A rigorous, field-theoretical approach to the contact mechanics of rough, elastic solids

Martin H. Müser*

(Dated: September 5, 2007)

A statistical field theory is formulated, which allows one to calculate the pressure distribution $\Pr(p)$ in a contact formed by an elastic body and a rigid counter face of arbitrary topography. It is a cumulant expansion, which contains Persson's contact mechanics theory as the leading-order term. Our approach provides a framework with which corrections can now be derived systematically. As an example, $\Pr(p)$ is calculated to high accuracy for exponentially repulsive solids. Non-Gaussian tails in $\Pr(p)$ can be rationalized for surfaces whose height spectra differ from colored noise.

PACS numbers: 81.40.Pq,46.55.+d

Contact mechanics plays a central role in tribology [1–3]. The dependence of frictional forces on the load can be a function of contact mechanics and the amount of wear that occurs under sliding depends critically on the distribution of contact pressures. However, despite its importance, for example, in the design of durable, low-weight and thus fuel-efficient aluminum engines [4], there is no generally accepted theory for contact mechanics. The lack of such a theory is mainly due to the multi-scale nature of surface topographies, which impedes our ability to predict contact pressures even for devices as small as microelectromechanical systems [5, 6].

Traditionally, contact mechanics calculations are based on so-called overlap models, which were pioneered by Greenwood and Williamson (GW) [7]. In overlap models, parabolic or elliptic hills (also called asperities) are identified and assumed to get flattened as they approach the counter face provided that their summit height in the undeformed surface is above a given threshold value. Depending on the distribution of asperity heights, radii of curvature, etc., and depending on the single-asperity contact mechanics, predictions are made, for example, for the relation between real contact area and external load. A valuable outcome of GW models is an explanation for the frequently observed linearity between load and true contact area, which occurs despite non-linear, single-asperity relations.

Despite their success and widespread use, overlap models suffer from limitations. First, it is often far from obvious what to define as an asperity, how to handle small asperities that reside on top of large asperities, and thus how to construct their statistics unambiguously. Second, overlap models neglect the elastic coupling between asperities [8]. They do not take into account that all asperities in the vicinity of a high asperity are being pushed down when that high asperity comes into contact with the counter face. This shortcoming produces contact areas that are too clustered as compared to numerically exact calculations, be it on self-affine surfaces [9] or on complex topographies of aluminum-silicon alloys [10].

An alternative approach to contact mechanics was suggested by Persson in 2001 [8]. Stress σ or pressure p were recognized as scale-dependent properties. The more microscopic details of the surface profiles are resolved the larger the (calculated/observed) fluctuations in p and thus the broader the pressure distribution $\Pr(p)$. Motivated by this insight, a diffusion equation for $\Pr(p)$ was derived, which describes the broadening of the magnification-dependent $\Pr(p,\eta)$ with increasing magnification η. In this theory, the diffusion coefficient D is proportional to the surface roughness on a wavelength that can just be resolved at a given magnification η. This parameterization leads to a broadening of $\Pr(p,\eta)$ with increasing η that is independent of the pressure p, which turns $\Pr(p)$ into a Gaussian. Constraints motivated by the observation that p cannot be negative for purely repulsive walls can be added a-posteriori, for example by adding an "image Gaussian" to the original Gaussian. This leads to a deviation from Gaussians at small values of p but leaves the tails at large p unaffected. A critical discussion of Persson theory was given by Manners and Greenwood [11].

When applied to contact mechanics, Persson theory and overlap models show, give or take, similar deviations from numerically exact calculations [9, 12, 13]. In the debate which approach is more accurate, one should keep in mind that neither one can be exact. The required input information, two-point height-difference correlation functions (HDCFs) in Persson theory and asperity statistics for GW models, is insufficient to fully reconstruct the stochastic properties of the surfaces, which in turn makes it impossible to predict rigorously $\Pr(p)$. However, an HDCF-type approach bears the promise that higher-order HDCFs can be included into it, whereby corrections could be added systematically until the desired accuracy is achieved. Providing a framework for doing this is the purpose of this work.

