$

We will refer to this model as CF-K.

The second approximation linearly extrapolates the energy release rate (Bonamy et al., 2008; Chopin et al., 2015; Ponson, 2016),

$$G(\theta) = G_{\mathrm{JKR}}(\tilde{a}_0) + \frac{\partial G_{\mathrm{JKR}}(\tilde{a}_0)}{\partial \tilde{a}_0}(a(\theta) - \tilde{a}_0) + \frac{G_{\mathrm{JKR}}(\tilde{a}_0)}{\tilde{a}_0} \sum_{n \in \mathbf{Z}} |n| e^{in\theta} \tilde{a}_n, \tag{34}$$

and solves for Eq. (13) We will refer to this model as CF-G.

An alternative first-order extrapolation,

$$G(\theta) = G_{\mathrm{JKR}}(a(\theta)) \left(1 + \frac{1}{a(\theta)} \sum_{n \in \mathbf{Z}} |n| e^{in\theta} \tilde{a}_n \right),$$

should in principle be more accurate, because $G_{\mathrm{JKR}}(a(\theta))$ is not linearized. However we observed that using the latter equation does not improve the prediction of the normal force and Eq. (34) is better suited for analytical models. We did the same observation with the analogous expression for the stress intensity factor.

Because these crack-front models do not possess an elastic energy, we cannot use Eq. (30) to compute the normal force. As an approximation, we neglect the effect of the undulations of the crack front on the normal force and insert the mean contact radius $\tilde{a}_0$ into the expression for a perfectly circular contact, Eq. (9). We will discuss in Section 5 that the crack shape affects the normal force only to second order.

4.6. Symmetries

Rice obtained the first-order variation of K, Eq. (23), from the symmetry of second derivatives (or the path independence of the work) with respect to the contact radius and a point force applied on the crack face. Similarly, the second derivatives with respect to the contact radius are symmetric (Gao and Rice, 1989; Leblond et al., 2012; Salvadori and Fantoni, 2014):

$$\frac{\delta^2 U_{\mathrm{el}}}{\delta a(\theta) \delta a(\theta_{\mathrm{P}})} = \frac{\delta^2 U_{\mathrm{el}}}{\delta a(\theta_{\mathrm{P}}) \delta a(\theta)} \tag{35}$$

For the energy release rate, it follows from Eq. (12) that $a(\theta)\delta G(\theta)/\delta a(\theta_{\mathrm{P}})$ should be symmetric.

The linear extrapolation of K, Eq. (32), violates this symmetry. Linearizing G, Eq. (34), fulfills the symmetry only for translational invariance in the crack propagation direction, $\partial G_{\mathrm{JKR}}(\tilde{a}_0)/\partial \tilde{a}_0 = 0$, i.e. in the case of a semi-infinite crack (Leblond et al., 2012).

4.7. Numerical implementation of the crack-front model

We describe the crack front with N collocation points (contact radii a_j) at equally spaced angles θ_j. Accordingly, the Fourier series is truncated for $|n| > N/2$ and $\tilde{a}_n$ is now the discrete Fourier transform of a_j:

$$\tilde{a}_n = \frac{1}{N} \sum_{j=0}^{N} a_j e^{-in\theta_j}. \tag{36}$$

For the crack-front model CF-E, we require that the equilibrium equation Eq. (29) is satisfied in each θ_j,

$$\frac{\partial U_{\mathrm{el}}}{\partial a_j} - \frac{2\pi}{N} a_j w(\theta_j, a_j) = 0, \tag{37}$$

with the discretized gradient of the elastic energy

$$\frac{\partial U_{\mathrm{el}}}{\partial a_j} = \frac{1}{N} \frac{\partial U_{\mathrm{JKR}}(a_j)}{\partial a_j} + \frac{2\pi}{N} G_{\mathrm{JKR}}(\tilde{a}_0) \sum_{n=-N/2}^{N/2-1} |n| \tilde{a}_n e^{in\theta_j} + \frac{\pi}{N} \frac{\partial G_{\mathrm{JKR}}(\tilde{a}_0)}{\partial \tilde{a}_0} \sum_{n=-N/2}^{N/2-1} |n| |\tilde{a}_n|^2. \tag{38}$$

We introduced the factor $2\pi/N$ between Eqs. (29) and (37) so that the left hand side of Eq. (37) is the gradient of the discretized total energy, with

$$U_{\mathrm{el}} = \frac{1}{N} \sum_{i=0}^{N-1} U_{\mathrm{JKR}}(a_i) + \pi G_{\mathrm{JKR}}(\tilde{a}_0) \sum_{n=-N/2}^{N/2-1} |n| |\tilde{a}_n|^2, \tag{39}$$

and

$$\Pi_{\mathrm{surf}} = \frac{2\pi}{N} \sum_{i=0}^{N-1} \int_0^{a_i} \mathrm{d}r \, r w(\theta_i, r). \tag{40}$$

For CF-G and CF-K we proceed similarly and require that Eqs. (13) and (33), respectively, are satisfied for each θ_j.

We are interested in stable equilibria, therefore it is important to use a minimization algorithm and not a root finder to solve for the equilibrium condition. When we simulate the indentation and retraction process, we use the solution obtained at the previous penetration as initial guess. This procedure mimics the quasi-static dynamics of the contact, where the crack front moves to the closest metastable state.

When the crack front moves to the next metastable state through an instability, we need a robust minimization algorithm. We use a trust-region Newton conjugate-gradient algorithm (Steihaug, 1983; Nocedal and Wright, 2006, Algorithm 7.2), where we *fix* the radius of the trust region. Since the work of adhesion field is the source of nonlinearity, the trust radius needs to be slightly smaller than the size of the heterogeneity. If the minimum contact radius is smaller than this value during a Newton step, the trust region is further reduced below the contact radius in order to avoid negative contact radii. Note that trust-region algorithms usually tune the radius according to the discrepancy between the quadratic subproblem and the actual potential. We omitted that feature because, as discussed above, some crack-front models do not possess an elastic energy. An alternative to fixing the trust radius would be to tune it according to gradients instead of energies. We observed that the minimization algorithm performed well even though CF-G and CF-K do not satisfy the symmetries associated with the existence of a potential.

5. Validation and comparison of the perturbation methods

We investigate which first-order crack-perturbation method — linear extrapolation of K (CF-K, Eq. (32)), linear extrapolation of G (CF-G, Eq. (34)) or quadratic extrapolation of the energy (CF-E, Eqs. (26)–(28)) — describes most accurately the contact area and the normal force. We use BEM simulations with $\ell^*_{\mathrm{coz}} \simeq 0.0007$ and pixel size $\ell^*_{\mathrm{pix}} = 0.00015625$ as a reference, the grid size is 32768×32768 without the padding region. As a test case, we consider the ray-shaped work of adhesion heterogeneity (see inset to Fig. 3a)

$$w(r,\theta) = w_m + \Delta w \cos(n_{\mathrm{rays}}\ \theta), \tag{41}$$

where we vary n_{rays} and Δw. We fix the value of the penetration to $\Delta^* = 1$.

Fig. 3a,b compares the contact geometry obtained from BEM and the different crack-front models for $\Delta w = 0.4$ and 4 rays (Fig. 3a) and 64 rays (Fig. 3b). To estimate the remaining deviations of BEM from the short-ranged limit, we also show a BEM simulation with a four times larger cohesive zone ($\ell^*_{\mathrm{coz}} \simeq 0.0028$) and pixel size ($\ell^*_{\mathrm{pix}} = 0.000625$, 8192×8192 pixels). For decreasing ℓ^*_{coz}, the contact radius increases and the normal force decreases and when the work of adhesion is uniform, the normal force converges towards the JKR solution (Fig. 3c). This convergence behavior is consistent with reports in Maugis (1992), Greenwood (1997) and Müser (2014).

For $n_{\mathrm{rays}} = 4$ (Fig. 3a), the contact shapes of all crack-front models are similarly close to the BEM results. While CF-K fits better in the peak and the valley of the undulation, CF-G and CF-E are closer to BEM between these two locations. For $n_{\mathrm{rays}} = 64$ (Fig. 3b), the crack-front models differ more clearly but the errors due to the finite interaction range in BEM are also relatively large. However, as we mentioned in the previous paragraph, the short-ranged limit has slightly higher contact radii than our reference simulation. Therefore we conclude that for large n_{rays}, the contact radius is significantly underestimated by CF-K.

The crack-front models unambiguously differ in the prediction of the normal force. The CF-E model best captures how the normal force depends on the heterogeneity wavelength (Fig. 3c) and amplitude (Fig. 3d). The deviations of CF-G from CF-E are small and disappear for small heterogeneities. CL-K has the largest discrepancies and predicts less adhesive normal forces. However, all errors are relatively small because the leading order contribution of Δw to the normal forces is of second order in BEM and in all the crack-front models. This scaling has been pointed out by Argatov (2021) and is visible in Fig. 3d, where we represent the normal force as a function of Δw^2.

We now discuss qualitatively how the crack shapes depend on Δw (Fig. 3e). In all crack-front models, the amplitude of the contact radius undulation is smaller than in BEM. This is consistent with the numerical simulations of Fares (1989) and the second order perturbation by Leblond et al. (2012). As in Fares' simulations, the relation between contact radius and work of adhesion (i.e. stress intensity factor) is more nonlinear for the minimum contact radius than for the maximum contact radius.

Leblond et al. (2012) pointed out that for a semi-infinite crack ($n_{\mathrm{rays}} \to \infty$), the mean contact radius depends only on the mean work of adhesion along the crack front. Leblond's statement follows from the existence of an elastic potential and hence is satisfied by construction in CF-E. In CF-G (Eq. (34)), the mean contact radius $\tilde{a}_0$ is independent of Δw at any n_{rays}; in fact, CF-E and CF-G are identical in the limit $n_{\mathrm{rays}} \to \infty$. CF-G conserves energy only in that limit. In contrast, in the CF-K model (Eq. (32)), $\tilde{a}_0$ decreases with increasing Δw. The reason is that $\tilde{a}_0$ depends only on the average fracture toughness: taking the average of Eq. (32) yields $\langle K_{\mathrm{c}} \rangle = K_{\mathrm{JKR}}(\tilde{a}_0)$. At fixed w_m, $\langle K_{\mathrm{c}} \rangle \propto \left\langle \sqrt{w} \right\rangle$ decreases with increasing heterogeneity. This contact radius offset is independent of n_{rays} and becomes large compared to the undulation when n_{rays} increases (Fig. 3b). Therefore, by predicting too small contact radii, CF-K violates energy conservation. In summary, for large n_{rays}, the contact area is more accurate in CF-G and CF-E than in CF-K. Because CF-K underestimates the contact area, it also predicts less adhesive normal forces.

The main benefit of CF-E is that it allows us to define rigorously the normal force by Eq. (30). Inserting the Taylor expansion of F_{JKR} into Eq. (31), only terms of second order in the amplitude of the contact radius remain.[1] $\partial^2 F_{\mathrm{JKR}}/\partial a^2$ and $\partial G_{\mathrm{JKR}}/\partial \Delta$ are both negative, so that small-scale fluctuations of the contact radius lead to a more negative (more adhesive) force. Hence, fluctuations of the work of adhesion along the contact line slightly increase adhesion, which is in agreement with the BEM results in Fig. 3d. However, this effect vanishes for small heterogeneities (Fig. 3c): for large n, $\tilde{a}_n \propto \Delta w/n$, so that $F(a) - F_{\mathrm{JKR}}(\tilde{a}_0) \propto \sum_n |n||\tilde{a}_n|^2 \to 0$.

To conclude, while all the crack-perturbation methods are valid first-order approximations, the ansatz possessing an elastic energy (CF-E) yields the best overall results. At large n_{rays} (small heterogeneities), the CF-G model and the CF-E model are equivalent. The CF-G equation is simpler and can be used in analytical theories for crack pinning (Démery and Ponson, 2014). For an amplitude of the heterogeneity of $\Delta w \leq 0.4 w_m$, the errors due to the finite interaction range in expensive BEM simulations are comparable to the errors of the first-order perturbation models.

[1] The first term in the Taylor expansion of F_{JKR} around $a(\theta) = \tilde{a}_0$ disappears by averaging over the perimeter.

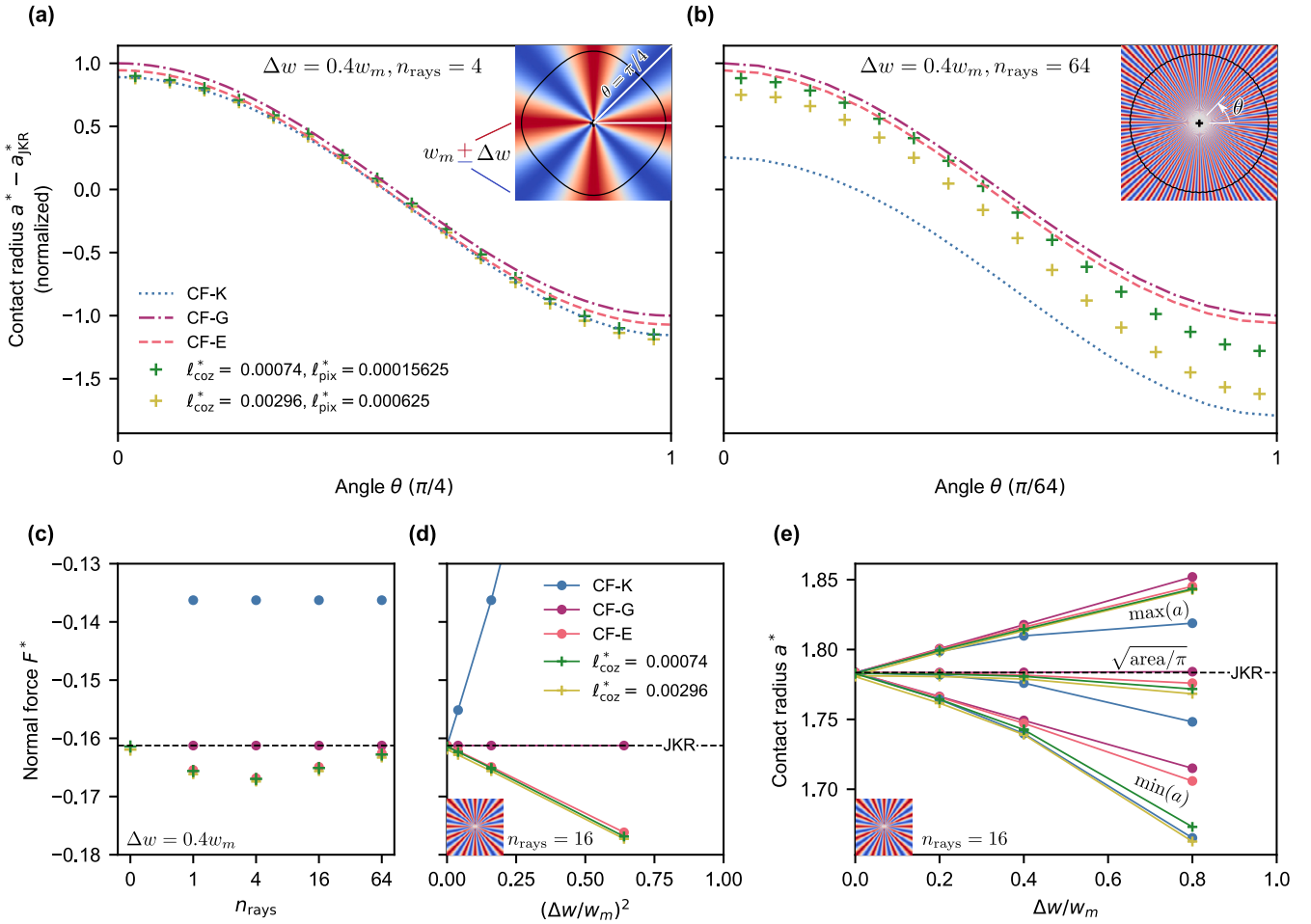


Fig. 3. Comparison of the different crack-front models with BEM simulations for ray-shaped work of adhesion heterogeneity. (a,b) Contact radii computed using BEM and the different CF models on the cosinusoidal work of adhesion landscapes illustrated in the insets. We show the contact radii $a(\theta)$ only over a half period of heterogeneity. We used the radii predicted by CF-G (Eq. (34)) to normalize the contact radius. Since the amplitude of the contact radius is roughly proportional to the wavelength of the heterogeneity, the magnification is higher for $n_{\text{rays}} = 64$ and the effect of the cohesive zone size in the BEM simulations is more apparent. (c) Normal force as a function of the number of rays. $n_{\text{rays}} = 0$ corresponds to a uniform work of adhesion, where it can be verified that BEM is close to the short-range limit (JKR). F^* is the normal force normalized by $\pi w_m R$, as discussed in Section 2. (d) Normal force as a function of the (squared) amplitude of the work of adhesion heterogeneity, for $n_{\text{rays}} = 16$. In the crack-front models CF-K and CF-G, the normal force is computed by inserting the mean contact radius into the JKR equation (Eq. (9)). In CF-E, the normal force is computed by taking the derivative of the elastic energy (Eq. (31)). (e) Contact area, minimum contact radius and maximum contact radius in BEM simulations and different crack-front models as a function of the amplitude of the work of adhesion heterogeneity.

6. Application to a random heterogeneity

To illustrate the usefulness of the crack-front model to study the adhesion hysteresis, we simulated the indentation and retraction of a sphere with randomly fluctuating work of adhesion. We compare the results of BEM and the CF-E model in Fig. 4. The crack-front simulation yields almost the same result as BEM, but requires only one minute instead of one day of computational time. (Both simulations were performed on a single core.)

The BEM simulation was discretized on a 1024 × 1024 grid with pixel size $\ell^*_{\text{pix}} = 0.005$. On the same grid, we generated a random Gaussian toughness field with mean $\langle K_c \rangle = \sqrt{2E' w_m}$ and standard deviation $K_{c,\text{rms}} = 0.2 \langle K_c \rangle$ (Fig. 4b). We chose a Gaussian fracture toughness field instead of a work of adhesion field because it is easier to avoid negative work of adhesion values. Because the work of adhesion is the square of a Gaussian field, the mean $\langle w \rangle$ exceeds the median w_m by a factor 1.04. The standard deviation $w_{\text{rms}} \simeq 0.39 w_m$. The Fourier spectrum is flat at wavelengths above the correlation length $\ell^*_{\text{het}} = 0.2$ and 0 below. The interaction range corresponds to a cohesive zone size $\ell^*_{\text{coz}} \simeq 0.012$.

We performed a crack-front simulation on the same work of adhesion field. In order to evaluate the equilibrium condition of the crack front (Eq. (29)), we interpolated the work of adhesion with bicubic splines between the grid points. We used 512 collocation points on the crack front so that the largest spacing between the collocation points (at the maximal penetration) is $\simeq \ell^*_{\text{het}}/10$.

In both simulations, we increased the penetration Δ^* in steps of 0.01 until the maximum penetration $\Delta^*_{\text{max}} = 1$ was reached and then decreased it until pull off. The starting penetration was chosen to be lower than the jump into contact instability.

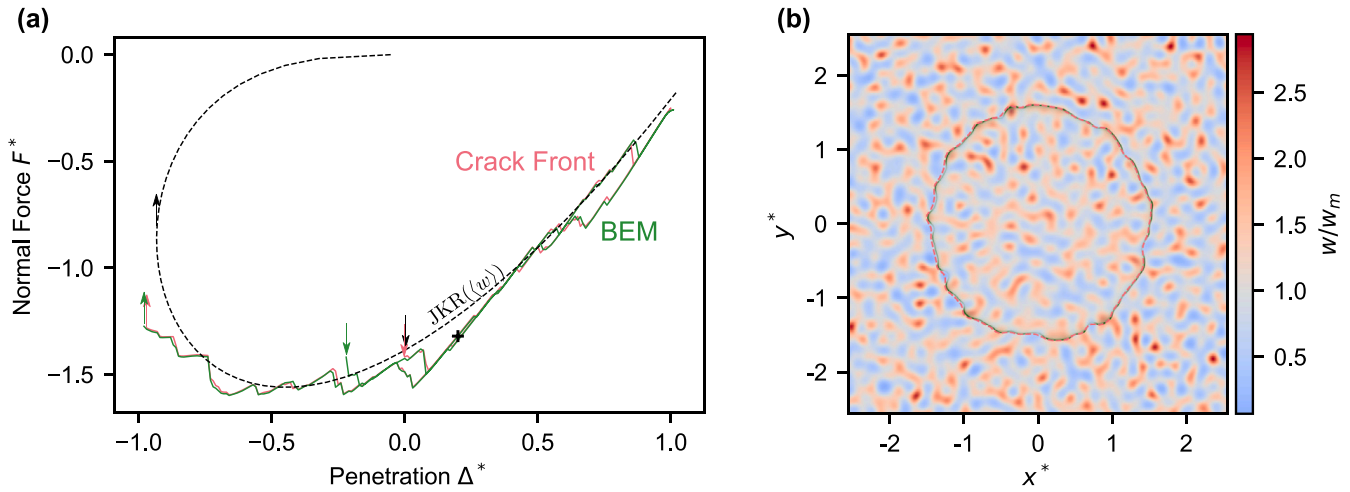


Fig. 4. (a) Force-penetration curves from a boundary-element method (BEM) and a crack-front simulation (CF-E) on the spatially random work of adhesion shown in panel (b). We also show the prediction by the JKR model for a homogeneous surface at the same average work of adhesion. The arrows indicate the jump into contact and the jump out of contact instabilities. (b) Contact areas at the penetration $\Delta^* = 0.2$, indicated by the cross in the force-penetration curve, on top of the work of adhesion field. The tensile pressures of the contact mechanics simulation are shown in green, so that the perimeter of the contact is indicated by the darkest green pixels. The pink dashed line is the contact perimeter calculated with the crack-front model. The work of adhesion field has a standard deviation $\simeq 0.4 w_m$ and a short wavelength cutoff at $\ell^*_{\mathrm{het}} = 0.2$. We used the median work of adhesion $w_m = \langle K_c \rangle^2 / 2E'$ for nondimensionalization.

In Fig. 4a, the force-penetration curves computed with BEM and the crack-front model nearly overlap. For reference, we also show the force-penetration curve resulting from a homogeneous work of adhesion (JKR model) having the same mean value. The homogeneous contact has one hysteresis loop corresponding to the jump into and the jump out of contact instabilities; the force is reversible at positive penetrations. In our simulations, the heterogeneities are energy barriers that pin the crack front and alter the normal force. The kinks in the force curve correspond to depinning instabilities that dissipate energy and lead to additional hysteresis loops (Joanny and de Gennes, 1984). The adhesion hysteresis caused by crack-front pinning will be discussed in more detail in an upcoming publication.

We now discuss the differences between the crack front and BEM results. In BEM, the jump into contact instability occurs too early: it converges much slower with interaction range than other quantities (Wu, 2010; Ciavarella et al., 2017; Wang et al., 2021). The remainder of the force-penetration curve, including depinning instabilities, is well converged. Other discrepancies in the force-penetration curves are due to the linearization in the crack-front model. The contact perimeters agree well, with the most significant deviations in the regions with low work of adhesion. This is consistent with what we observed in the previous section. The two simulations deviate more significantly at a few penetration values, where the instabilities occur at slightly different penetrations. An animation comparing the contact shape of BEM and the contact line during the whole indentation retraction process is provided in the supplementary material.

As mentioned at the beginning of this section, the crack-front simulations are computationally much cheaper than the BEM simulations. Furthermore, in BEM and in the crack-front model, the number of pixels in the linear dimension n scales with ℓ^{-1}_{het} (the total number of pixels in BEM is then $\sim n^2$). While in BEM, the computation time of the elastic deformations increases as $n^2 \log n$ (two dimensional fast Fourier transform), in the crack-front model it only increases as $n \log n$ (one dimensional fast Fourier transform). The crack-front model will enable us to simulate smaller heterogeneities and softer spheres than possible with traditional BEM approaches.

7. Conclusion

We described the adhesion of a sphere against a chemically heterogeneous surface by first-order crack perturbation and validated this model against a boundary-element method (BEM). We compared different variants of the first-order perturbation and found that the best approach is to approximate the energy quadratically. Linearizing the energy release rate is equivalent in the limit of small heterogeneity size, but linearizing the stress intensity factor underestimates adhesion in that limit.

By its efficiency, the crack-front model allows us to simulate orders of magnitude larger systems than possible with BEMs. This simplified model requires, however, that the work of adhesion heterogeneity is sufficiently small for the contact area to be nearly circular. The crack-front model can be adapted to describe the effect of surface roughness in the limit of small roughness amplitudes and soft materials. For contacts meeting these conditions, the crack-front model will allow to better understand the role of surface roughness in adhesion hysteresis, and to make theoretical predictions based on previous work on pinning of elastic lines by a random field (Démery and Ponson, 2014).

CRediT authorship contribution statement

Antoine Sanner: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Investigation, Data curation, Visualization. **Lars Pastewka:** Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank W. Beck Andrews, Ali Dhinojwala, Patrick Dondl, Andreas Greiner and Tevis D. B. Jacobs for useful discussion. We thank Sindhu Singh for implementing the constrained conjugate-gradients algorithm and Laura Mahoney for writing assistance. We are indebted to Mark O. Robbins for pointing us to the analogy to an elastic line. Funding was provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2193/1 – 390951807 and by the European Research Council (StG-757343). Numerical simulations were performed on bwForCluster NEMO (University of Freiburg, DFG grant INST 39/963-1 FUGG) and data was stored on bwSFS (University of Freiburg, DFG grant INST 39/1099-1 FUGG).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmps.2022.104781.

References

Amaral, L.A.N., Barabasi, A.L., Makse, H.A., Stanley, H.E., 1995. Scaling properties of driven interfaces in disordered media. Phys. Rev. E 52 (4087), http://dx.doi.org/10.1103/PhysRevE.52.4087.

Argatov, I.I., 2021. Controlling the adhesive pull-off force via the change of contact geometry. Philos. Trans. R. Soc. London, Ser. A 379, 20200392. http://dx.doi.org/10.1098/rsta.2020.0392.

Autumn, K., Liang, Y.A., Hsieh, S.T., Zesch, W., Chan, W.P., Kenny, T.W., Fearing, R., Full, R.J., 2000. Adhesive force of a single gecko foot-hair. Nature 405, 681–685. http://dx.doi.org/10.1038/35015073.

Barthel, E., 2008. Adhesive elastic contacts: JKR and more. J. Phys. D: Appl. Phys. 41, 163001. http://dx.doi.org/10.1088/0022-3727/41/16/163001.

Batrouni, G.G., Hansen, A., Schmittbuhl, J., 2002. Heterogeneous interfacial failure between two elastic blocks. Phys. Rev. E 65, 036126. http://dx.doi.org/10.1103/PhysRevE.65.036126.

Bazrafshan, M., de Rooij, M.B., Valefi, M., Schipper, D.J., 2017. Numerical method for the adhesive normal contact analysis based on a Dugdale approximation. Tribol. Int. 112, 117–128. http://dx.doi.org/10.1016/j.triboint.2017.04.001.

Bonamy, D., Santucci, S., Ponson, L., 2008. Crackling dynamics in material failure as the signature of a self-organized dynamic phase transition. Phys. Rev. Lett. 101, 045501. http://dx.doi.org/10.1103/PhysRevLett.101.045501.

Borodachev, N.M., 1991. Contact problem for an elastic half-space with a near-circular contact area. Sov. Appl. Mech. 27, 118–123. http://dx.doi.org/10.1007/BF00887799.

Briggs, G.A.D., Briscoe, B.J., 1977. The effect of surface topography on the adhesion of elastic solids. J. Phys. D: Appl. Phys. 10, 2453–2466. http://dx.doi.org/10.1088/0022-3727/10/18/010.

Bugnicourt, R., Sainsot, P., Dureisseix, D., Gauthier, C., Lubrecht, A.A., 2018. FFT-based methods for solving a rough adhesive contact: Description and convergence study. Tribol. Lett. 66, 29. http://dx.doi.org/10.1007/s11249-017-0980-z.

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16, 1190–1208. http://dx.doi.org/10.1137/0916069.

Campañá, C., Müser, M.H., 2006. Practical Green's function approach to the simulation of elastic semi-infinite solids. Phys. Rev. B 74, 075420. http://dx.doi.org/10.1103/PhysRevB.74.075420.

Carbone, G., Pierro, E., Recchia, G., 2015. Loading-unloading hysteresis loop of randomly rough adhesive contacts. Phys. Rev. E 92, 062404. http://dx.doi.org/10.1103/PhysRevE.92.062404.

Carlson, A., Bowen, A.M., Huang, Y., Nuzzo, R.G., Rogers, J.A., 2012. Transfer printing techniques for materials assembly and micro/nanodevice fabrication. Adv. Mater. 24, 5284–5318. http://dx.doi.org/10.1002/adma.201201386.

Chaudhury, M.K., Whitesides, G.M., 1991. Direct measurement of interfacial interactions between semispherical lenses and flat sheets of poly(dimethylsiloxane) and their chemical derivatives. Langmuir 7, 1013–1025. http://dx.doi.org/10.1021/la00053a033.

Chen, Y.L., Helm, C.A., Israelachvili, J.N., 1991. Molecular mechanisms associated with adhesion and contact angle hysteresis of monolayer surfaces. J. Phys. Chem. 95, 10736–10747. http://dx.doi.org/10.1021/j100179a041.

Chen, B., Shi, X., Gao, H., 2008. Apparent fracture/adhesion energy of interfaces with periodic cohesive interactions. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 464, 657–671. http://dx.doi.org/10.1098/rspa.2007.0240.

Chopin, J., Boudaoud, A., Adda-Bedia, M., 2015. Morphology and dynamics of a crack front propagating in a model disordered material. J. Mech. Phys. Solids 74, 38–48. http://dx.doi.org/10.1016/j.jmps.2014.10.001.

Ciavarella, M., Greenwood, J.A., Barber, J.R., 2017. Effect of Tabor parameter on hysteresis losses during adhesive contact. J. Mech. Phys. Solids 98, 236–244. http://dx.doi.org/10.1016/j.jmps.2016.10.005.

Creton, C., Leibler, L., 1996. How does tack depend on time of contact and contact pressure? J. Polym. Sci. B 34, 545–554. http://dx.doi.org/10.1002/(SICI)1099-0488(199602)34:3<545::AID-POLB13>3.0.CO;2-I.

Dalvi, S., Gujrati, A., Khanal, S.R., Pastewka, L., Dhinojwala, A., Jacobs, T.D.B., 2019. Linking energy loss in soft adhesion to surface roughness. Proc. Natl. Acad. Sci. U.S.A 116, 25484–25490. http://dx.doi.org/10.1073/pnas.1913126116.

Dapp, W.B., Müser, M.H., 2015. Contact mechanics of and Reynolds flow through saddle points: On the coalescence of contact patches and the leakage rate through near-critical constrictions. Europhys. Lett. 109, 44001. http://dx.doi.org/10.1209/0295-5075/109/44001.
Démery, A., Ponson, L., 2014. From microstructural features to effective toughness in disordered brittle solids. Europhys. Lett. 105, 34003. http://dx.doi.org/10.1209/0295-5075/105/34003.
Deneke, N., L. Chau, A., S. Davis, C., 2021. Pressure tunable adhesion of rough elastomers. Soft Matter 17, 863–869. http://dx.doi.org/10.1039/D0SM01754J.
Deng, W., Kesari, H., 2017. Molecular statics study of depth-dependent hysteresis in nano-scale adhesive elastic contacts. Modell. Simul. Mater. Sci. Eng. 25, 055002. http://dx.doi.org/10.1088/1361-651X/aa6ef8.
Deng, W., Kesari, H., 2019. Depth-dependent hysteresis in adhesive elastic contacts at large surface roughness. Sci. Rep. 9, 1639. http://dx.doi.org/10.1038/s41598-018-38212-z.
Fares, N., 1989. Crack fronts trapped by arrays of obstacles: Numerical solutions based on surface integral representation. J. Appl. Mech. 56, 837–843. http://dx.doi.org/10.1115/1.3176179.
Feng, J.Q., 2000. Contact behavior of spherical elastic particles: A computational study of particle adhesion and deformations. Colloids Surf. A 172, 175–198. http://dx.doi.org/10.1016/S0927-7757(00)00580-X.
Finnis, M., 2010. Interatomic Forces in Condensed Matter. Oxford University Press, New York, http://dx.doi.org/10.1093/acprof:oso/9780198509776.001.0001.
Fisher, D.S., 1983. Threshold behavior of charge-density waves pinned by impurities. Phys. Rev. Lett. 50, 1486–1489. http://dx.doi.org/10.1103/PhysRevLett.50.1486.
Fuller, K.N.G., Tabor, D., 1975. The effect of surface roughness on the adhesion of elastic solids. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 345, 327–342. http://dx.doi.org/10.1098/rspa.1975.0138.
Galin, L.A., 2008. Contact problems. In: Gladwell, G.M.L. (Ed.), Volume 155 of Solid Mechanics and Its Applications. Springer Netherlands, Dordrecht, http://dx.doi.org/10.1007/978-1-4020-9043-1.
Gao, H., Rice, J.R., 1987a. Nearly circular connections of elastic half spaces. J. Appl. Mech. 54, 627–634. http://dx.doi.org/10.1115/1.3173080.
Gao, H., Rice, J.R., 1987b. Somewhat circular tensile cracks. Int. J. Fract. 33, 155–174. http://dx.doi.org/10.1007/BF00013168.
Gao, H., Rice, J.R., 1989. A first-order perturbation analysis of crack trapping by arrays of obstacles. J. Appl. Mech. 56, 828–836. http://dx.doi.org/10.1115/1.3176178.
Ghanbarzadeh, A., Faraji, M., Neville, A., 2020. Deterministic normal contact of rough surfaces with adhesion using a surface integral method. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 476, 20200281. http://dx.doi.org/10.1098/rspa.2020.0281.
Giaquinta, M., Hildebrandt, S., 2004. Calculus of Variations I. Springer, Berlin; New York, http://dx.doi.org/10.1007/978-3-662-03278-7.
Gjerden, K.S., Stormo, A., Hansen, A., 2013. Universality classes in constrained crack growth. Phys. Rev. Lett. 111, 135502. http://dx.doi.org/10.1103/PhysRevLett.111.135502.
Gjerden, K.S., Stormo, A., Hansen, A., 2014. Local dynamics of a randomly pinned crack front: A numerical study. Front. Phys. 2 (66), http://dx.doi.org/10.3389/fphy.2014.00066.
Greenwood, J.A., 1997. Adhesion of elastic spheres. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 453, 1277–1297. http://dx.doi.org/10.1098/rspa.1997.0070.
Greenwood, J.A., 2017. Reflections on and extensions of the Fuller and Tabor theory of rough surface adhesion. Tribol. Lett. 65, 159. http://dx.doi.org/10.1007/s11249-017-0938-1.
Greenwood, J.A., Williamson, J.B.P., 1966. Contact of nominally flat surfaces. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 295, 300–319. http://dx.doi.org/10.1098/rspa.1966.0242.
Griffith, A.A., Taylor, G.I., 1921. VI. the phenomena of rupture and flow in solids. Philos. Trans. R. Soc. London, Ser. A 221, 163–198. http://dx.doi.org/10.1098/rsta.1921.0006.
Guduru, P.R., 2007. Detachment of a rigid solid from an elastic wavy surface: Theory. J. Mech. Phys. Solids 55, 445–472. http://dx.doi.org/10.1016/j.jmps.2006.09.004.
Guduru, P.R., Bull, C., 2007. Detachment of a rigid solid from an elastic wavy surface: Experiments. J. Mech. Phys. Solids 55, 473–488. http://dx.doi.org/10.1016/j.jmps.2006.09.007.
Hensel, R., Moh, K., Arzt, E., 2018. Engineering micropatterned dry adhesives: from contact theory to handling applications. Adv. Funct. Mater. 28, 1800865. http://dx.doi.org/10.1002/adfm.201800865.
Hertz, H., 1881. Über die Berührung fester elastischer Körper. J. Reine Angew. Math. 92, 156–171. http://dx.doi.org/10.1515/crll.1882.92.156.
Hockney, R.W., 1970. The potential calculation and some applications. In: Alder, B.A., Fernbach, S., Rotenberg, M. (Eds.), Methods in Computational Physics, Vol. 9. Academic Press, New York, pp. 135–211.
Hulikal, S., Bhattacharya, K., Lapusta, N., 2017. The relation between a microscopic threshold-force model and macroscopic models of adhesion. Acta Mech. Sin. 33, 508–515. http://dx.doi.org/10.1007/s10409-016-0630-y.
Hwang, D.G., Trent, K., Bartlett, M.D., 2018. Kirigami-inspired structures for smart adhesion. ACS Appl. Mater. Interfaces 10, 6747–6754. http://dx.doi.org/10.1021/acsami.7b18594.
Irwin, G.R., 1957. Analysis of stresses and strains near the end of a crack transversing a plate. J. Appl. Mech. 24, 361–364. http://dx.doi.org/10.1115/1.4011547.
Joanny, J.F., de Gennes, P.G., 1984. A model for contact angle hysteresis. J. Chem. Phys. 81, 552–562. http://dx.doi.org/10.1063/1.447337.
Johnson, K.L., 1985. Contact Mechanics. Cambridge University Press.
Johnson, K.L., Kendall, K., Roberts, A.D., 1971. Surface energy and the contact of elastic solids. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 324, 301–313. http://dx.doi.org/10.1098/rspa.1971.0141.
Karp, J.M., Langer, R., 2011. Dry solution to a sticky problem. Nature 477, 42–43. http://dx.doi.org/10.1038/477042a.
Kendall, K., 2001. Molecular Adhesion and Its Applications: The Sticky Universe. Springer, Berlin; New York, http://dx.doi.org/10.1007/b100328.
Kesari, H., Doll, J.C., Pruitt, B.L., Cai, W., Lew, A.J., 2010. Role of surface roughness in hysteresis during adhesive elastic contact. Philos. Mag. Lett. 90, 891–902. http://dx.doi.org/10.1080/09500839.2010.521204.
Kesari, H., Lew, A.J., 2011. Effective macroscopic adhesive contact behavior induced by small surface roughness. J. Mech. Phys. Solids 59, 2488–2510. http://dx.doi.org/10.1016/j.jmps.2011.07.009.
Kwak, M.K., Jeong, H.E., Suh, K.Y., 2011. Rational design and enhanced biocompatibility of a dry adhesive medical skin patch. Adv. Mater. 23, 3949–3953. http://dx.doi.org/10.1002/adma.201101694.
Larkin, A.I., Ovchinnikov, Y.N., 1979. Pinning in type II superconductors. J. Low. Temp. Phys. 34, 409–428. http://dx.doi.org/10.1007/BF00117160.
Lazarus, V., 2011. Perturbation approaches of a planar crack in linear elastic fracture mechanics; A review. J. Mech. Phys. Solids 59, 121–144. http://dx.doi.org/10.1016/j.jmps.2010.12.006.
Lebihain, M., Ponson, L., Kondo, D., Leblond, J.B., 2021. Effective toughness of disordered brittle solids: A homogenization framework. J. Mech. Phys. Solids 153, 104463. http://dx.doi.org/10.1016/j.jmps.2021.104463.
Leblond, J.B., Patinet, S., Frelat, J., Lazarus, V., 2012. Second-order coplanar perturbation of a semi-infinite crack in an infinite body. Eng. Fract. Mech. 90, 129–142. http://dx.doi.org/10.1016/j.engfracmech.2012.03.002.
Legrand, L., Patinet, S., Leblond, J.B., Frelat, J., Lazarus, V., Vandembroucq, D., 2011. Coplanar perturbation of a crack lying on the mid-plane of a plate. Int. J. Fract. 170, 67–82. http://dx.doi.org/10.1007/s10704-011-9603-0.
Li, Q., Pohrt, R., Popov, V.L., 2019. Adhesive strength of contacts of rough spheres. Front. Mech. Eng. 5 (7), http://dx.doi.org/10.3389/fmech.2019.00007.

