

Theory of the effects of multiscale surface roughness and stiffness on static friction

J. B. Sokoloff

*Physics Department and Center for Interdisciplinary Research in Complex Systems, Northeastern University,
Boston, Massachusetts 02115, USA*

(Received 14 July 2005; published 6 January 2006)

It is shown on the basis of simple scaling arguments that an interface between two three-dimensional elastic solids, consisting of completely flat disordered surfaces, which interact with interatomic hard core interactions, will be in a weak pinning regime, and hence exhibit negligibly small static friction. It is argued, however, that the presence of roughness on multiple length scales can lead to much larger friction (i.e., static friction coefficients not too much smaller than 1), as is characteristic of most solid surfaces. This approach suggests a possible way of understanding why coatings of materials with high elastic constants are often excellent lubricants.

DOI: [10.1103/PhysRevE.73.016104](https://doi.org/10.1103/PhysRevE.73.016104)

PACS number(s): 81.40.Pq, 68.35.Af, 46.55.+d

I. INTRODUCTION

Surface roughness and micron and submicron scale roughness are shown to have a crucial effect on static friction between two three-dimensional elastic solids that interact at their interface primarily through interatomic hard core interaction, which will be the case for sufficiently high load [1]. Although plasticity will likely play an important role in many circumstances [2], it is important to understand static friction for an elastic solid as a first step toward understanding the microscopic mechanisms for friction. Furthermore, the elastic solid model is clearly valid in the low load regime. It is shown in this paper that most such surfaces under typical loads would have negligible static friction if it were not for the presence of roughness on multiple length scales.

Static friction is often easier to study than kinetic friction, and since in many applications, static and kinetic friction are generally of the same order, we often gain some insights into kinetic friction in the slow sliding speed limit from such studies [3,4]. In Refs. [3,5] it was argued that a flat disordered interface between two macroscopic size surfaces which do not interact chemically will exhibit effectively no static friction for interface interaction per unit interface area small compared to the shear elastic constant and high friction for interface interaction above this value. This frictionless regime is known as the weak pinning regime and the large friction regime is known as the strong pinning regime. In Ref. [5], it was suggested that this could provide possible mechanisms for lubrication. The lubricants in this model work by switching the interface from the strong to weak pinning regime by increasing the number of atoms in contact at the interface, and hence spreading the load of normal force pushing the solids together over more points of contact. In fact, it will be argued later in this paper that friction between surfaces that interact primarily by hard core interactions is primarily due to the presence of multiscale surface roughness on most surfaces [6]. Without such roughness, the static friction would typically be negligibly small. The idea that increasing the number of atoms in contact can actually reduce, rather than increase, the force of friction appears to contradict the prevailing assumption that increasing the area of contact increases the friction [2]. Reference [7] also chal-

lenged the assumption that friction is proportional to the area of contact on the grounds that the real area of contact is generally not well defined when the interface is viewed on the atomic scale. Nevertheless, they were able to obtain Amontons's law (that the friction depends only on the force pushing the surfaces together and not on the apparent interface area) without invoking this concept. In fact, it will be seen that, at least for static friction, it is possible to explain Amontons's law by a very simple argument that does not involve the real area of contact at all. Another possible mechanism for reducing friction, which was not discussed in Ref. [5], is to coat the surfaces with a sufficiently thick coating of a material with large elastic constants, as the transition from strong to weak pinning will occur at a larger value of the interfacial interaction if the elastic constants are greater [8]. In this paper, static friction will be discussed from the point of view of collective pinning and multistability theory for nonchemically interacting surfaces, which are pushed together by a sufficiently large normal force so that the interfacial interaction is dominated by hard core interaction between the atoms in contact at the interface. This model can be applied to nonmetallic solids and some oxide coated metals, with saturated bonds at the interface, so that there will be no chemical bonding across the interface. The discussion will focus on how collective pinning and multistability arguments are affected by surface roughness, in particular by the existence of asperities on several length scales.