Starting point of our theory will be the formal definition of the pressure distribution $\Pr(p)$ as

$$\Pr(p) = \langle \delta\{p - \sigma(\mathbf{r})\} \rangle \tag{1}$$

where $\langle \dots \rangle$ stands for an average over the full interface and $\sigma(\mathbf{r})$ for the stress at the lateral position $\mathbf{r}$. Writing $\Pr(p)$ as $(1/2\pi) \int_{-\infty}^{+\infty} dk \, \langle \exp[ik\{p - \sigma(\mathbf{r})\}] \rangle$, allows one to apply the concept of cumulant expansions, i.e., the

exponential function can be expanded into a Taylor series, the expectation values of the various moments of the stress are taken, and the resulting expressions can be cast back into the exponential function up to the desired order [14]. To leading-order one obtains (see the mathematical appendix in EPAPS Document No. [] for additional intermediate steps in the algebra throughout the manuscript)

$$\Pr(p) = \frac{\exp\left\{-(p-p_0)^2/2\langle\Delta\sigma^2\rangle\right\}}{\sqrt{2\pi\,\langle\Delta\sigma^2\rangle}}\left\{1+\mathcal{O}(K_3)\right\} \quad (2)$$

Here, we have introduced the mean macroscopic pressure $p_0 = \langle\sigma\rangle$, the local fluctuation in stress $\Delta\sigma(\mathbf{r}) = \sigma(\mathbf{r}) - p_0$, its second moment $\langle\Delta\sigma^2\rangle$, and the third-order cumulant $K_3 = \langle\Delta\sigma^3\rangle - p_0\langle\Delta\sigma^2\rangle$.

At this point, we have essentially arrived at the solution of Persson's diffusion equation for $\Pr(p)$. The remaining steps are: Parameterize the stress fluctuation under the assumption that it is similar to the one that one would have under full contact (see the treatment below for clarification), add the mirror Gaussian in the case of repulsive walls. Then conserve the norm, $\int_0^\infty dp\,\Pr(p)$ must be unity, by adding a δ-function contribution at $p = 0$, and interpret the prefactor to the δ-function as the relative surface area that is not in contact with the counter face.

The new feature of Eq. (2) is that it provides a recipe for how to systematically include corrections in the form of higher-order cumulants. Expressions similar to those occurring in Eq. (2) can be derived systematically to arbitrary order in the cumulants, i.e., knowledge of the various moments of the stress allow one to reconstruct the pressure distribution. We thus need a scheme with which one can calculate the various moments or cumulants of the stress. In simple harmonic systems, the gradient of the displacement u can be related to the stress σ so that the knowledge of the various moments of ∇u will allow us to calculate the moments in σ, e.g.,

$$\langle\sigma^*(\mathbf{q})\sigma(\mathbf{q})\rangle = (qE')^2\,\langle u^*(\mathbf{q})u(\mathbf{q})\rangle/4, \quad (3)$$

where $E' = E/(1-\nu^2)$, E being the bulk elastic modulus and ν the Poisson ratio [8]. Similar relations hold for higher-order moments of the stress field. Before proceeding with the theory, we want to clarify that we assume spatial isotropy, the small-slope approximation, and that all the elasticity is mapped into one solid and all the roughness into the counter face [2]. Including anharmonicity is desirable but would require more complicated treatments. Phase field modeling may be a promising avenue for a future perturbative generalization of the present theory [15]. As of now, we will merely be concerned with purely harmonic solids.