Liu, S., Wang, Q., Liu, G., 2000. A versatile method of discrete convolution and FFT (DC-FFT) for contact analyses. Wear 243, 101–111. http://dx.doi.org/10.1016/S0043-1648(00)00427-0.

Love, A.E.H., 1929. IX. The stress produced in a semi-infinite solid by pressure on part of the boundary. Philos. Trans. R. Soc. Lond. Ser. A 228, 377–420. http://dx.doi.org/10.1098/rsta.1929.0009.

Maugis, D., 1992. Adhesion of spheres: the JKR-DMT transition using a dugdale model. J. Colloid Interface Sci. 150, 243–269. http://dx.doi.org/10.1016/0021-9797(92)90285-T.

Maugis, D., 2010. Contact, Adhesion and Rupture of Elastic Solids. Springer, Berlin; New York, http://dx.doi.org/10.1007/978-3-662-04125-3.

Medina, S., Dini, D., 2014. A numerical model for the deterministic analysis of adhesive rough contacts down to the nano-scale. Int. J. Solids Struct. 51, 2620–2632. http://dx.doi.org/10.1016/j.ijsolstr.2014.03.033.

Middleton, A.A., 1992. Asymptotic uniqueness of the sliding state for charge-density waves. Phys. Rev. Lett. 68, 670–673. http://dx.doi.org/10.1103/PhysRevLett.68.670.

Monti, J., McGuiggan, P.M., Robbins, M.O., 2019. Effect of roughness and elasticity on interactions between charged colloidal spheres. Langmuir 35, 15948–15959. http://dx.doi.org/10.1021/acs.langmuir.9b02161.

Monti, J.M., Sanner, A., Pastewka, L., 2021. Distribution of gaps and adhesive interaction between contacting rough surfaces. Tribol. Lett. 69, 80. http://dx.doi.org/10.1007/s11249-021-01454-6.

Muller, V.M., Yushchenko, V.S., Derjaguin, B.V., 1980. On the influence of molecular forces on the deformation of an elastic sphere and its sticking to a rigid plane. J. Colloid Interface Sci. 77, 91–101. http://dx.doi.org/10.1016/0021-9797(80)90419-1.

Müser, M.H., 2014. Single-asperity contact mechanics with positive and negative work of adhesion: influence of finite-range interactions and a continuum description for the squeeze-out of wetting fluids. Beilstein J. Nanotechnol. 5, 419–437. http://dx.doi.org/10.3762/bjnano.5.50.

Müser, M.H., 2016. A dimensionless measure for adhesion and effects of the range of adhesion in contacts of nominally flat surfaces. Tribol. Int. 100, 41–47. http://dx.doi.org/10.1016/j.triboint.2015.11.010.

Müser, M.H., Dapp, W.B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T.A., Persson, B.N.J., Harris, K., Bennett, A., Schulze, K., Rohde, S., Ifju, P., Gregory Sawyer, W., Angelini, T., Esfahani, H.A., Kadkhodaei, M., Akbarzadeh, S., Wu, J.J., Vorlaufer, G., Vernes, A., Solhjoo, S., Vakis, A.I., Jackson, R.L., Xu, Y., Streator, J., Rostami, A., Dini, D., Medina, S., Carbone, G., Bottiglione, F., Afferrante, L., Monti, J., Pastewka, L., Robbins, M.O., Greenwood, J.A., 2017. Meeting the contact-mechanics challenge. Tribol. Lett. 65, 118. http://dx.doi.org/10.1007/s11249-017-0900-2.

Nocedal, J., Wright, S.J., 2006. Numerical optimization, second ed. In: Springer Series in Operations Research, Springer, New York.

Pastewka, L., Robbins, M.O., 2014. Contact between rough surfaces and a criterion for macroscopic adhesion. Proc. Natl. Acad. Sci. U.S.A 111, 3298–3303. http://dx.doi.org/10.1073/pnas.1320846111.

Pastewka, L., Robbins, M.O., 2016. Contact area of rough spheres: Large scale simulations and simple scaling laws. Appl. Phys. Lett. 108, 221601. http://dx.doi.org/10.1063/1.4950802.

Pastewka, L., Sharp, T.A., Robbins, M.O., 2012. Seamless elastic boundaries for atomistic calculations. Phys. Rev. B 86, 075459. http://dx.doi.org/10.1103/PhysRevB.86.075459.

Patinet, S., Alzate, L., Barthel, E., Dalmas, D., Vandembroucq, D., Lazarus, V., 2013. Finite size effects on crack front pinning at heterogeneous planar interfaces: Experimental, finite elements and perturbation approaches. J. Mech. Phys. Solids 61, 311–324. http://dx.doi.org/10.1016/j.jmps.2012.10.012.

Persson, B.N.J., 2002a. Adhesion between an elastic body and a randomly rough hard surface. Eur. Phys. J. E 8, 385–401. http://dx.doi.org/10.1140/epje/i2002-10025-1.

Persson, B.N.J., 2002b. Adhesion between elastic bodies with randomly rough surfaces. Phys. Rev. Lett. 89, 245502. http://dx.doi.org/10.1103/PhysRevLett.89.245502.

Persson, B.N.J., Albohr, O., Tartaglino, U., Volokitin, A.I., Tosatti, E., 2004. On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion. J. Phys.: Condens. Matter 17, R1–R62. http://dx.doi.org/10.1088/0953-8984/17/1/R01.

Persson, B.N.J., Scaraggi, M., 2014. Theory of adhesion: Role of surface roughness. J. Chem. Phys. 141, 124701. http://dx.doi.org/10.1063/1.4895789.

Persson, B.N.J., Tosatti, E., 2001. The effect of surface roughness on the adhesion of elastic solids. J. Chem. Phys. 115, 5597–5610. http://dx.doi.org/10.1063/1.1398300.

Piccolroaz, A., Mishuris, G., Movchan, A., 2007. Evaluation of the Lazarus–Leblond constants in the asymptotic model of the interfacial wavy crack. J. Mech. Phys. Solids 55, 1575–1600. http://dx.doi.org/10.1016/j.jmps.2007.02.001.

Pohrt, R., Popov, V.L., 2015. Adhesive contact simulation of elastic solids using local mesh-dependent detachment criterion in boundary elements method. Facta Univ. 13, 3–10.

Polonsky, I.A., Keer, L.M., 1999. A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear 231, 206–219. http://dx.doi.org/10.1016/S0043-1648(99)00113-1.

Ponson, L., 2016. Statistical aspects in crack growth phenomena: How the fluctuations reveal the failure mechanisms. Int. J. of Fract. 201, 11–27. http://dx.doi.org/10.1007/s10704-016-0117-7.

Ponson, L., Bonamy, D., 2010. Crack propagation in brittle heterogeneous solids: Material disorder and crack dynamics. Int. J. Fract. 162, 21–31. http://dx.doi.org/10.1007/s10704-010-9481-x.

Rey, V., Anciaux, G., Molinari, J.F., 2017. Normal adhesive contact on rough surfaces: Efficient algorithm for FFT-based BEM resolution. Comput. Mech. 60, 69–81. http://dx.doi.org/10.1007/s00466-017-1392-5.

Rice, J.R., 1985a. First-order variation in elastic fields due to variation in location of a planar crack front. J. Appl. Mech. 52, 571–579. http://dx.doi.org/10.1115/1.3169103.

Rice, J.R., 1985b. Three-dimensional elastic crack tip interactions with transformation strains and dislocations. Int. J. Solids Struct. 21, 781–791. http://dx.doi.org/10.1016/0020-7683(85)90081-2.

Rice, J., 1989. Weight function theory for three-dimensional elastic crack analysis. In: Wei, R., Gangloff, R. (Eds.), Fracture Mechanics: Perspectives and Directions (Twentieth Symposium). American Society for Testing and Materials, Philadelphia, USA, pp. 29–57. http://dx.doi.org/10.1520/STP18819S.

Robbins, M.O., Joanny, J.F., 1987. Contact angle hysteresis on random surfaces. Europhys. Lett. 3, 729–735. http://dx.doi.org/10.1209/0295-5075/3/6/013.

Rosso, A., Krauth, W., 2002. Roughness at the depinning threshold for a long-range elastic string. Phys. Rev. E 65, 025101. http://dx.doi.org/10.1103/PhysRevE.65.025101.

Rosso, A., Le Doussal, P., Wiese, K.J., 2007. Numerical calculation of the functional renormalization group fixed-point functions at the depinning transition. Phys. Rev. B 75, 220201. http://dx.doi.org/10.1103/PhysRevB.75.220201.

Salvadori, A., Fantoni, F., 2014. Weight function theory and variational formulations for three-dimensional plane elastic cracks advancing. Int. J. Solids Struct. 51, 1030–1045. http://dx.doi.org/10.1016/j.ijsolstr.2013.11.029.

Schmittbuhl, J., Hansen, A., Batrouni, G.G., 2003. Roughness of interfacial crack fronts: Stress-weighted percolation in the damage zone. Phys. Rev. Lett. 90, 045505. http://dx.doi.org/10.1103/PhysRevLett.90.045505.

Schmittbuhl, J., Roux, S., Vilotte, J.P., K, Jorgen. Måloy., 1995. Interfacial crack pinning: Effect of nonlocal interactions. Phys. Rev. Lett. 74, 1787–1790. http://dx.doi.org/10.1103/PhysRevLett.74.1787.

Sneddon, I.N., 1946. Boussinesq's problem for a flat-ended cylinder. Math. Proc. Cambridge Philos. Soc. 42, 29–39. http://dx.doi.org/10.1017/S0305004100022702.

Stanley, H.M., Kato, T., 1997. An FFT-based method for rough surface contact. J. Tribol. 119, 481–485. http://dx.doi.org/10.1115/1.2833523.
Steihaug, T., 1983. The conjugate gradient method and trust regions in large scale optimization. SIAM J. Numer. Anal. 20, 626–637. http://dx.doi.org/10.1137/0720042.
Stormo, A., Gjerden, K.S., Hansen, A., 2012. Onset of localization in heterogeneous interfacial failure. Phys. Rev. E 86, 025101. http://dx.doi.org/10.1103/PhysRevE.86.025101.
Tabor, D., 1977. Surface forces and surface interactions. J. Colloid Interface Sci. 58, 2–13. http://dx.doi.org/10.1016/0021-9797(77)90366-6.
Tada, H., Paris, P.C., Irwin, G.R., 2000. The Stress Analysis of Cracks Handbook, third ed. ASME Press, New York.
Tanguy, A., Vettorel, T., 2004. From weak to strong pinning I: A finite size study. Eur. Phys. J. B 38, 71–82. http://dx.doi.org/10.1140/epjb/e2004-00101-6.
Thomson, R., Hsieh, C., Rana, V., 1971. Lattice trapping of fracture cracks. J. Appl. Phys. 42, 3154–3160. http://dx.doi.org/10.1063/1.1660699.
Violano, G., Afferrante, L., 2021. Roughness-induced adhesive hysteresis in self-affine fractal surfaces. Lubricants 9 (7), http://dx.doi.org/10.3390/lubricants9010007.
Vollebregt, E.A.H., 2014. The bound-constrained conjugate gradient method for non-negative matrices. J. Optim. Theory Appl. 162, 931–953. http://dx.doi.org/10.1007/s10957-013-0499-x.
Wang, A., Müser, M.H., 2017. Gauging Persson theory on adhesion. Tribol. Lett. 65, 103. http://dx.doi.org/10.1007/s11249-017-0886-9.
Wang, A., Zhou, Y., Müser, M.H., 2021. Modeling adhesive hysteresis. Lubricants 9 (17), http://dx.doi.org/10.3390/lubricants9020017.
Wei, Z., He, M.F., Zhao, Y.P., 2010. The effects of roughness on adhesion hysteresis. J. Adhes. Sci. Technol. 24, 1045–1054. http://dx.doi.org/10.1163/016942409X12584625925222.
Wu, J.J., 2010. The jump-to-contact distance in atomic force microscopy measurement. J. Adhes 86, 1071–1085. http://dx.doi.org/10.1080/00218464.2010.519256.
Xia, S., Ponson, L., Ravichandran, G., Bhattacharya, K., 2012. Toughening and asymmetry in peeling of heterogeneous adhesives. Phys. Rev. Lett. 108, 196101. http://dx.doi.org/10.1103/PhysRevLett.108.196101.
Xia, S., Ponson, L., Ravichandran, G., Bhattacharya, K., 2015. Adhesion of heterogeneous thin films II: Adhesive heterogeneity. J. Mech. Phys. Solids 83, 88–103. http://dx.doi.org/10.1016/j.jmps.2015.06.010.
Yu, Z., Cheng, H., 2018. Tunable adhesion for bio-integrated devices. Micromachines 9 (529), http://dx.doi.org/10.3390/mi9100529.
Zappone, B., Rosenberg, K.J., Israelachvili, J., 2007. Role of nanometer roughness on the adhesion and friction of a rough polymer surface and a molecularly smooth mica surface. Tribol. Lett. 26, 191. http://dx.doi.org/10.1007/s11249-006-9172-y.
Zhou, Y., 2000. Critical Dynamics of Contact Lines. (Ph.D. thesis). Johns Hopkins University, Baltimore, Maryland, USA.
Zhou, Y., Robinson, A., Steiner, U., Federle, W., 2014. Insect adhesion on rough surfaces: Analysis of adhesive contact of smooth and hairy pads on transparent microstructured substrates. J. R. Soc. Interface 11, 20140499. http://dx.doi.org/10.1098/rsif.2014.0499.

Publication IV

Applied Surface Science Advances 7 (2022) 100190

Contents lists available at ScienceDirect

Applied Surface Science Advances

journal homepage: www.sciencedirect.com/journal/applied-surface-science-advances

Scale-dependent roughness parameters for topography analysis

Antoine Sanner [a,b], Wolfram G. Nöhring [a], Luke A. Thimons [c], Tevis D.B. Jacobs [c], Lars Pastewka [*,a,b]

[a] Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, Freiburg 79110, Germany
[b] Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, Freiburg 79110, Germany
[c] Department of Mechanical Engineering and Materials Science, University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, Pennsylvania 15261, USA

ARTICLE INFO

Keywords:
Surface roughness
Autocorrelation function
Spectral analysis
Variable bandwidth method
Tip convolution artifacts

ABSTRACT

The failure of roughness parameters to predict surface properties stems from their inherent scale-dependence; in other words, the measured value depends on how the parameter was measured. Here we take advantage of this scale-dependence to develop a new framework for characterizing rough surfaces: the Scale-Dependent Roughness Parameters (SDRP) analysis, which yields slope, curvature, and higher-order derivatives of surface topography at many scales, even for a single topography measurement. We demonstrate the relationship between SDRP and other common statistical methods for analyzing surfaces: the height-difference autocorrelation function (ACF), variable bandwidth methods (VBMs) and the power spectral density (PSD). We use computer-generated and measured topographies to demonstrate the benefits of SDRP analysis, including: novel metrics for characterizing surfaces across scales, and the detection of measurement artifacts. The SDRP is a generalized framework for scale-dependent analysis of surface topography that yields metrics that are intuitively understandable.

1. Introduction

Surface roughness is primarily characterized in terms of scalar parameters; especially common are the root-mean-square (rms) height and slope, which are the rms deviations from the mean height and mean slope. Some variant of these quantities is computed by all surface topography instruments, and they are often reported to describe surface topography in publications. These quantities are useful for describing the amplitude of spatial fluctuations in height and slope across the measured topography. However, a core issue with these roughness parameters is that all of them explicitly depend on the scale of the measurement [1]: The rms height depends on the lateral size (largest scale) of the measurement; the rms slope depends on the resolution (smallest scale) of the measurement. A direct demonstration of this effect on real-world measurements can be found in Refs. [2,3]. We note that some standardized expressions for obtaining these values, such as Rq from ISO 4287 [4] include high- and low-frequency filtering. These values are still strongly scale-dependent, but now the relevant scale is the size of the filter rather than the size of the measurement.

The scale dependence of these values is typically a signature of the multi-scale nature of surface topography. A simple illustration is given in a classic article by Benoit Mandelbrot on the length of coastlines [5]. Mandelbrot illustrated, that the length L_{coast} of a coastline depends on the length of the yardstick ℓ used to measure it. A smaller yardstick picks up finer details and hence leads to longer coastlines. For (self-affine) fractals [6], the functional relationship $L_{\mathrm{coast}}(\ell)$ is a power-law whose exponent characterizes the fractal dimension of the coastline. In the case of a surface topography measurement, ℓ corresponds to the resolution of the scientific instrument used to measure the topography and the property corresponding to the length of a coastline is the true surface area $S(\ell)$ of the topography. We have in prior work directly demonstrated that $S(\ell)$ (and also the rms slope and curvature) scales with measurement resolution ℓ [2,3,7]. This scaling of the surface area has, for example, direct relevance to adhesion between soft surfaces [7]. Many surfaces do not behave as ideal fractals, but nearly all surfaces exhibit some form of size dependence of the roughness parameters discussed above. This is because processes that shape surfaces, such as fracture [8–10], plasticity [11–14] or erosion [15], all lead to multi-scale, fractal-like topography over a range of length scales.

Here, we suggest a route to generalize these (and other) geometric properties of measured topography to explicitly contain a notion of measurement scale. We define the individual roughness parameter as a

* Corresponding author.
E-mail address: lars.pastewka@imtek.uni-freiburg.de (L. Pastewka).

https://doi.org/10.1016/j.apsadv.2021.100190
Received 30 June 2021; Received in revised form 30 September 2021; Accepted 26 October 2021
Available online 13 November 2021

function of scale ℓ over which it is measured, leading to curves identifying the value of the parameter as a function of ℓ. Unlike above, where ℓ was restricted to the resolution of the instrument or some fixed filter cutoff, we now broaden the concept of this scale ℓ to refer to any size over which a parameter is computed. For a given topography scan, it can range from the pixel size up to the scan size. We term the resulting curves the *scale-dependent roughness parameters (SDRPs)* and outline the relationship to three common characterization techniques: the height-difference autocorrelation function (ACF), the variable bandwidth method (VBM) and the power spectral density (PSD). SDRPs are useful because they are easily interpreted: While it is difficult to attach a geometric meaning to a certain value of the PSD (where even units can be unclear [1]), the slope and curvature both have simple geometric interpretations. Since slope and curvature are also the primary ingredients for modern theories of contact between rough surfaces [16–24], SDRPs are directly connected to functional properties of rough surfaces. Finally, we illustrate below how SDRPs can be used to estimate tip-radius artifacts in contact-based measurements, such as scanning probe microscopy and stylus profilometry.

2. Analysis methodology

2.1. Computing typical roughness parameters in real space

Surface topography is commonly described by a function $h(x, y)$, where x and y are the coordinates in the plane of the surface. This is sometimes called the Monge representation of a surface, which is an approximation as it excludes overhangs (reentrant surfaces). A real measurement does not yield a continuous function but height values

$$h_{kl} = h(x_k, y_l) \tag{1}$$

on a set of discrete points x_k and y_l. Measurements are often taken on equidistant samples where $x_k = k\Delta x$ and $y_l = l\Delta y$, where Δx and Δy are the distance between the sample-points in their respective directions. Furthermore $k \in [0, N_x - 1]$ and $l \in [0, N_y - 1]$ where $N_x \times N_y$ is the total number of sample points.

Topographies are often random such that h_{kl} is a random process and its properties must be described in a statistical manner. Starting with Longuet-Higgins [25,26] and Nayak [27], many authors have discussed this random process model of surface roughness, yet the most commonly used roughness parameters have remained simple.

We will illustrate the following concepts using the one-dimensional case, i.e. for line scans or profiles. In many real scenarios, even areal topographic measurements are interpreted as a series of line scans. This is for example the case in atomic force microscopy (AFM), where a topographic map is stitched together from a series of adjacent line scans. Because of temporal (instrumental) drift, these line scans may not be perfectly aligned and the "scan"-direction is then the preferred direction for statistical evaluation. In the following mathematical development, we will implicitly assume that all values are obtained by averaging over these consecutive scans, but we will not write this average explicitly in the equations that follow. Extension to true two-dimensional topography maps of the ideas presented here is straightforward and briefly discussed in Appendix A.

The most straightforward statistical property is the root-mean-square (rms) height,

$$h_{\mathrm{rms}} = \langle h_k^2 \rangle^{1/2} \equiv \langle h^2(x) \rangle^{1/2}, \tag{2}$$

where the average $\langle \cdot \rangle$ is taken over all indices k. (We will omit the explicit index k in the following equations.) The rms height measures the amplitude of height fluctuations on the topography, where the midline is defined as $h = 0$. In addition to the height fluctuation, we can also quantify the amplitude of slopes,

$$h'_{\mathrm{rms}} = \left\langle \left(\frac{D}{Dx} h(x) \right)^2 \right\rangle^{1/2}, \tag{3}$$

where D/Dx is a discrete derivative in the x-direction.

A common way to compute discrete derivatives on experimental data is to use a *finite-differences* approximation. Finite-differences approximate the height $h(x)$ locally as a polynomial (a Taylor series expansion). The first derivative can then be computed as

$$\frac{\partial}{\partial x} h(x) \approx \frac{D}{Dx} h(x) = \frac{h(x + \Delta x) - h(x)}{\Delta x}. \tag{4}$$

This expression is called the first-order right-differences scheme. We will use the symbol D for the discrete derivatives, and the term "order" here refers to the truncation order, or how fast the error decays with grid spacing Δx: it drops linearly with decreasing Δx in this scheme. Another interpretation is that the truncation order gives the highest exponent of the polynomial used to interpolate between the points x and $x + \Delta x$. The derivative of a linear interpolation is constant between these points and given by Eq. (4).

We can also quantify the amplitude of higher derivatives,

$$h^{(\alpha)}_{\mathrm{rms}} = \left\langle \left(\frac{D^\alpha}{Dx^\alpha} h(x) \right)^2 \right\rangle^{1/2}, \tag{5}$$

where $\alpha = 2$ yields the rms curvature. A discrete formulation of the second derivative is

$$\frac{D^2}{Dx^2} h(x) = \frac{h(x + \Delta x) - 2h(x) + h(x - \Delta x)}{\Delta x^2}. \tag{6}$$

This expression is called the second-order central-differences approximation. Again, this can be interpreted as fitting a second-order polynomial to the three points $x - \Delta x$, x, and $x + \Delta x$, and interpreting the (constant) second derivative of this polynomial as the approximate second derivative of the discrete set of data points. The third derivative is given by

$$\frac{D^3}{Dx^3} h(x) = \frac{h(x + 2\Delta x) - 3h(x + \Delta x) + 3h(x) - h(x - \Delta x)}{\Delta x^3}, \tag{7}$$

which again can be interpreted in terms of fitting a cubic polynomial to (four) collocation points.

We can generally write the discrete derivative as a weighted sum over the collocation points x_k,

$$\frac{D^\alpha}{Dx^\alpha} h(x_k) = \frac{1}{\Delta x^\alpha} \sum_{l=-\infty}^{\infty} c_l^{(\alpha)} h(x_{k+l}). \tag{8}$$

The values $c_l^{(\alpha)}$ are called the *stencil* of the derivative operator and α indicates the order of the derivative. For the above derivatives,

$$c_0^{(1)} = -1, c_1^{(1)} = 1, \tag{9}$$

$$c_0^{(2)} = -2, c_{\pm 1}^{(2)} = 1 \quad \text{and} \tag{10}$$

$$c_0^{(3)} = 3, c_1^{(3)} = -3, c_{-1}^{(3)} = -1, c_2^{(3)} = 1, \tag{11}$$

and all other $c_l^{(\alpha)}$s are zero. Higher-order derivatives lead to wider stencils.

2.2. Computing scale-dependent roughness parameters in real space

The discrete derivatives of the preceding section are all defined on the smallest possible scale that is given by the sample spacing Δx and have an overall width of $\alpha \Delta x$. It is straightforward to attach an explicit scale to these derivatives, by evaluating Eq. (8) over a sample spacing

$\eta\Delta x$ (with integer η) rather than Δx,

$$\frac{D^{\alpha}_{(\eta)}}{D_{(\eta)}x^{\alpha}}h(x) \equiv \frac{1}{(\eta\Delta x)^{\alpha}}\sum_{l=-\infty}^{\infty} c_l^{(\alpha)} h(x_{k+\eta l}). \tag{12}$$

We will call the factor η the *scale factor*. The corresponding derivative is measured at the *distance scale* $\ell = \alpha\eta\Delta x$.

Fig. 1 a illustrates this concept. For a simple right-differences scheme as given by Eq. (4), the scale-dependent first derivative is simply the slope of the two points at distance ℓ. For the second derivative given by Eq. (6), we fit a quadratic function through three points with overall spacing ℓ and the curvature of this function is the scale-dependent second derivative.

We now define the *scale-dependent roughness parameter (SDRP)* as

$$h^{(\alpha)}_{\mathrm{SDRP}}(\alpha\eta\Delta x) = \left\langle \left(\frac{D^{\alpha}_{(\eta)}}{D_{(\eta)}x^{\alpha}}h(x) \right)^2 \right\rangle^{1/2}_{\mathrm{domain}}. \tag{13}$$

This new function defines a series of descriptors for the surface that are analogous to the traditional rms slope ($h^{(1)}_{\mathrm{SDRP}} \equiv h'_{\mathrm{SDRP}}$) and to the rms curvature ($h^{(2)}_{\mathrm{SDRP}} \equiv h''_{\mathrm{SDRP}}$); but instead of being a single scalar value, each represents a curve as a function of the distance scale $\ell = \alpha\eta\Delta x$.

The distance scale ℓ is only clearly defined for the stencils of lowest truncation order. For the n-th derivative, those can be interpreted as fitting a polynomial of order n to $n+1$ data points (see Fig. 1a). The n-th derivative of this polynomial is then a constant over the width of the stencil; this width must then equal the distance scale ℓ. Higher truncation orders can be interpreted as fitting a polynomial of order $m > n$ to $m+1$ data points. The n-th derivative is not constant over the stencil and it is not clear what the corresponding length scale is. We will here only use stencils of lowest truncation order where the distance scale is clear.

For non-periodic topographies we need to take care to only include derivatives that we can actually compute, i.e. where the stencil remains in the domain of the topography. This is indicated by the subscript "domain" in Eq. (13).

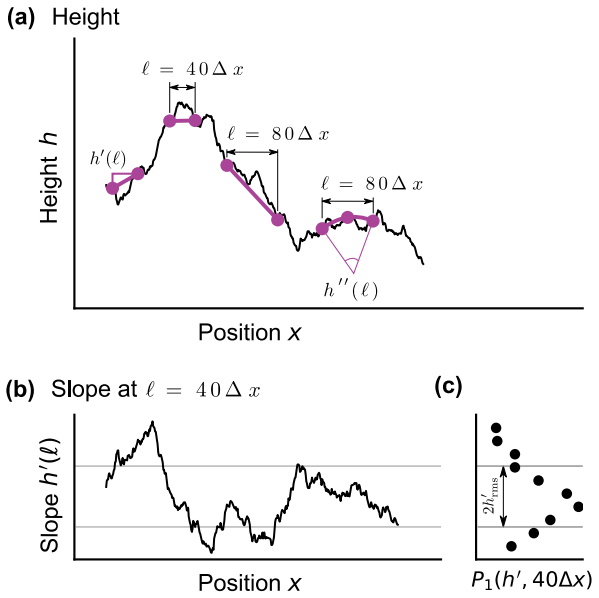


Fig. 1. Illustration of the basic idea behind the scale-dependent roughness parameters. **(a)** Example line scan showing the computation of slopes $h'(\ell)$ and curvatures $h''(\ell)$ from finite differences. A scale can be attached to this computation by computing these finite differences at different distances ℓ, shown for $\ell = 40\Delta x$ and $\ell = 80\Delta x$ where Δx is the sample spacing. Similarly, the curvature at a finite scale ℓ is given by fitting a quadratic function through three points spaced at a distance $\ell/2$. **(b)** Local slope, obtained at a distance scale of $\ell = 40\Delta x$ for the line scan shown in panel (a). The slope is defined for each sample point since we can compute it for overlapping intervals. **(c)** Distribution of the local slope obtained from the slope profile shown in panel (b). The rms slope for this length scale is the width of this distribution.

2.3. Beyond root-mean-square parameters: Computing the full distribution

The rms value, such as the one defined in Eq. (13), characterizes the amplitude of fluctuations, or the width of the underlying distribution function. Rather than looking at this single parameter, we can also determine the full scale-dependent distribution. Formally we can write this distribution as

$$P_{\alpha}(\chi;\eta) = \left\langle \delta\left(\chi - \frac{D^{\alpha}_{(\eta)}}{D_{(\eta)}x^{\alpha}}h(x) \right) \right\rangle \tag{14}$$

where $\delta(x)$ is the Dirac-δ function and χ the value of the derivative (of order α) that we are interested in. In any practical (numerical) determination of the distribution, we broaden the δ-function into individual bins and count the number of occurrences of a certain derivative value.

To illustrate this concept on the example of the slope ($\alpha = 1$), Fig. 1b shows the scale-dependent derivative at $\ell = 40\Delta x$ of the line scan shown in Fig. 1a. The distribution function of the slopes at this scale, $P_1(h', 40\Delta x)$, is then obtained by counting the occurence of a certain slope value. The resulting distribution is shown in Fig. 1c.

The rms parameters defined in the previous section are the square roots of the second moments of this distribution,

$$h^{(\alpha)}_{\mathrm{SDRP}}(\alpha\eta\Delta x) = \left[\int d\chi\, \chi^2 P_{\alpha}(\chi;\eta) \right]^{1/2}. \tag{15}$$

The second moment characterizes the underlying distribution fully only if this distribution is Gaussian. We will see below that, for example, scanning probe artifacts introduce deviations from Gaussianity that we can easily detect once we have the full distribution function.

In summary, these probability distributions of arbitrary derivatives (such as slope, curvature, or higher-order functions) serve as an additional set of descriptors for a surface. The distributions are themselves scale dependent, but can be used to compute a wide variety of scale-dependent statistical parameters, including higher cumulants, such as skewness or kurtosis.

3. Analysis: Relationship of scale-dependent roughness parameters to other methods

3.1. Relationship to the autocorrelation function

A common way of analyzing the statistical properties of surface topography is the *height-difference* autocorrelation function, which will be designated here as ACF or $A(\ell)$. (See Ref. [28] for an authoritative discussion of properties and use.) The ACF is defined as

$$\begin{aligned} A(\ell) &= \frac{1}{2}\left\langle [h(x+\ell) - h(x)]^2 \right\rangle \\ &= \left\langle \frac{1}{2}h^2(x) + \frac{1}{2}h^2(x+\ell) - h(x)h(x+\ell) \right\rangle. \end{aligned} \tag{16}$$

Note that some authors call $2A(\ell)$ the *structure function* and use the term ACF for the bare height autocorrelation function $\langle h(x)h(x+\ell)\rangle$ [29,30]. The height ACF and the height-difference ACF are related by

$$A(\ell) = h^2_{\mathrm{rms}} - \langle h(x)h(x+\ell)\rangle. \tag{17}$$

The ACF has the limiting properties $A(0) = 0$ and $A(\ell \to \infty) = h^2_{\mathrm{rms}}$.

Equation (16) resembles the finite-differences expression for the first derivative, Eq. (4). Indeed, we can rewrite the ACF as

$$A(\eta\Delta x) = \frac{1}{2}\left\langle \left[\frac{D_{(\eta)}}{D_{(\eta)}x}h(x)\right]^2 \right\rangle (\eta\Delta x)^2 \tag{18}$$

using the scale-dependent derivative. The scale-dependent rms slope then becomes

$$h'_{\mathrm{SDRP}}(\ell) = [2A(\ell)]^{1/2}/\ell. \tag{19}$$

The height-difference ACF can hence be used to compute the scale-dependent slope introduced above.

We now show that we can also express higher-order derivatives in terms of the ACF. Using the stencil of the second derivative given in Eq. (6), the scale-dependent second derivative can be written as

$$h''_{\mathrm{SDRP}}(\ell) = \frac{4}{\ell^2}\left\langle [h(x+\ell/2) - 2h(x) + h(x-\ell/2)]^2 \right\rangle^{1/2}. \tag{20}$$

We can rewrite this expression as

$$\begin{aligned} h''_{\mathrm{SDRP}}(\ell) &= \frac{4}{\ell^2}\langle 6h^2(x) - 8h(x)h(x+\ell/2) \\ &\quad +2h(x)h(x+\ell)\rangle^{1/2}. \end{aligned} \tag{21}$$

and use Eq. (17) to introduce the ACF into this expression, yielding

$$h''_{\mathrm{SDRP}}(\ell) = 4[8A(\ell/2) - 2A(\ell)]^{1/2}/\ell^2. \tag{22}$$

Similarly, the scale-dependent third derivative from the stencil given in Eq. (7) becomes

$$h'''_{\mathrm{SDRP}}(\ell) = \frac{27}{\ell^3}[30A(\ell/3) - 12A(2\ell/3) + 2A(\ell)]^{1/2}. \tag{23}$$

We can therefore relate the scale-dependent root-mean-square slope, curvature, or any other higher-order derivative to the ACF.

In summary, we have shown that the commonly used ACF function can be thought of a specific case of the SDRP analysis: being equivalent to the finite-differences calculation of scale-dependent slope. We further showed that the ACF function can be used as one method to compute higher-order SDRPs.

3.2. Relationship to the variable-bandwidth method

We now introduce an alternative way to arrive at SDRPs based on a different notion of scale. Notice that the discussion leading up to Eq. (13) does not involve the length L of the line scan. This length is only relevant when it comes to determining an upper limit for the stencil length $\ell = \alpha\eta\Delta x$, which is the notion of scale in a measurement based on Eq. (13). Alternatively, we could interpret L as the relevant scale, and study scale-dependent roughness by varying L. This interpretation leads to a class of methods which have been referred to as scaled windowed variance methods [31] or variable bandwidth methods (VBMs). Members of this class of methods differ only in the way that the data is detrended and have been given a variety of names including: bridge method (attributed to Mandelbrot); roughness around the mean height (MHR) [32] (sometimes termed VBM [33]); detrended fluctuation analysis (DFA) [34,35]; and roughness around the rms straight line (SLR) [32].

In all cases, one performs multiple roughness measurements on the same specimen (or the same material) but with different scan sizes L. Plotting the rms height h_{rms} from these measurements versus scan size L, or the rms slope h'_{rms} versus scan resolution (the smallest measurable scale) yields insights into the multi-scale nature of surface topography. An example of an experimental realization of this idea is the classic paper by Sayles & Thomas [36]. The present authors have used and discussed this approach in the past to characterize the topography of diamond thin films [2,3].

These methods can be generalized for the analysis of single measurements. Consider a line scan $h(x_k)$ of length L. The scan is partitioned into $\zeta \geq 1$ segments of length $\ell(\zeta) = L/\zeta$ (with $\ell \leq L$ now being the relevant scale). The dimensionless number ζ, which we call the *magnification*, defines the scale. Some authors use sliding windows rather than exclusive segments [13,32].

The VBM considers the rms height fluctuations in each of the segments, i.e. one computes the standard deviation of the height $h_{\mathrm{VBM},i}(\zeta)$ within segment i at magnification ζ, and then takes the average over all i to compute a scale-dependent $h_{\mathrm{VBM}}(\zeta)$. Some authors (including our own prior work [14]) have tilt corrected the individual segments, i.e. each segment is detrended by subtracting the corresponding mean height and slope (obtained by linear regression of the data in the segment) before computing $h_{\mathrm{VBM},i}(\zeta)$; this approach is called the DFA [34,35] while without tilt correction it is called MHR. In the bridge method, the connecting line between the first and last point in each segment is used for detrending (see e.g. Ref. [33]).

These VBMs are extremely similar to the SDRP. When computing the slope in the SDRP, we compute it by simply connecting the two boundary points at $x = i\ell(\zeta)$ and $x = (i+1)\ell(\zeta)$ with a straight line, as is done in the bridge method. This is distinct from DFA, which uses all data points between the two boundary points and fits a straight line using linear regression. Detrending can be generalized to higher-order polynomials, but this has not been reported in the literature. The relationship between SDRP and VBMs with detrending of order 1 and 2 is conceptually illustrated in Fig. 2.

In DFA, the trend line is simply used as a reference for the computation of fluctuations around it. The coefficients of the detrending polynomial can also be used to analyse how the slope and curvature of the surface depend on scale [37,38]. This yields an alternative measure of the scale-dependent rms slope, $h'_{\mathrm{VBM}}(\zeta)$, obtained at magnification ζ or distance scale $\ell = L/\zeta$. $h'_{\mathrm{VBM}}(\zeta)$ is simply the standard deviation of slopes obtained within all segments i at a certain magnification ζ. We show in the examples below that this scale-dependent slope is virtually identical to the slope obtained from the SDRP.