II. STATIC FRICTION DUE TO HARD CORE INTERACTIONS

Consider two surfaces in contact which are disordered so that those atoms which are in contact are randomly distributed over the interface. Let us also assume that the atoms in contact at the interface interact only with hard core interactions. This could occur either because the surface atoms are chemically inert and there is negligible adhesion, or because the surfaces are being pushed together with a sufficient load so that the hard core interactions dominate. Let P denote the load or normal force per unit interface area, a , the mean atomic spacing, and c , the fraction of the surface atoms of

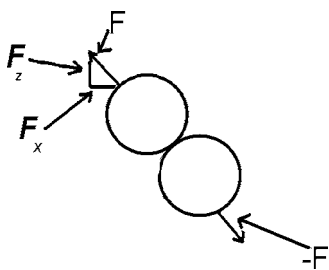


FIG. 1. This figure illustrates how the hard core interaction between a pair of atoms, one belonging to each of the surfaces in contact, can both support the load and give rise to static friction between the surfaces. Since the force F between the pair of atoms can have both a component normal to the interface, F_z , which contributes to the normal force supporting the load, and a component along the interface F_x , the mean value of F_x must be proportional to the mean value of F_z .

one surface that are in contact with the second surface. Then, each of the atoms in contact must contribute on the average to the normal force, a force of order Pa^2/c . Since the force due to the hard core interaction between a pair of atoms acts along the line joining the atoms, for most relative positions of the atoms, it has a component along the interface, as illustrated in Fig. 1. In the strong pinning regime, each surface atom will sink into an interface potential minimum at the expense of the elastic forces holding it in place. Such a minimum will generally occur at an interstitial region on the second surface. If we now attempt to slide the surfaces relative to each other, each of the atoms in contact with the second surface will now, as it gets pulled out of its potential minimum, exert a component of its hard core interaction with the second surface parallel to the interface directed so as to oppose the attempted sliding motion. This is identified with the static friction. Since each of these atoms must also provide a component Pa^2/c normal to the interface on the average, it is clear the static friction is proportional to the load. This is the case because the contribution to the load and the static friction for each pair of atoms in contact is provided by the same hard core force acting between the atoms. The proportionality constant μ_s is identified with the coefficient of static friction which is not too much smaller than 1. This accounts for Amontons's law without the need to assume that the friction is proportional to an ill-defined area of real contact. In the weak pinning limit, the component along the interface of the hard core force is random, and hence, for an infinite interface area and, hence, an infinite number of interface atoms, the components along the interface of the hard core forces cancel, resulting in effectively no static friction in the thermodynamic or macroscopic solid limit.

III. EFFECTS OF ROUGHNESS (I.E., ASPERITIES ON SEVERAL LENGTH SCALES)

Let us first consider two flat disordered surfaces. For two perfectly flat surfaces in contact pushed together with sufficiently weak load, each containing N surface atoms, the mean force of static friction will be the sum of the components of these hard core forces parallel to the surface, which

have random magnitudes and directions. The net static friction, which is the sum of these components, is proportional to $N^{1/2}$ because they do not add together coherently. This is known as the weak pinning regime [9–11]. It was shown in Refs. [3,9–11] that for sufficiently weak interaction across the interface (which will be true at low loads), the system will be in this regime. (Although mica is probably the only material with complete atomic scale flatness, bare mica surfaces actually exhibit relatively high friction because they interact with each other quite strongly.) At higher loads, the atoms from each surface can be pushed sufficiently far into the regions between atoms on the second surface (resulting in the atoms from the second surface being pushed apart), as this will minimize the repulsive energy. When we attempt to slide the surfaces relative to each other, there will be a net component of hard core repulsive forces along the surface opposing the sliding, i.e., static friction. In this case, the net force of static friction will be proportional to N . For models for surfaces with single length scale roughness, like the Greenwood-Williamson model [2], since there are of the order of 10^8 atoms at an interface between two micron size asperities, since $N^{1/2}$ is 10^4 , the static friction is a factor of 10^4 smaller in the weak pinning than in the strong pinning regime. From Eq. (A3) in Appendix I, which reproduces some of the results of Ref. [5], we find that for flat surfaces, the load per unit area P below which the interface is in the weak pinning regime is given by $P \approx c^{1/2} K$, where c is the fraction of surface atoms in contact with the substrate and K is the shear elastic constant. Since $K \approx 10^{11} N/m^2$ for most solid materials, unless c is extremely small, we find on the basis of this argument that any flat surfaces in contact that interact with only hard core interactions will be in the weak pinning regime.