For a given normal load and interaction between the two solids, the normal position $z(\mathbf{r})$ of the elastic solid's surface will depend on the rigid substrate's topography $h(\mathbf{r})$. To streamline the representation, we will sometimes write $z(\mathbf{r}) = z_0 + u(\mathbf{r})$, where z_0 will be called the centroid. $u(\mathbf{r})$ plays the role of a displacement field with $\langle u(\mathbf{r})\rangle = 0$. Furthermore, we will chose the coordinate system such that $\langle h(\mathbf{r})\rangle = 0$. Tildes will indicate Fourier transforms, e.g., $\tilde{u}(\mathbf{q}) = (1/A)\int d^2r\exp(i\mathbf{qr})u(\mathbf{r})$, and $\mathbf{q}$ is a wave vector compatible with the in-plane periodic boundary conditions. A is the "macroscopic" contact area. If the system is sufficiently well behaved, it will be possible to expand the $\tilde{u}(\mathbf{q})$ into a power series of the $\tilde{h}(\mathbf{q})$. The leading-order, symmetry-allowed terms (wave vectors in each term on the r.h.s. of the following equation have to add up to $\mathbf{q}$) have the form

$$\tilde{u}(\mathbf{q}) = G_1(\mathbf{q})\tilde{h}(\mathbf{q}) + \sum_{\mathbf{q}'} G_2(\mathbf{q},\mathbf{q}')\tilde{h}(\mathbf{q}-\mathbf{q}')\tilde{h}(\mathbf{q}') + \ldots, \quad (4)$$

where the expansion coefficients $G_n(\mathbf{q},\mathbf{q}',...)$ will in general depend on the normal load and the interaction between the surfaces, see the treatment for a model system further below.

Given Eq. (4), the second moment or second-order cumulant of $\tilde{u}(\mathbf{q})$ can be written as

$$\begin{aligned}\langle|\tilde{u}(\mathbf{q})|^2\rangle &= |G_1(\mathbf{q})|^2\left\langle\tilde{h}^*(\mathbf{q})\tilde{h}(\mathbf{q})\right\rangle \\ &+ G_1^*(\mathbf{q})G_2(\mathbf{q},\mathbf{q}')\left\langle\tilde{h}^*(\mathbf{q})\tilde{h}(\mathbf{q}-\mathbf{q}')\tilde{h}(\mathbf{q}')\right\rangle + \text{c.c} \\ &+ \ldots \end{aligned} \quad (5)$$

Note that the term related to the third power in $\tilde{h}$ along with many other not explicitly-mentioned terms disappear for colored noise, that is, when $\tilde{h}(\mathbf{q})$ and $\tilde{h}(\mathbf{q}')$ are independent random variables for all $\mathbf{q} \neq \mathbf{q}'$. Thus, any theory (such as Persson's) that truncates after the first term on the r.h.s. of Eq. (5) and that furthermore neglects the higher-order cumulants on the r.h.s. of Eq. (2) will be less accurate for surfaces whose height spectra show correlation than for surfaces whose topography can be characterized as colored noise.

To close our theory, it is necessary to find an expression for the expansion coefficients $G_n(\mathbf{q})$. These are load and system-specific and can thus be derived only if the interaction between the elastic solid and its counter face are known, because one needs this interaction to minimize the total energy E of the system. Generally, the interaction between the solids will be a functional $\mathcal{V}[g]$ of the gap $g(\mathbf{r}) = z_0 + u(\mathbf{r}) - h(\mathbf{r})$ between elastic solid and counter face. The existence of such a functional could be proven in a similar way as that for any other functional, e.g., as those used in either density functional theory of electrons or classical fluids. One can thus write

$$E = \mathcal{V}[g] + \frac{A}{4}\sum_{\mathbf{q}} qE'\,|\tilde{u}(\mathbf{q})|^2 + \int d^2r\,p_0 z(\mathbf{r}) \quad (6)$$

where the term containing E' reflects the elastic energy and p_0 is the external pressure. The solution for the displacement field can be obtained by requirering that the

functional derivative $\delta E/\delta z(\mathbf{r})$ vanishes for all $\mathbf{r}$. For adhesive interactions, there can be more than one solution.