We can use this idea to extend the DFA to higher-order derivatives. Rather that fitting a linear polynomial in each segment, we detrend using a higher-order polynomial. For extracting a scale-dependent rms curvature, we fit a second-order polynomial to the segment and interpret twice the coefficient of the quadratic term as the curvature. The standard deviation of this curvature over the segments then gives the scale-dependent second derivative, $h''_{\mathrm{VBM}}(\zeta)$. Fig. 2 illustrates this concept, again in comparison to the SDRP that for the second-order derivative fits a quadratic function through just three collocation points.

An alternative route of thinking about VBMs is that they use a stencil whose number of coefficients equals the segment length. The stencil can be explicitly constructed from least squares regression (at each scale) of

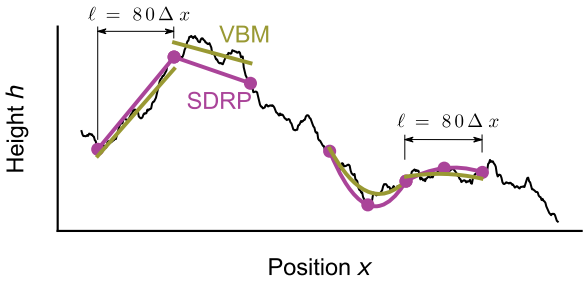


Fig. 2. Illustration of the computation of scale-dependent roughness parameters from the variable bandwidth method (VBM). While in finite differences, the slope is computed between two points at distance ℓ, in the VBM we fit a trend line to a segment of width ℓ. Similarly for the second derivative, the finite-differences estimation fits a quadratic function through three points while in the VBM we fit a quadratic trend line through all data points in an interval of length ℓ.

the polynomial coefficients. The closest equivalent to the SDRP would then be the respective VBM that uses sliding (rather than exclusive) segments. However, even in this case, a remaining difference is that SDRP uses stencils of identical number of coefficients at each scale. In the examples provided in Section 4, we use a VBM that uses nonoverlapping segments.

In summary, we have shown that the various methods for computing scale-dependent height (such as VBM, DFA, and others) can be thought of as a special case of SDRP analysis: where the scale-dependent detrending only occurs for at most linear trend lines. We have then shown how those analyses can be extended to define yet another method for computing SDRPs.

3.3. Relationship to the power spectral density

Finally we outline a fourth way to arrive at SDRPs using the power spectral density (PSD), another common tool for the statistical analysis of topographies [1]. Underlying the PSD is a Fourier spectral analysis, which approximates the topography map as the series expansion

$$h(x) = \sum_n a_n \phi_n(x), \tag{24}$$

where $\phi_n(x)$ are called *basis functions*. The Fourier basis is given by

$$\phi_n(x) = \exp(iq_n x), \tag{25}$$

with $q_n = 2\pi n/L$ where L is the lateral length of the sample. The inverse of Eq. (24) gives the expansion coefficients a_n which are typically computed using a fast Fourier-transform algorithm. The PSD is then obtained as [1]

$$C^{1D}(q_n) = L|a_n|^2. \tag{26}$$

Fourier spectral analysis is useful because a notion of scale is embedded in the definition Eq. (25): The wavevectors q_n describe plane waves with *wavelength* $\lambda_n = 2\pi/q_n$.

This basis leads to spectral analysis of surface topography and derivatives are straightforwardly computed from the derivatives of the basis functions,

$$\frac{\partial}{\partial x}\phi_n(x) = iq_n\phi_n(x) \quad \text{and} \tag{27}$$

$$\frac{\partial^2}{\partial^2 x}\phi_n(x) = -q_n^2\phi_n(x). \tag{28}$$

We can write the Fourier-derivative generally as

$$\frac{\partial^\alpha}{\partial x^\alpha}\phi_n(x) = \mathscr{D}_\alpha(q_n)\phi_n(x) \tag{29}$$

with $\mathscr{D}_1(q_n) = iq_n$ for the first derivative and $\mathscr{D}_2(q_n) = -q_n^2$ for the second derivative. The $\mathscr{D}_\alpha(q_n)$ are complex numbers that we will call the *derivative coefficients*.

The rms amplitude of fluctuations can be obtained in the Fourier picture from Parseval's theorem, that turns the real-space average in Eq. (5) into a sum over wavevectors,

$$h^{(\alpha)}_{\text{rms}} = \left[\sum_n |\mathscr{D}_\alpha(q_n)a_n|^2\right]^{1/2}. \tag{30}$$

The notion of a scale-dependence can be introduced in the Fourier picture by removing the contribution of all wavevectors $|q_n| > q_c$ larger than some characteristic wavevector q_c, i.e. setting the corresponding expansion coefficients a_n to zero. This means there are no longer short wavelength contributions to the topography. We will refer to this process as *Fourier filtering*. Fourier filtering can be used to introduce a scale-dependent roughness parameter, e.g.

$$h^{(\alpha)}_{\text{PSD}}(q_c) = \left[\sum_n |\mathscr{D}^{\text{F}}_\alpha(q_n;q_c)|^2 C^{1D}(q_n)\right]^{1/2} \tag{31}$$

with $\mathscr{D}^{\text{F}}_\alpha(q_n;q_c) = \Theta(q_c - |q_n|)\mathscr{D}_\alpha(q_n)$ that we call the Fourier-filtered derivative and $\Theta(x)$ is the Heaviside step function. Note that we have expressed Eq. (31) in terms of the PSD, which is typically obtained using a windowed topography if the underlying data is nonperiodic. In the examples that we show in Section 4, we applied a Hann window before computing the scale-dependent derivatives from the PSD.

We now show that Fourier-filtering and finite-differences are related concepts. We first interpret the finite-differences scheme in terms of a Fourier analysis. We apply the finite differences operation to the Fourier basis Eq. (25). This yields

$$\frac{D^\alpha_{(\eta)}}{D_{(\eta)}x^\alpha}\phi_n(x_k) = \mathscr{D}^{\text{s}}_\alpha(q_n;\eta)\phi_n(x_k) \tag{32}$$

with

$$\mathscr{D}^{\text{s}}_\alpha(q_n;\eta) = \frac{1}{(\eta\Delta x)^\alpha}\sum_{l=-\infty}^{\infty} c_l^{(\alpha)}\exp(iq_n\eta l\Delta x). \tag{33}$$

Note that the right hand side of Eq. (32) is fully algebraic, i.e. it no longer contains derivative operators, and the $\mathscr{D}^{\text{s}}_\alpha(q_n;\eta)$ are (complex) numbers. Inserting these derivative coefficients into Eq. (31) yields Eq. (13). We have therefore unified the description of (scale-dependent) derivatives in the Fourier basis and finite-differences in terms of the derivative coefficients $\mathscr{D}_\alpha$.

The remaining question is how the scale ℓ used to compute the finite-differences relates to the wavevector q_c used in Fourier-filtering. Fig. 3 shows $\mathscr{D}^{\text{F}}_1(\lambda_c)$ and $D^{\text{s}}_1(\ell)$ for different values of ℓ and λ_c. The location of the maximum of these derivative coefficients agrees if $\eta\Delta x \equiv \ell/\alpha = \lambda_c/2 = \pi/q_c$. For first derivatives ($\alpha = 1$), $\ell = \Delta x$. This is the Nyquist sampling theorem, which states that the shortest wavelength we can resolve is $\lambda_c = 2\Delta x$. This means to compare SDRP, VBM and PSD, we need to choose a filter cutoff of $q_c = \alpha\pi/\ell$ in the latter. Note that in the SDRP, the (soft) cutoff emerges implicitly from the finite-difference formulation.

In summary, we have shown that the SDRPs, which were defined in real-space in Section 2.2, can be equivalently computed in frequency-space using the PSD. The results should be the same, however, frequency-space calculations have the shortcomings that nonperiodic topographies need to be windowed, and a filter cutoff needs to be applied.

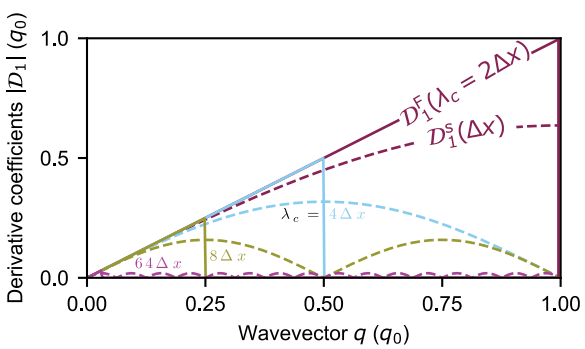


Fig. 3. Derivative coefficients for finite differences $\mathscr{D}^{\text{s}}_1$ and the Fourier-filtered derivative $\mathscr{D}^{\text{F}}_1$ for different distance scales ℓ. The coefficients agree at small wavevectors q. The maximum of the coefficient agrees if the filter wavelength $\lambda_c = 2\ell$, corresponding to the Nyquist sampling theorem (see text).

3.4. Summary notes about comparison to other methods

In this section, we have shown that three of the most common methods for roughness characterization, namely ACF, VBM, and PSD, are highly related to the SDRP. In fact, some form of SDRP can be computed using any of these three methods, instead of using the original definition, Eq. (13), with approximately equivalent results (see numerical experiments in the next section). Intuitively, the SDRP can be thought of as a general framework for analysis, which contains ACF, VBMs and PSD as special cases.

4. Discussion: Application of scale-dependent roughness parameters, and advantages over other methods

4.1. Application to a synthetic self-affine surface

We first apply the concepts presented above to a synthetic self-affine topography. The topography has been first presented in Ref. [1] and consists of three virtual "measurements" of a large (65 536 ×65 536 pixels) self-affine topography generated with a Fourier-filtering algorithm [1,39]. This algorithm works as follows: We superpose sine waves with uncorrelated random phases and amplitudes scaled according to a power-law. On the pixel at position $\vec{x}_{ij} = (x_i, y_j)$, the height can be written as

$$h_{ij} = \sum_{\substack{k,l\geq 0 \\ |\vec{q}_{kl}|<q_s}} A_{kl}\sin\left(\vec{q}_{kl}\cdot\vec{x}_{ij} + \phi_{kl}\right), \tag{34}$$

where $\vec{q}_{kl} = 2\pi/L(k,l)$ is the wavevector and L is the period of the topography. The phases ϕ_{kl} are uncorrelated and uniformly distributed between 0 and 2π. The amplitudes A_{kl} are uncorrelated Gaussian random variables with variance proportional to $|\vec{q}_{kl}|^{-2-2H}$. The sum runs only over wavevectors smaller than the short-wavelength cutoff $q_s = 2\pi/\lambda_s$. The (two-dimensional) PSD of the surface is the square of the amplitudes A_{kl} and is 0 for wavelengths below λ_s. We generated the surface with Hurst exponent $H = 0.8$, cutoff wavelength $\lambda_s = 10$ nm, pixel size $\Delta x = \Delta y = 2$ nm and physical size $L = 131$ μm. This surface was subsampled in three blocks of 500 × 500 pixels at overall lateral sizes of 100 μm × 100 μm, 10 μm × 10 μm and 1 μm × 1 μm to emulate measurement at different resolution. Each of these virtual measurements is nonperiodic and independently tilt-corrected. The data for the three subsampled topographies is available online [40].

Fig. 4a shows the topography map of these three emulated measurements. The measurements zoom subsequently into the center of the topography. The one-dimensional PSDs (C^{1D}, Fig. 4b) of the three topographies align well, showing zero power below the cutoff wavelength of λ_s. Note that unlike most authors (with few exceptions in geophysics [41–45]) or even our own prior work, we display the PSD as a function of wavelength $\lambda = 2\pi/q$ where q is the wavevector; this facilitates comparison with the real-space techniques introduced above, and also wavelengths are more intuitively understandable than wavevectors. Since the topography is self-affine, the PSD scales as $C^{1D} \propto \lambda^{1+2H}$ as indicated by the solid line.

The ACF (or rather its square root) is shown in Fig. 4c. The ACF and all other scale-dependent quantities reported below are obtained from averages over adjacent line scans, i.e. from one-dimensional profiles rather than two-dimensional area scans. This is compatible with how C^{1D} is computed (see Ref. [1]). The ACFs from the three measurements line up and follow $\sqrt{A} \propto \ell^H$ (see solid black line in Fig. 4c). Note that the ACF does not drop to zero for $\ell < \ell_s \equiv \lambda_s/2$ as the PSD did. This behavior becomes clearer by inspecting the scale-dependent slope $h'_{\mathrm{SDRP}}(\ell) = \sqrt{2A(\ell)}/\ell$ that saturates at a constant value for $\ell < \ell_s$. This is the true rms slope that is computed when all scales are considered. For large ℓ, the rms slope scales as $h'_{\mathrm{SDRP}} \propto \ell^{H-1}$ (solid black line in Fig. 4d).

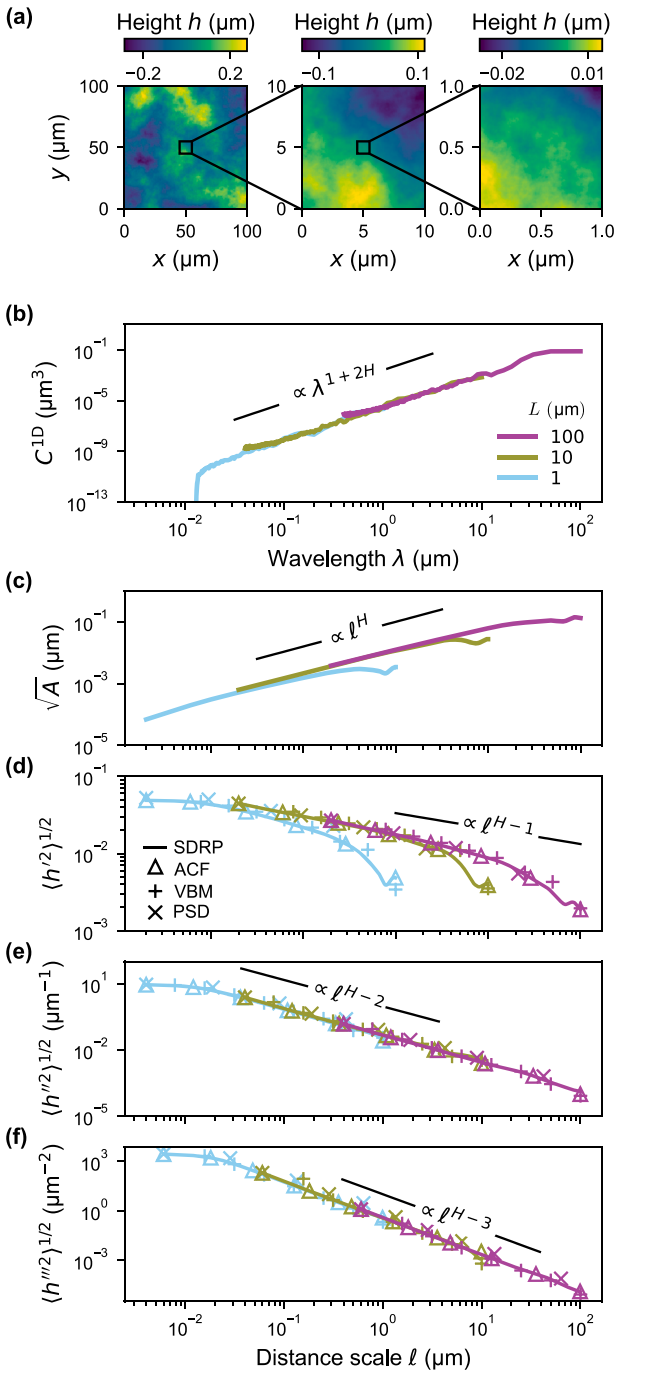


Fig. 4. Example of scale dependent-roughness parameters for an ideal self-affine surface with Hurst exponent $H = 0.8$. **(a)** A large surface was subsampled in three topographies of 500 × 500 pixels at different resolution. **(b)** Individual PSDs displayed as a function of wavelength $\lambda = 2\pi/q$, where q is the wavevector. **(c)** Square root of the ACF displayed as a function of distance scale ℓ. **(d)** Scale-dependent rms slope. **(e)** Scale-dependent rms curvature. **(f)** Third derivative as an example of how this method can be used to go beyond traditional analysis. Color is used to distinguish the three different topographies. The figure shows results from the four SDRPs (finite-differences based SDRP, ACF, VBM and PSD), showing that the results agree. The solid black lines in panels (b-f) shows the power-law scaling of an ideal self-affine topography with Hurst exponent H. Note that the deviations from power-law scaling at large scales are especially visible in panels (c) and (d) because of the smaller range of values on the y-axis.

Finally, we display the scale-dependent curvature $h''_{\mathrm{SDRP}}(\ell)$ in Fig. 4e. Like the rms slope, the curvature saturates for $\ell < \ell_s$ to the "true" small-scale value of the curvature. The curvatures of the three individual measurements again line up and follow $h''_{\mathrm{SDRP}}(\ell) \propto \ell^{H-2}$ because of the self-affine character of the overall surface. The rms curvature computed from the ACF (Eq. (22)) is strictly only applicable to periodic topographies, but in our numerical experiments the ACF agrees with the original definition of the SDRPs (Eq. (13)) within the thickness of the line. The errors occur at large distance scales and can, in principle, lead to negative values of h''^2_{SDRP}, but we have not seen this occurring in the numerical data presented in this paper.

In our derivation above we have presented alternative routes for obtaining scale-dependent roughness parameters from the VBM and PSD. The plusses (+) in Figs. 4d and e show the rms slope and curvature obtained using the VBM, while the crosses (x) show the results obtained using the PSD. They align well with the respective parameters obtained from the SDRP and only deviate at large scales. In summary, all three routes (ACF, VBM, PSD) for obtaining SDRPs are equally valid and lead to results that are consistent with those computed using the original definition (Eq. (13)). The advantage of the SDRP, ACF and the VBM over the PSD is that they are directly (without windowing) applicable to nonperiodic data.

We have now demonstrated four independent ways of obtaining scale-dependent slopes, curvatures and higher-order derivatives. We would like to point out that all four routes constitute novel uses of the underlying analysis methodology. Our primary tool in what follows will be the SDRP; however, we have demonstrated that the ACF, VBM and PSD yield equivalent results. The broader importance of using scale-dependent slopes and curvatures over the "bare" ACF, VBM or PSD is that it is straightforward to interpret the meaning of these parameters. We all have intuitive understanding of the meaning of slopes and curvatures, whereas it is difficult to ascribe a geometric meaning to a value of the PSD (that can even differ in unit, see discussion in Ref. [1]).

4.2. *Detecting tip artifacts in simulated topography measurements*

We now turn to another example, the analysis of tip artifacts. This will exploit a power of the SDRP, namely the fact that we can compute the full underlying distribution of arbitrary derivatives outlined in Section 2.3.

Fig. 5a shows two computer generated nonperiodic topographies of size 0.1 μm × 0.1 μm. The first topography is pristine and was generated using the Fourier-filtering algorithm mentioned above. As in the previous example, we ensured the scan is not periodic by taking a section of a larger (0.5 μm) periodic scan. The second topography contains tip artifacts and was obtained from the pristine surface using the following nonlinear procedure: For every location (x_i, y_i) on the topography we lower a sphere with radius R_{tip} (here 40 nm) towards a position (x_i, y_i, z_i) until the sphere touches the pristine topography anywhere. The resulting z-position z_i of the sphere is then taken as the "measured" height of the topography. This topography was discussed in Ref. [1] and the data files are available at Ref. [46]. The two curves underneath the maps in Fig. 5a are cross-sections through the middle of the respective topography.

It is clear from simply looking at the data in Fig. 5a that the scanning

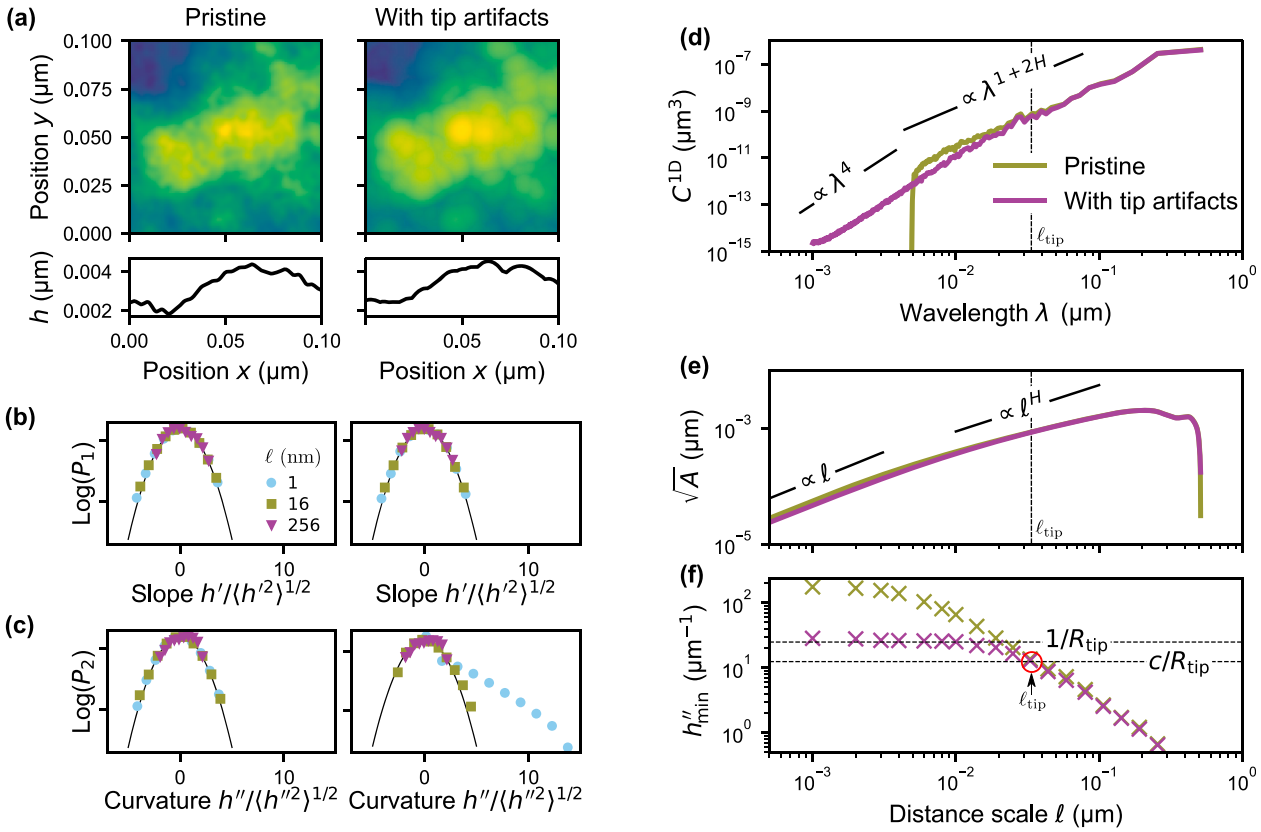


Fig. 5. Scale-dependent roughness parameters for the analysis of tip artifacts. **(a)** A computer-generated "pristine" topography was scanned with a virtual tip of R_{tip} = 40 nm radius. The bottom row shows cross-sectional profiles of the maps shown above. The artifacted map and profile show clear blunting of the peaks and cusps in the valleys (see text for more discussion). **(b)** Distribution of slopes at distance scales ℓ = 1 nm (circles •), 16 nm (squares ■) and 256 nm (triangles ▼). **(c)** Distribution of curvatures at these scales. Both slopes and curvatures are obtained in the x-direction. The left plots in (b) and (c) show the computed values for the pristine surface, while the right plots in these panels show the values for the tip-artifacted measurement. Black solid lines show the normal distribution. **(d)** PSDs and **(e)** ACFs of both topographies. **(f)** The plot of minimum curvature h''_{min} (see text) shows a clear deviation between the pristine and artifacted measurement that starts at approximately the point where the scale-dependent minimum curvature equals the radius of the tip.

probe smoothens the peaks of the topography. Indeed the curvature near the peaks must be equal to $-1/R_{\mathrm{tip}}$. Conversely, the valleys look like cusps that originate from the overlap of two spherical bodies. These cusps are sharp and should lead to large (in theory unbounded, but in practice bounded by resolution and noise) positive values of the curvature. Church & Takacs [47,48] have pointed out that tip artifacts should lead to PSD $C^{1D}(q) \propto q^{-4}$, which is precisely a result of the cusps in the topography. (The Fourier transform of a triangle scales as q^{-2}, such that the PSD $\propto q^{-4}$.) We have demonstrated in Refs. [1,3] numerically that this is indeed the case.

We are now in a position to more precisely look at the effect of tip radius. Fig. 5b shows the scale-dependent slope distribution $P_1(h', \ell)$, normalized by the rms slope at the respective scale. The black solid line shows a Gaussian distribution (of unit width) for reference. It is clear that both our pristine topography (left columns) and the topography with tip-radius artifacts (right column) follow a Gaussian distribution for the scale dependent slopes across scales from 1 nm to 256 nm shown in the figure.

The situation is different for the scale-dependent curvature, shown in Fig. 5c. While the pristine surface (left column) follows a Gaussian distribution, the topography with tip-radius artifacts is only Gaussian for larger scales (ℓ = 16 nm and 256 nm). There is a clear deviation at the smallest scales, showing an exponential distribution for positive curvature values, corroborrating the empirical discussion above that cusps leads to large positive values for the curvature. As argued above and in Refs. [47,48], these cusps lead to a PSD $\propto q^{-4} \propto \lambda^4$. Fig. 5d shows the PSDs of both topographies. The artifacted surface indeed crosses over to $C^{1D} \propto \lambda^4$ at a wavelength of $\lambda \sim 20 - 40$ nm.

We note this cross-over to λ^4 is subtle and difficult to detect in measured data. Other measures, such as the ACF shown in Fig. 5e, are unsuitable to detect these artifacts. The region where the $C^{1D} \propto \lambda^4$ shows up as a linear region in the square root of the ACF, $\sqrt{A} \propto \ell$. The exponent of 1 from that region is too close to the exponent of $H = 0.8$ to be clearly distinguishable. The present authors have previously suggested a tip-radius reliability cutoff [1,2], where the scale-dependent rms curvature was compared to the tip curvature. Based on similar ideas, we now propose an additional metric that is intended to more accurately detect the onset of the tip-radius artifact.

Rather than computing the width of the distribution as do the rms measures, we now ask the question of what is the *minimum* curvature value found at a specific scale ℓ. We therefore evaluate

$$h''_{\mathrm{min}}(\ell) = -\min_k \left[\frac{D^2_{(\ell)}}{D_{(\ell)}x^2} h(x_k) \right]. \tag{35}$$

The crosses in Fig. 5f show this quantity for the pristine and the artifacted surface. It is clear that at small scales the curvature of the pristine surface is larger than the artifacted one. Additionally, the artifacted surface settles to $h''_{\mathrm{min}}(\ell) \approx 1/R_{\mathrm{tip}}$ as $\ell \to 0$. This is a clear indicator that the curvature of the peaks on the artifacted surface is given by the tip radius and that in principle, the tip radius can be deduced from h''_{min}. However, in real AFM data, h''_{min} has no well defined $\ell \to 0$ limit because there are noise sources not considered in our simulated measurement. The tip radius then needs to be determined from auxiliary measurements (see next section).

For each tip radius and surface topography, there is a critical length scale ℓ_{tip} below which AFM data is unreliable. We estimate ℓ_{tip} by numerically solving

$$h''_{\mathrm{min}}(\ell_{\mathrm{tip}}) = c/R_{\mathrm{tip}} \tag{36}$$

for ℓ_{tip} using a bisection algorithm. The empirical factor c needs to be close to or slightly smaller than unity. Fig. 5f shows this condition as a dashed horizontal line. Note that ℓ_{tip} depends both on the tip radius and the curvature of the measured surface: measurements on rough surfaces have more tip artifacts than measurements on smooth surfaces because a tip that can conform to the valleys of a smooth surface may not be able to sample the valleys on a rougher surface. We also indicate the scale ℓ_{tip} in the ACF (Fig. 5e) and in the PSD. The factor $c = 1/2$ was chosen such that λ_{tip} marks the crossover from artifacted $C^{1D} \propto \lambda^4$ to the pristine $C^{1D} \propto \lambda^{1+2H}$. We will use the same factor when analyzing experimental data in the next section, where there is no "pristine" measurement available for comparison.

Our proposed measure is useful because it can be robustly and automatically carried out on large sets of measurements; by contrast, the detection of $C^{1D} \propto \lambda^4$ is difficult because fitting exponents requires data over at least a decade in length and carries large errors [49].

4.3. Application to an experimental measurement

As a final example, we turn to an experimental analysis of an ultrananocrystalline diamond (UNCD) film that has been described in detail in Ref. [2]. Fig. 6a shows a single representative AFM scan of that surface available online at Ref. [50]. The peaks have rounded tips similar to the synthetic scan shown in Fig. 5a. The curvature distribution (Fig. 6b) also has a similar characteristic to our synthetic topography (Fig. 5c). At large scales, the distribution is approximately Gaussian (shown by the solid black line). At smaller scales, we see deviations to higher curvature

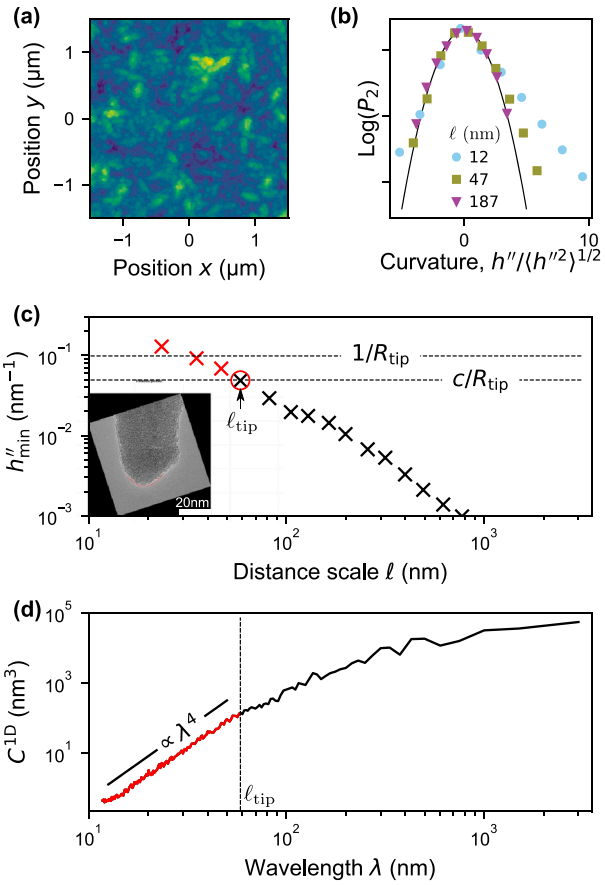


Fig. 6. AFM measurement of an ultrananocrystalline diamond film. **(a)** AFM measurement showing the smoothing of peaks similar to the emulated scans shown in Fig. 5a. **(b)** Normalized curvature distribution at distance scales ℓ = 12 nm (circles •), 47 nm (squares ■) and 187 nm (triangles ▼). ℓ = 12 nm corresponds to a scale factor $\eta = 1$. **(c)** We use the peak curvature h''_{min} (see text) to estimate the scale ℓ_{tip} below which the AFM data is unreliable (highlighted in red). The empirical constant $c = 1/2$. Inset: TEM image of the AFM tip. Fitting a parabola to the tip yields a radius R_{tip} of 10 nm. **(d)** Power spectral density (PSD) of the measurement. The black solid line shows scaling with λ^4 that indicates tip artifacts.

values, indicative of the cusps that are characteristic of tip artifacts. We also see a minor deviation to higher negative curvatures, which we attribute to additional instrumental noise that contributes to small-scale features of the data.

These negative curvatures prevent the conclusive determination of the tip radius from the scale-dependent tip curvature (Fig. 6c). Unlike the synthetic surfaces, the scale-dependent tip curvature $h''_{\mathrm{min}}(\ell)$ (Fig. 5f) does not saturate to a specific value at small distances ℓ. Instead, we determined the radius of AFM tip from auxiliary transmission electron microscopy (TEM) measurements (Fig. 6c inset). For the measured $R_{\mathrm{tip}} = 10$ nm, we can identify the region where $h''_{\mathrm{min}}(\ell) > 1/(2R_{\mathrm{tip}})$ as unreliable, leading to a lateral length-scale of around $\ell_{\mathrm{tip}} \approx 60$ nm below which the data is no longer reliable. The PSD (Fig. 6d) shows λ^4 scaling below the characteristic wavelength ℓ_{tip}.

After having looked at tip-radius effects on single measurements, we now turn to applying SDRPs to the full experimental dataset from Ref. [2], where a total of 126 individual measurements from three different instruments, a stylus profilometer, an AFM and a TEM, were combined to extract the power spectrum of the surface over eight orders of magnitude. (The dataset is available online at Ref. [51].) Fig. 7 shows the PSD, ACF, rms slope and rms curvature for each individual measurement as well as an average curve representative of the whole surface. For each tip-based measurement (stylus and AFM), we computed the critical scale ℓ_{tip} using Eq. (36) as above and excluded data on scales below ℓ_{tip}. The good overlap of the AFM data with the TEM data confirms that this procedure removed tip artifacts. The full data set (see Ref. [2]) shows clear regions where the PSD $C^{1\mathrm{D}} \propto q^{-4}$.

As shown in Fig. 7, all four methods can be used to stitch together the data from a large set of measurements to obtain the resulting SDRP of the underlying physical surface. The ACF (Fig. 7b) and rms slope h'_{rms} (Fig. 7c) of the TEM measurements curve down at large ℓ, an effect also seen (but less pronounced) in our synthetic data of Fig. 4c and d. This is a consequence of tilt correction, that enforces zero slope at the size of the overall measurement, hence forcing h'_{rms} to drop towards zero. While more sophisticated schemes for tilt correction could be devised to eliminate this long-wavelength artifact, the rms curvature h''_{rms} is free of this artifact as it is unaffected by local tilt of the measurement. We therefore advocate that it is important to look at a combination of these scale-dependent analysis techniques rather than relying on a single one.

5. Summary & conclusions

First, we demonstrated the calculation of scale-dependent parameters using a finite-differences scheme, with a variable distance scale. We have termed this characterization of first- and higher-order derivatives the *scale-dependent roughness parameter (SDRP)* analysis. Then we showed that the commonly used height-difference autocorrelation function (ACF) can be interpreted as the scale-dependent root-mean-square (rms) finite-difference slope. This leads to a straightforward generalization of the ACF for higher derivatives, yielding for example a scale-dependent rms curvature. We have further generalized this analysis to compute distributions of roughness parameters as a function of scale; the curvature distribution analysis can be used to identify tip-radius artifacts. To connect our analysis to conventional techniques, we have shown how equivalent scale-dependent roughness parameters can be computed not only from the ACF, but also from the the variable bandwidth method and the power spectral density. We demonstrated the successful use of these analyses to describe computer-generated and experimentally measured surfaces. Additional work is ongoing to apply this analysis on a wide range of surfaces, for example surfaces with facets [3] or periodic structures.

In summary, we proposed a novel SDRP analysis, and demonstrated how this is a generalization of commonly used roughness metrics. We suggest that this SDRP approach serves to harmonize competing roughness descriptors, but also offers advantages over those other methods, especially in terms of ease of calculation, intuitive interpretability, and detection of artifacts.

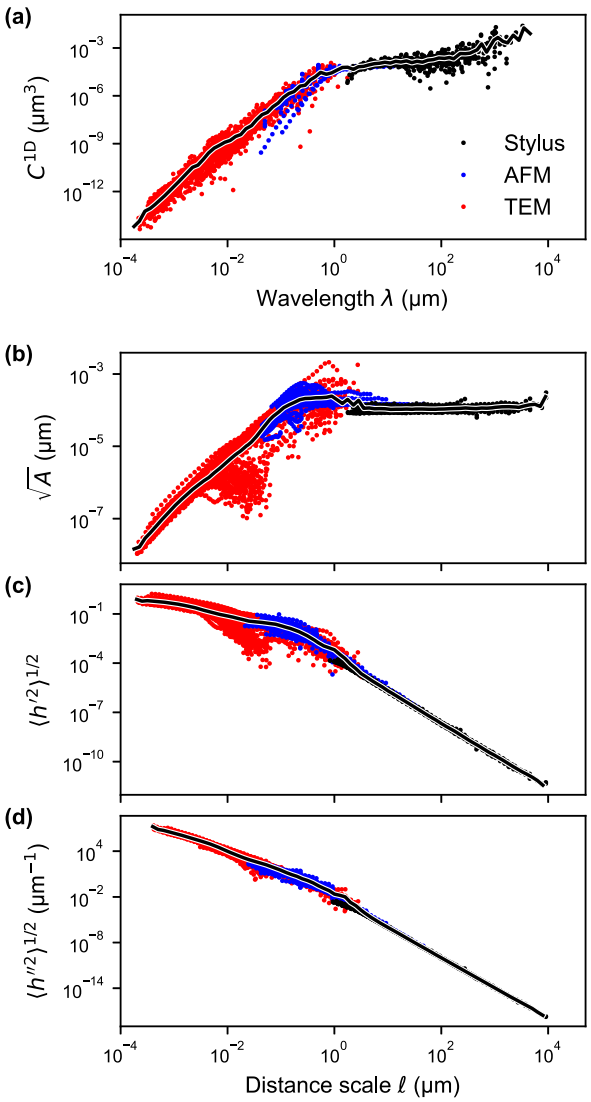


Fig. 7. Topography measurements of an ultrananocrystalline diamond film are combined across eight orders of magnitude of scales using the PSD (a), ACF (b), rms slope (c) and rms curvature (d). For each of these measures, we obtain a curve representative of the surface by averaging over all the individual measurements (solid black line). For tip-based methods, AFM and stylus profilometry, we used the criterion proposed in Section 4.2 to exclude unreliable scales.

Data availability statement

All the underlying synthetic and experimental scans are available at https://contact.engineering. Individual datasets are referenced throughout the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

LP thanks the Deutsche Forschungsgemeinschaft (project EXC-2193/1 – 390951807) and the European Research Council (StG-757343) for

funding this work. TDBJ acknowledges funding from the U.S. National Science Foundation under award number CMMI-1727378.