Of course, no surfaces are perfectly smooth, and in fact, there can be roughness on several length scales. Let us assume then that there are n orders of length scales, which we represent as follows: We divide the surface into M_n asperities, a fraction c_n of which are in contact with the second surface. In the discussion in this paragraph, the asperities are treated as rigid “hills” on the surface. Effects of distortion of the asperities is treated later in this section. At each pair of asperities in “contact,” the contact is likely to only occur at selected isolated regions, which we may refer to as $n-1$ order asperities. The interface between a pair of $n-1$ order asperities can be divided up into $n-2$ order asperities as well. The surfaces can never be truly self-affine [13], however, because when we reach atomic dimensions at $n=0$ order, this subdivision into smaller and smaller length scales terminates. In more detail, we divide an N -atom surface into M_n asperities of which a fraction c_n are in contact with the second surface, which for purposes of obtaining a qualitative understanding of the problem, can be replaced by a flat substrate, as described above. We then divide the areas of contact of each of these contacting asperities into M_{n-1} asperities, of which a fraction c_{n-1} are in contact. We then divide the area of contact between two contacting asperities into M_{n-2} asperities, a fraction c_{n-2} of which are in contact, etc., until we have done n sublevels of this subdivision. The area of contact of a zeroth level (i.e., smallest) asperity will contain of the order of $N_0 = N / (M_0 M_2 \dots M_n)$ atoms, a fraction c_a

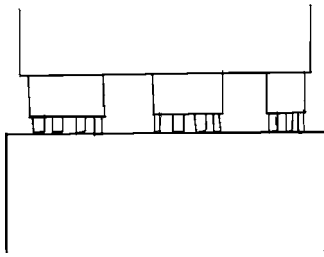


FIG. 2. This is a schematic illustration of the asperity hierarchy on the top surface sliding on a flat substrate (i.e., the bottom block). (Real asperities have arbitrary shapes, as opposed to the square shapes shown in this schematic representation.) Each asperity of a given order has a number of (smaller) asperities of one order lower on its surface. In turn, each of these lower order asperities has a number of (smaller) asperities of one order lower. This continues until we reach the zeroth order asperity, whose surface consists of atoms, although only three orders of asperities are illustrated here.

of which are in contact. It is these atoms at the zeroth (i.e., the final) order of asperities which support the load.

If the (n_1) st order set of asperities are in the weak pinning regime, the static friction acting on it is reduced by a factor $(c_{n_1}M_{n_1})^{1/2}$, because by the above arguments, the static friction forces from these asperities act incoherently, and if the atoms at the interfaces of zeroth order miniasperities in contact are in the weak pinning regime, the static friction is reduced from the strong pinning regime value (i.e., μ_s not too much smaller than 1) by a factor $(c_a N_0)^{1/2}$. Thus the mechanism is proposed as a possible way to explain why coatings of stiff materials are good lubricants [8].

The arguments in the above paragraphs are only correct if each asperity is completely rigid, as we have assumed that the elastic forces which oppose displacements of the points of contact with the substrate resulting from the forces that the substrate exerts on them are due to the bulk solid. Since each asperity has some height, however, it can distort so as to move its interface with the substrate closer to its potential minimum, even without distorting the bulk solid. Thus, whereas a particular substrate with completely rigid asperities might be in the weak pinning regime, if the asperities are able to distort by a sufficient amount, it might be in the strong pinning regime. This may explain why it is that although the estimates given earlier in this section indicate that most solids should be in the weak pinning regime, this is not consistent with the magnitudes of the friction coefficients that are observed for most solids.