Whenever $\mathcal{V}[g]$ is sufficiently well-behaved, one can expand the energy density into a power series of the gap function and its derivatives. Unfortunately, this is not the case for hard wall interactions, which commonly form the basis for contact mechanics theories. While $\mathcal{V}[g]$ takes a simple form in that case, specifically, $\mathcal{V}[g] = 0$ if $g(\mathbf{r}) \geq 0$ everywhere and infinity otherwise, it is not possible to expand it into a power series of the Fourier coefficients $\tilde{g}(\mathbf{q})$. Consequently, the $\tilde{u}(\mathbf{q})$'s are non-analytical functions of the $\tilde{h}(\mathbf{q})$'s. We find numerically that $G_1(\mathbf{q}) = 1$ if $q\tilde{h}(\mathbf{q})$ is less or equal a load-dependent threshold value while else $G_1(\mathbf{q}) < 1$.

The problems discussed in the previous paragraph should not prevent one from pursuing our theory. First, real systems do not repel in a discontinuous fashion. In most realistic molecular simulations, repulsion is modeled via exponential functions. Second, one can construct differentiable potentials that mimics the behavior of hard walls. Local exponential repulsion between walls show precisely the desired characteristics, because normal forces increase exponentially fast with overlap and decrease exponentially as the surfaces retract from one another. Conclusions drawn for exponentially repulsive walls will thus generalize to a significant degree to that of hard walls. We wish to note that exponential repulsion has been used extensively in modeling tribological phenomena [16–18] and that pressure profiles resemble that of hard wall interactions quite closely when parameterized appropriately [19].

Finding coefficients G_n for any well-behaved $\mathcal{V}[g]$ can be done in a fashion that is similar to the one, which we will pursue here for exponential repulsion:

$$\begin{aligned} \mathcal{V}[g] &= v_0 \int d^2r\, e^{-g(\mathbf{r})/\zeta} \\ &= v_0\, e^{-z_0/\zeta} \int d^2r\, e^{\{h(\mathbf{r})-u(\mathbf{r})\}/\zeta}. \end{aligned} \tag{7}$$

v_0 is a constant prefactor of unit energy per surface element. Its precise value turns out irrelevant for the pressure distribution as long as it is positive (increasing its value would merely decrease z_0 at fixed p_0), and ζ is a parameter of unit length. One can interpret ζ as a length-scale over which interactions are smeared out and thus associate it loosely with the inverse magnification used in Persson theory.

In our model, p_0 satisfies the equation

$$p_0 = \frac{\mathcal{V}[u]}{\zeta A}, \tag{8}$$

whenever $u(\mathbf{r})$ minimizes E, which can be shown by requiring $\partial E/\partial z_0$ to disappear. This can be exploited in the minimization procedure, which in principle can be done to arbitrary accuracy by expanding the integrand of Eq. (7) into a power series, which could then be followed by an iterative solution for the $\tilde{u}(\mathbf{q})$. Here, we will content ourselves with a second-order cumulant expansion of $\mathcal{V}[g]$ so that

$$\begin{aligned} \frac{E}{A} \;\approx\; v_0 \exp\left\{ -\frac{z_0}{\zeta} + \sum_{\mathbf{q}} \frac{\left|\tilde{h}(\mathbf{q}) - \tilde{u}(\mathbf{q})\right|^2}{2\zeta^2} \right\} \\ + \sum_{\mathbf{q}} \frac{E' q \left|\tilde{u}(\mathbf{q})\right|^2}{4} + p_0 z_0 \end{aligned} \tag{9}$$

At that level of approximation, one can solve exactly for the $\tilde{u}(\mathbf{q})$ to yield the leading expansion coefficient

$$G_1(\mathbf{q}) = \frac{1}{1 + \zeta q E'/2p_0}. \tag{10}$$

$G_1(\mathbf{q})$ has the meaningful property that it predicts less good contact for large q and small p_0.