Appendix A. Generalization to two dimensions

We here briefly outline the generalization of the SDRP to two-dimensions. The main difference is that in two dimensions the derivative becomes the (discrete) gradient $(Dh/Dx, Dh/Dy)$, the curvature becomes the Hessian $(D^2h/Dx^2, D^2h/Dy^2, D^2h/DxDy)$ and higher-order derivatives contain additional cross terms. All averages are carried out over areas, not line scans. We can then for example define a scale-dependent *gradient* as

$$h'^{,2D}_{\mathrm{SDRP}}(\ell) = \left\langle \left(\frac{D_{(\ell/\Delta x)}}{D_{(\ell/\Delta x)x}} h(x,y) \right)^2 + \left(\frac{D_{(\ell/\Delta y)}}{D_{(\ell/\Delta y)y}} h(x,y) \right)^2 \right\rangle^{1/2}, \tag{A.1}$$

where the average $\langle \cdot \rangle$ now runs over the area. We note that in two-dimensions the situation may arise, where the scale factors $\eta_x = \ell / \Delta x$ and $\eta_y = \ell / \Delta y$ are no longer integer; this in particular happens if the aspect ratio of the individual pixel is not unity, $\Delta x \neq \Delta y$. In this case the additional (numerical) complexity arises, that one needs to interpolate between data points to measure the derivatives at the same distance scale in x- and y-direction.

References

[1] T.D.B. Jacobs, T. Junge, L. Pastewka, Quantitative characterization of surface topography using spectral analysis, Surf. Topogr. Metrol. Prop. 5 (2017) 013001.
[2] A. Gujrati, S.R. Khanal, L. Pastewka, T.D.B. Jacobs, Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales, ACS Appl. Mater. Interf. 10 (2018) 29169.
[3] A. Gujrati, A. Sanner, S.R. Khanal, N. Moldovan, H. Zeng, L. Pastewka, T.D. B. Jacobs, Comprehensive topography characterization of polycrystalline diamond coatings, Surf. Topogr. Metrol. Prop. 9 (2021) 014003.
[4] 4287 ISO, Geometrical product specifications (GPS)–Surface texture: Profile method, Standard, Geneva: International Organization for Standardization, 1997.
[5] B. Mandelbrot, How long is the coast of Britain? Statistical self-similarity and fractional dimension, Science 156 (1967) 636.
[6] B.B. Mandelbrot, The fractal geometry of nature, W. H. Freeman, 1982.
[7] S. Dalvi, A. Gujrati, S.R. Khanal, L. Pastewka, A. Dhinojwala, T.D.B. Jacobs, Linking energy loss in soft adhesion to surface roughness, Proc. Natl. Acad. Sci. USA 116 (2019) 25484.
[8] B.B. Mandelbrot, D.E. Passoja, A.J. Paullay, Fractal character of fracture surfaces of metals, Nature 308 (1984) 721.
[9] L. Ponson, D. Bonamy, E. Bouchaud, Two-dimensional scaling properties of experimental fracture surfaces, Phys. Rev. Lett. 96 (2006) 035506.
[10] D. Bonamy, L. Ponson, S. Prades, E. Bouchaud, C. Guillot, Scaling exponents for fracture surfaces in homogeneous glass and glassy ceramics, Phys. Rev. Lett. 97 (2006) 135504.
[11] E. Nadgorny, J. Schwerdtfeger, F. Madani-Grasset, V. Koutsos, E.C. Aifantis, M. Zaiser, Evolution of self-affine surface roughness in plastically deforming KCl single crystals. Proceedings of International conference on Statistical Mechanics of Plasticity and Related Instabilities – PoS(SMPRI2005) volume 023, 2006, p. 012.
[12] J. Schwerdtfeger, E. Nadgorny, F. Madani-Grasset, V. Koutsos, J.R. Blackford, M. Zaiser, Scale-free statistics of plasticity-induced surface steps on KCl single crystals, J. Stat. Mech. – Theory Exp. 2007 (2007) L04001.
[13] S. Sandfeld, M. Zaiser, Deformation patterns and surface morphology in a minimal model of amorphous plasticity, J. Stat. Mech. – Theory Exp. 2014 (2014) P03014.
[14] A.R. Hinkle, W.G. Nöhring, R. Leute, T. Junge, L. Pastewka, The emergence of small-scale self-affine surface roughness from deformation, Sci. Adv. 6 (2020). Eaax0847.
[15] B.N.J. Persson, On the fractal dimension of rough surfaces, Tribol. Lett. 54 (2014) 99.
[16] A.W. Bush, R.D. Gibson, T.R. Thomas, The elastic contact of a rough surface, Wear 35 (1975) 87.
[17] B.N.J. Persson, Elastoplastic contact between randomly rough surfaces, Phys. Rev. Lett. 87 (2001) 116101.
[18] B.N.J. Persson, Theory of rubber friction and contact mechanics, J. Chem. Phys. 115 (2001) 3840.
[19] B.N.J. Persson, E. Tosatti, The effect of surface roughness on the adhesion of elastic solids, J. Chem. Phys. 115 (2001) 5597.
[20] B.N.J. Persson, Adhesion between elastic bodies with randomly rough surfaces, Phys. Rev. Lett. 89 (2002) 245502.
[21] S. Hyun, L. Pei, J.-F. Molinari, M.O. Robbins, Finite-element analysis of contact between elastic self-affine surfaces, Phys. Rev. E 70 (2004) 026117.
[22] C. Campañá, M.H. Müser, Contact mechanics of real vs. randomly rough surfaces: a Green's function molecular dynamics study, EPL 77 (2007) 38005.
[23] M.H. Müser, Rigorous field-theoretical approach to the contact mechanics of rough elastic solids, Phys. Rev. Lett. 100 (2008) 055504.
[24] L. Pastewka, M.O. Robbins, Contact between rough surfaces and a criterion for macroscopic adhesion, Proc. Natl. Acad. Sci. USA 111 (2014) 3298.
[25] M.S. Longuet-Higgins, The statistical analysis of a random, moving surface, Philos. Trans. R. Soc. Lond. A 249 (1957) 321.
[26] M.S. Longuet-Higgins, Statistical properties of an isotropic random surface, Philos. Trans. R. Soc. Lond. A 250 (1957) 157.
[27] P.R. Nayak, Random process model of rough surfaces, J. Lubr. Technol. 93 (1971) 398.
[28] A. Wang, M.H. Müser, On the usefulness of the height-difference-autocorrelation function for contact mechanics, Tribol. Int. 123 (2018) 224.
[29] R.S. Sayles, T.R. Thomas, The spatial representation of surface roughness by means of the structure function: a practical alternative to correlation, Wear 42 (1977) 263.
[30] T.R. Thomas, Rough surfaces, World Scientific, 1998.
[31] M.J. Cannon, D.B. Percival, D.C. Caccia, G.M. Raymond, J.B. Bassingthwaighte, Evaluating scaled windowed variance methods for estimating the hurst coefficient of time series, Physica A 241 (1997) 606.
[32] G. Moreira, J.d. Silva, S. Kamphorst, On the fractal dimension of self-affine profiles, J. Phys. A 27 (1994) 80794089.
[33] J. Schmittbuhl, J.-P. Vilotte, S. Roux, Reliability of self-affine measurements, Phys. Rev. E 51 (1995) 131.
[34] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E 49 (1994) 1685.
[35] C.K. Peng, S. Havlin, H.E. Stanley, A.L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos 5 (1995) 82.
[36] R.S. Sayles, T.R. Thomas, Surface topography as a nonstationary random process, Nature 271 (1978) 431.
[37] P. Podsiadlo, M. Wolski, G.W. Stachowiak, Novel directional blanket covering method for surface curvature analysis at different scales and directions, Tribol Lett 65 (2017) 2.
[38] P. Podsiadlo, M. Wolski, G.W. Stachowiak, Directional signatures of surface texture, Tribol Lett 67 (2019) 109.
[39] S.B. Ramisetti, C. Campañá, G. Anciaux, J.-F. Molinari, M.H. Müser, M.O. Robbins, The autocorrelation function for island areas on self-affine surfaces, J. Phys. Condens. Matter 23 (2011) 215004.
[40] T.D.B. Jacobs, T. Junge, L. Pastewka, Self-affine synthetic surface, 2020. https://contact.engineering/go/867nv/.
[41] A. Sagy, E.E. Brodsky, G.J. Axen, Evolution of fault-surface roughness with slip, Geology 35 (2007) 283.
[42] T. Candela, F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, E.E. Brodsky, Roughness of fault surfaces over nine decades of length scales, J. Geophys. Res. – Sol. Ea. 117 (2012).B08409.
[43] E.E. Brodsky, J.D. Kirkpatrick, T. Candela, Constraints from fault roughness on the scale-dependent strength of rocks, Geology 44 (2016) 19.
[44] T. Candela, E.E. Brodsky, The minimum scale of grooving on faults, Geology 44 (2016) 603.
[45] C.A. Thom, E.E. Brodsky, R.W. Carpick, G.M. Pharr, W.C. Oliver, D.L. Goldsby, Nanoscale roughness of natural fault surfaces controlled by scale-dependent yield strength, Geophys. Res. Lett. 44 (2017) 9299.
[46] T.D.B. Jacobs, T. Junge, L. Pastewka, Synthetic self-affine topography scanned with a virtual spherical probe, 2021. https://contact.engineering/go/v9qwe/.
[47] E.L. Church, P.Z. Takacs, Instrumental Effects in Surface Finish Measurement. Proc. SPIE 1009, Surface Measurement and Characterization, 1989, pp. 46–55.
[48] E.L. Church, P.Z. Takacs, Effects of the Nonvanishing Tip Size in Mechanical Profile Measurements. Proc. SPIE 1332, Optical Testing and Metrology III: Recent Advances in Industrial Optical Inspection, 1991, pp. 504–514.
[49] A. Clauset, C. Shalizi, M. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (2009) 661.
[50] A. Sanner, L.A. Thimons, W.G. Nöhring, T.D.B. Jacobs, L. Pastewka, Unpolished UNCD (measured with 10.11 nm AFM tip), 2021. https://contact.engineering/go/mvttv/.
[51] A. Gujrati, S.R. Khanal, L. Pastewka, T.D.B. Jacobs, Ultrananocrystalline diamond (UNCD) (2020).https://contact.engineering/go/8sc7t/.

Publication V

Experimental Mechanics (2021) 61:1109–1120
https://doi.org/10.1007/s11340-021-00733-6

RESEARCH PAPER

Hard-material Adhesion: Which Scales of Roughness Matter?

L. A. Thimons[1] · A. Gujrati[1] · A. Sanner[2,3] · L. Pastewka[2,3] · T. D. B. Jacobs[1]

Received: 9 December 2020 / Accepted: 9 May 2021 / Published online: 1 July 2021

Abstract

Background Surface topography strongly modifies adhesion of hard-material contacts, yet roughness of real surfaces typically exists over many length scales, and it is not clear which of these scales has the strongest effect. Objective: This investigation aims to determine which scales of topography have the strongest effect on macroscopic adhesion.

Methods Adhesion measurements were performed on technology-relevant diamond coatings of varying roughness using spherical ruby probes that are large enough (0.5-mm-diameter) to sample all length scales of topography. For each material, more than 2000 measurements of pull-off force were performed in order to investigate the magnitude and statistical distribution of adhesion. Using sphere-contact models, the roughness-dependent *effective* values of work of adhesion were measured, ranging from 0.08 to 7.15 mJ/m^2 across the four surfaces. The data was more accurately fit using numerical analysis, where an interaction potential was integrated over the AFM-measured topography of all contacting surfaces.

Results These calculations revealed that consideration of nanometer-scale plasticity in the materials was crucial for a good quantitative fit of the measurements, and the presence of such plasticity was confirmed with AFM measurements of the probe after testing. This analysis enabled the extraction of geometry-independent material parameters; the *intrinsic* work of adhesion between ruby and diamond was determined to be 46.3 mJ/m^2. The range of adhesion was 5.6 nm, which is longer than is typically assumed for atomic interactions, but is in agreement with other recent investigations. Finally, the numerical analysis was repeated for the same surfaces but this time with different length-scales of roughness included or filtered out.

Conclusions The results demonstrate a critical band of length-scales—between 43 nm and 1.8 μm in lateral size—that has the strongest effect on the total adhesive force for these hard, rough contacts.

Keywords Adhesion · Nanocrystalline diamond · Multi-scale surface roughness · Range of adhesion · Pull-off force

Introduction

All real surfaces exhibit roughness, which has profound effects on surface properties. This includes the mechanics of interfaces: adhesion [1, 2], contact stiffness [3–5], wetting [6], and friction [7]. Various analytical models have been developed to describe the dependence of functional properties on the geometry of the rough surface. The classic Greenwood and Williamson [8] multiasperity model for contact between rough surfaces was extended by Fuller and Tabor [1] and Maugis [9] to include adhesion. Further progress was made in connecting contact properties and roughness by Bush, Gibson, and Thomas (BGT) [10] and Rumpf [11, 12]. These models approximate real-world roughness using simpler mathematical functions and typically associate properties with a single geometric parameter, such as the root-mean square (RMS) height. More recent models have attempted to explicitly account for the hierarchical, multi-scale nature of roughness. Mandelbrot began characterizing surfaces as fractal-like and self-affine using spectral analysis [13]. Later, Persson [14] created a theory for rubber friction which draws quantitative connections between fractal roughness and contact properties, including adhesion [15]. These multi-scale models are expansions

✉ T. D. B. Jacobs
tjacobs@pitt.edu

[1] Department of Mechanical Engineering and Materials Science, University of Pittsburgh, Pittsburgh, PA 15261, USA

[2] Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany

[3] Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany

that start from the limit of conforming contacts. They balance the adhesive energy gained when making contact with the stored elastic deformation energy, and are expected to work well for soft, conformal contacts.

A simpler approach is sufficient for hard contacts, in which the elastic energy required for creating a conforming contact is much larger than the interfacial adhesive energy gain: An interaction potential is integrated over the undeformed contacting geometry. This approach can be applied to simple analytical geometries, such as spheres [16, 17], and can also be applied to more complex geometries, including rough surfaces, sharp tips, etc. [18–20]. Pioneering work by Delrio et al. [18] showed that long-range Casimir forces contribute strongly to surface adhesion of ultra-flat (RMS height of 2–10 nm) micromachined surfaces at separation distances up to tens of nanometers.

Three critical questions remain for describing the roughness-dependent adhesion of surfaces with multi-scale topography. First, can the classic analytical models such as those proposed by Fuller and Tabor, Rumpf, and Maugis, be applied to describe their behavior? Second, is the importance of long-range adhesive interactions limited to ultra-flat surfaces (as found in [18]) or is it generalizable to real-world coatings that are rough over many length scales? Third, and most generally, which size-scales contribute most strongly to adhesion?

The purpose of the present work is to investigate these questions using adhesion tests of some of the hardest materials: a ruby sphere on polycrystalline diamond substrates. Diamond coatings are technologically relevant in a number of applications [21], including medical devices [22, 23], tool coating [24], face seals [25], and microelectromechanical systems (MEMS)[26]. The roughness of these films can be controlled by varying the growth condition or by polishing. Thus, we can test substrates with varying roughness but nominally identical surface chemistry. This allows us to isolate the effects of topography on adhesion. The surface topography [27] of these materials and their adhesion to soft PDMS [28] has been extensively characterized in prior publications. The present investigation examines their adhesion to a hard material: ruby.

Many adhesion studies have used atomic force microscopy (AFM) or colloidal AFM to characterize surface topography and then perform tip-based adhesion tests on the measured surface [29–32]. Such investigations provide valuable information on the atomic-scale parameters governing nanoscale adhesion. However, the small size of the contact limits their applicability in understanding the contribution of multi-scale roughness to macroscale adhesion. The present investigation overcomes this limitation by using AFM to characterize the topography, while using a large 0.5-mm-diameter sphere to measure adhesion.

Methods

Experimental Adhesion and Topography Measurements

Adhesion tests were carried out between ruby spheres and polycrystalline diamond coatings using a MEMS-based force sensing probe (FT-MA02, FemtoTools, Buchs, Switzerland). The 0.5-mm-diameter spheres (B0.50R, Swiss Jewel, Philadelphia, PA) were pre-polished to an ultra-smooth (RMS height < 1 nm) finish using a slurry of ruby particles (0.05 μm). The spheres were glued to the tips of the force-probes to create a sphere-on-flat geometry for the test. The substrates comprised four different polycrystalline diamond coatings, which were grown by hot-filament chemical vapor deposition (HF-CVD) and are boron-doped for electrical conductivity. The substrates have varying grain size, and are denoted microcrystalline diamond (MCD), nanocrystalline diamond (NCD), ultrananocrystalline diamond (UNCD), and a polished form of UNCD. The deposition and surface topography of the diamond coatings are characterized in Ref. [33].

For this work, the topography of the spheres and substrates were measured using atomic force microscopy (AFM) (Dimension V, Bruker, Billerica, MA). Measurements were made using diamond-like carbon (DLC)-coated probes (Tap DLC300, Mikromasch, Watsonville, CA) in tapping mode. Scans with lateral size of 2.5 μm (512 × 512 pixels) were performed on each of the four substrates and on the ruby spheres.

Representative images for the surface topography of the substrates, and of the polished spheres are shown in Fig. 1. Using the AFM measurements of 2.5-micron lateral size, the root-mean-square height, slope, and curvature of the surfaces

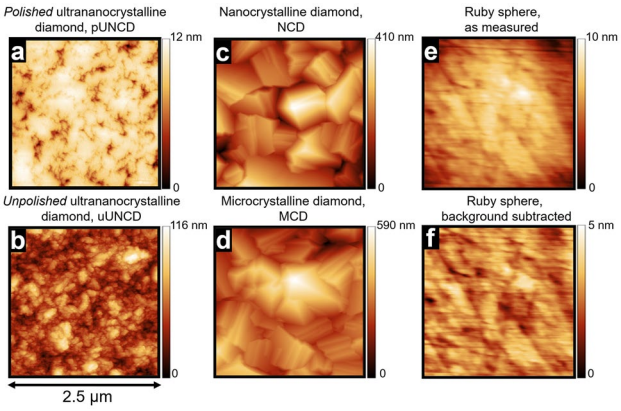


Fig. 1 AFM measurements of the four polycrystalline diamond substrates (**a-d**) and one instance of a ruby sphere (**e**). The ruby sphere is also shown with its spherical geometry subtracted (**f**) to allow for direct comparison of roughness against the other substrates

are given in Table 1. The surface topography of the polycrystalline diamond films has been extensively measured in Ref. [27]. The numerical analysis included here, however, uses only the AFM-based topography measurements.

The MCD and NCD have the largest roughness magnitude of the four substrates, as is shown by the vertical scale bar of the AFM images in Fig. 1. These two surfaces also show clear faceting due to the grain structure. The unpolished UNCD is significantly smoother than the MCD and NCD, and the faceting is not apparent at these size scales. The features that are visible have been shown to correspond to clusters of much smaller grains [27]. The polished UNCD is the smoothest of the four surfaces and shows no obvious grain structure. Similarly, the ruby tip images show a very smooth surface; while scratches are visible from the polishing process, the peak-to-valley roughness of this ruby sphere is smaller than all other surfaces. These AFM images, along with four more measurements from different sample areas of the various materials, form the basis of the numerical analysis that was performed.

The adhesion testing was performed using a custom micromechanical tester in a controlled-environment vacuum chamber on a vibration-isolation table. Dry air was flowed into the chamber prior to testing until the relative humidity was less than 1% (below the minimum reading of the humidity sensor). Dry air was flowed in for the duration of the test at low flow rates to ensure consistently low humidity levels.

A three-axis slip-stick piezoelectric stage provides closed-loop motion control and real-time x–y-z position data. For each individual adhesion measurement, the sphere was brought into contact with the substrate, loaded to a 10-μN preload (corresponding to a nominal Hertz stress of 135 MPa), and then withdrawn at a rate of 30 nm/s. The 10-μN preload occurs before the test and represents the minimum force required for the probe to find the point of contact. After finding contact the tip is lowered slowly onto the substrate up to a preload of

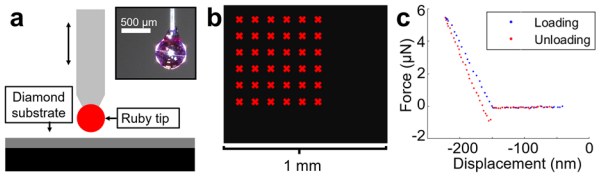


Fig. 2 The tester consisted of a ruby sphere glued to a MEMS-based force probe, as shown schematically (**a**) and in an optical image (inset). Arrays of measurements (**b**) were performed on a single sample, with one measurement per location. A typical force–displacement curve is shown in (**c**), with the most-negative point recorded as the pull-off force

5μN. Then, the force required to pull the sphere off of the surface is recorded as $F_{pull\text{-}off}$, with a force resolution of 30 nN. A single "test" comprised an array of individual adhesion measurements (typically 1-by-1 mm), for a total of 400 individual measurements per test. Tests were performed in immediate sequence on all four samples in randomized order, without opening the chamber or modifying test conditions. Six such sequences were performed with different spheres and different substrate samples, to ensure repeatability and reproducibility of results. A schematic of a typical test setup is shown in Fig. 2.

Numerical Analysis of Results

The experimental data was fitted using a cohesive zone model, using an exponential interaction potential with energy U given by

$$U(r) = -W_{\text{adh,int}} \exp\left(\frac{-r}{\rho}\right) \tag{1}$$

with a hard-wall repulsion at $r = 0$. Here, $W_{adh,int}$ is the intrinsic work of adhesion, r is the distance between interacting bodies, and ρ is the "range of adhesion" [34], which

Table 1 RMS roughness values calculated from AFM data only (top) and also from the full-spectrum roughness data, (bottom) which includes topography data from stylus, AFM, and TEM measurements of the same surfaces

AFM Data Only		Polished UNCD	Unpolished UNCD	NCD	MCD
	RMS height (nm)	3.2 ± 1.6	17.4 ± 2.1	97.1 ± 11.2	107.1 ± 12.0
	RMS slope	0.04 ± 0.01	0.39 ± 0.05	0.51 ± 0.07	0.49 ± 0.07
	RMS curvature (nm^{-1})	$2.0\times10^{-3} \pm 1.3\times10^{-3}$	$4.8\times10^{-2} \pm 3.8\times10^{-3}$	$3.9\times10^{-2} \pm 1.1\times10^{-2}$	$2.8\times10^{-2} \pm 6.6\times10^{-3}$
Stylus, AFM & TEM Data [11]		Polished UNCD	Unpolished UNCD	NCD	MCD
	RMS height (nm)	4.2 ± 0.8	17.4 ± 1.3	97.2 ± 11.7	101.2 ± 8.0
	RMS slope	0.31 ± 0.08	1.17 ± 0.28	0.92 ± 0.10	0.85 ± 0.10
	RMS curvature (nm^{-1})	1.99 ± 0.35	6.32 ± 1.20	5.91 ± 1.83	5.04 ± 1.45

describes the characteristic length scale of the adhesive interaction. Note that U and W are energies per unit surface area. Since we do not know the exact nature of the atomic-scale interaction between the two surfaces, this approach of using an empirical exponential interaction potential is a pragmatic approach to simplify the mathematical calculations. Similar results would be obtained using more complicated functional forms, e.g. based on instantaneous or retarded dispersion interactions, electrostatic interaction, or others. Distinguishing precisely between competing functional forms from our macroscopic experiments would be difficult. The functional form given in Eq. (1) allows separate fitting of the intrinsic work of adhesion (U at distance $r=0$) and the range of adhesion ρ, which yields the strength and length-scale of the interaction. This interaction potential has been widely used, including in the recent Contact Mechanics Challenge [34].

This interaction potential can be converted into a cohesive law (stress-distance relationship) for two interacting bodies,

$$p(r) = -\frac{dU}{dr} = -\frac{W_{\text{adh,int}}}{\rho}\exp\left(\frac{-r}{\rho}\right) \tag{2}$$

where p is the (compressive) pressure acting between the surfaces. For the present analysis, this potential was applied to each pixel pairing between the substrate and the tip, resulting in the following calculated force F_{calc} between the two contacting surfaces at separation d:

$$F_{\text{calc}}(d) = -\sum_{x,y}\frac{W_{\text{adh,int}}}{\rho}\exp\left(\frac{-(g_{x,y}+d)}{\rho}\right)A_{\text{pix}} \tag{3}$$

Here, A_{pix} is the area for a single surface pixel and the sum runs over all pixels in x and y. Note that $g_{\text{x,y}}$ in Eq. (3) is the *difference* of the topography maps of the ruby sphere and the diamond coating, while $g_{\text{x,y}}+d$ is the gap between the two interacting surfaces. The calculation can only be carried out for distances d where the gap $g_{\text{x,y}}+d$ is non-negative everywhere and the surfaces do not interpenetrate. The calculated adhesion values were thus found by summing the interaction potential pixel-by-pixel over every pixel pair of the two scans. The pull-off force is the minimum value of the force-separation curve $F_{\text{calc}}(d)$ that is found at the point of closest approach $d=$ -min $g_{\text{x,y}}$.

Due to random topography variation, there were sometimes significant contributions to adhesion from near the edges of the AFM scans. Therefore, for all substrate-tip combinations, the scans were stitched together so that there were no longer edge contributions to the adhesive interaction between the rough surface and a sphere of 0.5-mm diameter (< 0.5% change from additional stitching). The stitching was done by mirroring the surface scans horizontally and vertically to ensure that all edges matched up. This was needed because real topography measurements are not periodic. Similar to the experiments, the tip was brought into contact with the substrate in many locations over a square array.

The above *rigid* analysis was supplemented by an elastic and a plastic analysis. For the elastic analysis, we computed surface deformation using a Fast-Fourier-transform-accelerated boundary element technique [2, 35]. The hard-wall constraint was realized using an L-BFGS-B optimizer [36]. Note that we do not report the results of the elastic analysis explicitly here, but it was carried out to rule out the influence of elastic surface deformation. From this elastic analysis, we generally found that the surface pressure was exceeding common hardness values in most of the contact area such that a purely plastic analysis is appropriate.

For the plastic analysis, we use a simple bearing-area approach. This assumes that the harder surface plastically deforms the softer surface on all points that penetrate, and that the pressure in the contact area is equal to the hardness H. The penetration of the tip is then such that the number of contacting (and hence plastically deformed) surface pixels is sufficient to support the preload, $N_{\text{contact}} = F/HA_{\text{pix}}$. A preload of 10 μN, chosen to match the experimental preload, was used to determine the amount of plastic deformation of the softer surface. No plastic deformation then occurs during pull-off; the pull-off force is simply a result of the plastically deformed geometry. Note that we did not employ a combination of elastic and plastic contact, but similar plasticity models were used in elasto-plastic contexts in Refs. [37–39]. For the plastic analysis, surfaces were brought into contact up to the specified preload, and the deformed surface at this preload was then used in a rigid pull-off calculation.

Experimental Results

The topography can be used to compute roughness metrics such as root-mean-square (RMS) height, slope, and curvature, which are commonly used as inputs for rough-contact models. Table 1 shows the results of these calculations when performed only using the AFM measurements from this investigation (top) as compared to the same parameters that are computed when all of the many scales of roughness are included (bottom, using the full multi-scale spectral analysis from Ref. [27]). Table 1 shows the values that might serve as inputs to classical models, such as those of Maugis [9] or Rumpf [11]. Table 1 also serves to underscore just how widely varying these parameters can be when measured at different length scales. For example, the root-mean-square slope, a parameter that has been identified as important for multi-scale roughness models, varies for polished UNCD from 0.04 to 0.31 depending on how it is measured.

Figure 3 shows the distributions of values for adhesion force that have been measured on these four substrates. More than 2000 adhesion tests have been performed on every surface, with at least 6 different spheres. Each color on the histogram represents a new ruby sphere on a new sample of the diamond substrate. The mean adhesion of the polished UNCD was far higher than the unpolished version of the same material. Both UNCD surfaces showed higher adhesion than either the MCD or the NCD. The mean (median) pull-off forces of all four surfaces are 0.11 μN (0.06 μN), 0.16 μN (0.12 μN), 0.4 μN (0.31 μN), and 8.8 μN (1.75 μN), for MCD, NCD, unpolished UNCD, and polished UNCD respectively. Due to the large skew in the data, the mean value is shifted away from the peak of the distribution toward higher adhesion values.

All four of the distributions can be fit using a log-normal distribution. This is more difficult to see in the linear plots, given the skew of a large number of events with small pull-off force.

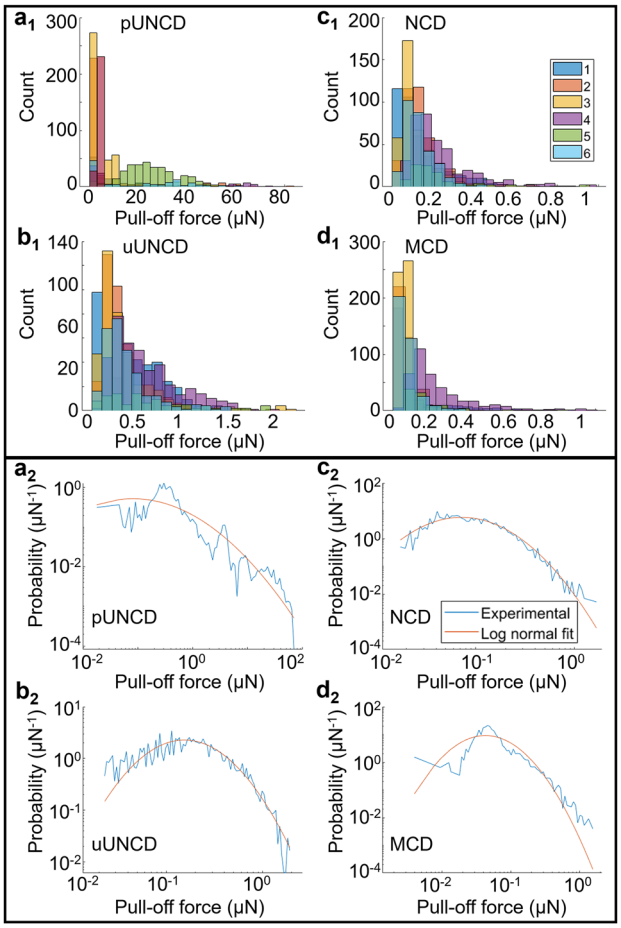


Fig. 3 Experimental pull-off data is shown for the four substrates across different testing sessions (different colors) on a linear scale (a_1-d_1). The same data from the top panel is combined into single datasets and shown on log–log plots (a_2-d_2), with log-normal fits shown (red lines)

Plotted on a log–log scale (Fig. 3, bottom), the log-normal distribution is shown to accurately fit the data over at least two decades of adhesive force. In all cases, the low end of the distribution is cut off at 30 nN, as this represents the sensitivity of the force probe. Similar shapes for adhesion distributions have been reported previously for measurements in various contexts, including: centrifugal adhesion studies of particle adhesion in powders [40, 41]; biological samples and cell adhesion [42, 43] and many other studies using AFM adhesion measurements [44–46]. The origin of this distribution shape is not yet clear. While the fit is good for a log-normal distribution, there are other distributions—such as half of a gaussian distribution an inverse-gaussian distribution— that also give good qualitative fits. Further investigation is required to ascertain the origin of the shape of these distributions.

Discussion

Effective Work of Adhesion and the Application of Classical Rough-adhesion Models

A standard method of analyzing adhesion in rough spheres is to use classical sphere-contact models (such as JKR or DMT [47]) to extract an *effective* work of adhesion $W_{adh,eff}$, and then to use standard roughness models (such as those described in the first paragraph of the introduction) to relate $W_{adh,eff}$ to standard roughness parameters. Following the procedure of Grierson et al. [48], with material parameters of ruby (elastic modulus E = 365 GPa and Poisson ratio ν=0.29) and diamond (E = 1010 GPa, ν=0.22), a nominal sphere diameter of 0.5 mm, the Tabor parameter is determined to be 0.81. This falls in the transition region between the DMT and JKR models. Using Maugis' analysis for the transition region between JKR and DMT, and an approximate equilibrium spacing of 0.3 nm, the analysis yields values of $W_{adh,eff}$ = 0.08, 0.13, 0.32, and 7.15 mJ/m^2 for MCD, NCD, unpolished UNCD, and polished UNCD, respectively. The surface chemistry is assumed to be similar for all of these HF-CVD diamond coatings, and therefore this difference is attributed primarily to surface topography.

It is clear from these measured values of effective work of adhesion, along with the values of RMS parameters shown in Table 1, that there are no simple relationships between RMS parameters and effective work of adhesion. Attempts to fit this data using simple analytical models [1, 9, 11] were unsuccessful, regardless of which roughness parameters were used (AFM-based or multi-scale). One potential explanation for why these models fail here is that the pull-off force for these hard materials is most dependent on the behavior of the uppermost contact points. These contacts represent the extreme-value statistics of the distribution of surface heights. They do not follow the central limit theorem

and are likely not described by the many models based on Gaussian statistics.

Intrinsic Work of Adhesion and Range of Adhesion

Instead of a single-asperity model extracting the effective work of adhesion for each substrate, a numerical analysis can be performed using the combined roughness of the sphere and substrate (see Methods). Like the experiments, the calculations were repeated for an array of 20×20 contacts on each substrate. Each numerical calculation yields a computed pull-off force F_{calc} for a specific choice of input values for $W_{adh,int}$ and ρ and a specific contact location on the rough surface (Fig. 4a). Then a fitting routine can be applied to all data to extract the best-fit values of those material parameters.

For numerical tractability, the analysis was only performed over a square of size of 12.5 μm rather than the 1-mm size scale of the experiments. Initially, the calculations were performed assuming rigid and/or elastic deformation only. Elastic calculations were virtually indistinguishable from the rigid calculations and we concluded that elasticity does not play a significant role in these contacts. Additionally, the micronewton-scale adhesive forces measured experimentally could not be explained with adhesion models based on rigid or elastically deforming surfaces, thus indicating that plastic deformation is likely occurring in these contacts.

We incorporated plastic deformation of the ruby tip with hardness of $H = 25$ GPa [49, 50] into our model using a penetration hardness model (see Methods). We first determine the plastic deformation in the softer sphere at a preload of 10 μN, identical to the average load used to find contact in the experiments. Pull-off calculations using the plastically deformed topography of the sphere were able to accurately reproduce the micronewton scale of pull-off forces from the experimental data. The possibility of plastic flow in similar hard materials has been reported in nanopillars [51] and nanoparticles [52]. A more in-depth analysis of the role of plasticity in these contacts is included in the following section.

The numerical analysis can be fit to the mean values of the experimental data from all four substrates in order to extract best-fit values for intrinsic work of adhesion and the range of adhesion. We note that while $W_{adh,int}$ simply rescales the computed pull-off force, the dependence of F_{calc} on ρ is nonlinear and depends on the specific topography. The range of adhesion ρ is extracted by analyzing the *ratios* of pull-off forces between materials, since this cancels the (unknown) intrinsic work of adhesion $W_{adh,int}$ in our model equations. Figure 4b shows the ratio of the pull-off force of polished UNCD, NCD and MCD with respect to unpolished UNCD. The solid horizontal lines are the experimental results and the data points represent calculations carried out at various values of ρ (x-axis). The error bars represent the variation over the contact points. Only for a range of adhesion of approximately $\rho = 5$ nm do all three lines cross the experimental results *simultaneously*. This means that while different values of ρ (with modified values of $W_{adh,int}$) can describe individual experiments, a simultaneous fit yields a range of adhesion around 5 nm. Note that the increase in pull-off force for small values for range of adhesion is due to the finite pixel size. Once the range of adhesion was fit, the data was scaled by a factor $W_{adh,int}$ to match the magnitude of the experiments. A second relative error minimization was performed to find the best-fit value for $W_{adh,int}$ at the best-fit range of adhesion. Figure 4c shows the computed pull-off results calculated at various points on the rough topography as a function of range of adhesion ρ. The work of adhesion used in this plot is the value that yields the best possible final fit.

It is clear from Fig. 4 that the range of adhesion strongly affects the values of adhesion force. Rougher surfaces, like MCD and NCD, are less strongly affected and can be fit over a wider band of values for ρ. Smoother surfaces, such as the polished UNCD, are more influenced by changes in ρ because the increasing range of adhesion enables more of the substrate to contribute to adhesion. This can be seen in Fig. 4c as a steeper slope for the smoothest polished UNCD surface and for the unpolished UNCD. The majority of the adhesion contribution to the rougher surfaces (MCD, NCD) comes from just one or two asperities, and therefore larger values for range of adhesion do not lead to such significant contributions to the area of interaction.

There is only one combination of parameters that enables the best fit for all samples. The fit was evaluated by computing and minimizing the mean relative error (MRE) between

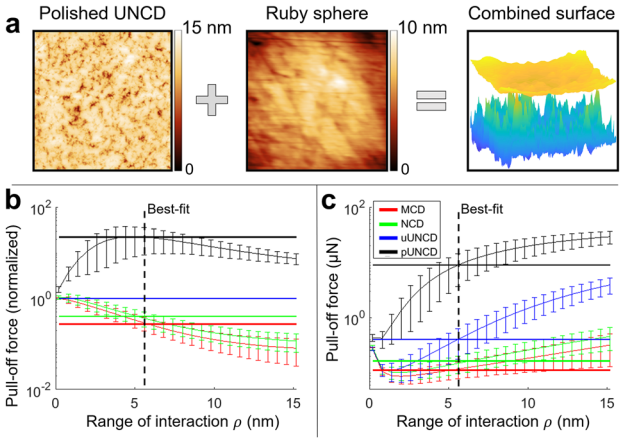


Fig. 4 Computed pull-off forces were calculated by integrating an interaction potential over the combined roughness of the sphere and substrate (**a**). The best-fit value of range of adhesion was found by fitting to ratios of pull-off force (**b**), to eliminate the absolute value of work of adhesion. Finally, the absolute values of pull-off force were matched (**c**) by finding the best-fit value of intrinsic work of adhesion

the fitted and measured adhesion. The uncertainty in the fit was computed for all values around the best-fit value with MRE < 0.1. This match between calculated and experimental data was used to extract values for work of adhesion of 46.3 ± 3.5 mJ/m^2 and range of adhesion of 5.6 ± 0.5 nm. Previous adhesion measurements on rough contacts between diamond and other hard materials report similar values for the intrinsic work of adhesion by accounting for surface roughness [20, 53, 54].