Let us now consider the distortions of the asperities that occur in response to the substrate potential. What we will do now is to assume that there exist a bunch of smallest asperities (which will be considered the lowest or zeroth level) which are in contact with the substrate. There are several groups of these that are assumed to be attached to a bunch of larger next or first order asperities. Groups of these first order asperities are then attached to larger asperities which are known as second order asperities. This hierarchy continues until we reach an “asperity” of width equal to that of the whole interface. This hierarchy of asperities is illustrated schematically in Fig. 2. Consider the zeroth, the lowest order (i.e., the smallest), asperity. Let it have a height of order L'_0

and a width of order L_0 . To find its distortion resulting from the sum of the substrate potential energies of all of the atoms of the asperity which are in contact with the substrate, we must minimize the sum of its elastic and substrate potential energies. The substrate potential energy is given by $V_0(L_0/a)f_0(\Delta x_0/a)$, where V_0 is the amplitude of the interaction of a single atom with the substrate, resulting primarily from hard core repulsions between the atoms, Δx_0 is the amount that the surface in contact with the substrate slides under the influence of the substrate potential as the asperity distorts while all higher level asperities remain in an arbitrary rigid configuration and $f_0(\Delta x_0/a)$ is a function of order unity which gives the variation of the substrate potential with Δx_0 for fixed, undistorted asperities of higher order (i.e., larger size in the present context). (Clearly, each of the zeroth order asperities has a different function; f_0 denotes a generic function describing the interface potential energy for a typical zeroth order asperity.) Each function clearly must possess multiple minima. We are assuming here that the surface of the asperity in contact with the substrate is in the weak pinning limit. The factor (L_0/a) , which is of the order of the square root of the number of atoms in this surface expresses this fact. If the surface of the asperity in contact with the substrate is in the strong pinning limit instead, this factor will be replaced by $(L_0/a)^2$, the number of atoms at the interface. Treating this asperity as an elastic three-dimensional solid in contact with the substrate, we find from the discussion in Appendix I that the interface between the substrate and this asperity is in the weak pinning limit if $P_0 < c_a^{1/2}K$, where P_0 is the mean load per unit interface area supported by this asperity and c_a is the fraction of the surface atoms of this asperity that are in contact with the substrate. Assume that a fraction c_0 of the zeroth order asperities have atoms belonging to them in contact with the substrate. Let c_1 represent the fraction of next order (first order) asperities whose zeroth order asperities are in contact with the substrate, c_2 , the fraction of second order asperities whose first order asperities have their zeroth order asperities in contact with the substrate, etc., up to n th order. Then $P_0 = P/(c_0 c_1 c_2 \cdots c_n)$, where P is the load per unit apparent area of the surface of the whole solid. Then, we conclude that the criterion for the atoms at the interface between the zeroth order asperity and the substrate to be in the weak pinning regime is that $P < (c_a^{1/2} c_0 c_1 c_2 \cdots c_n)K$. We see from this inequality that the more fractal the surface is, the more difficult it is for the zeroth order asperity to be in the weak pinning regime. The cost in elastic energy due to the shear distortion of the asperity can be determined by the following scaling argument: The elastic energy density for shear distortion of the asperity is proportional to $(\partial u_x / \partial z)^2$, where u_x represents the local displacement due to the distortion, the x direction is along the interface and the z direction is normal to it. The u_x must scale with Δx_0 and the dependence of u_x on z has a length scale L'_0 . Thus the elastic strain energy of the asperity is of the order of $(1/2)L_0^2 L_0' K (\Delta x_0 / L_0')^2$, where K is the shear elastic constant and $(\Delta x_0 / L_0')$ is the average shear strain and L_0 is the mean width of the asperity. Minimizing the sum of these expressions for elastic and substrate potential energy, we obtain

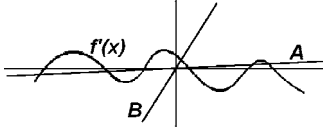


FIG. 3. This figure illustrates the solution of Eqs. (1), (2), and (3), for Δx_0 , Δx_1 , and Δx_n , respectively. $f'(x)$ is a schematic illustration of the functions f'_0 , f'_1 , and f'_n , and x denotes Δx_0 , Δx_1 , or Δx_n , respectively. Lines A and B represent the line $y = (Ka^3/V_0)(L_n/L'_n)x$, for $Ka^3/V_0)(L_n/L'_n) < 1$ and $Ka^3/V_0)(L_n/L'_n) > 1$, respectively. For the situation illustrated by line A, there are multiple solutions (i.e., multistability), while for the situation illustrated by line B, there is only one (monostability).