While we have not gone yet beyond harmonic corrections, it is worth discussing the effect that they and deviations from colored noise would have if these corrections were mapped onto the diffusion equation representation: First, higher-order corrections would induce a stress-dependent drift term to the diffusion equation and perhaps more importantly, it would also render D stress dependent. (In addition, following Pawula's theorem [20], one may need to add an infinite number of terms to the drift/diffusion equation for an *exact* theory.) Assuming that fluctuations in stress are large upon a change in magnification η and following the laws of large numbers, one may assume that $D \propto \sigma$ for large σ. Such a diffusion coefficient would then result in an exponential rather than in a Gaussian tail. Exponential tails were indeed identified by Campaña and Müser in GFMD calculations for experimental (non-colored noise) surface topographies [13].

The diffusion equation representation will also have to be modified at small p when repulsion is long ranged. Stress would not vanish completely and thus the hard-wall δ-function peak in $\Pr(p)$ at $p = 0$ will be broadened. Persson uses the earlier-mentioned mirror Gaussians to enforce $\Pr(p < 0) = 0$. This method is identical to setting D to zero for $p \leq 0$. For continuously decaying repulsion, the cutoff in D will have to be smoothed. If one wishes to switch D from zero to its default value over a finite pressure range so that D and its first two derivatives are continuous functions of p, then the correction factor $\{1 - \theta(p_\mathrm{c} - p)\cos(\pi p/p_\mathrm{c})\}$ appears to be the most simple choice. where $\theta(\dots)$ is the Heaviside step function. This way D remains unaltered above the cutoff stress p_c and thus does not noticeably affect width and tail of $\Pr(p)$ but only the way in which the $\delta(p)$-peak of $\Pr(p)$ broadens.

It is now possible to compare pressure histograms that are based on the new theory to numerical simulations,

which are obtained with Green's function molecular dynamics (GFMD) [19]. The system that was investigated has roughness on wavelengths $32 \leq \lambda \leq 1024$, the Hurst roughness exponent is $H = 0.3$, and the root-mean-square slope is 0.031. Furthermore, $E' = 8/3$ and $p_0 = 0.004$ in the same units. ζ is varied between 0.001 and 4. The unit for length scale is defined by the distance between two grid points in the numerical solution. Except for replacing hard-wall with exponential repulsion, we proceed in the same way as in Ref. 13. Fig. 1 summarizes the comparison between theory and simulation. Values of p_c used for our calculations were $p_c = 0.0027$ for all curves with $\zeta \geq 0.25$ and $p_c = 0.0082[0.0068]$ for $\zeta = 0.01[0.001]$. These three values for p_c are the only adjustable coefficients used in Fig. 1.

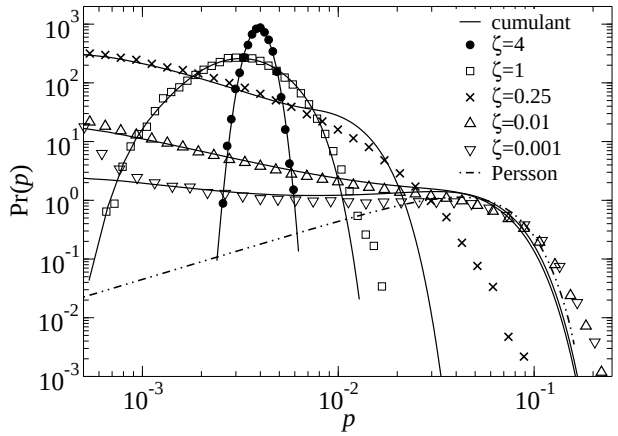

FIG. 1: Pressure distribution $\Pr(p)$ obtained for a fractal surface with different values of ζ. Symbols show numerical data, full lines correspond to calculations based on our approach. The dashed line corresponds to the original theory by Persson, to which our theory reduces in the limit $\zeta \to 0$.