The measured range of adhesion is much longer than is expected for typical atomic interactions such as covalent bonds or van der Waals forces, which are typically considered to have a range of adhesion around 0.3 to 0.6 nm [55]. However, there is prior nanoscale literature that supports a larger-than-expected value for range of adhesion. Using DLC-coated AFM tips, Grierson et al. have measured a range of adhesion between DLC and UNCD of 4–5 nm [56]. While for spherical (parabolic) tips the pull-off force does not depend on range of adhesion [57], their measurements exploited the non-parabolic shapes of worn tips, where pull-off force does depend on range of adhesion. In separate experiments also involving AFM pull-off measurements, Jiang et al. have measured a range of adhesion between UNCD and PMMA of 1.5–2.5 nm [20]. Similarly, in nanoindentation experiments adhesive forces were found to act over distances of 1.5–4.5 nm [58]. As mentioned in the introduction, the presence of long-range forces has also been observed by DelRio et al. [18] in adhesion experiments involving silicon micro-cantilevers. The experiments showed contributions to adhesion from distances up to tens of nanometers.

The origin of these large values for range of adhesion is still in dispute. Previously proposed explanations involve electrostatic interactions due to contact charging [59–61], capillary adhesion [62–64], and Casimir forces [18]. First, electrostatic interactions have been proposed as a possible explanation because of the well-known phenomena of contact charging [60, 65–68]. While the detailed physical mechanism is still in discussion [69], the results are a net charge between the two materials that can result in measurable electrostatic interactions. In the present testing, these long-range electrostatic forces would be expected to be seen as measurable forces observed before and after contact. This can manifest as a tilting of the "out-of-contact" region of the force–displacement curve or as an earlier-than-expected snap-into-contact event as the charged sphere attracts the uncharged substrate in a new location. While such long-range interactions *have* been observed in other, unrelated testing where ruby tips were brought into contact with non-conductive substrates, the present substrates were boron-doped for conductivity, and the tester and substrates were electrically grounded to the vacuum chamber. The measured force curves in the present testing were similar to that shown in Fig. 2d, with no interaction forces observed until contact was initiated. Additionally, the operation of a static-reducing ionization bar had no measurable effect on adhesive forces. Therefore, contact charging is not expected to have played a significant role in the present results.

A second common explanation for longer-than expected values for range of adhesion is capillarity. Water bridges across a contact can increase the area of interaction of a rough contact and are known to significantly increase the adhesive force. The relative humidity determines the presence and size of these capillary bridges, which in turn affect the adhesive force. The present testing was carried out in a dry atmosphere (< 1% RH). This is insufficient to eliminate all water from the contact, but will limit its contribution. He et al. [70] showed that, even for hydrophilic surfaces, capillary necks could not form below a relative humidity of 20–40%. Numerical analyses [71] also suggests that capillary formation should not play a role in adhesion at low humidity. Therefore, capillarity is not expected to be the dominant factor in explaining the effect of topography on adhesion.

A third common explanation for large values for range of adhesion is retarded dispersion, or Casimir, interactions. These forces arise due to the finite speed of electromagnetic interactions and typically act over ranges larger than a few nanometers, even up to tens of nm [18]. The present investigation is consistent with these findings, since the large micronewton adhesive forces cannot be explained without considering longer-range interactions. Furthermore, the smoother surfaces show a stronger contribution from these longer-range interactions, while these interactions play a less important role for the rougher surfaces, with fewer, sharper asperities in contact. However, the interaction potential used here does not explicitly account for any specific attraction mechanism. It is an empirical potential that elucidates the strength and length-scale of the interaction. In this case, those parameters are consistent with Casimir forces, but further investigation would be required to conclusively demonstrate the physical origin.

The Role of Plasticity in Adhesion of These Contacts

To specifically verify that plastic deformation can occur in these contacts, an additional investigation was performed with AFM imaging performed on the ruby sphere (softer material) in the exact location of contact before and after an array test was performed. The standard adhesion test setup does not permit this precise knowledge of test location; therefore, an alternate custom micromechanical test setup was used, with a cantilever based force sensor, but otherwise similar setup. A ruby sphere was polished, pre-imaged in the AFM (Fig. 5b), and then used in an array test of adhesion against an NCD substrate. This testing repeated

the 400 measurements in an array of locations from a typical adhesion experiment, but used a preload of 20 μN (the minimum load of the alternate test apparatus). The apex of the sphere was imaged again after the test (Fig. 5c) and the location was matched to the pre-test image. The AFM imaging presented clear evidence of indentations at the tip of the sphere. The indentations were approximately 150 – 300 nm in lateral size, and approximately 2 – 10 nm in depth. These indentations were scattered across the tip of the sphere, with single indents concentrated around a region of multiple overlapping indents.

The numerical modelling (recalculated at a preload of 20 μN) predicted deformations with a depth of approximately 2 nm and edge lengths of approximately 50 nm (Fig. 5a). The measurements are in reasonable agreement with predictions. The region of overlapping indentations makes it difficult to determine the size of a single indentation. However, there do appear to be single indentations scattered around that region. The measured deformation for what appear to be single indents had depths ranging from single nanometers to nearly 10 nm and lateral sizes for measured deformation of approximately 100 – 500 nm. The computed deformations are for a single adhesion test, while the experimental deformations correspond to the cumulative effect of 400 adhesion tests against different contact points. Therefore, the overall scales of deformations compare favorably, and likely indicate that the tallest asperities on the substrates are serving to permanently indent the polished spheres.

Determining the Most Relevant Length-scales of Roughness

The numerical analysis in this investigation has demonstrated that adhesive interactions act over a length scale of greater than 5 nm, and that plasticity serves to increase

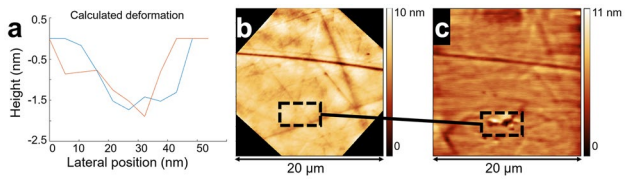


Fig. 5 The computed deformation for a single adhesion test is shown in panel (**a**) with the red and blue lines representing x and y direction line scans respectively. The actual deformation of the ruby sphere after an array of adhesion tests is measured using AFM images taken at the sphere apex before (**b**) and after (**c**) testing. The images have been precisely located at the apex of the tip where contact took place, and fiducial markers have been used to orient the image. The after-test image confirm the presence of permanent deformation, as is assumed in the numerical modeling, seems to be in order-of-magnitude agreement with what would be expected after 400 tests in different locations, each with the deformation shown in (**a**)

contact area above the predictions of rigid or elastic calculations. These two factors may limit the impact of certain length-scales of topography on the total macroscopic adhesion.

To check the influence of different roughness scales, we repeated the pull-off force calculations on a variety of virtual surfaces, with different scales of roughness filtered out or included. This was accomplished by taking advantage of the all-scale measurements performed on the same substrates in Ref. [27] and combining them with the AFM measurements performed here. Since we do not have multi-scale measurements taken in the exact same location, we used the statistics of the random roughness to add smaller- and larger-scale roughness to the measurements. Specifically, we started with an AFM image of the surface of the type shown in Fig. 1, then we superimposed artificially generated roughness that was created using a Fourier-filtering algorithm [72, 73] based on the measured PSD for that particular substrate. Therefore, these virtual surfaces are representative of the true multi-scale topography of each substrate. Then, from these multi-scale "master" surfaces, we filtered out different scales of roughness. Finally, we performed the numerical calculations on each of the filtered surfaces to compute the pull-off force and determine the sensitivity to different scales of roughness. The detailed approach of creating and filtering these surfaces is described in the next paragraph.

To add small-scale roughness, we first stitched the 512×512 pixel AFM scan using mirror images, leading to a 1024×1024 periodic topography. This stitched surface was first Fourier interpolated on an 8192×8192 grid (0.625 nm pixel size) and parts of the spectrum with wavelength smaller than $\lambda_T = 20$ nm were cut out. A randomly rough surface that follows the substrate PSD for wavelengths $\lambda < \lambda_T$ and has a constant roll-off above λ_T was added to this interpolated topography. Features below the varying cutoff wavelengths λ_S are then filtered out to check their effect on the calculated pull-off force. To add large-scale roughness, the AFM scan was again stitched to create a periodic topography, and this was stitched multiple times to yield an 8192×8192 (20 μm linear size) grid. Fourier components at wavelengths bigger than $\lambda_T = 1$ μm were cut out. A randomly rough surface with spectrum following the substrate PSD for wavelengths $\lambda > \lambda_T$ and zero below λ_T was added to this topography. Features with wavelength above the varying cutoff wavelengths λ_L are then filtered out to check how they affected the calculated pull-off force.

Figure 6 shows the change in pull-off that occurs when different length scales of roughness are filtered out. In Fig. 6a, which shows the effect of small-scale roughness, the leftmost datapoints represent the pull-off force computed on the unfiltered surface. As the short-wavelength cutoff (x axis) gets larger, more and more small-scale roughness is removed from

the surface. Thus, the surface is perfectly smooth below this cutoff; the rightmost data points approach the pull-off force from a perfectly flat plane. The value of pull-off force remains constant (within 10%) until the cutoff wavelength reaches 43 nm. Removing roughness *above* this size scale has a strong effect on the adhesive force, but removing roughness *below* this size has almost no effect. Figure 6b shows a similar calculation, but now with long-wavelength roughness filtered out; thus, the rightmost datapoint represents a nearly unfiltered surface. As the long-wavelength cutoff decreases, shorter and shorter wavelengths of roughness are removed from the surface, along with all wavelengths above the cutoff. The results show that, with removal of roughness with wavelength above approximately 1.8 microns, the adhesion remains constant (within 10%). Taken together, the two plots in Fig. 6 demonstrate that there is almost no effect on adhesion from roughness with lateral length scales smaller than 43 nm or larger than 1.8 microns. The critical finding of this analysis is that there is a certain band of length scales of roughness, 43 nm to 1.8 microns, that most-strongly affects adhesion in these materials; roughness outside of this band plays a secondary role in adhesion.

The explanation for this critical band of scales of topography may be different for the large and small scales. The unimportance of large-scale topography is likely linked to the area that is interacting with the sphere. Given a range of adhesion ρ, a sphere of diameter D will interact with a flat surface within a disk of radius $r = \sqrt{D\rho}$. For $D = 500$ nm and $\rho = 5.6$ nm we obtain a radius of $r = 1.7$ μm, almost exactly the wavelength above which large-scale topography no longer matters. This shows that macroscopic pull-off forces are strongly affected by finite-size effects, and that the magnitude of pull-off forces will depend strongly on the sphere radius. This also means that the scales of roughness that matter are determined by the macroscopic contact geometry, as long as sphere radius R is much larger than typical scales of the roughness.

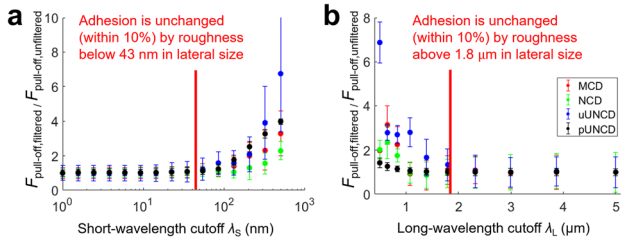


Fig. 6 The contribution to pull-off force from various length scales can be directly demonstrated by recalculating pull-off force after filtering out small (**a**) and large (**b**) scales of roughness. Specifically, the pull-off force calculated from the filtered surfaces is normalized by the pull-off force calculated from the unfiltered surfaces. In panel (**a**), the x-axis indicates a short-wavelength cutoff, where all roughness below this size scale has been removed. A value near 1 indicates that there is almost no effect on pull-off force of filtering out roughness below that size scale. In panel (**b**), the x-axis indicates a long-wavelength cutoff, where all roughness above this size is removed. Here, a value of 1 indicates no contribution to pull-off force from roughness above that size scale

For the unimportance of small-scale topography, there are two effects that enhance each other: the large range of adhesion, and the effect of plasticity. The large range of adhesion (5.6 nm), which was determined from the numerical analysis, indicates that topography variations below this scale have a reduced contribution to adhesion. For example, for rigid surfaces with a sinusoidal gap of amplitude 2 nm, a range of adhesion of 0.5 nm would mean that only the contacting peaks contribute to adhesion and the rest of the surface is irrelevant; while a range of adhesion of 5 nm would mean that all portions of the surface are adhering, with only small differences in relative contributions from different locations. An additional explanation is the effect of plasticity. The small-scale roughness has the highest local slope, and thus the highest surface stress. This means that the small-scale roughness will cause deformation, which smooths out these scales earlier than other scales, and reduces their contribution to macroscale adhesion.

Implications of the Present Findings

The results demonstrated that, for the macroscale adhesion of extremely stiff materials, the very smallest scales of roughness do not determine adhesion. This is in stark contrast to the adhesion of nanoscale contacts of hard materials [19] and to the adhesion of macroscale contacts of soft solids [28], both of which show a critical influence of smallest-scale roughness. While the present work draws on extensive roughness characterization at all scales using stylus profilometry, AFM, and TEM, in the end the AFM-scale roughness data (which covers the critical band of length scales discussed in the prior section) was sufficient to describe adhesion in these contacts. The introduction of smaller-scale roughness, as measured in the TEM, had little influence on the predicted adhesion. This means that parameters like RMS slope and curvature, that are most strongly influenced by the smallest-scale roughness, are less important for these hard-material contacts.

In these measurements, the larger scales of roughness were also less significant. This implies that measurements based on stylus profilometry, which is resolution-limited by the micron-scale radius of the tip, are not sufficient to predict and describe adhesion of these materials. It also implies that a simple scalar parameter such as RMS height is insufficient to determine macroscale adhesion. We look forward to investigating this point further, with the goal to understand the generalizability of this result beyond the current experimental setup.

Another key result of the calculations for hard materials is that adhesion is dominated by the asperities at the very high end of the height distribution. This leads to highly variable extremes in adhesion that can far exceed common predictions based on the average asperity height. The adhesion distributions appear to be log-normal, with a long tail, which strongly impacts the mean adhesion value and leads to rare but significant ultra-high-adhesion events. This has strong implications for real-world applications, such as MEMS devices, which must overcome such surface forces and the threat of stiction. The shape of the measured distributions would suggest that any moving parts should be significantly overdesigned to ensure they can overcome the long-tail events.

Another important finding is that the experimental results were unable to be fit without the inclusion of plasticity. The assumption is that at some length scale, the contact pressure will overcome the hardness of one of the materials. Ruby, in our experiments, is the softer material. Whether this results in plastic flow or fracture, the contact area should evolve to support the preload applied and will be significantly larger than predictions from elastic models. In recent work in both SEM [74] and TEM [51, 52] experiments, plasticity in nanoscale ruby and diamond samples has been reported. Simple experiments were performed to confirm the presence of small-scale plasticity on the ruby tip. These findings are supported by prior work demonstrating connections between nanoscale plasticity and large-scale properties [75]. The nanometer-scale deformation is likely an important factor behind the presence of a small-wavelength cutoff in the roughness that affects adhesion.

The *effective* work of adhesion (that includes the effect of topography) of these surfaces varies by almost two orders of magnitude, from 0.08 to 7.15 mJ/m^2. These values for $W_{adh,eff}$ are calculated from the overall sphere geometry using spherical contact-mechanics models. Given the wide variability of the adhesion force between interfaces of identical large-scale geometry, it is not surprising that the effective work of adhesion varies so much. These variations, however, are not explained by simple analytical models, such as those based on a Gaussian distribution of asperity heights, nor those based on a balance between elastic and adhesive energy.

Common (elasto-)adhesion theories balance the elastic energy required for deformation with the interfacial energy (intrinsic work of adhesion) gained during contact [2, 15, 76–78]. In our case, the interfaces are so stiff that the deformation energy vastly exceeds any energy gain from making contact and we expect no pull-off force (or no stickiness [2]) in the "thermodynamic" limit of large surface areas and vanishing range of adhesion. In our case, the pull-off force is then determined by the interfacial stress carried by the intermolecular potential between the two surfaces and we can simply compute it by summing up these stress contributions (as we did in our numerical calculations); or in other words, the interface does not separate like a crack [76]. The explanation for the appreciable pull-off force is tightly linked to the long range (5.6 nm) of interaction extracted from this analysis.

Conclusions

By combining detailed measurements of topography, thousands of mm-scale adhesion measurements, and numerical integration of an interaction potential, we computed both the intrinsic material parameters governing adhesion as well as the contributions to adhesion from multi-scale topography. The intrinsic work of adhesion between ruby and polycrystalline diamond was found to be 46.3 mJ/m^2 while the range of adhesion was 5.6 nm. This large value for range of adhesion, along with the requirement for plasticity in the calculations, leads to a diminished role of small-scale topography on the macroscale adhesion of these hard contacts. While prior work on soft-material adhesion on the same substrates [28] demonstrated the important role of single-digit-nm topography on adhesion, the same is not true for the present measurements of hard-material adhesion. In fact, based on this analysis incorporating plasticity and the large range of adhesion, it has been demonstrated that there is a critical band of length scales of topography—43 nm to 1.8 μm—which plays the most significant role in macroscale adhesion for these hard materials.

Acknowledgements We thank Nathaniel Miller for useful discussion and feedback on the manuscript. TDBJ, AG, and LAT acknowledge support from the National Science Foundation under award CMMI-1727378. AS and LP acknowledge funding by the Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy (project EXC-2193/1 – 390951807) and by the European Research Council (Starting Grant 757343). Use of the NanoFabrication and Characterization Facility (NFCF) in the Petersen Institute for Nano Science and Engineering (PINSE) is acknowledged.

Data Availability All data from this investigation is freely available for download. The surface topography data can be accessed using the following five URLS: (1) https://contact.engineering/go/btpax/ (2) https://contact.engineering/go/4r74d/ (3) https://contact.engineering/go/jqcfh/ (4) https://contact.engineering/go/9bcrf/ (5) https://contact.engineering/go/gneqk/ and the other measured data can be accessed using the following https://doi.org/10.18117/x77q-fg63.

Conflict of Interests The authors declare that they have no potential conflicts of interest regarding this research.

References

1. Fuller KNG, Tabor D (1975) The Effect of Surface Roughness on the Adhesion of Elastic Solids. Proc R Soc Lond A 345:327–342
2. Pastewka L, Robbins MO (2014) Contact between rough surfaces and a criterion for macroscopic adhesion. Proc Natl Acad Sci USA 111(9):3298–3303
3. Berthoud P, Baumberger T (1998) Shear Stiffness of a Solid-Solid Multicontact Interface. Proc Math Phys Eng Sci 454(1974):1615–1634
4. Benz M, Rosenberg KJ, Kramer EJ, Israelachvili JN (2006) The deformation and adhesion of randomly rough and patterned surfaces. J Phys Chem B 110(24):11884–11893
5. Akarapu S, Sharp T, Robbins MO (2011) Stiffness of contacts between rough surfaces. Phys Rev Lett 106(20):204301
6. Nosonovsky M, Bhushan B (2008) Biologically inspired surfaces: Broadening the scope of roughness. Adv Funct Mater 18(6):843–855
7. Persson BNJ (2005) Sliding Friction. Springer
8. Greenwood JA, Williamson JBP (1966) Contact of nominally flat surfaces. Proc R Soc London Ser A Math Phys Sci 295(1442):300–319
9. Maugis D (1995) On the contact and adhesion of rough surfaces. J Adhes Sci Technol 10(2):161–175
10. Bush AW, Gibson RD, Thomas TR (1975) The Elastic Contact of a Rough Surface. Wear 35:87–111
11. Rabinovich YI, Adler JJ, Ata A, Singh RK, Moudgil BM (2000) Adhesion between nanoscale rough surfaces. I. Role of asperity geometry. J Colloid Interface Sci 232(1):10–16
12. Rabinovich YI, Adler JJ, Ata A, Singh RK, Moudgil BM (2000) Adhesion between nanoscale rough surfaces: II. Measurement and comparison with theory. J Colloid Interface Sci 232(1):17–24
13. Mandelbrot BB, Passoja DE, Paullay AJ (1984) Fractal character of fracture surfaces of metals. Nature 308(5961):721–722
14. Persson BNJ (2001) Theory of rubber friction and contact mechanics. J Chem Phys 115(8):3840–3861
15. Persson BNJ, Tosatti E (2001) The effect of surface roughness on the adhesion of elastic solids. J Chem Phys 115(12):5597-5610
16. Bradley RS (1932) The cohesive force between solid surfaces and the surface energy of solids. London, Edinburgh Dublin Philos Mag J Sci 13(86):853-862
17. Greenwood JA (1997) Adhesion of Elastic Spheres. Proc R Soc Lond A 453:1277–1297
18. Delrio FW, DeBoer MP, Knapp JA, Reedy ED, Clews PJ, Dunn ML (2005) The role of van der Waals forces in adhesion of micromachined surfaces. Nat Mater 4:629–634
19. Jacobs TDB et al (2013) The effect of atomic-scale roughness on the adhesion of nanoscale asperities: A combined simulation and experimental investigation. Tribol Lett 50(1):81–93
20. Jiang Y, Turner KT (2016) Measurement of the strength and range of adhesion using atomic force microscopy. Extrem Mech Lett 9:119–126
21. May PW (2000) Diamond thin films : a 21st-century material. Phil Trans R Soc Lond A 358:473–495
22. Mochalin VN, Turcheniuk K (2017) Biomedical Applications of Nanodiamond (Review). Nanotechnology 28:252001
23. Narayan R (2013) Diamond-Based Materials for Biomedical Applications. Woodhead Publishing
24. Sheikh-Ahmad J, Davim JP (2012) Machining Technology for Composite Materials.Woodhead Publishing
25. Sumant AV et al (2005) Ultrananocrystalline diamond film as a wear-resistant and protective coating for mechanical seal applications. Tribol Trans 48(1):24–31
26. Auciello O et al (2007) Are diamonds a MEMS' best friend? IEEE Microw Mag 8(6):61–75
27. Gujrati A et al (2021) Comprehensive topography characterization of polycrystalline diamond coatings. Surf Topogr Metrol Prop 9(014003)
28. Dalvi S, Gujrati A, Khanal SR, Pastewka L, Dhinojwala A, Jacobs TDB (2019) Linking energy loss in soft adhesion to surface roughness. Proc Natl Acad Sci USA 116(51):25484–25490
29. Kumar A, Staedler T, Jiang X (2013) Role of relative size of asperities and adhering particles on the adhesion force. J Colloid Interface Sci 409:211–218
30. Laitinen O, Bauer K, Niinimäki J, Peuker UA (2013) Validity of the Rumpf and the Rabinovich adhesion force models for alumina substrates with nanoscale roughness. Powder Technol 246:545–552
31. Zou Y, Jayasuriya S, Manke CW, Mao G (2015) Influence of Nanoscale Surface Roughness on Colloidal Force Measurements. Langmuir 31:10341–10350
32. Ramakrishna SN, Clasohm LY, Rao A, Spencer ND (2011) Controlling adhesion force by means of nanoscale surface roughness. Langmuir 27(16):9972–9978
33. Gujrati A, Khanal SR, Pastewka L, Jacobs TDB (2018) Combining TEM, AFM, and Profilometry for Quantitative Topography Characterization Across All Scales. ACS Appl Mater Interfaces 10(34):29169–29178
34. Müser MH et al (2017) Meeting the Contact-Mechanics Challenge. Tribol Lett 65(118)
35. Pastewka L, Robbins MO (2016) Contact area of rough spheres: Large scale simulations and simple scaling laws. Appl Phys Lett 108(221601)
36. Byrd RH, Lu P, Nocedal J, Zhu C (1995) A Limited Memory Algorithm for Bound Constrained Optimization. SIAM J Sci Comput 16(5):1190–1208
37. Almqvist A, Sahlin F, Larsson R, Glavatskih S (2007) On the dry elasto-plastic contact of nominally flat surfaces. Tribol Int 40(4):574–579
38. Pérez-Ràfols F, Larsson R, Almqvist A (2016) Modelling of leakage on metal-to-metal seals. Tribol Int 94:421–427
39. Weber B, Suhina T, Junge T, Pastewka L, Brouwer AM, Bonn D (2018) Molecular probes reveal deviations from Amontons' law in multi-asperity frictional contacts. Nat Commun 9(1):1–7
40. Salazar-Banda GR, Felicetti MA, Gonçalves JAS, Coury JR, Aguiar ML (2007) Determination of the adhesion force between particles and a flat surface, using the centrifuge technique. Powder Technol 173(2):107–117
41. Lam KK, Newton JM (1991) Investigation of applied compression on the adhesion of powders to a substrate surface. Powder Technol 65(1–3):167–175
42. Chen Y, Busscher HJ, van der Mei HC, Norde W (2011) Statistical analysis of long- and short-range forces involved in bacterial adhesion to substratum surfaces as measured using atomic force microscopy. Appl Environ Microbiol 77(15):5065–5070
43. Sztilkovics M et al (2020) Single-cell adhesion force kinetics of cell populations from combined label-free optical biosensor and robotic fluidic force microscopy. Sci Rep 10(1):1–13

44. Tolias P et al (2018) Adhesive force distributions for tungsten dust deposited on bulk tungsten and beryllium-coated tungsten surfaces. Nucl Mater Energy 15:55–63
45. Tormoen GW, Drelich J, Nalaskowski J (2005) A distribution of AFM pull-off forces for glass microspheres on a symmetrically structured rough surface. J Adhes Sci Technol 19:215–234
46. Götzinger M, Peukert W (2004) Particle Adhesion Force Distributions on Rough Surfaces. Langmuir 20(20):5298–5303
47. Jacobs TDB, Mate CM, Turner KT, Carpick RW (2013) Understanding the Tip–Sample Contact: An Overview of Contact Mechanics from the Macro- to the Nanoscale, in (Dalia G. Yablon, ed.) Scanning Probe Microscopy in Industrial Applications: Nanomechanical Characterization. John Wiley & Sons, Inc.
48. Grierson DS, Flater EE, Carpick RW (2005) Accounting for the JKR-DMT transition in adhesion and friction measurements with atomic force microscopy. J Adhes Sci Technol 19(3–5):291–311
49. Krell A, Schädlich S (2001) Nanoindentation hardness of submicrometer alumina ceramics. Mater Sci Eng A 307(1–2):172–181
50. Krell A, Blank P (1995) Grain size dependence of hardness in dense submicronmeter alumina. J Am Ceram Soc 78:1118–1120
51. Bu YQ, Wang P, Nie AM, Wang HT (2020) Room-temperature plasticity in diamond. Sci China Technol Sci 64:32–36
52. Issa I et al (2018) Room temperature plasticity and phase transformation of nanometer-sized transition alumina nanoparticles under pressure. Acta Mater 150:308–316
53. Mastrangelo CH, Hsu CH (1992) Simple experimental technique for the measurement of the work of adhesion of microstructures. Tech Dig IEEE Solid-State Sens Actuator Work 208–212
54. Sumant AV, Grierson DS, Gerbi JE, Carlisle JA, Auciello O, Carpick RW (2007) Surface chemistry and bonding configuration of ultrananocrystalline diamond surfaces and their effects on nanotribological properties. Phys Rev B - Condens Matter Mater Phys 76(23):1–11
55. Atkins P, de Paula J (2006) Physical Chemistry for the Life Sciences, Oxford University Press
56. Grierson DS, Liu J, Carpick RW, Turner KT (2013) Adhesion of nanoscale asperities with power-law profiles. J Mech Phys Solids 61(2):597–610
57. Maugis D (1992) Adhesion of spheres: The JKR-DMT transition using a dugdale model. J Colloid Interface Sci 150(1):243–269
58. Wang M, Liechti KM, White JM, Winter RM (2004) Nanoindentation of polymeric thin films with an interfacial force microscope. J Mech Phys Solids 52(10):2329–2354
59. Kovalev AE, Gorb SN (2012) Charge contribution to the adhesion performance of polymeric microstructures. Tribol Lett 48(1):103–109
60. Feshanjerdi M, Malekan A (2019) Contact electrification between randomly rough surfaces with identical materials. J Appl Phys 125(165302)
61. Tang T, Hui CY, Jagota A (2006) Adhesive contact driven by electrostatic forces. J Appl Phys 99(054906)
62. de Boer MP, de Boer PC (2007) Thermodynamics of capillary adhesion between rough surfaces. J Colloid Interface Sci 311(1):171–185
63. DelRio FW, Dunn ML, de Boer MP (2008) Capillary adhesion model for contacting micromachined surfaces. Scr Mater 59(9):916–920
64. Sedighi M, Svetovoy VB, Palasantzas G (2016) Capillary-force measurement on SiC surfaces. Phys Rev E 93(062803)
65. Muller VM, Aleinikova IN, Shcherbina GI, Toporov YP, Derjaguin BV (1994) The influence of contact electrification on the adhesion of dielectric elastic spheres subjected to external loads before detachment. Prog Surf Sci 45(1–4):199–222
66. Wang Y et al (2015) The Instability of Angstrom-Scale Head-Disk Interface Induced by Electrostatic Force. IEEE Trans Magn 51(11):1–4
67. Burgo TAL, Silva CA, Balestrin LBS, Galembeck F (2013) Friction coefficient dependence on electrostatic tribocharging. Sci Rep 3:2384
68. Waitukaitis SR, Lee V, Pierson JM, Forman SL, Jaeger HM (2014) Size-dependent same-material tribocharging in insulating grains. Phys Rev Lett 112(21):1–5
69. Lacks DJ, Sankaran RM (2011)Contact electrification of insulating materials. J Phys D: Appl Phys 44 (453001)
70. He M, Szuchmacher Blum A, Aston DE, Buenviaje C, Overney RM, Luginbühl R (2001) Critical phenomena of water bridges in nanoasperity contacts. J Chem Phys 114(3):1355–1360
71. Bazrafshan M, de Rooij MB, Schipper DJ (2018) Adhesive force model at a rough interface in the presence of thin water films: The role of relative humidity. Int J Mech Sci 140:471–485
72. Ramisetti SB, Campañá C, Anciaux G, Molinari JF, Müser MH, Robbins MO (2011) The autocorrelation function for island areas on self-affine surfaces. J Phys Condens Matter 23(215004)
73. Jacobs TDB, Junge T, Pastewka L (2017) Quantitative characterization of surface topography using spectral analysis. Surf Topogr Metrol Prop 5(013001)
74. Regan B et al (2020) Plastic Deformation of Single-Crystal Diamond Nanopillars. Adv Mater 32(1906458)
75. Barenblatt GI, Monteiro PJM (2010) Scaling laws in nanomechanics. Phys Mesomech 13(5–6):245–248
76. Johnson KL, Kendall K, Roberts AD (1971) Surface energy and the Contact of Elastic Solids. Proc R Soc Lond A Math Phys Sci 324(1558):201–313
77. Muller VM, Yushenko VS, Derjaguin BV (1980) On the Influence of Molecular Forces on the Deformation of an Elastic Sphere. J Colloid Interface Sci 77(5):157-167
78. Persson BN, Bucher F, Chiaia B (2002) Elastic contact between randomly rough surfaces: Comparison of theory with numerical results. Phys Rev B 65(18):184106

Publication VI

Tribology Letters (2021) 69:80
https://doi.org/10.1007/s11249-021-01454-6

ORIGINAL PAPER

Distribution of Gaps and Adhesive Interaction Between Contacting Rough Surfaces

Joseph M. Monti[1,2] · **Antoine Sanner**[3,4] · **Lars Pastewka**[3,4]

Received: 30 January 2021 / Accepted: 5 May 2021 / Published online: 21 May 2021

Abstract
Understanding the distribution of interfacial separations between contacting rough surfaces is integral for providing quantitative estimates for adhesive forces between them. Assuming non-adhesive, frictionless contact of self-affine surfaces, we derive the distribution of separations between surfaces near the contact edge. The distribution exhibits a power-law divergence for small gaps, and we use numerical simulations with fine resolution to confirm the scaling. The characteristic length scale over which the power-law regime persists is given by the product of the rms surface slope and the mean diameter of contacting regions. We show that these results remain valid for weakly adhesive contacts and connect these observations to recent theories for adhesion between rough surfaces.

Keywords Surface roughness · Contact mechanics · Adhesion

1 Introduction

Contact between nominally flat, rough surfaces has been the subject of study for countless experimental, analytical, and numerical investigations over the past century [1–31]. A universal theme found in most cases is that contact is limited to the peaks or asperities of the rough topography and the real area of contact A_{rep} is much less than the apparent projected area A_0. Substantial progress has been made in determining the relationship between A_{rep} and the applied normal force F in non-adhesive, frictionless systems assuming elastic response. In such cases, the proportionality $A_{\text{rep}} = \kappa_{\text{rep}} F/h_0' E^*$ is found when $A_{\text{rep}} \lesssim 0.1 A_0$. Here, h_0' is the root mean square (rms) slope of the rough topography, E^* is the elastic contact modulus, and the dimensionless constant $\kappa_{\text{rep}} \approx 2$ [2, 6, 7]. This result implies a load-independent mean compressive stress, $\sigma_{\text{rep}} = h_0' E^*/\kappa_{\text{rep}}$ in contacting regions. Since flattening a region with slope h_0' introduces a strain of order h_0' into the surface, σ_{rep} reflects the stress required to flatten the rough topography.

Recently, interest has turned to systems including attractive interactions that lead to macroscopic adhesion. Two opposite limits exist in the classical literature on the adhesion of smooth spheres: in the Derjaguin–Muller–Toporov (DMT) [32] limit, weak attractive forces between solids do not alter the geometry of contact, but do reduce the global mean pressure; in the opposite limit, known as the Johnson–Kendall–Roberts (JKR) [33] limit, strong attractive forces significantly change the structure of the contact edge and can lead to contact hysteresis. For the contact of spheres, the Tabor and Maugis parameters [34] describe the continuous transition between the two limits in terms of the relative strengths of adhesive and elastic parameters.

Working in a DMT-like limit, Pastewka and Robbins developed a theory to predict the onset of stickiness in contacts of self-affine rough surfaces [23] that was later independently confirmed by Müser [35]. They split the total normal force as $F = F_{\text{rep}} + F_{\text{att}}$ into the sum of a repulsive contribution

✉ Lars Pastewka
lars.pastewka@imtek.uni-freiburg.de

Joseph M. Monti
jmonti3@jhu.edu

1 Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA

2 Present Address: Sandia National Laboratories, Albuquerque, USA

3 Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany

4 Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany

$F_{rep} > 0$ and an attractive contribution $F_{att} < 0$. Repulsive and attractive contributions originate from repulsive surface patches of total area A_{rep} and attractive surface patches of area A_{att} (see Fig. 1). The total force is then given by

$$F = \sigma_{rep} A_{rep} - \sigma_{att} A_{att}, \tag{1}$$

where σ_{att} is the mean stress in the attractive patches that is roughly constant and of order $\sigma_{att} = w/\Delta r$, where w is the work of adhesion and Δr the range of the attractive interaction.

Like in the purely non-adhesive limit, the geometry of contact is fractal in the DMT-like limit with proportionality between A_{rep} and the contact perimeter P_{rep} given by

$$P_{rep} = \pi A_{rep}/d_{rep}, \tag{2}$$

where the mean contact diameter d_{rep} is approximately constant [7, 23, 36]. Note that Eq. (2) holds generally for any geometric object, but d_{rep} varies with A_{rep} in most cases. Short-ranged attractive interactions generate narrow bands of approximately constant width d_{att} located around contact regions (see Fig. 1). If d_{att} is small, then the total area contributing to attractive forces $A_{att} = P_{rep} d_{att}$, which can be related to A_{rep} *via* mutual proportionality with P_{rep} given by Eq. (2). This means, that for non-sticky interfaces, the (repulsive) contact area is given by the expression

$$A_{rep} = \kappa \frac{F}{h_0' E^*} \tag{3}$$

with an effective $1/\kappa = 1/\kappa_{rep} - 1/\kappa_{att}$. The adhesive interaction hence increases the effective value of the dimensionless constant κ. A macroscopic force is required to separate the two surface when $|F_{att}| \approx F_{rep}$ or equivalently $\kappa_{att} \approx \kappa_{rep}$. Interfaces that require a macroscopic force for separation are called "sticky".

This theory depends sensitively upon the distribution of interfacial separations or gaps, that is assumed to be unaltered from the non-adhesive scenario. Previous work has primarily focused on the behavior of the mean gap $\bar{g}$, which is commonly found to be exponentially related to the normal load in non-adhesive contact as $F \propto \exp\left(-\bar{g}/\gamma h_0\right)$, where h_0 is the rms surface height and γ is a dimensionless constant of order unity [8, 9, 12–17, 21, 24, 29]. Almqvist et al. [16], the contact mechanics challenge [29] and Wang and Müser [37] have reported distributions of interfacial separations, but these works have either not focused on the behavior at small gaps that is important for understanding short-ranged adhesion or indirectly reported it through analysis of percolation in Reynolds flow.

In this paper, we derive the distribution of interfacial separations in the vicinity of contacting regions and show numerically that our expression holds even in the weakly adhesive limit. This distribution can be used to compute the total attractive contribution to the force and hence the force–area relationship for weakly adhesive interfaces.

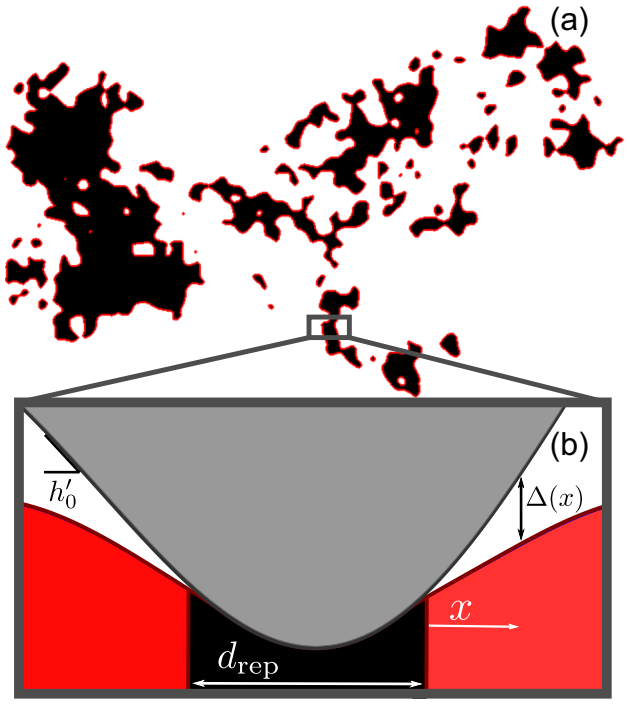


Fig. 1 **a** Contact map for a self-affine surface with $H = 0.8$ and $\lambda_{min} = 16a_0$. A_{rep} is the sum over all pixels shown in black, and the contact perimeter P_{rep} is marked in red. **b** Schematic of a contact region created by a contacting asperity, showing the mean contact diameter d_{rep}. The gap $\Delta(x)$ between surfaces grows as the lateral distance $x^{3/2}$ (Color figure online)

2 Simulation Methods

In our simulations, we invoke the standard mapping that allows the contact of two rough, elastic solids to be treated as contact between an initially flat, elastic solid and a rough, rigid surface [38]. If the elastic properties of the two original surfaces are encoded by the Young's moduli E_1 and E_2 and Poisson's ratios ν_1 and ν_2, the combined elastic response is given by the elastic contact modulus

$$\frac{1}{E^*} = \frac{1-\nu_1^2}{E_1} + \frac{1-\nu_2^2}{E_2}. \tag{4}$$

The roughness profile of each periodic, $L \times L$ surface with nominal area $A_0 = L^2$ is described by a self-affine fractal between an upper cutoff length scale λ_{max} and a lower cutoff λ_{min}, where length scales are given in terms of the pixel size a_0. This means that the power spectral density (PSD)

$C(\mathbf{q})$ of the isotropic, self-affine roughness depends only on wavevector magnitude $q = |\mathbf{q}|$ and satisfies

$$C(q) = \begin{cases} C_0 & \text{if } q \le q_1 \\ C_0 \left(\frac{q}{q_1}\right)^{-2(1+H)} & \text{if } q_1 < q \le q_2 \\ 0 & \text{if } q > q_2 \end{cases}, \tag{5}$$

where $q_1 = 2\pi/\lambda_{\max}$ and $q_2 = 2\pi/\lambda_{\min}$ are the wavevector magnitudes corresponding to the roughness cutoffs and H is the Hurst exponent ($0 < H < 1$) that determines correlations in the roughness. The resolution of the calculations is given by $\lambda_{\min}/a_0$, while the ratio $L/\lambda_{\max}$ controls the surface "representativity" [18]. The distribution of heights $P(h)$ for a Gaussian self-affine rough surface with mean height $\bar{h}$ and $L/\lambda_{\max} \gg 1$ has (by construction) Gaussian form. We use $\lambda_{\min} \ge 32a_0$ for the results presented here to ensure that the contact edge is sufficiently resolved. We use a Fourier-filtering algorithm to create such self-affine surfaces for our calculations [36, 39].

Performing the mapping described above results in a rough surface that is the incoherent sum of the rough profiles of the original surfaces. For the combination of two profiles that have identical statistical properties, the commonly utilized statistical measures of roughness—the rms height $h_0 = \sqrt{\langle h^2 \rangle}$, rms slope $h_0' = \sqrt{\langle |\nabla h|^2 \rangle}$, and rms curvature $h_0'' = \sqrt{\langle |\nabla^2 h|^2 \rangle}/2$—of the combined surface each increase by a factor of $\sqrt{2}$. We therefore here work in the limit of a rigid rough surface contacting a flat deformable elastic half-space and all statistical properties are to be interpreted for a combined surface. Note that $\langle \cdot \rangle$ denotes the spatial average over the domain where the topography function $h(x, y)$ is defined.

We use a static boundary element method to compute the linear elastic deformation induced by normal contact on an isotropic half-space (Refs. [40–42]). In order for linear elasticity to be a good approximation, the surface slope must be small. We use $h_0' = 0.1$ to ensure that this approximation is justified, but note that real surfaces may have values $h_0' \approx 1$ or larger [31, 43, 44].

Our simulations assume frictionless contact with a non-interpenetration constraint. Pixels on the surface of the elastic solid are considered to be in contact when they bear a compressive pressure; each contacting pixel contributes an area a_0^2 to the total contact area. For non-adhesive calculations, we solve the interpenetration constraint using a constrained conjugate-gradient optimizer [45]. For adhesive calculations, we assume an interaction energy of

$$v(g) = -w \exp(-g/\rho), \tag{6}$$

that depends on the gap g, where ρ is the interaction range. (Note that the range defined in Ref. [23] is $\Delta r \approx 1.36\rho$ for this potential.) The overall adhesive energy is then given by

$$E_{\text{att}} = \int dxdy\, v(g(x, y)), \tag{7}$$

where $g(x, y)$ is the local gap at position x, y. This attractive interaction is minimized with an interpenetration constraint realized through the constrained non-linear conjugate-gradient algorithm of Ref. [46]. Similar boundary element methods have been used extensively to study rough contacts [11, 18–23, 25–28, 30]. The attractive interaction is identical to the one employed in the contact mechanics challenge [29].

3 Theory

The distribution of interfacial separations g between a rough surface and an undeformed elastic solid with surface at $h = 0$ is equal to the (Gaussian) distribution of heights,

$$p(g) = \frac{1}{\sqrt{2\pi}h_0} \exp\left[-\frac{(g - \bar{g})^2}{2h_0^2}\right] \quad \text{for} \quad g \ge 0, \tag{8}$$

where $\bar{g}$ is the initial mean surface height. The width of the distribution is the same as for the rough surface itself and scales with the long wavelength cutoff, $h_0 \sim \lambda_{\max}^H$.

When the solids are pushed together under load, the elastic solid deforms and the mean interfacial separation $\bar{g}$ shrinks. We now write the gap distribution as the additive decomposition

$$p(g) = p_c(g) + p_n(g) + p_f(g), \tag{9}$$

where $p_c(g)$ contains the distribution of the gaps g within the contacting area, $p_n(g)$ the contribution from near the contact edge, and $p_f(g)$ the contribution from farther distances from the contact edge. Since $p(g)$ is normalized, the probability of contact for a given contact fraction $c = A_{\text{rep}}/A_0$ is

$$p_c(g) = c\delta(g), \tag{10}$$

where δ is the Dirac δ-function.

The next contribution $p_n(g)$ comes from small separations near the contact edge. As shown in numerical calculations in Refs. [7, 23], the contact edge has a total perimeter $P_{\text{rep}} = \pi A_{\text{rep}}/d_{\text{rep}}$ with a constant d_{rep}. (This means that both area and contact edge are fractal objects—see Fig. 1a—and suggests that their fractal dimensions are identical.) We can write this contribution as

$$p_n(g) = \frac{P_{\text{rep}}}{A_0} \int_0^{\xi} dx\, \delta(g - \Delta(x)) \tag{11}$$

where $\Delta(x)$ describes how the mean gap between the two contacting surfaces varies as a function of distance x from the contact edge (see Fig. 1b). Note that we take the integral in Eq. (11) out to a characteristic length ξ that is close

enough to the contact edge such that the number of points contributing to the gap distribution is still proportional to P_{rep}.

The arguments that lead to Eq. (11) rely on the geometry of the contact patches formed between contacting self-affine surfaces. Refs. [7, 23] showed that in this case $P_{\text{rep}} \propto A_{\text{rep}}$, compared to $P_{\text{rep}} \propto A_{\text{rep}}^{1/2}$ as expected for simple contact shapes like circles. The larger perimeter scaling exponent arises because fractal contact regions are not compact, and a significant perimeter contribution comes from regions inside the convex hull enclosing individual patches. As patches become larger, they simply contain more non-contacting regions. The presence of non-contacting regions makes the contact geometry appear locally rectangular. The deformation cross-section for each line segment drawn through a single continuous contact diameter looks like that of a cylindrical indenter, rather than a spherical one. Within Euclidean geometry, a rectangle with constant thickness along its minor axis has the property $P_{\text{rep}} \propto A_{\text{rep}}$, if the rectangle is thin enough such that the contributions to P_{rep} from the shorter sides can be neglected. The characteristic width of the contacting rectangle, and hence the contact diameter for the cylinder, is d_{rep}.

Working with this analogy, we now derive an analytic prediction for the distribution of gaps produced by non-adhesive contact of smooth surfaces. Assuming a non-adhesive cylindrical contact [47, 48], we have

$$\Delta(x) = \frac{4h_0' d_{\text{rep}}}{3}\left(\frac{x}{d_{\text{rep}}}\right)^{3/2} = g_0\left(\frac{x}{d_{\text{rep}}}\right)^{3/2}, \tag{12}$$

where h_0' is the local slope of the cylinder at the contact edge. It can be shown generally for non-adhesive contact that the local separation $\Delta(x) \propto x^{3/2}$ for small lateral distances x from the contact edge [38]. The characteristic scale for the interfacial separation at the contact edge is $g_0 = 4h_0' d_{\text{rep}}/3$, the prefactor in Eq. (12). Inserting Eqs. (12) into (11) yields

$$p_n(g) = \frac{2P_{\text{rep}} d_{\text{rep}}}{3A_0 g_0}\left(\frac{g_0}{g}\right)^{1/3} = \frac{2\pi c}{3g_0}\left(\frac{g_0}{g}\right)^{1/3}, \tag{13}$$

where Eq. (2) was used for the right-hand equality. This is our prediction for the gap distribution near the contact edge. For small g, the distribution diverges as $g^{-1/3}$.

We note that $\int_0^\infty dg\, p(g) = 1$, but taking the integral out to infinity for the contribution to $p(g)$ given by Eq. (13) diverges. The contribution $p_n(g)$ to the gap distribution can therefore only be valid up to $g \sim g_0$, the characteristic gap that is reached a distance d_{rep} from the contact edge. The "far" contribution $p_f(g)$ looks like the undeformed distribution of heights given by Eq. (8) (see also Refs. [16, 29]).

Almqvist et al. [16] have observed a divergence of the gap distribution for small g but have not quantified the exponent. Pastewka and Robbins [23] have used this divergence for their theory of "stickiness" but have not provided extensive numerical evidence for its validity. We now supplement these observations with additional high-resolution numerical data.

4 Results

We performed simulations with sufficiently fine resolution ($\lambda_{\text{min}} = 32a_0$ and $128a_0$) to test Eq. (13). Figure 2 shows distributions of interfacial separations for non-contacting grid points. The distributions were normalized to the respective non-contacting fractional area, $1 - c$, and divided by the prefactor of Eq. (13) to collapse all data points onto a single $(g/g_0)^{-1/3}$ power-law. We used the numerically measured value of d_{rep} (see Ref. [23] on details of how to compute d_{rep}) for the data collapse. The power-law regime emerges for both $H = 0.3$ (Fig. 2a) and $H = 0.8$ (Fig. 2b).

The data only collapse over a limited range of gaps. The divergence at small gap predicted by Eq. (13) is always cut off at a minimum length scale, below which $p(g)$ is uniformly distributed. In our simulations, the cutoff scale is $\sim 10^{-3} - 10^{-2}a_0$; the threshold for saturation of $p(g)$ at small g/g_0 is inversely proportional to the resolution a_0/λ_{min}. For $H = 0.3$ (Fig. 2a), contributions to $p(g)$ from near-contact and far-from-contact overlap (i.e., $p_n(g)$ and $p_f(g)$ have similar magnitude) for $\lambda_{\text{min}} = 32a_0$ (solid symbols). The power-law regime only clearly emerges over 1–2 decades for $H = 0.3$ if the resolution of the calculation is improved by increasing λ_{min}, as is shown for $\lambda_{\text{min}} = 128a_0$ (open symbols). For $H = 0.8$, on the other hand, the power-law regime extends over a much larger range of gaps even for our "coarse" calculations with $\lambda_{\text{min}} = 32a_0$. As for $H = 0.3$, increasing the short-wavelength cutoff to $\lambda_{\text{min}} = 128a_0$ extends the power-law to smaller gaps.

One can integrate $p_n(g)$ to calculate the cumulative fractional area $c_n(g_c)$ in the near-contacting region closer than a cutoff gap g_c. We find

$$c_n(g_c) = \int_0^{g_c} dg\, p_n(g) = \pi c\left(\frac{g_c}{g_0}\right)^{2/3}, \tag{14}$$

which for $g_c = g_0$ yields $c_n/c = \pi$. This means the area in the near-contact region is ~ 3 times the area in the contacting regions. The area in the near-contact region is equal to the contact area when $g_c \approx 0.18g_0$. The breakdown of the power-law region at small g occurs at $g/g_0 \approx 10^{-3}$ or smaller (for $H = 0.8$), meaning that the area within the roll-off region at small g is at most about 3% of the contact area.

For large gaps the power-law regime is cut off by a Gaussian gap distribution that reflects the distribution of undeformed or weakly deformed parts of the surface (see also

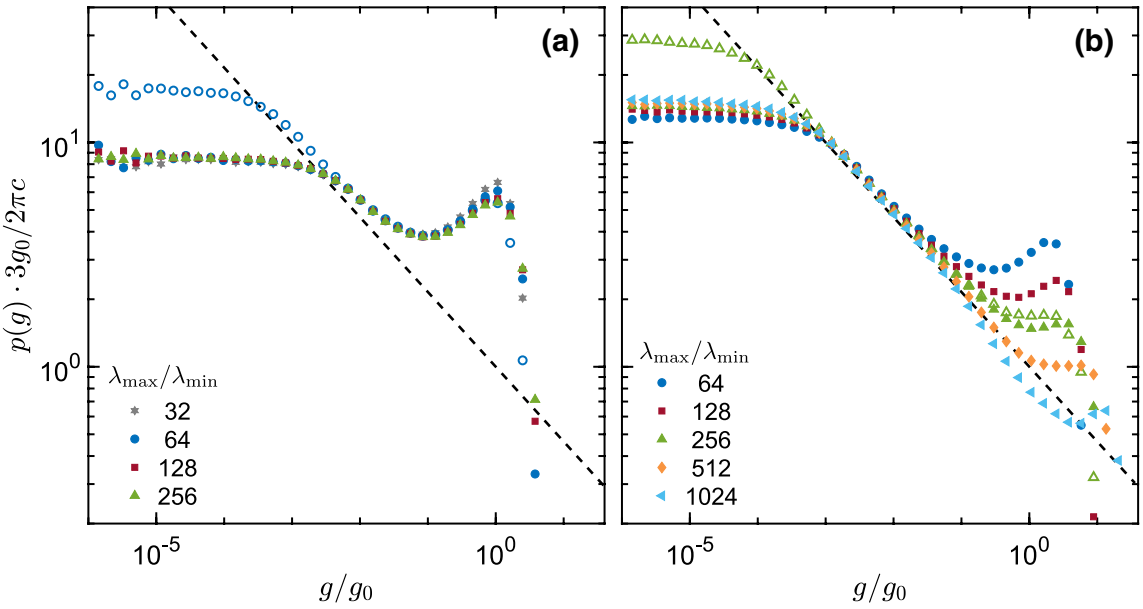


Fig. 2 The probability distribution of interfacial separations for $H = 0.3$ (**a**) and $H = 0.8$ (**b**) normalized to $1 - c$ and divided by the prefactor in Eq. (13) for the ratios $\lambda_{max}/\lambda_{min}$ indicated in the legend, with $\lambda_{min} = 32a_0$ (solid symbols) and $\lambda_{min} = 128a_0$ (open symbols, color matches $\lambda_{max}/\lambda_{min}$ in the legend). Here, $A/A_0 \approx 0.03$ for $\lambda_{min} = 32a_0$ and $A/A_0 \approx 0.04$ for $\lambda_{min} = 128a_0$. The power-law $p_n(g)$ from Eq. (13) is shown as a dashed black line

Eq. (8)). For both $H = 0.3$ and $H = 0.8$, the uptick in $p(g)$ in the range $g/g_0 \sim 1 - 10$ is the peak of this Gaussian height distribution. Since $h_0 \propto (\lambda_{max}/\lambda_{min})^H$, the power-law regime extends much further for $H = 0.8$ (Fig. 2b) than for $H = 0.3$ (Fig. 2a). Increasing $\lambda_{max}/\lambda_{min}$ shifts the Gaussian peak out to larger g/g_0 (most prominently for $H > 0.5$), and extends the range over which $p(g) \approx p_n(g)$. However, for the largest $\lambda_{max}/\lambda_{min}$ we studied, $p(g) < p_n(g)$ for $0.1 \lesssim g/g_0 \lesssim 1$. This suggests that the cutoff of the integral in Eq. (11) may be up to an order of magnitude smaller than g_0, about equal to the point where the near-contact area matches the contact area. Still larger calculations may be required to verify this result.

The crucial question for adhesive theories is how applicable our results are for gap distributions in the presence of attractive interactions. We quantify the strength of the attractive interaction by the value of $1/\kappa_{att}$ (see Ref. [23] and discussion below), since interfaces become sticky for $1/\kappa_{att} \gtrsim 1/2$ [23]. For the data collapse, we use the values of d_{rep} measured in the non-adhesive calculations.

Figure 3 shows the gap distributions for adhesive calculations as w increases (with constant ρ), using surfaces with $H = 0.8$ and $\lambda_{min} = 128a_0$. To facilitate the comparison between non-adhesive and adhesive calculations, the adhesive simulations are conducted with the constraint that the mean gap is identical to the non-adhesive simulation ($1/\kappa_{att} = 0$). This choice of constraint means that the contact areas are not equal, particularly for sticky surfaces. For non-sticky surfaces (up to $1/\kappa_{att} \approx 0.2$), the gap distribution

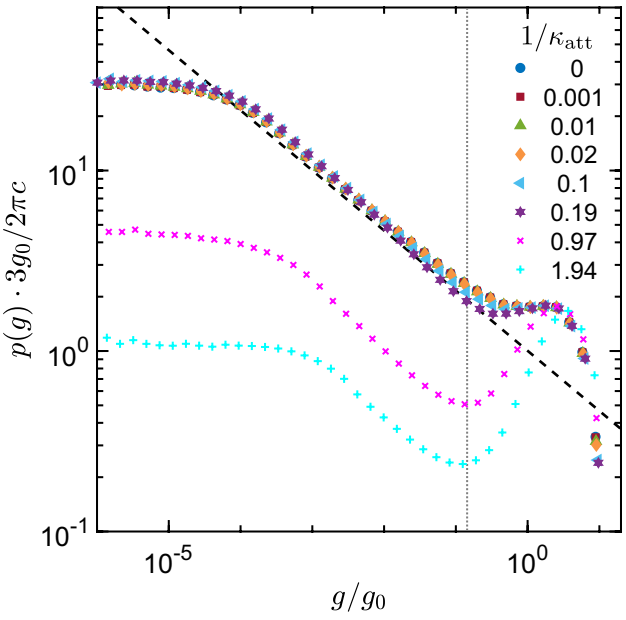


Fig. 3 Comparison of the interfacial probability distributions for adhesive contact with increasing w, for $H = 0.8$, $\lambda_{min} = 128a_0$, $\lambda_{max}/\lambda_{min} = 256$, and $L/\lambda_{max} = 2$. The corresponding non-adhesive contact distribution from Fig. 2b is replotted ($1/\kappa_{att} = 0$, $A_{rep}/A_0 \approx 0.04$), and the adhesive distributions are normalized by g_0 and c obtained from the non-adhesive result. The power-law $p_n(g)$ from Eq. (13) is shown as a dashed black line. The dotted gray line corresponds to $g/g_0 = \rho/g_0$ with $\rho = 4a_0$.

follows the predicted power-law over the same range as the non-adhesive result. At $1/\kappa_{\text{att}} \approx 0.2$ ($A_{\text{rep}}/A_0 \approx 0.06$) there is a slight deviation toward smaller gaps. The sticky cases with $1/\kappa_{\text{att}} \approx 1$ and $1/\kappa_{\text{att}} \approx 2$ ($A_{\text{rep}}/A_0 \approx 0.12$ and 0.16, respectively) clearly deviate from our prediction for the divergence. These sticky interfaces appear to still exhibit a regime where $p(g) \propto g^{-1/3}$, but the range of the power-law regime becomes narrower as $1/\kappa_{\text{att}}$ increases. This means that the prefactor from Eq. (13) no longer captures the intensity of the divergence and much of the non-contacting area is pushed out toward larger gaps. This is also reflected by the increase of the peak at larger gaps. The characteristic gap below which the distribution rolls-off and becomes constant appears to increase slightly in the sticky limit.

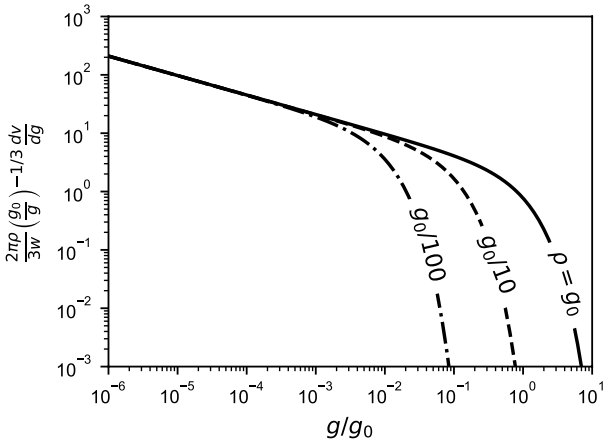


Fig. 4 Integrand of Eq. (16), i.e., non-dimensionalized contribution to the overall attractive force as a function of gap g/g_0 for different ranges ρ of the adhesive interaction

5 Discussion and Conclusions

The behavior of the near-contact interfacial separation is an important consideration in the context of adhesive contact because it determines the attractive contribution to the force,

$$\frac{F_{\text{att}}}{A_0} = \int_0^\infty dg\, p(g) \frac{dv}{dg}. \tag{15}$$

Assuming that the contribution from $p_f(g)$ is negligible, i.e., that our potential is sufficiently short-ranged, the attractive pressure per unit contact area is then given by

$$\frac{F_{\text{att}}}{A_{\text{rep}}} = \frac{2\pi}{3} \int_0^\infty \frac{dg}{g_0} \left(\frac{g_0}{g}\right)^{1/3} \frac{dv}{dg} = -\frac{2\pi}{3} \frac{w}{(g_0^2\rho)^{1/3}}, \tag{16}$$

where we have used the interaction law $v(g)$ from Eq. (6). Note that Eq. (16) defines the value of the dimensionless constant $1/\kappa_{\text{att}} = F_{\text{att}}/h_0' E^* A_{\text{rep}}$, which we used to quantify the strength of adhesion in Fig. 3 and which can be used to determine the effective range of adhesion Δr (see Ref. [23]).

This expression can also be used to quantify what "short-ranged" adhesion means for rough surfaces and, therefore, to determine the limits of the theories of Pastewka and Robbins [23] and Müser [35]. Figure 4 shows the normalized integrand of Eq. (16) as a function of the normalized gap g/g_0. The integrand depends on the range of the interaction potential ρ. For $\rho \approx 0.1 g_0$ and below, the main contribution to the integrand and thereby the attractive force comes from gaps with $g/g_0 < 0.1$ where the power-law holds. The roll-off region at small gap contributes negligibly to this integral. This means that the adhesion range is short enough if $\rho \lesssim 0.1 g_0$. The calculations presented here were carried out with $\rho = 4a_0$ (as is typical for attractive interactions, e.g., Refs. [49, 50]) and our adhesive calculations have $g_0 > 25a_0$, which means the range is sufficiently small. Interactions with a larger range (e.g., electrostatic interactions) interact with the full topography of the surface and require corrections to the expression for F_{att}. The same is true for extremely smooth surfaces with small values of g_0. This defines bounds for theory outlined in Refs. [23, 35].

Besides leaving the gap distribution unmodified, the DMT-like limit also implies that the repulsive area A_{rep} equals the typical definition of the contact area, namely, vanishing gaps $g = 0$. The latter definition makes sense in continuum theories (like the present work), but not for models that consider the full intermolecular interaction, which exhibits soft repulsion (e.g., Ref. [23]), or for those that include thermal fluctuations (e.g., Refs. [51, 52]). Like in the JKR model [33], these definitions of contact no longer agree for sticky interfaces. In addition, JKR-like contacts separate as $x^{1/2}$ and not $x^{3/2}$ near the contact edge [53], leading to a gap distribution $p_n(g) \propto g$ that is clearly distinct from what we observe in our calculations. Furthermore, in the sticky limit, the contact geometry changes substantially and is not expected to give rise to the simple proportional contact area–perimeter relationship used in our calculations. This is especially true for soft solids, which can deform to fill in interior non-contacting regions without large elastic energy penalties, thereby increasing the contact area at the expense of the overall contact perimeter.

Nevertheless, the present calculations are a powerful demonstration of the universal emergence of the $g^{-1/3}$ divergence in the distribution of gaps between elastically stiff rough surfaces. Since this behavior is a direct consequence of the fractal character of the contacting interfaces as manifested in the proportionality between perimeter and contact area, our results are another indirect demonstration of this aspect of the contact geometry. The distribution is unaltered by weak adhesive interactions, giving additional support for

the DMT-like approximation that underlies the adhesive theories of Pastewka and Robbins [23] and Müser [35].

Acknowledgements This work has emerged from numerous enlightening interactions with Mark Robbins over the last decade. All of us were inspired by him and JMM and LP will forever be thankful for having been given the opportunity to closely work with Mark. We also thank Martin Müser for useful discussions and Sindhu Singh for implementing the optimization algorithm of Ref. [46].

Author Contributions JMM and LP devised the study. JMM and AS carried out and analyzed contact mechanics calculations. JMM wrote the first manuscript draft. All authors contributed to editing and finalizing the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. We thank the DAAD for support for a short visit of JMM to Freiburg and the US National Science Foundation (Grant DMR-1411144 and DMR-1929467), the European Commission (Marie-Curie IOF-272619), and the Deutsche Forschungsgemeinschaft (Grants PA 2023/2 and EXC 2193/1—390951807) for funding. Computations were carried out on the Johns Hopkins University Homewood High Performance Cluster, the BlueCrab cluster at the Maryland Advanced Research Computing Center, and NEMO at the University of Freiburg (DFG Grant INST 39/963-1 FUGG).

Code Availability We used the CONTACT.ENGINEERING ecosystem for all calculations. The code is available at https://github.com/ContactEngineering under the terms of the MIT license.

Data Availibility Data are available from the authors upon request.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Greenwood, J.A., Williamson, J.B.P.: Contact of nominally flat surfaces. Proc. R. Soc. Lond. A **295**(1442), 300–319 (1966)
2. Bush, A.W., Gibson, R.D., Thomas, T.R.: The elastic contact of a rough surface. Wear **35**(1), 87–111 (1975)
3. Dieterich, J.H., Kilgore, B.D.: Direct observation of frictional contacts: new insights for state-dependent properties. Pure Appl. Geophys. **143**(1–3), 283–302 (1994)
4. Dieterich, J.H., Kilgore, B.D.: Imaging surface contacts: power law contact distributions and contact stresses in quartz, calcite, glass and acrylic plastic. Tectonophysics **256**(1–4), 219–239 (1996)
5. Persson, B.N.J.: Theory of rubber friction and contact mechanics. J. Chem. Phys. **115**(8), 3840–3861 (2001)
6. Persson, B.N.J.: Elastoplastic contact between randomly rough surfaces. Phys. Rev. Lett. **87**(11), 116101-1–116101-4 (2001)
7. Hyun, S., Pel, L., Molinari, J.-F., Robbins, M.O.: Finite-element analysis of contact between elastic self-affine surfaces. Phys. Rev. E **70**(2), 1–12 (2004)
8. Pei, L., Hyun, S., Molinari, J.-F., Robbins, M.O.: Finite element modeling of elasto-plastic contact between rough surfaces. J. Mech. Phys. Solids **53**(11), 2385–2409 (2005)
9. Benz, M., Rosenberg, K.J., Kramer, E.J., Israelachvili, J.N.: The deformation and adhesion of randomly rough and patterned surfaces. J. Phys. Chem. B **110**(24), 11884–11893 (2006)
10. Hyun, S., Robbins, M.O.: Elastic contact between rough surfaces: effect of roughness at large and small wavelengths. Tribol. Int. **40**(10–12), 1413–1422 (2007)
11. Campañá, C., Müser, M.H.: Contact mechanics of real vs. randomly rough surfaces: a Green's function molecular dynamics study. Europhys. Lett. **77**(3), 38005 (2007)
12. Persson, B.N.J.: Relation between interfacial separation and load: a general theory of contact mechanics. Phys. Rev. Lett. **99**(12), 125502 (2007)
13. Yang, C., Persson, B.N.J.: Molecular dynamics study of contact mechanics: contact area and interfacial separation from small to full contact. Phys. Rev. Lett. **100**(2), 024303 (2008)
14. Yang, C., Persson, B.N.J.: Contact mechanics: contact area and interfacial separation from small contact to full contact. J. Phys. **20**(21), 215214 (2008)
15. Lorenz, B., Persson, B.N.J.: Interfacial separation between elastic solids with randomly rough surfaces: comparison of experiment with theory. J. Phys. **21**(1), 015003 (2009)
16. Almqvist, A., Campañá, C., Prodanov, N., Persson, B.N.J.: Interfacial separation between elastic solids with randomly rough surfaces: comparison between theory and numerical techniques. J. Mech. Phys. Solids **59**(11), 2355–2369 (2011)
17. Akarapu, S., Sharp, T., Robbins, M.O.: Stiffness of contacts between rough surfaces. Phys. Rev. Lett. **106**(20), 204301 (2011)
18. Yastrebov, V.A., Anciaux, G., Molinari, J.-F.: Contact between representative rough surfaces. Phys. Rev. E **86**(3), 035601 (2012)
19. Pohrt, R., Popov, V.L.: Normal contact stiffness of elastic solids with fractal rough surfaces. Phys. Rev. Lett. **108**(10), 104301 (2012)
20. Prodanov, N., Dapp, W.B., Müser, M.H.: On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections, and reference data. Tribol. Lett. **53**(2), 433–448 (2013)
21. Pastewka, L., Prodanov, N., Lorenz, B., Müser, M.H., Robbins, M.O., Persson, B.N.J.: Finite-size scaling in the interfacial stiffness of rough elastic contacts. Phys. Rev. E **87**(6), 062809 (2013)
22. Yastrebov, V.A., Anciaux, G., Molinari, J.-F.: The contact of elastic regular wavy surfaces revisited. Tribol. Lett. **56**(1), 171–183 (2014)
23. Pastewka, L., Robbins, M.O.: Contact between rough surfaces and a criterion for macroscopic adhesion. Proc. Natl. Acad. Sci. USA **111**(9), 3298–3303 (2014)
24. Prodanov, N., Dapp, W.B., Müser, M.H.: On the contact area and mean gap of rough, elastic contacts: dimensional analysis, numerical corrections, and reference data. Tribol. Lett. **53**(2), 433–448 (2014)
25. Yastrebov, V.A., Anciaux, G., Molinari, J.-F.: From infinitesimal to full contact between rough surfaces: evolution of the contact area. Int. J. Solids Struct. **52**, 83–102 (2015)

26. Pastewka, L., Robbins, M.O.: Contact area of rough spheres: large scale simulations and simple scaling laws. Appl. Phys. Lett. **108**(22), 221601 (2016)
27. Yastrebov, V.A., Anciaux, G., Molinari, J.-F.: On the accurate computation of the true contact-area in mechanical contact of random rough surfaces. Tribol. Int. **114**, 161–171 (2017)
28. Yastrebov, V.A., Anciaux, G., Molinari, J.-F.: The role of the roughness spectral breadth in elastic contact of rough surfaces. J. Mech. Phys. Solids **107**, 469–493 (2017)
29. Müser, M.H., Dapp, W.B., Bugnicourt, R., Sainsot, P., Lesaffre, N., Lubrecht, T.A., Persson, B.N.J., Harris, K., Bennett, A., Schulze, K., Rohde, S., Ifju, P., Sawyer, W.G., Angelini, T., Ashtari Esfahani, H., Kadkhodaei, M., Akbarzadeh, S., Wu, J.J., Vorlaufer, G., Vernes, A., Solhjoo, S., Vakis, A.I., Jackson, R.L., Xu, Y., Streator, J., Rostami, A., Dini, D., Medina, S., Carbone, G., Bottiglione, F., Afferrante, L., Monti, J., Pastewka, L., Robbins, M.O., Greenwood, J.A.: Meeting the contact-mechanics challenge. Tribol. Lett. **65**(4), 1–18 (2017)
30. Weber, B., Suhina, T., Junge, T., Pastewka, L., Brouwer, A.M., Bonn, D.: Molecular probes reveal deviations from Amontons' law in multi-asperity frictional contacts. Nat. Commun. **9**(1), 888 (2018)
31. Dalvi, S., Gujrati, A., Khanal, S.R., Pastewka, L., Dhinojwala, A., Jacobs, T.D.B.: Linking energy loss in soft adhesion to surface roughness. Proc. Natl. Acad. Sci. USA **116**(51), 25484–25490 (2019)
32. Derjaguin, B.V., Muller, V.M., Toporov, Y.U.P.: Effect of contact deformation on the adhesion of particles. J. Colloid Interface Sci. **52**(3), 105–108 (1975)
33. Johnson, K.L., Kendall, K., Roberts, A.D.: Surface energy and the contact of elastic solids. Proc. R. Soc. Lond. A **324**, 301–313 (1971)
34. Maugis, D.: Adhesion of spheres: the JKR-DMT transition using a Dugdale model. J. Colloid Interface Sci. **150**(1), 243–269 (1992)
35. Müser, M.H.: A dimensionless measure for adhesion and effects of the range of adhesion in contacts of nominally flat surfaces. Tribol. Int. **100**, 41–47 (2016)
36. Ramisetti, S.B., Campañá, C., Anciaux, G., Molinari, J.-F., Müser, M.H., Robbins, M.O.: The autocorrelation function for island areas on self-affine surfaces. J. Phys. **23**(21), 215004 (2011)
37. Wang, A., Müser, M.H.: Percolation and Reynolds flow in elastic contacts of isotropic and anisotropic, randomly rough surfaces. Tribol. Lett. **69**(1), 1 (2020)
38. Johnson, K.L.: Contact Mechanics. Cambridge University Press, Cambridge (1985)
39. Jacobs, T.D.B., Junge, T., Pastewka, L.: Quantitative characterization of surface topography using spectral analysis. Surf. Topogr. **5**(1), 013001 (2017)
40. Stanley, H.M., Kato, T.: An FFT-based method for rough surface contact. J. Tribol. **1**(July), 2–6 (1997)
41. Campañá, C., Müser, M.H.: Practical Green's function approach to the simulation of elastic semi-infinite solids. Phys. Rev. B **74**(7), 75420 (2006)
42. Pastewka, L., Sharp, T.A., Robbins, M.O.: Seamless elastic boundaries for atomistic calculations. Phys. Rev. B **86**, 075459 (2012)
43. Gujrati, A., Khanal, S.R., Pastewka, L., Jacobs, T.D.B.: Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales. ACS Appl. Mater. Interf. **10**(34), 29169–29178 (2018)
44. Gujrati, A., Sanner, A., Khanal, S., Moldovan, N., Zeng, H., Pastewka, L., Jacobs, T.D.B.: Comprehensive topography characterization of polycrystalline diamond coatings. Surf. Topogr. **9**(1), 014003 (2021)
45. Polonsky, I.A., Keer, L.M.: A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear **231**(2), 206–219 (1999)
46. Bugnicourt, R., Sainsot, P., Dureisseix, D., Gauthier, C., Lubrecht, A.A.: FFT-based methods for solving a rough adhesive contact: description and convergence study. Tribol. Lett. **66**(1), 29 (2018)
47. Baney, J.M., Hui, C.-Y.: A cohesive zone model for the adhesion of cylinders. J. Adhes. Sci. Technol. **11**(3), 393–406 (1997)
48. Yang, F., Cheng, Y.-T.: Revisit of the two-dimensional indentation deformation of an elastic half-space. J. Mater. Res. **24**(06), 1976–1982 (2009)
49. Grierson, D.S., Liu, J., Carpick, R.W., Turner, K.T.: Adhesion of nanoscale asperities with power-law profiles. J. Mech. Phys. Solids **61**(2), 597–610 (2013)
50. Thimons, L.A., Gujrati, A., Sanner, A., Pastewka, L., Jacobs, T.D.B.: Hard material adhesion: which scales of roughness matter? Exp. Mech. (2021). https://doi.org/10.1007/s11340-021-00733-6
51. Cheng, S., Robbins, M.O.: Defining contact at the atomic scale. Tribol. Lett. **39**(3), 329–348 (2010)
52. Zhou, Y., Wang, A., Müser, M.H.: How thermal fluctuations affect hard-wall repulsion and thereby hertzian contact mechanics. Front. Mech. Eng. **5**, 67 (2019)
53. Tada, H., Paris, P.C., Irwin, G.R.: The Stress Analysis Of Cracks Handbook, 3rd edn. ASME Press, New York (2000)

Publication VII

Surface Topography: Metrology and Properties

OPEN ACCESS

RECEIVED
24 September 2020

REVISED
16 December 2020

ACCEPTED FOR PUBLICATION
11 February 2021

PUBLISHED
17 March 2021

PAPER

Comprehensive topography characterization of polycrystalline diamond coatings

Abhijeet Gujrati[1], **Antoine Sanner**[2,3], **Subarna R. Khanal**[1], **Nicolaie Moldovan**[4], **Hongjun Zeng**[4,5], **Lars Pastewka**[2,3,*] **and Tevis D. B. Jacobs**[1,*]

[1] Department of Mechanical Engineering and Materials Science, University of Pittsburgh, 3700 O'Hara St, Pittsburgh, PA, 15261, United States of America
[2] Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany
[3] Cluster of Excellence livMatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany
[4] Alcorix Co., 14047 Franklin Ct., Plainfield, IL 60544, United States of America
[5] Aqua via Rock LLC, 712 Muirhead Ct., Naperville, IL 60565, United States of America
[*] Authors to whom any correspondence should be addressed.