Fig. 1 supports the claim that the theory presented here allows one to derive systematic corrections to Persson theory, as for large ζ theory and simulations agree flawlessly. When ζ decreases, the second-order cumulant expansion cannot be accurate any longer and consequently deviations between theory and simulations arise. Surprisingly, at very small values of ζ the curvature of $\ln \Pr(p)$ is only off by about 20% and thus more accurate than at intermediate values of ζ, which is why the relatively good agreement at small ζ must be fortuitous. This argument is supported by the observation that the stress-autocorrelation function obtained numerically at low loads does not match Persson's parameterization of the diffusion coefficient for small relative contact areas [9]. Unfortunately, the similarly good predication for the curvature of $\ln \Pr(p)$ made by overlap models must likewise be seen as fortuitous, because their contact autocorrelation functions are also in contradiction with computational results [9, 10].

In conclusion, we have presented a field theory for the contact mechanics of a harmonic solid pressed against a rough, rigid counter face. The theory starts from the formal definition of the pressure distribution and does not contain any *uncontrolled* approximations. The leading-order term corresponds to Persson's contact mechanics theory. Harmonic corrections greatly improve the agreement between theory and numerically exact calculations. It remains to be seen if including a few more higher-order terms than those currently used will suffice to address the complicated contact mechanics of real-life surfaces.

* mmuser@uwo.ca;

Acknowledgment: The author is grateful to Carlos Campaña for providing him with the stress distribution functions of exponentially repulsive walls. Enlightening discussions with him and Mark Robbins as well as financial support from NSERC are also gratefully acknowledged.

[1] F. P. Bowden and D. Tabor. Wiley, New York, 1956.
[2] K. L. Johnson. Cambridge University Press, Cambridge, 1966.
[3] B. N. J. Persson, O. Albohr, U. Tartaglino, A. I. Volokitin, and E. Tosatti. *J. Phys.: Condens. Matter*, 17:R1, 2005.
[4] R. L. Deuis, C. Subramanian, and J. M. Yellup. *Compos. Sci. Technol.*, 57:415, 1997.
[5] Y. P. Zhao, L. S. Wang, and T. X. Yu. *J. Adhes. Sci. Technol.*, 17:519, 2003.
[6] E. E. Flater, A. D. Corwin, M. P. de Boer, and R. W. Carpick. *Wear*, 260:580, 2006.
[7] J. A. Greenwood and J. B. P. Williamson. *Proc. R. Soc. London*, A295:300, 1966.
[8] B. N. J. Persson. *J. Chem. Phys.*, 115:3840, 2001.
[9] S. Hyun and M. O. Robbins. *Tribol. Int.*, page doi:10.1016/j.triboint.2007.02.003, in press.
[10] C. Campañá, M. H. Müser, C. Denniston, Y. Qi, and T. Perry. *J. Appl. Phys.*, page submitted, 2007.
[11] W. Manners and J. A. Greenwood. *Wear*, 261:600, 2006.
[12] S. Hyun, L. Pei, J.-F. Molinari, and M. O. Robbins. *Phys. Rev. E*, 70:026117, 2004.
[13] C. Campañá and M. H. Müser. *Europhys. Lett.*, 77:38005, 2007.
[14] H. Risken. *The Fokker-Planck Equation*. Springer, Berlin, 1984.
[15] K. R. Elder, M. Katakowski, M. Haataja, and M. Grant. *Phys. Rev. Lett.*, 88:245701, 2002.
[16] V. Zaloj, M. Urbakh, and J. Klafter. *Phys. Rev. Lett.*, 82:4823, 1999.
[17] M. H. Müser, L. Wenning, and M. O. Robbins. *Phys. Rev. Lett.*, 86:1295, 2001.
[18] C. Campaña. *Phys. Rev. B*, 75:155419, 2007.
[19] C. Campañá and M. H. Müser. *Phys. Rev. B*, 74:075420, 2006.
[20] R. F. Pawula. *Phys. Rev.*, 162:186, 1967.