E-mail: lars.pastewka@imtek.uni-freiburg.de **and** tjacobs@pitt.edu

Keywords: surface topography, surface roughness, power spectral density (PSD), transmission electron microscopy (TEM), diamond coatings

Supplementary material for this article is available online

Abstract

The surface topography of diamond coatings strongly affects surface properties such as adhesion, friction, wear, and biocompatibility. However, the understanding of multi-scale topography, and its effect on properties, has been hindered by conventional measurement methods, which capture only a single length scale. Here, four different polycrystalline diamond coatings are characterized using transmission electron microscopy to assess the roughness down to the sub-nanometer scale. Then these measurements are combined, using the power spectral density (PSD), with conventional methods (stylus profilometry and atomic force microscopy) to characterize all scales of topography. The results demonstrate the critical importance of measuring topography across all length scales, especially because their PSDs cross over one another, such that a surface that is rougher at a larger scale may be smoother at a smaller scale and vice versa. Furthermore, these measurements reveal the connection between multi-scale topography and grain size, with characteristic scaling behavior at and slightly below the mean grain size, and self-affine fractal-like roughness at other length scales. At small (subgrain) scales, unpolished surfaces exhibit a common form of residual roughness that is self-affine in nature but difficult to detect with conventional methods. This approach of capturing topography from the atomic- to the macro-scale is termed *comprehensive topography characterization*, and all of the topography data from these surfaces has been made available for further analysis by experimentalists and theoreticians. Scientifically, this investigation has identified four characteristic regions of topography scaling in polycrystalline diamond materials.

1. Introduction

Surface topography controls surface properties of carbon coatings. For example, prior measurements of diamond-like carbon (DLC) coatings show that tribological behavior [1] and adhesion [2] are strongly affected by surface texture, all the way down to the nanoscale. The surface topography of diamond coatings [3] affects their performance, including their friction [4], wear [5], adhesion [6, 7], and biocompatibility [8]. Diamond is an important material that is used in many industrial applications [9] such as for mechanical seals [10], MEMS devices [11], biomedical applications [12], seals and bearings [13], and nuclear fusion [14] because it has low friction and wear [15, 16], and because it is robust and chemically inert so that it can be operated in corrosive environments.

Numerical models have been proposed to describe the effect of surface roughness on contact and adhesion [17]. Initially, the classic Greenwood-Williamson [18] and Fuller-Tabor [19] models described contact area and adhesion based on the average height of the

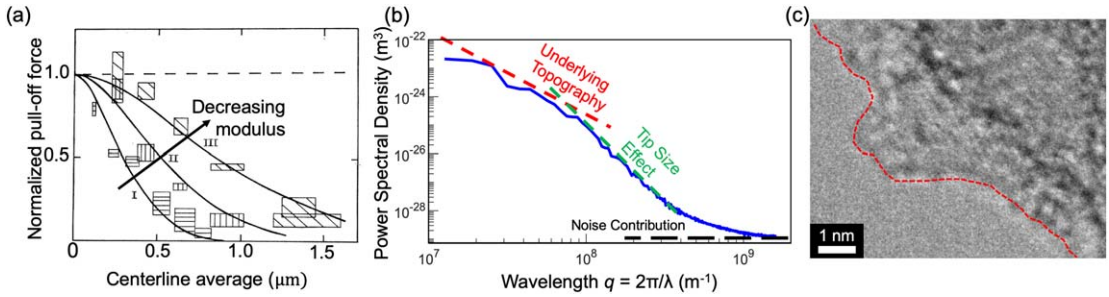


Figure 1. Prior work has been done to study roughness and its effect on contact properties. Fuller and Tabor [19] showed that roughness is related to adhesion (a). However, in practice, artifacts from the imaging technique can strongly influence the measurement of surface roughness, as demonstrated by the power spectral density (see main text) of topography, shown here for an ultrananocrystalline diamond surface (b). Small-scale topography can be more accurately captured using cross-section electron microscopy (c), as shown here for the same material from panel (b). Panel (a) is reproduced from [19], copyright 1975, Royal Society, with addition of arrow and text 'Decreasing modulus'. Panel (b) is reproduced from [31], Copyright 2017, IEEE.

roughness (figure 1(a)). Recently, the multi-scale nature of roughness has been included in the modeling of surface properties [20–29]. In particular, it has been shown that a critical quantity controlling contact area, adhesion, friction, and wear is the root-mean-square slope of the surface h'_{rms} and that this quantity, for a surface with multi-scale roughness, is strongly influenced by small-scale features [22, 30]. Therefore, understanding and prediction of the topography-dependent surface properties of carbon coatings requires multi-scale characterization of roughness, down to the atomic scale.

Because conventional methods for topography measurement cannot capture these small scales, then researchers must choose either to ignore the small-scale topography, or to extrapolate it from the single-scale measurements at larger scales. Because many real-world surfaces exhibit hierarchical topography [32–36], this extrapolation is often done by assuming that the material is self-similar or self-affine. *Self-similarity* implies that the topography is statistically indistinguishable at all magnifications ζ; in other words, if the lateral length scale L is rescaled to ζL, then the measured height h is rescaled to ζh. *Self-affinity* is related, but characterized by the Hurst exponent H; where rescaling the length to ζL and the height to $\zeta^H h$ yields statistically indistinguishable surfaces. Mathematically, the root-mean-square slope h'_{rms} of a self-affine, randomly-rough surface depends strongly on the smallest scales at which the roughness exists [22, 30]. For real surfaces, it is often assumed that surfaces are self-affine down to the atomic scale, but this assumption is largely untested.

Instead of ignoring or extrapolating for small-scale topography, new approaches are required for characterizing topography across all length scales. While stylus profilometry and atomic force microscopy (AFM) are indispensable tools in characterizing surface topography, these tip-based techniques are unable to provide the small-scale topography of rough surfaces because the radius of the scanning tip introduces artifacts [37]. This limits the range of reliability of the roughness measurements [30] (figure 1(b)). Optical techniques, such as scanning white-light interferometry or laser confocal microscopy, suffer from diffraction-limited lateral resolution and optical-transfer-function artifacts, and are thus similarly incapable of measuring the smallest-scale topography [38]. The result of this is that conventional measurements of surface topography are incomplete, and computed surface metrics (such as h'_{rms}) are unreliable, depending explicitly on the lateral resolution of the measurement. Even very advanced methods of *analyzing* surface topography, such as those in [39], are limited in their effectiveness by the range of size-scales in the underlying topography measurement. Instead, cross-section electron microscopy provides a reliable method to characterize surface topography down to the Ångström-scale [40] (figure 1(c)). Further, the small-scale topography can be stitched together with the medium- and large-scale topography using the power spectral density (PSD), to provide a comprehensive statistical description of surface topography at all size scales [32]. The PSD is a mathematical tool which separates the contribution to roughness from different length scales λ, and it is commonly represented as a function of wavevector $q = 2\pi/\lambda$.

This method of combining many different measurements, from the atomic to the macroscale, is termed *comprehensive topography characterization* and was applied here to investigate the surface roughness of four different varieties of diamond coatings, namely ultrananocrystalline diamond (UNCD), polished UNCD (pUNCD), nanocrystalline diamond (NCD), and microcrystalline diamond (MCD).

2. Methods

Thin films of the diamond materials were deposited (Advanced Diamond Technologies, Romeoville, IL) using a tungsten hot-filament chemical vapor

Table 1. Deposition parameters for various forms of polycrystalline diamond.

Diamond type	CH_4/H_2 Ratio	Pressure (Torr)	Filament temperature (°C)	Filament power (KW)	Polished
MCD	1.5%	25	2460	15.4	N
NCD	2.9%	10	2505	15.0	N
UNCD	4.7%	5	2550	15.1	N
pUNCD	4.7%	5	2550	15.1	Y

deposition (HF-CVD) system with parameters as described in [41]. To improve electrical conductivity, all materials were boron doped with a B/C ratio of 3000 ppm. All materials were deposited to a thickness of 2 μm on polished silicon wafers after the wafers were sonicated with slurries containing suspended diamond nanoparticles. The deposition parameters are listed in table 1.

The smallest-scale topography measurements were made using transmission electron microscopy, following the techniques described in [40]. For the UNCD, NCD, and MCD, the 'wedge deposition technique' was used, whereas for pUNCD, the 'surface-preserving cross-section technique' was used. These techniques are described in detail in [40]. Briefly, the wedge deposition technique involves depositing the diamond film directly onto TEM-ready silicon thin-wedge substrates. The surface-preserving cross-section technique utilizes conventional methods for the preparation of TEM cross-sections (sectioning, grinding, polishing, dimple-grinding, and ion etching) with process modifications to ensure that the original surface is preserved. The samples were imaged using a TEM (JEOL JEM 2100F, Tokyo, Japan) operated at 200 keV. The images were taken using magnification levels from 5000× to 600 000×.

The profiles were extracted from the TEM images by using custom Matlab scripts to trace the outermost boundary of the material. Before tracing, the images were rotated such that the boundary was approximately horizontal. The vast majority of the measured surfaces were functions, i.e. for every point on the x-axis, there was exactly one point on the y-axis; in other words, the measured topographies were not reentrant. However, there were some cases where two adjacent points were captured with identical or decreasing horizontal position (locally reentrant). For these cases, the latter point was removed to restore non-reentrant behavior so that the topography can be described by a function as required by the calculation of the PSD. They were only observed in 12 out of 160 TEM profiles, and even then, only in a small number of points per profile, therefore the process of removing such points should not affect the accuracy of the analysis. Further, these locally reentrant points are attributed to imperfect rotation of the TEM profiles, rather than to truly reentrant features on the surface.

The medium-scale topography was measured using an atomic force microscope (Dimension V, Bruker, Billerica, MA) in tapping mode with diamond-like carbon-coated probes (Tap DLC300, Mikromasch, Watsonville, CA). For all substrates, square measurements were taken with the following lateral sizes: 3 scans each at 100 nm, 500 nm, and 5 μm; 1 scan each at 250 nm and 1 μm. The scanning speed was maintained at 1 μm s^{-1} for all scans. Each scan had 512 lines, with 512 data points per line, corresponding to pixel sizes in the range of 0.2 to 98 nm. The wear of the AFM tip was minimized using the best practices described in [42]. Specifically, the values of free-air amplitude and amplitude ratio, which is the ratio of the amplitude of AFM probe tip vibration when performing a scan to the amplitude when vibrating in free air, were kept in the range of 37–49 nm and 0.15–0.3, respectively. Though AFM provides a two-dimensional description of the surface topography, the data were analyzed as a series of line scans. This practice maintained consistency with the other techniques, which yield one-dimensional measurements, and also eliminated artifacts due to instrumental drift in the slow-scan axis.

The largest scales of topography were measured using a stylus profilometer (Alpha Step IQ, KLA Tencor, Milpitas, CA) with a 5-μm diamond tip. Measurements were collected at a scanning speed of 10 μm s^{-1}, with data points every 100 nm. A total of 8 measurements were taken on each substrate, with 2 measurements each at scan sizes of 0.5, 1, 2, and 5 mm. All measurements were corrected using a parabolic fit to remove the tilt of the sample and the bowing artifact from the tool. For the UNCD and pUNCD, the larger scan sizes exhibited consistent non-parabolic trends due to instrument artifacts. These artifacts were corrected by taking reference scans on polished silicon wafers and subtracting the averaged reference profiles from the measurements.

Finally, the PSD was used to combine all measurements from a single surface into one averaged curve that describes the topography of that surface. The PSD is the Fourier transform of the autocorrelation function of a line scan with height $h(x)$, which is mathematically equivalent to the square of the amplitude of $\tilde{h}(q)$; i.e., $C(q) = L^{-1}|\tilde{h}(q)|^2$, q is the wavevector and L is the length of the scan. All data were collected and analyzed as 1D line scans, enabling the calculation of the one-dimensional PSD, denoted here as C (designated C^{1D} in [30, 32]). These calculations follow the standards established in [30] for computing and reporting PSDs.

3. Results

All topography data sets collected from this investigation are freely available for download and analysis [43–46].

3.1. Multi-scale topography measurement

Representative images of all three techniques are shown for the four materials in figure 2. The stylus profilometry data are shown with decreasing scan size in figure 2(a). It is clear that the roughness on the MCD has the largest amplitude (RMS height) while the pUNCD surface shows the smallest. Going from larger scans (figure 2(a_1)) to smaller scans (figure 2 (a_3)), the amplitude of measured topography decreases for all four diamond species. For example, while a 5-mm scan of MCD spans a vertical range of 646 nm, a 0.5-mm scan of the same surface spans just 391 nm. In order to interpret the stylus data correctly, an estimate of the tip radius is needed because the tip introduces artifacts at and below this size scale. Figure 2(a_4) shows a scanning electron microscopy (SEM) image of the stylus tip. Fitting a circle to the tip yields a radius of $R = 5.1$ μm. The exact point where tip artifacts become dominant [37] will be estimated using the PSD in the following section.

The AFM measurements are shown as topography maps of increasing resolution in figures 2(b)–(e). The NCD and MCD clearly show grain structure and faceting at scan sizes of 5 μm and 1 μm (figure 2($d_{1\text{-}2}$) and ($e_{1\text{-}2}$)). The UNCD sample (figure 2($c_{1\text{-}2}$)) also shows strong texture, but the size indicates that these are multi-grain clusters. The pUNCD shows topographic features, but no grain structure. At the highest resolution (figures 2(b_3)–(e_3)), all features look relatively smooth. The smoothness of the features is likely related to tip artifacts (caused by convolution between tip curvature and topography). Figures 2(b_4)–(e_4) show TEM images of the AFM tips after they were used to image the surfaces. The tip radii were measured as $R = 17$ nm for pUNCD, $R = 36$, 47 and 31 nm for UNCD, NCD and MCD respectively using the same procedure used for the stylus tip. As with the stylus tip, a circle was fitted to the tip profile for the extraction of the radius (figures 2(b_4), (c_4), (d_4), and (e_4)). In cases where the tip apex did not appear perfectly circular, a best-fit circle was fitted to the region of the tip that makes contact with the substrate.

Finally, the surfaces were analyzed using side-view TEM (figures 2(f)–(i)). Once again, the pUNCD surfaces had the lowest amplitude of topography and the MCD had the highest. The NCD and MCD materials showed clear faceting from the individual crystallites. At the scales accessible by the TEM, the UNCD surface also shows faceting (see figures 2(g_2), and (g_3)). However, the smooth pUNCD surface shows no indication of faceting but rather a smoothly varying surface topography. Surprisingly, despite significant differences in topography at larger scales, the smallest-scale topography is nearly identical between the microcrystalline diamond, the nanocrystalline diamond, and the ultrananocrystalline diamond. This finding is further discussed in the next paragraph.

To further investigate the similarities in roughness at the smallest scales, representative images of the three unpolished surfaces are shown in greater detail in figure 3. In all cases, significant roughness is visible on the scale of Angstroms to nanometers. This small-scale topography is visible even on a single facet of a single grain of the NCD and MCD materials. The atomic lattice of the diamond is clearly visible in the TEM whenever a grain is aligned with a zone axis lying near to the imaging axis. In these cases, the lattice is observed in many areas to extend to within 1 nm of the surface. Therefore, it is not simply a rough and potentially amorphous surface layer that is sitting on a flat diamond facet, rather the diamond crystal itself exhibits significant roughness at the small scale.

3.2. Computing topography metrics, in real-space and in frequency-space

Now, a more quantitative analysis of the topography data is presented, and scalar roughness parameters are computed for the various surfaces. First, the root-mean-square (RMS) height h_{rms}, RMS slope h'_{rms}, and RMS curvature h''_{rms} are computed in real-space from each line scan by numerically integrating the squared height data (or its derivatives) over the scan length L [32]:

$$h_{\text{rms}}^2 = \frac{1}{L}\int_0^L h^2(x)\,dx,\; h_{\text{rms}}'^2 = \frac{1}{L}\int_0^L \left(\frac{dh}{dx}\right)^2 dx,$$
$$h_{\text{rms}}''^2 = \frac{1}{L}\int_0^L \left(\frac{d^2h}{dx^2}\right)^2 dx \tag{1}$$

using the trapezoidal rule (equation (4) of [32]). Figure 4 shows the computed roughness parameters as a function of scan size L (for RMS height, which depends on the larger-scale features) and measurement resolution l (for RMS slope and curvature, which depend on smaller-scale features). Note that for TEM data, the 'measurement resolution' is different from the size of a pixel in the camera ('pixel size') (which can be sub-atomic at the highest magnifications). The measurement resolution l is determined from the point spacing of the extracted profiles, as shown in figure 2. Computing the RMS height as a function of size is equivalent to an analysis of the surfaces' self-affine properties using a variable bandwidth method [47].

Figure 4(a) shows that h_{rms} increases with L for all the surfaces studied here at small L, i.e. L less than 1 μm. There is a crossover to constant (independent of L) h_{rms} at a scan size of 1–10 μm. All of the surfaces studied here show this transition, which corresponds to the thickness of these coatings (2 μm). The amplitude of the pUNCD surface is much smaller than the

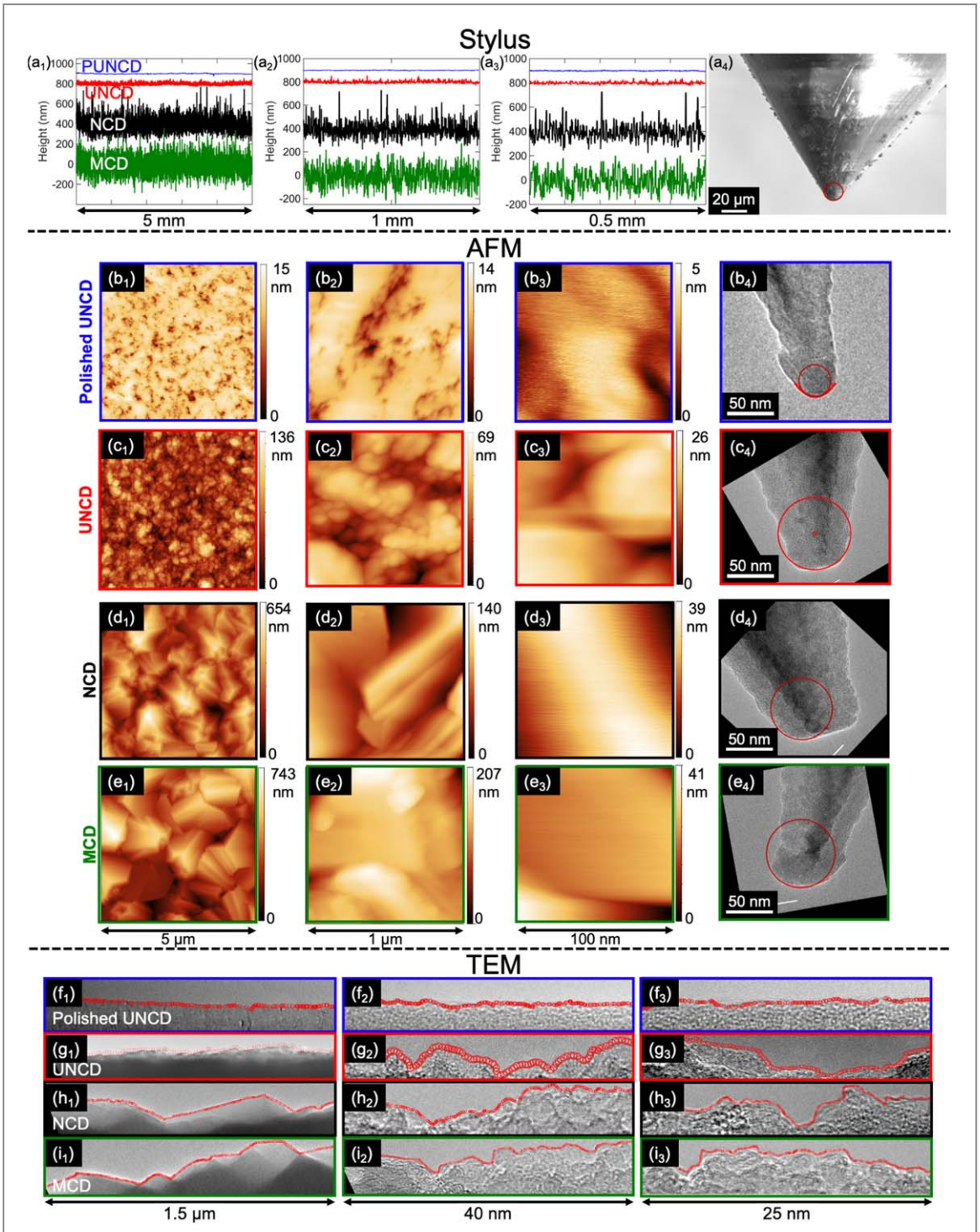


Figure 2. Comprehensive topography characterization is achieved by performing more than 50 multi-resolution measurements on each material using stylus profilometry (top section), AFM (middle section) and side-view TEM (bottom section). Representative data from each technique at various magnifications are presented for all four materials. For stylus and AFM, the radius of the scanning tip shown in panels (a_4), (b_4), (c_4), (d_4) and (e_4) is measured using electron microscopy.

other three surfaces. NCD and MCD roll off to the same constant value. The values in figure 4(a) at large L correspond to the observation of the amplitude in stylus profilometry shown in figure 2, with pUNCD being the 'smoothest' and MCD being the 'roughest' surface at large scales.

From the analysis of ideal self-affine surfaces [22, 30, 32], it can be shown that $h_{\rm rms} \propto L^H$ while $h'_{\rm rms} \propto l^{H-1}$ and $h''_{\rm rms} \propto l^{H-2}$. Therefore, in figure 4(a), the small-scale data has been fit with a power-law function (solid lines), and the extracted values are used to determine the expected trends in figures 4(b), (c) (dashed lines). As argued above and first described in [36, 48], it is apparent from figure 4 that no single value of RMS height, RMS slope, and RMS curvature can be defined, because these parameters depend on L or l. This demonstrates a key difficulty that impedes efforts to link surface function to a single scalar roughness parameter; these

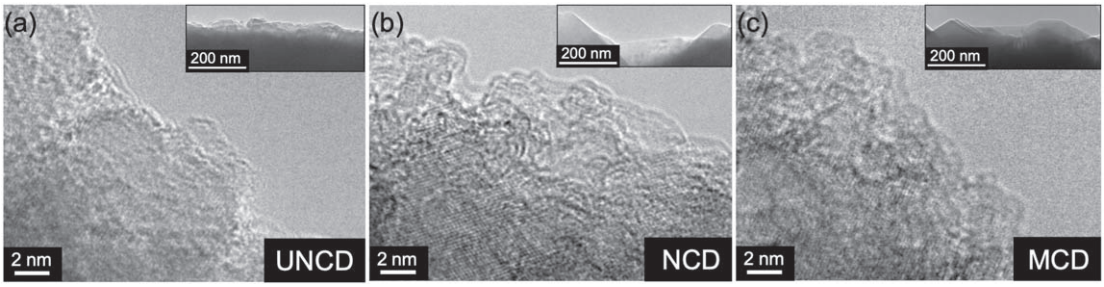


Figure 3. Significant small-scale roughness is observed on all of the as-deposited diamond materials. The unpolished UNCD (left), NCD (middle), and MCD (right) demonstrate similar roughness at the Angstrom-to-single-digit-nanometer scale, which is superimposed on larger-scale topography that varies significantly between materials (insets). This small-scale topography is not detectable with conventional topography measurements.

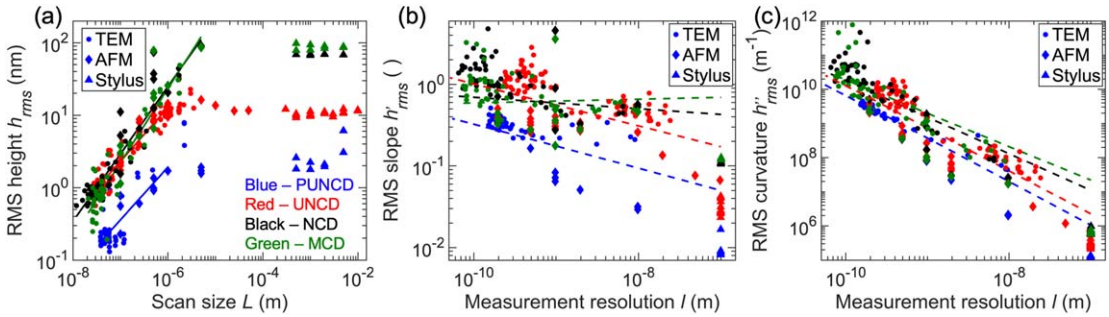


Figure 4. Scalar roughness parameters show systematic variation with length scale. The RMS height (a) is presented as a function of scan size L; while the RMS slope (b) and RMS curvature (c) are shown as a function of resolution l (see main text). The solid lines in (a) show the best-fit power-law exponent H to the data at small scan sizes. This yields a measurement of the Hurst exponent from the variable-bandwidth model [47], corresponding to $H_{\mathrm{VBM}} = 0.73$ and 0.74 for pUNCD and UNCD, respectively and $H_{\mathrm{VBM}} = 0.93$ and 1.02 for NCD and MCD, respectively. The dashed lines in (b) and (c) show the power-law dependency *expected* from the fits of panel (a), $H - 1$, and $H - 2$, respectively (see main text).

common roughness parameters (including RMS roughness) are scanning-length dependent and do not describe an intrinsic property of the material.

As an alternative to analyzing the real-space measurements individually (as done in figure 4), the measurements can all be combined in frequency space to create a single PSD, denoted $C(q)$, that yields a complete statistical description of topography for each surface, as shown in figure 5. This analysis was carried as follows: first, the PSD of each measurement was computed, following the procedures laid out in [30, 32]. Second, a reliability cutoff [30, 32] due to tip artifacts was calculated for each PSD based on the measured tip radius (figure 2), and all data below this size scale was deemed unreliable and removed [32]. Third, all of the reliable portions of the many individual PSDs were combined by computing the arithmetic average of all measurements in logarithmically-spaced bins. The result is a single whole-surface PSD that describes the material across all length scales. There are no fitting parameters in this analysis; rather the PSD serves to separate the different size scales of topography, and the various techniques agree within experimental uncertainty. The only exception is for pUNCD, where the small-size stylus data lies below the AFM and TEM data, causing a dip around $q = 10^7\ \mathrm{m}^{-1}$. This is believed to be an instrumental artifact, rather than resulting from the real topography. Overall, the value of these comprehensive PSDs is that they can be used in analytical and numerical models (such as [22, 25–27, 49–52]) to understand and predict surface properties (e.g. [17]).

To compute scale-invariant scalar roughness parameters, the full stitched-together PSD was used to compute RMS height, RMS slope, and RMS curvature as:

$$(h_{\mathrm{rms}})^2 = \frac{1}{\pi}\int_0^\infty C(q)\,dq,\ (h'_{\mathrm{rms}})^2 = \frac{1}{\pi}\int_0^\infty q^2 C(q)\,dq,$$
$$(h''_{\mathrm{rms}})^2 = \frac{1}{\pi}\int_0^\infty q^4 C(q)\,dq \tag{2}$$

Table 2 shows the computed RMS parameters. Also computed from the whole-surface PSD is the area ratio, which is the increase in full surface area, per unit apparent area, due to the roughness. The area ratio is a critical parameter in investigations of soft-material adhesion, including [17], and it is calculated as

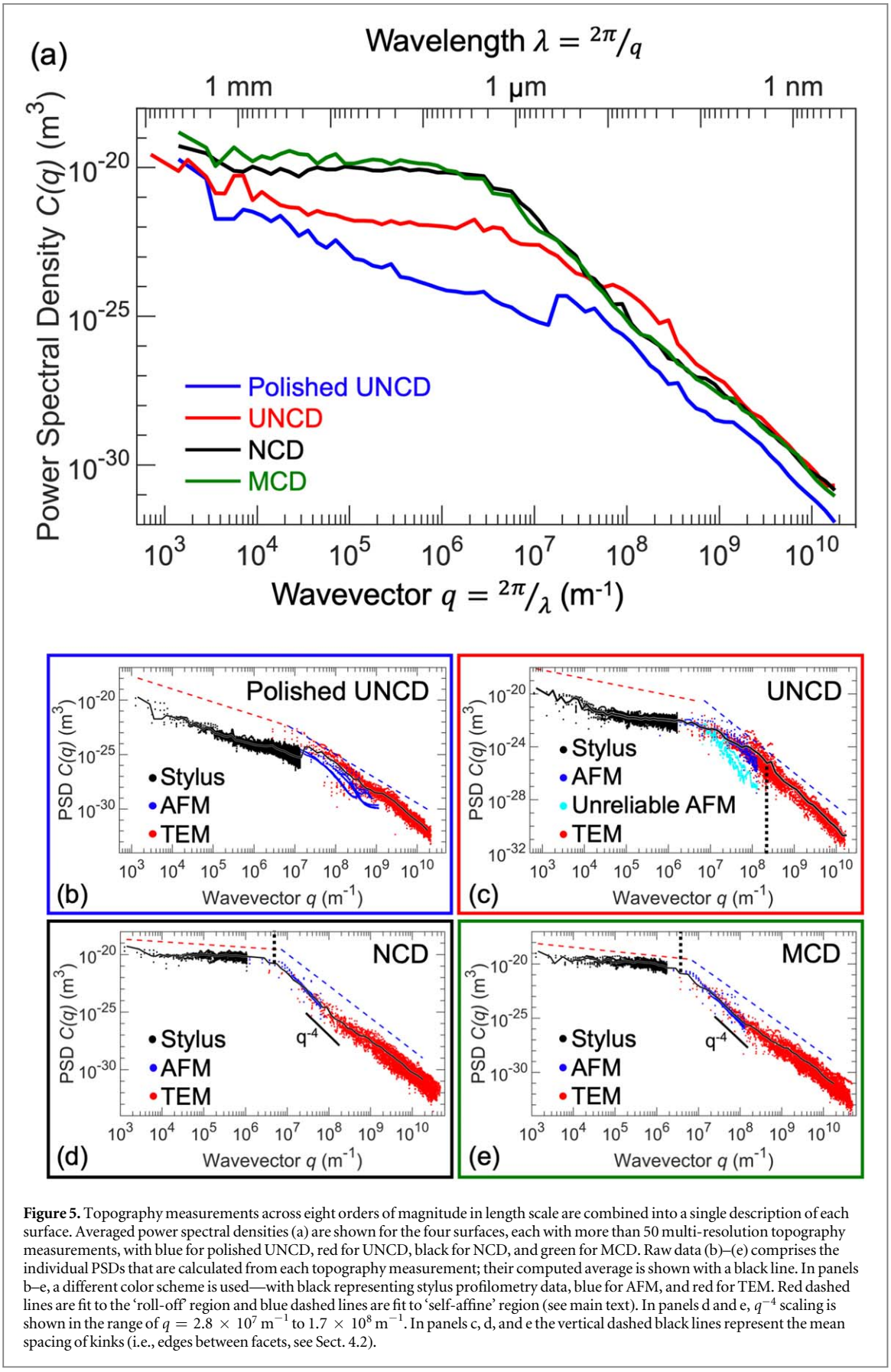


Figure 5. Topography measurements across eight orders of magnitude in length scale are combined into a single description of each surface. Averaged power spectral densities (a) are shown for the four surfaces, each with more than 50 multi-resolution topography measurements, with blue for polished UNCD, red for UNCD, black for NCD, and green for MCD. Raw data (b)–(e) comprises the individual PSDs that are calculated from each topography measurement; their computed average is shown with a black line. In panels b–e, a different color scheme is used—with black representing stylus profilometry data, blue for AFM, and red for TEM. Red dashed lines are fit to the 'roll-off' region and blue dashed lines are fit to 'self-affine' region (see main text). In panels d and e, q^{-4} scaling is shown in the range of $q = 2.8 \times 10^7\ \mathrm{m}^{-1}$ to $1.7 \times 10^8\ \mathrm{m}^{-1}$. In panels c, d, and e the vertical dashed black lines represent the mean spacing of kinks (i.e., edges between facets, see Sect. 4.2).

described in [17]. While MCD is the roughest in terms of RMS height, the unpolished UNCD is the steepest in terms of RMS slope. However, while table 2 shows the mathematically correct or 'true' values of RMS parameters for a surface, any individual application or surface property may depend only on a certain range

Table 2. One-dimensional roughness parameters for nanodiamond substrates computed from the whole-surface PSD for these materials.

	Polished UNCD	UNCD	NCD	MCD
RMS height	4.2 ± 0.8 nm	17.4 ± 1.3 nm	97.2 ± 11.7 nm	101.2 ± 8.0 nm
RMS slope	0.31 ± 0.03	1.17 ± 0.28	0.92 ± 0.10	0.85 ± 0.10
RMS curvature	1.99 ± 0.35 nm^{-1}	6.32 ± 1.20 nm^{-1}	5.91 ± 1.83 nm^{-1}	5.04 ± 1.45 nm^{-1}
Area Ratio	1.07	1.69	1.47	1.42

of length-scales. In that case, scale-dependent parameters would be recomputed by integrating only across the relevant size scales.

4. Discussion

4.1. Evaluating the fractal nature of diamond coatings, and the meaning of Hurst exponents

The Hurst exponent (which can be related to the fractal dimension, as described in [53]), can be calculated from the PSD, which is commonly separated (somewhat arbitrarily) into the 'self-affine' region, where the topography appears to be described by a power law relationship of $C \propto q^{\beta}$ where β is the power-law exponent, and the 'roll-off' region, where the PSD appears to be flatter. The Hurst exponent H is typically extracted from the self-affine region as $H = (\beta - 1)/2$ [53, 54]. Using this procedure, and the region of the curves between $q = 6.3 \times 10^6$ m^{-1} and 1.8×10^{10} m^{-1}, the resulting value was $H = 0.62 \pm 0.09$ and 0.77 ± 0.06 for pUNCD and UNCD respectively; and $H = 0.89 \pm 0.04$ and 0.87 ± 0.03 for NCD and MCD respectively.

There are two alternative methods of extracting the Hurst exponent: from the real-space data using the variable bandwidth method (VBM) [55]; and from the roll-off region by assuming that the full PSD is described by Fractional Gaussian Noise (FGN) [53], a hypothesis put forth by some of us in [32]. The VBM is nothing more than an analysis of the functional dependence of RMS height h_{rms} as a function of scan size L (that scales as $h_{rms}(L) \propto L^H$ for self-affine surfaces), as shown in figure 4(a). Note that even for an individual scan, the RMS height could be computed over a subsection of that scan (yielding an estimate of the Hurst exponent for a single realization of the topography [33, 47, 55]) but here full-size scans were used for calculation. As shown in figure 4(a), RMS height $h_{rms}(L)$ can be accurately fit with a power-law form over the range from the smallest size ($L \sim 10$ nm) up to approximately $L = 1 - 10$ μm. Over this region, the Hurst exponents were $H_{VBM} = 0.73 \pm 0.18$ for pUNCD; $H_{VBM} = 0.74 \pm 0.05$ for UNCD; $H_{VBM} = 0.93 \pm 0.09$ for NCD and $H_{VBM} = 1.02 \pm 0.10$ for MCD. The Hurst exponents from this method (H_{VBM}) yield similar results to the above Hurst exponents (H) that are extracted from the more common method of fitting the 'self-affine' region of the PSD. The VBM is a useful technique because it is simpler to perform—requiring only a straightforward calculation of the root-mean-square height, calculated in real-space, from a series of multi-resolution measurements. Additionally, it can be readily performed on reentrant surfaces, while the calculation of the PSD requires surfaces to be functions (one height value for each lateral position).

The second alternative method for computing the Hurst exponent uses the 'roll-off' region and the fractional Gaussian noise (FGN) approach. Here, the Hurst exponent is given by $H_{FGN} = (\alpha + 1)/2$ [32, 53], where α is the scaling exponent $C(q) \propto q^{\alpha}$ at low q. This results in $H_{FGN} = 1.10 \pm 0.04$ for pUNCD; $H_{FGN} = 0.82 \pm 0.04$ for UNCD; and $H_{FGN} = 0.62 \pm 0.04$ and 0.70 ± 0.05 for NCD and MCD, respectively. A prior paper by the present authors ([32]) speculated that there may be a connection between H_{FGN} and H (from the self-affine region). This observation would be extremely useful as it suggests that the small-scale behavior could be predicted from large-scale measurements. Unfortunately, when these four different surfaces are compared, there is no clear relationship that emerges.

Even in the traditional 'self-affine' portion of the PSD, a range of values can be extracted for Hurst exponent for a single surface depending on the window used for fitting H. This is particularly true for MCD and NCD where Hurst exponents can be calculated by dividing the 'self-affine' region into two parts. For q in the range of 2.8×10^7 to 1.7×10^8 m^{-1}, the extracted values are $H_{larger\lambda} = 1.27 \pm 0.27$ and 1.32 ± 0.09 for NCD and MCD, respectively; for q in the range of 1.7×10^8 to 1.8×10^{10} m^{-1}, the extracted Hurst exponents are $H_{smaller\lambda} = 0.75 \pm 0.04$ and 0.78 ± 0.05 for NCD and MCD, respectively. Indeed, the roughness in these two portions of the curve seems to be qualitatively different, with the upper portion having a scaling behavior near $C \propto q^{-4}$, corresponding to $H = 1.5$. The origin of these differences in scaling behavior between different length scales is discussed in detail in the next section.

Because there can be so much variability in the measurement of a single surface, the whole practice of assuming self-affinity and assigning a single Hurst exponent to describe a surface must be done with caution. It is mathematically convenient to assume self-affinity as this simplifies numerical and analytical models, and it is common practice in experiments to use assumptions of self-affinity to extrapolate to small scales where the topography is not easily measured.

However, at least for the diamond materials investigated here, the best-fit value for H depends on the region over which it is measured, and it is strongly influenced by other factors such as grain size (see next section). Instead, where possible, it is preferable to measure surface topography across all size scales and to use the whole-surface PSD as the primary descriptor for the surface, rather than any scalar parameter.

4.2. The effect of grain size on topography

As discussed in the previous section, the larger-grain-size materials (NCD and MCD) demonstrate a region where the PSD scaling is similar to q^{-4} in the larger-wavelength portion of the 'self-affine' region (see solid black lines in figures 5(d) and (e)). This scaling is characteristic of 'kinks' in the real-space line scan, such as sharp peaks or valleys. Note that such kinks can arise in topography as an artifact of the nonvanishing tip radius [30, 37], so one must first rule out this as the cause. However, in the present work, this q^{-4} scaling is clearly observed both in the reliable portion of the AFM measurement and in the TEM measurement, both of which are free from tip-based artifacts. Therefore, this q^{-4} scaling in the PSDs of the MCD and NCD is a feature of the measured topography, rather than emerging from an artifact. This behavior of the PSD corresponds to kinks in the surface topography that are directly observable in the TEM imaging, as shown in figures 2 and 3.

Our hypothesis to describe this local q^{-4} scaling is that it is characteristic of topography at scales approximately equal to the grain size of the material. At these scales, features are dominated by the crystal facets and kinks between them: adjacent grains with different orientations will create concave kinks where the grains meet, while edges between crystal facets will create convex kinks. It is therefore assumed that typical topography line scans will pass through approximately two kinks (one concave and one convex) per grain. At sizes much below the grain size, the topography of MCD and NCD reverts to random, self-affine behavior, showing roughness very similar to UNCD (see figure 3), where the PSD scales as q^{-1-2H} (see blue dashed lines in figures 5(b)–(e)). This finding agrees with prior work [56] on fracture surfaces in sandstone. In that work, stylus profilometry measurements showed a transition around the grain size; however sub-grain features were mostly inaccessible there due to the aforementioned tip artifacts. Computer-generated profiles were used to verify that facets can cause q^{-4}-scaling at a wavevector related to the average kink spacing, designated l_k. The mathematical basis for this hypothesis linking kink spacing to q^{-4} scaling in the PSD is given in the appendix.

In order to demonstrate how the PSD is affected by the superposition of facets from a characteristic grain size and random roughness below that size, artificial one-dimensional surfaces were created that were composed of: a superposition of triangular peaks (figure 6(a_1)); self-affine random roughness (figure 6(a_2)); and the summation of those two into a single surface (figure 6(a_3)). The piecewise linear surface (figure 6(a_1)) has kinks with uncorrelated heights drawn from a Gaussian distribution and lateral distances between kinks drawn from a Rayleigh distribution. The surface is scaled in order to have an RMS slope of 1 and an average kink spacing l_k. The small-scale self-affine random roughness (figure 6(a_2)) is generated using a Fourier-filtering algorithm [30, 57] with a Hurst exponent $H = 0.8$ and RMS-slope of 1.2.

When the PSDs (figure 6(b)) are computed from these surfaces, there is a transition from a flat PSD at the largest sizes to scaling as $C \propto q^{-4}$, and the transition point occurs at a wavelength of $\lambda_k = 4l_k$, which corresponds to $q_k = \frac{2\pi}{\lambda_k} = \frac{\pi}{2l_k}$. (The transition is somewhat gradual; the choice of this particular value for the transition point is motivated from the mathematical consideration in the appendix). Importantly, the summed-surface PSD follows the kinked-surface PSD and displays q^{-4} scaling at larger scales, and then transitions to self-affine scaling at smaller scales, in this case corresponding to a Hurst exponent H of 0.8. To ensure that these results were not unique to the particular way that the kinked surface was generated, this analysis was repeated using kinked surfaces with uniform and exponential distributions of kink spacings, as well as a surface with slopes alternating between -1 and 1. These analyses are shown in the supplemental section 1 (available online at stacks.iop.org/STMP/9/014003/mmedia), but the results are similar, differing only in the sharpness of transitions. The key finding from this analysis is that the kink spacing, which corresponds approximately to the grain size, introduces a signature in the scaling of the topography that causes deviations from the commonly assumed fractal-like self-affine scaling.

The PSD of the summed surface (figure 6(b)) reproduces the three regions that are visible in the PSDs of MCD and NCD: flat behavior at small q (large sizes); scaling like q^{-4} at intermediate values; and self-affine scaling ($H \sim 0.8$) at large q (small sizes). However, this does not yet explain the self-affine scaling behavior that is observed above the grain size in polished and unpolished UNCD. To account for this, another synthetic surface was created that is similar to the first, but this time with spatially correlated kink heights (figure 6(c)). The spatial correlation of the kink heights is enforced using the method of [56] as follows: a self-affine random surface was generated using the Fourier-filtering algorithm; then facets were computed by interpolating linearly between random points (the kinks); finally this faceted surface was summed with the same self-affine random roughness as before (figure 6(a_2)). While this does not substantially alter the behavior at small wavelengths, this adds in an

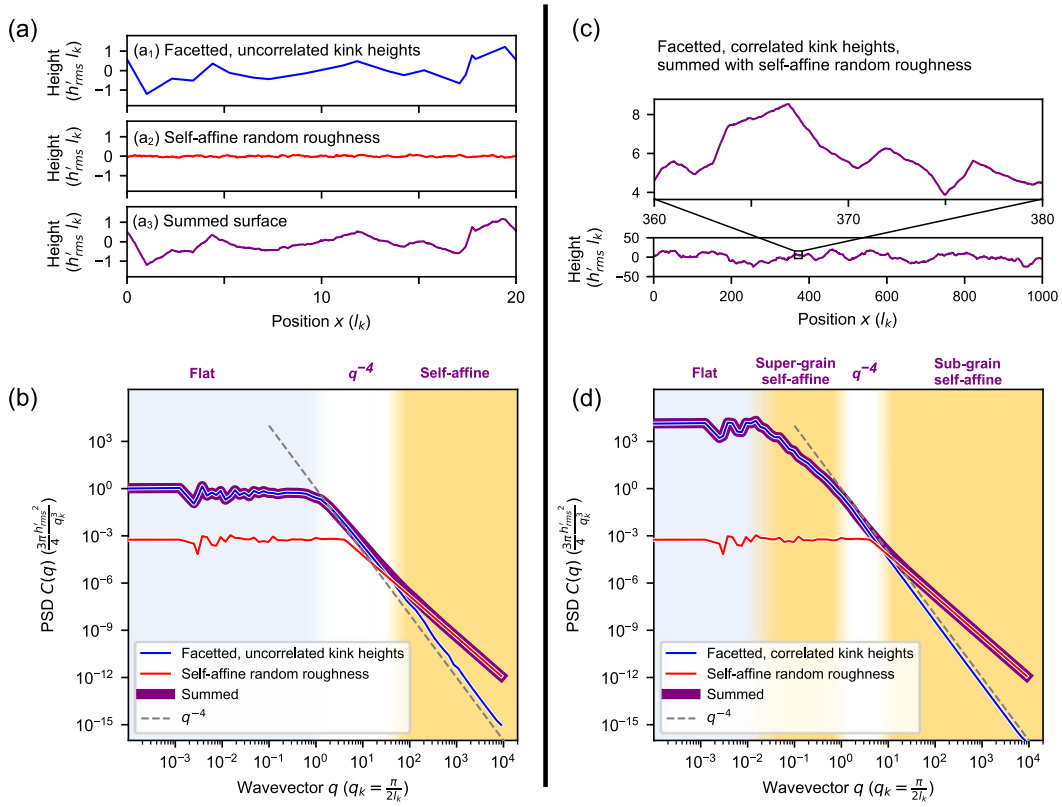


Figure 6. Computer-generated profiles support the experimental finding of characteristic regions of topography scaling for polycrystalline surfaces. Panel (a) shows a representative section of the three types of computer-generated profiles investigated: a faceted profile with random, uncorrelated kink heights and spacings (a_1), a self-affine random roughness (a_2); and the sum of these profiles (a_3). Their PSDs, averaged over logarithmically-spaced bins, are shown in (b). Three characteristic scaling regions (see main text) of the PSD for the summed surface are indicated by background color and labels above the graph. Panel (c) shows a surface created by summing a faceted profile with correlated heights and the small-scale self-affine random roughness (same as a_2). The resulting PSD (d) shows four characteristic scaling regions. The RMS slope of the faceted profiles h'_{rms} has been used to normalize all PSDs and profiles.

additional self-affine region above the grain size (figure 6(d)).

To verify this proposed link between q^{-4} scaling of the PSD and the kink spacing of the material, an analysis of kink spacing and grain size was performed on the real materials using the AFM and TEM measurements. For the MCD and NCD, concave kinks between grains were readily visible, and therefore were quantified using feature sampling, as is used in metallographic analysis [58] (supplemental section 2.1). The mean lineal intercept between concave kinks, which is of the same order as the grain diameter [58], was measured as 839 ± 68 nm for MCD and 647 ± 42 nm for NCD. Because the mean kink spacing used in the mathematical analysis involves both convex and concave kinks, it was computed as half of this value, i.e. 419 ± 34 nm for MCD and 323 ± 21 nm for NCD. These values can be converted to frequency-space using $q_k = \frac{\pi}{2l_k}$, as discussed above, and the values calculated are $q_k = 3.7 \times 10^6$ m^{-1} for MCD and $q_k = 4.9 \times 10^6$ m^{-1} for NCD. The same approach could not be applied to UNCD where the kink spacing was at or below the reliability cut-off due to tip artifacts. Therefore, the mean kink spacing for UNCD was estimated as half of the grain size, which was computed by averaging a sample of 20 different grains observed in the TEM (see supplemental section 2.2). The mean grain size of the UNCD was determined to be 14 ± 3 nm, which corresponds to a kink spacing for UNCD of 7 ± 2 nm and $q_k = 2.2 \times 10^8$ m^{-1}. The mean kink spacings of these three materials are indicated as vertical bars in figures 5(c)–(e). Indeed, the MCD and NCD demonstrate that the scaling transitions to $C \sim q^{-4}$ right around the mean kink spacing. For UNCD, the mean kink spacing is too small to observe a clear q^{-4} scaling regime; however, the self-affine scaling behavior is confirmed for sizes much larger than the characteristic grain size. The remainder of the paper will frame results in terms of grain size (again, approximately twice the value of kink spacing), since grain size is more widely measured and reported for polycrystalline materials.

In summary, based on the experimental measurements of these four polycrystalline diamond surfaces, and based on the computed PSDs of artificially generated surfaces, four characteristic regimes of topography scaling have been identified. (1) At the smallest size scales (if significantly smaller than the grain size) power-law scaling may be observed that is characteristic of self-affine

random roughness, e.g., $C \sim q^{-1-2H}$, corresponding to H in the range of 0.6–0.9. (2) At size scales similar to and slightly smaller than the average grain size, the PSD displays characteristic scaling of $C \sim q^{-4}$ due to grain facets and kinks. (3) At sizes larger than the grain size, but smaller than the film thickness, there is another region of power-law scaling corresponding to random roughness ($C \sim q^{-1-2H}$). (4) Finally, at sizes larger than the film thickness, the PSD flattens out, with scaling in the range of q^0 to q^{-1}. These four regions of topography scaling accurately describe the polycrystalline diamond surfaces and computer-generated surfaces investigated here. These four regions and their boundaries may be broadly generalizable to other materials, but further investigation is required. For instance, the boundary between regions (3) and (4) corresponded to the thickness of the diamond coatings in these measurements, but this thickness was not varied to explicitly investigate this connection. The future application of comprehensive topography characterization on other materials will elucidate the applicability of these four regions to other polycrystalline materials.

5. Conclusions

First, these results further establish a multi-resolution approach that is designated 'comprehensive topography characterization,' which combines multiple different techniques at multiple magnifications for the same surface. Then the power spectral density can be used to combine all measurements into one statistical description of the surface. Because typical roughness metrics are inherently scale-dependent and incomplete, this paper provides a method to understand roughness at all scales, including the specific scale over which it may be relevant in a given device. Particularly the measurement of small-scale roughness may be extremely important to predict and tailor surface properties such as adhesion, friction and wear. Furthermore, all topography measurements from this publication have been made publicly available, [43–46] so that other experimentalists may compare results and so that computational modelers may use it in models to predict properties of diamond materials. The purpose of this data and the underlying approach is to advance the field towards the goal of fundamental, predictive understanding of the performance of rough surfaces.

Second, these results show that the surface roughness of polycrystalline diamond materials varies significantly with scale, with surfaces that are smoother when measured at the large-scale showing roughness that is identical or even higher when measured at the smaller scales. Furthermore, while self-affine scaling ($H \sim 0.6 - 0.9$) is observed over some length-scales, the grain size introduces a signature into the power spectral density, showing q^{-4} scaling behavior at scales approximately equal to the grain size. All unpolished surfaces show identical self-affine scaling at the smallest scale. This is a signature of small-scale roughness that is superimposed on the crystalline facets and is not observable with conventional (AFM, stylus) measurement techniques. Altogether, four characteristic regions of topography scaling were observed; these are expected to be applicable to all unpolished polycrystalline diamond films and may apply more broadly to other polycrystalline materials.

Acknowledgments

Sample preparation was performed in the Fischione Instruments Electron Microscopy Sample Preparation Laboratory at the University of Pittsburgh. The authors acknowledge the use of two facilities at the University of Pittsburgh: the Materials Micro-Characterization Laboratory (MMCL) in the Department of Mechanical Engineering and Materials Science; and the Nanoscale Fabrication and Characterization Facility (NFCF) in the Petersen Institute of Nano Science and Engineering (PINSE). The authors thank A. Baker for supplemental TEM imaging. LP acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (project EXC-2193/1–390951807) and by the European Commission (ERC-StG-757343). Funding for TDBJ and the University of Pittsburgh work was provided by the National Science Foundation under award number CMMI-1727378.

Data availability statement

The data that support the findings of this study are openly available at the following URLs:
https://contact.engineering/go/wcqj3/;
https://contact.engineering/go/8sc7t/;
https://contact.engineering/go/cjy6s/;
https://contact.engineering/go/mz7z5/.

Appendix. A mathematical basis for the q^{-4} scaling of the PSD slightly below the grain size

The analysis begins with a triangular peak $y_p(x)$:

$$y_p(x) = \begin{cases} h\dfrac{x - x_l}{x_c - x_l}, & x_l \leqslant x < x_c \\ h\left(1 - \dfrac{x - x_c}{x_r - x_c}\right), & x_c \leqslant x \leqslant x_r \\ 0, & else \end{cases} \quad \text{(A1)}$$

The Fourier transformation of the peak profile $y_p(x)$ is:

$$\tilde{y}_p(\mathrm{q}) = \frac{h}{q^2}\, e^{-iqx_c}\left(\frac{1-\mathrm{e}^{-iq(x_r-x_c)}}{x_r-x_c} + \frac{1-\mathrm{e}^{iq(x_c-x_l)}}{x_c-x_l}\right) \tag{A2}$$

For large q, the PSD of the peak $C(q) \propto |\tilde{y}(q)|^2$ oscillates with an amplitude decaying asq^{-4}. For $q \ll \min\left(\frac{\pi}{2(x_r-x_c)}, \frac{\pi}{2(x_c-x_l)}\right)$, $\tilde{y}_p(q) \simeq \frac{h}{2}(x_r - x_l)\, e^{-iqx_c}$, so the PSD is flat.

A piecewise linear function $y(x)$ with kinks at (x_k, h_k) can be written as a superposition of these peaks. Its Fourier transform is:

$$\tilde{y}(q) = \sum_{k=0}^{n-1} \frac{h_k}{q^2} e^{-iqx_k}\left(\frac{1-\mathrm{e}^{-iq(x_{k+1}-x_k)}}{x_{k+1}-x_k} + \frac{1-\mathrm{e}^{iq(x_k-x_{k-1})}}{x_k-x_{k-1}}\right) \tag{A3}$$

$h_n = h_0$ ensures continuity at the periodic boundary.

The PSD of the piecewise linear function has the same features as the PSD of the triangular peak: for large enough values of q, $C(q) \propto q^{-4}$; for $q \ll \min_k \frac{\pi}{2\,(x_{k+1}-x_k)}$, $\tilde{y}(q) \simeq \sum_{k=0}^{n-1} \frac{h_k(x_{k+1}-x_k)}{2} e^{-iqx_k}$ and the PSD is flat. Figure 6(b) shows that for the profiles considered, the PSD changes between flat and $\propto q^{-4}$ around $q_k = \frac{\pi}{2l_k}$, with l_k the mean kink spacing, rather than the maximum kink spacing. Similarly, the PSD in figure 6(d) changes between the super-grain self-affine scaling and $\propto q^{-4}$ around the same wavevector $q_k = \frac{\pi}{2l_k}$. This is close to the value $q_k \simeq \frac{\pi}{3l_k}$ determined in [56] for the same Hurst exponent of the super-grain self-affine regime, $H = 0.8$.

The different exponents of the PSD of faceted versus self-affine surfaces correspond to different behaviors of the scale-dependent RMS slope with increasing resolution. The scale-dependent RMS slope can be computed from the PSD $C(q) = q^{\alpha}$ using equation (2): $h'_{\mathrm{rms}}(q_{\max}) = \left(\int_{q_{\min}}^{q_{\max}} dq q^2 q^{\alpha}\right)^{\frac{1}{2}}$. As $q_{\max} \to \infty$, $h'_{\mathrm{rms}}(q_{\max})$ converges to a finite value for $\alpha < -3$, but diverges for $\alpha \geqslant -3$, the former case corresponding to a faceted surface and the latter to a self-affine random surface with $H \leqslant 1$. Since the slope between two kinks is constant by definition, the RMS slope of a faceted profile is finite and reaches its limit value once the smallest resolved length is below the kink spacing. In contrast to that, the RMS slope of an ideal self-affine random surface with $H \leqslant 1$ increases indefinitely with resolution, as smaller-scale features with increasing slope are resolved.

ORCID iDs

Abhijeet Gujrati https://orcid.org/0000-0001-7744-5743
Nicolaie Moldovan https://orcid.org/0000-0001-5715-4957
Hongjun Zeng https://orcid.org/0000-0002-4415-1367
Lars Pastewka https://orcid.org/0000-0001-8351-7336
Tevis D. B. Jacobs https://orcid.org/0000-0001-8576-914X

References

[1] Al-Azizi A A, Eryilmaz O, Erdemir A and Kim S H 2014 Nano-texture for a wear-resistant and near-frictionless diamond-like carbon *Carbon* **73** 403–12
[2] Bernal R A, Chen P, Schall J D, Harrison J A, Jeng Y-R and Carpick R W 2018 Influence of chemical bonding on the variability of diamond-like carbon nanoscale adhesion *Carbon* **128** 267–76
[3] Ţălu Ş, Bramowicz M, Kulesza S, Ghaderi A, Dalouji V, Solaymani S, Kenari M F and Ghoranneviss M 2016 Fractal features and surface micromorphology of diamond nanocrystals *J. Microsc.* **264** 143–52
[4] Schade A, Rosiwal S M and Singer R F 2007 Influence of surface topography of HF-CVD diamond films on self-mated planar sliding contacts in dry environments *Surf. Coatings Technol.* **201** 6197–205
[5] Kovalchenko A M, Elam J W, Erdemir A, Carlisle J A, Auciello O, Libera J A, Pellin M J, Gruen D M and Hryn J N 2011 Development of ultrananocrystalline diamond (UNCD) coatings for multipurpose mechanical pump seals *Wear* **270** 325–31
[6] Jacobs T D B, Ryan K E, Keating P L, Grierson D S, Lefever J A, Turner K T, Harrison J A and Carpick R W 2013 The effect of atomic-scale roughness on the adhesion of nanoscale asperities: a combined simulation and experimental investigation *Tribol. Lett.* **50** 81–93
[7] Ryan K E, Keating P L, Jacobs T D B, Grierson D S, Turner K T, Carpick R W and Harrison J A 2014 Simulated adhesion between realistic hydrocarbon materials: Effects of composition, roughness, and contact point *Langmuir* **30** 2028–37
[8] Alcaide M, Papaioannou S, Taylor A, Fekete L, Gurevich L, Zachar V and Pennisi C P 2016 Resistance to protein adsorption and adhesion of fibroblasts on nanocrystalline diamond films: The role of topography and boron doping *J. Mater. Sci. Mater. Med.* **27** 90
[9] May P W 2000 Diamond thin films: A 21st-century material *Philos. Trans. R. Soc. London. Ser. A Math. Phys. Eng. Sci.* **358** 473–95
[10] Otschik J, Schrüfer A, Thelke J, Kirchhof M and Schmaderer S 2010 Diamond seal faces—the benefits *World Pumps.* **2010** 18–20
[11] Auciello O *et al* Are diamonds a MEMS' best friend? *IEEE Microw. Mag.* **8** 2007 61–75
[12] Nistor P A and May P W 2017 Diamond thin films: Giving biomedical applications a new shine *J. R. Soc. Interface* **14** 20170382
[13] Zeng H, Jarvik R, Catausan G, Moldovan N and Carlisle J 2016 Diamond coated artificial cardiovascular devices *Surf. Coatings Technol.* **302** 420–5
[14] Biener J *et al* 2009 Diamond spheres for inertial confinement fusion *Nucl. Fusion* **49** 112001
[15] Konicek A R, Grierson D S, Gilbert P U P A, Sawyer W G, Sumant A V and Carpick R W 2008 Origin of ultralow friction and wear in ultrananocrystalline diamond *Phys. Rev. Lett.* **100** 235502
[16] Pastewka L, Moser S, Gumbsch P and Moseler M 2011 Anisotropic mechanical amorphization drives wear in diamond *Nat. Mater.* **10** 34–8
[17] Dalvi S, Gujrati A, Khanal S R, Pastewka L, Dhinojwala A and Jacobs T D B 2019 Linking energy loss in soft adhesion to surface roughness *Proc. Natl Acad. Sci.* **116** LP–25425490
[18] Greenwood J A and Williamson J B P 1966 Contact of nominally flat surfaces *Proc. R. Soc. London. Ser. A. Math. Phys. Sci.* **295** 300–19

[19] Fuller K N G and Tabor D 1975 The effect of surface roughness on the adhesion of elastic solids *Proc. R. Soc. London. A. Math. Phys. Sci.* **345** 327–42
[20] Persson B N J and Tosatti E 2001 The effect of surface roughness on the adhesion of elastic solids *J. Chem. Phys.* **115** 5597
[21] Persson B N J 2001 Theory of rubber friction and contact mechanics *J. Chem. Phys.* **115** 3840–61
[22] Persson B N J, Albohr O, Tartaglino U, Volokitin A I and Tosatti E 2004 On the nature of surface roughness with application to contact mechanics, sealing, rubber friction and adhesion *J. Phys. Condens. Matter* **17** R1–62
[23] Persson B N J 2001 Elastoplastic contact between randomly rough surfaces *Phys. Rev. Lett.* **87** 116101
[24] Hyun S, Pei L, Molinari J-F and Robbins M O 2004 Finite-element analysis of contact between elastic self-affine surfaces *Phys. Rev. E* **70** 26117
[25] Campañá C and Müser M H 2007 Contact mechanics of real vs. randomly rough surfaces: A Green's function molecular dynamics study *Europhys. Lett.* **77** 38005
[26] Pastewka L and Robbins M O 2014 Contact between rough surfaces and a criterion for macroscopic adhesion *Proc. Natl Acad. Sci.* **111** LP–3293303
[27] Müser M H *et al* 2017 Meeting the contact-mechanics challenge *Tribol. Lett.* **65** 118
[28] Prodanov N, Dapp W B and Müser M H 2014 On the contact area and mean gap of rough, elastic contacts: Dimensional analysis, numerical corrections, and reference data *Tribol. Lett.* **53** 433–48
[29] Yastrebov V A, Anciaux G and Molinari J-F 2015 From infinitesimal to full contact between rough surfaces: Evolution of the contact area *Int. J. Solids Struct.* **52** 83–102
[30] Jacobs T D B, Junge T and Pastewka L 2017 Quantitative characterization of surface topography using spectral analysis *Surf. Topogr.: Metrol. Prop.* **5** 013001
[31] Gujrati A, Khanal S R and Jacobs T D B 2017 A method for quantitative real-time evaluation of measurement reliability when using atomic force microscopy-based metrology *2017 IEEE 17th Int. Conf. Nanotechnology, NANO* **2017** 135–8
[32] Gujrati A, Khanal S R, Pastewka L and Jacobs T D B 2018 Combining TEM, AFM, and profilometry for quantitative topography characterization across all scales *ACS Appl. Mater. Interfaces* **10** 29169–78
[33] Hinkle A R, Nöhring W G, Leute R, Junge T and Pastewka L 2020 The emergence of small-scale self-affine surface roughness from deformation *Sci. Adv.* **6** eaax0847
[34] Candela T, Renard F, Klinger Y, Mair K, Schmittbuhl J and Brodsky E E 2012 Roughness of fault surfaces over nine decades of length scales *J. Geophys. Res. Solid Earth* **117** B08409
[35] Mandelbrot B B, Passoja D E and Paullay A J 1984 Fractal character of fracture surfaces of metals *Nature* **308** 721–2
[36] Sayles R S and Thomas T R 1978 Surface topography as a nonstationary random process *Nature* **271** 431–4
[37] Church E L and Takacs P Z 1991 Effects of the nonvanishing tip size in mechanical profile measurements *34th Annual International Technical Symposium on Optical and Optoelectronic Applied Science and Engineering* 1332 *(San Diego, CA, 1 January 1991)* 504–14
[38] Stover J C 1995 *Optical Scattering: Measurement and Analysis, Second, SPIE, 1000 20th Street (Bellingham, WA 98227-0010 USA)*
[39] Wolski M, Podsiadlo P, Stachowiak G W, Holmberg K, Laukkanen A, Ronkainen H, Gee M, Nunn J, Gachot C and Li L 2017 Multiscale characterisation of 3D surface topography of DLC coated and uncoated surfaces by directional blanket covering (DBC) method *Wear.* **388–389** 47–56
[40] Khanal S R, Gujrati A, Vishnubhotla S B, Nowakowski P, Bonifacio C S, Pastewka L and Jacobs T D B 2018 Characterization of small-scale surface topography using transmission electron microscopy *Surf. Topogr.: Metrol. Prop.* **6** 045004
[41] Zeng H, Konicek A R, Moldovan N, Mangolini F, Jacobs T D B, Wylie I, Arumugam P U, Siddiqui S, Carpick R W and Carlisle J A 2015 Boron-doped ultrananocrystalline diamond synthesized with an H-rich/Ar-lean gas system *Carbon* **84** 103–17
[42] Vahdat V and Carpick R W 2013 Practical method to limit tip–sample contact stress and prevent wear in amplitude modulation atomic force microscopy *ACS Nano.* **7** 9836–50
[43] Gujrati A, Khanal S, Pastewka L and Jacobs T D B 2020 Polished Ultrananocrystalline Diamond (Version 1), Contact. Engineering. (https://contact.engineering/go/wcqj3/)
[44] Gujrati A, Khanal S R, Pastewka L and Jacobs T D B 2020 Ultrananocrystalline diamond (UNCD) (Version 1), Contact. Engineering. (https://contact.engineering/go/8sc7t/)
[45] Gujrati A, Khanal S, Pastewka L and Jacobs T D B 2020 Nanocrystalline Diamond (Version 1), Contact.Engineering. (https://contact.engineering/go/cjy6s/)
[46] Gujrati A, Khanal S, Pastewka L and Jacobs T D B 2020 Microcrystalline Diamond (Version 1), Contact.Engineering. (https://contact.engineering/go/mz7z5/)
[47] Mandelbrot B B 1985 Self-affine fractals and fractal dimension *Phys. Scr.* **32** 257–60
[48] Thomas T R and Sayles R S 1978 Some problems in the tribology of rough surfaces *Tribol. Int.* **11** 163–8
[49] Persson B N J and Scaraggi M 2014 Theory of adhesion: Role of surface roughness *J. Chem. Phys.* **141** 124701
[50] Persson B N J 2002 Adhesion between elastic bodies with randomly rough surfaces *Phys. Rev. Lett.* **89** 245502
[51] Bush A W, Gibson R D and Thomas T R 1975 The elastic contact of a rough surface *Wear* **35** 87–111
[52] Putignano C, Afferrante L, Carbone G and Demelio G 2012 The influence of the statistical properties of self-affine surfaces in elastic contacts: A numerical investigation *J. Mech. Phys. Solids* **60** 973–82
[53] Gneiting T and Schlather M 2004 Stochastic models that separate fractal dimension and the hurst effect *SIAM Rev.* **46** 269–82
[54] Orey S 1970 Gaussian sample functions and the Hausdorff dimension of level crossings *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* **15** 249–56
[55] Schmittbuhl J, Vilotte J-P and Roux S 1995 Reliability of self-affine measurements *Phys. Rev.* E **51** 131–47
[56] Boffa J M, Allain C and Hulin J P 1998 Experimental analysis of fracture rugosity in granular and compact rocks *Eur. Phys. J. Appl. Phys.* **2** 281–9
[57] Ramisetti S B, Campañá C, Anciaux G, Molinari J F, Müser M H and Robbins M O 2011 The autocorrelation function for island areas on self-affine surfaces *J. Phys. Condens. Matter* **23** 215004
[58] Russ J C and Dehoff R T 2000 *Practical Stereology, Second* (US, Boston, MA: Springer)

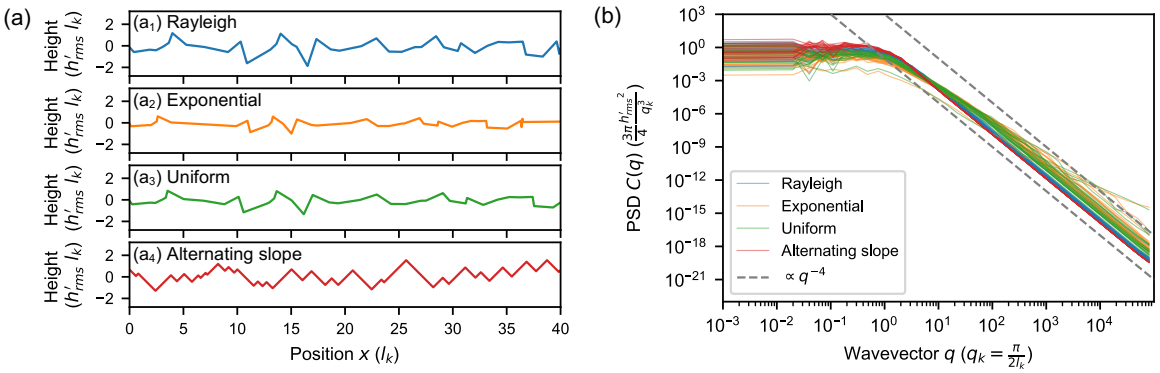
Supplemental Information for: Comprehensive topography characterization of polycrystalline diamond coatings

Abhijeet Gujrati[a], Antoine Sanner[b,c], Subarna Khanal[a], Nicolaie Moldovan[d], Hongjun Zeng[d,e], Lars Pastewka[b,c], Tevis D. B. Jacobs[a]

[a]Department of Mechanical Engineering and Materials Science, University of Pittsburgh, 3700 O'Hara St, Pittsburgh, PA, 15261, USA
[b]Department of Microsystems Engineering, University of Freiburg, Georges-Köhler-Allee 103, 79110 Freiburg, Germany
[c]Cluster of Excellence *liv*MatS, Freiburg Center for Interactive Materials and Bioinspired Technologies, University of Freiburg, Georges-Köhler-Allee 105, 79110 Freiburg, Germany
[d] Alcorix Co., 14047 Franklin Ct., Plainfield, IL 60544
[e] Aqua via Rock LLC, 712 Muirhead Ct., Naperville, 60565

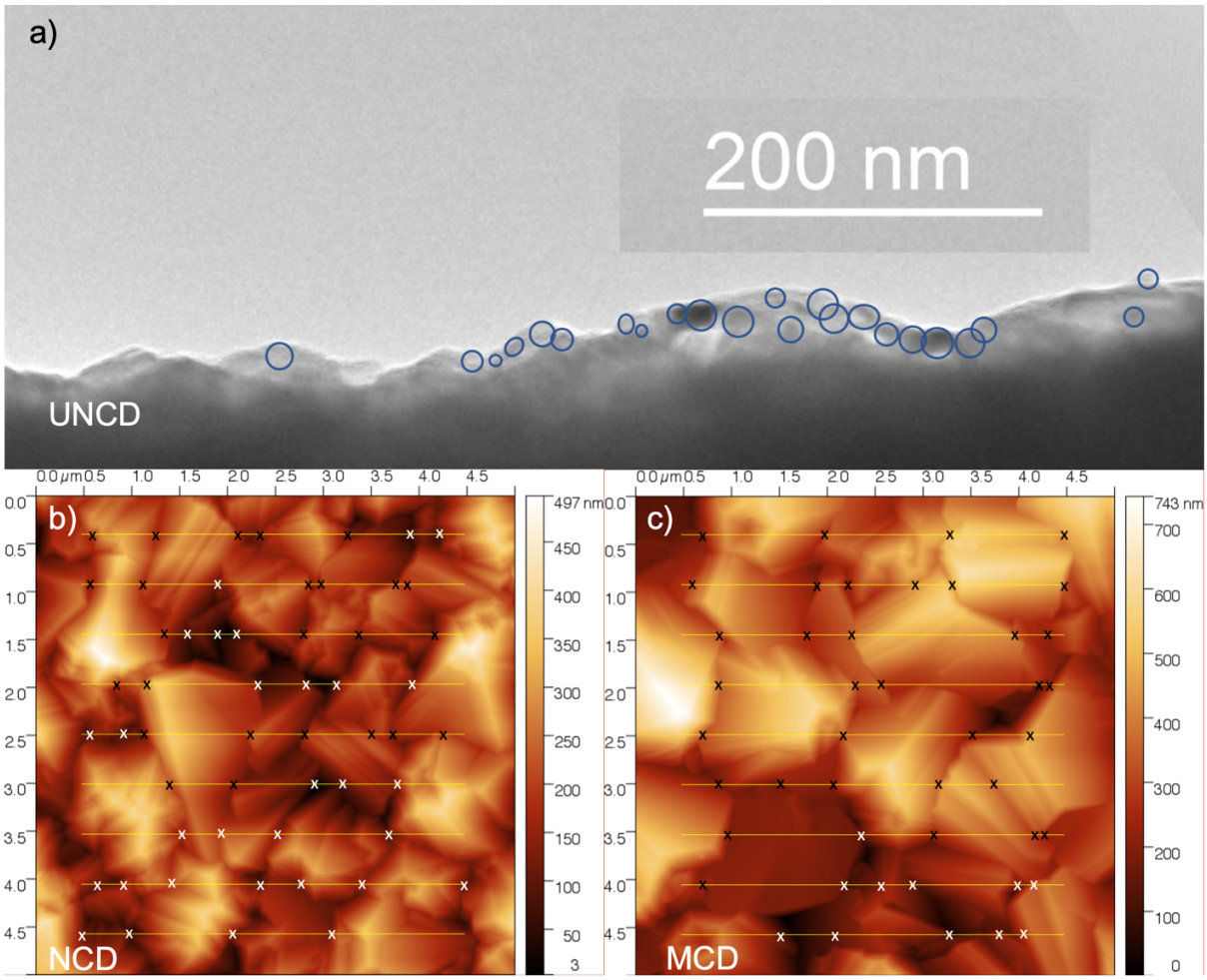
Supplemental Section 1: Alternate ways of generating faceted random surfaces

We show different ways to generate the faceted surfaces and how this affects the result. First, we replace the Rayleigh distribution of the kink distances used in the main text by an exponential and a uniform distribution (Fig. S1a$_2$ and a$_3$, respectively). Additionally, we generate a faceted surface following a different approach (Fig. S1a$_4$). The slope is alternating between –1 and 1, the height of each peak is drawn from a Gaussian distribution and the distance between each valley to its closest peak is drawn from a Rayleigh distribution. The distance to the other peak results from the height difference and the slope. Fig. S1b shows the PSDs of 20 realizations of these profiles, each having 200 kinks resolved on 10^7 pixels. For the uniform and exponential distribution of kink spaces, there is a broad transition from the flat regime to the proper q^{-4} scaling, due to the high probability of small kink distances. One realization of the exponential and uniform distribution has a minimum kink distance smaller than the pixel size, so that the q^{-4} regime is never reached. The profiles with alternating slopes have a sharp transition from flat to q^{-4} scaling similar to the profiles with random slopes and kinks spaced according to a Rayleigh distribution.



Supp. Figure 1: Faceted surfaces were generated using four different methods (a) with relatively little difference in their computed PSDs (b). Faceted surfaces were created (a_1 to a_3) with kink heights drawn from a Gaussian distribution and distances between kinks drawn from different distributions. Note that for a fixed RMS slope, the presence of close kinks (locally high slopes) in the surfaces with uniform and exponential distribution yields a smaller RMS height than for the Rayleigh distribution. Additionally, faceted surfaces were created with slope alternating between -1 and 1 (a_4); see text for details on the generation. PSDs are shown (b) for 20 realizations of each type of surface.

Supplemental Section 2: Measuring the grain size and mean kink spacing



Supp Figure 2: Calculation of mean kink spacing and grain size using TEM and AFM. TEM imaging (a) was used to determine the grain size for UNCD, where blue circles outline contrast variations that indicate the presence of a grain. The average diameter of these circles were used as a measure of grain diameter. For NCD (b) and MCD (c), 4-µm lines (yellow) were superimposed on 5-µm x 5-µm AFM images. A black or white 'x' was used to mark each intersection of a line with a grain boundary.

MCD and NCD

To get the mean kink spacing, first the mean lineal intercept was calculated following methods described in Ref. [1]. Specifically, an array of 9 lines of 4-µm length was super imposed on a 5-µm x 5-µm AFM scan. Then the number of intersections of these lines with grain boundaries were counted and denoted by the variable P (Fig. S2b, c). This gave the mean lineal intercept by using L/P, where L is length of all superimposed lines, which is $9 \times 4\ \mu m = 36\ \mu m$. This method was repeated in a perpendicular direction over 3 MCD and 3 NCD images and the average gave a mean lineal intercept of 839 ± 68 nm for MCD and 647 ± 42 nm for NCD. Finally, since this analysis includes only concave kinks, the mean kink spacing is calculated by dividing the mean lineal intercept by 2, to obtain 419 ± 34 nm for MCD and 323 ± 21 nm for NCD in real-space and $q_k =$

3.7×10^6 m^{-1} for MCD and $q_k = 4.9 \times 10^6$ m^{-1} for MCD in frequency-space (shown as vertical dashed black lines in Fig. 5d and e in main text).

Polished and unpolished UNCD

The same AFM-based procedure could not be applied for UNCD because the grain size approaches the reliability cut-off due to tip artifacts. So, instead a different approach was used; circles were superimposed (ovals if needed) on crystal grains that are visible as defined regions of TEM-contrast variation (Fig. S2a). Then, grain diameter was calculated for all circles (for ovals, using the mean of the two diameters) in calibrated units. Mean kink spacing was calculated by averaging grain diameters and then dividing by 2. The average grain diameter was calculated for UNCD as 14 ± 3 nm, so the mean kink spacing was 7 ± 2 nm in real-space and $q_k = 2.2 \times 10^8$ m^{-1} in frequency-space (shown as vertical dashed black lines in Fig. 5c in main text).

[1] J.C. Russ, R.T. Dehoff, Practical Stereology, Second, Springer US, Boston, MA, 2000. https://doi.org/10.1007/978-1-4615-1233-2.