$$\Delta x_0/a \approx (V_0/Ka^3)(L'_0/L_0)f'_0(\Delta x_0/a), \quad (1)$$

since f'_0 , the derivative of f_0 with respect to its argument, it is of order one, from the definition of f_0 . Let us follow a line of reasoning like that of Ref. [14], a modified version of which is given in Appendix II. For $(V_0/Ka^3)(L'_0/L_0)$ below a certain value of order one, for small V_0/Ka^3 , Eq. (1) can have only one solution for Δx_0 . The reason for this is illustrated in Fig. 3. Under such circumstances, the average kinetic friction, in the limit as the sliding velocity approaches zero, is zero. For a surface with an infinite number of asperities, distributed uniformly in space, it was shown in Ref. [14] that the static friction is zero as well. A modified version of this argument, which points out that for a surface with a finite number of asperities the static friction is nonzero, but smaller by a factor of a/L_0 compared to what it would be if the contributions of the asperities to static friction acted coherently is provided in Appendix II. If this asperity is in the strong pinning limit instead, we replace the factor of (L_0/a) by $(L_0/a)^2$ to account for this and as a result, the factor (L'_0/L_0) gets replaced by (L'_0/a) , which could easily make the asperity satisfy the criterion for multistability, and consequently, the friction from these asperities will no longer be reduced by the factor a/L_0 . For a load per unit area P , assumed to be primarily due to hard core interactions, we may assume $V_0 \approx Pa^3/c$, where c is the fraction of the surface atoms which are in contact with the substrate. By the above arguments, $c = c_a c_0 c_1 c_2 \cdots c_n$. Then, we see that the criterion for the zeroth order asperity to be multistable is $P > c_0 c_1 c_2 \cdots c_n K$. If the criterion for weak pinning for the zeroth order asperity surface is not satisfied, the criterion for monostability of this asperity gets changed from the above inequality to $(V_0/Ka^3)(L'_0/a) < 1$, which is more difficult to satisfy since L'_0/a can be considerably greater than 1.

At the next level, we have an asperity surface in contact with the substrate which consists of a collection of the lowest level (i.e., the smallest) asperities discussed in the previous paragraph. Assuming this asperity to be in the weak pinning regime, the potential of interaction with the substrate, which is the sum of all of the interactions of the substrate with the lowest order asperities, which cover a first order asperity, is of order $V_0(L_0/a)(L_1/L_0)f_1(\Delta x_1/a)$. Here L_1 represents the width of this order asperity, Δx_1 represents a displacement of the lower surface of this level asperity for fixed (i.e., undistorted) configurations of all higher order asperities, and f_1

denotes one of the functions which describes the interface potential energy of one of the first order asperities. It has at least one minimum and runs over a range of magnitude one as its argument runs over a range of order one. The elastic energy is of the order of $(1/2)L'_1 L_1^2 K(\Delta x_1/L'_1)^2$, by the argument given above Eq. (1), where L'_1 is the height of the body of the first order asperity, which is assumed to be much larger than L'_0 . Minimizing the sum of these two energies, we obtain

$$\Delta x_1/a \approx (V_0/Ka^3)(L'_1/L_1)f'_1(\Delta x_1/a). \quad (2)$$

Again, we conclude, based on the arguments presented in Ref. [14], which are summarized in Appendix II, that the static friction is reduced by a factor of L_0/L_1 below what it would be if the contributions to the static friction from each of the miniasperities at this level acted coherently. If the zeroth order asperities attached to this first order asperity are in the strong pinning regime, the factor of L_1/L_0 in the equation for the interaction of this asperity with the substrate is replaced by $(L_1/L_0)^2$, and hence, the right hand side of Eq. (2) has the factor L'_1/L_1 replaced by L'_1/L_0 , which can make the solutions to this equation for Δx_1 multistable.

Continuing this procedure, we find that the displacement of the n th level miniasperity is found by solving

$$\Delta x_n/a \approx (V_0/Ka^3)(L'_n/L_n)f'_n(\Delta x_n/a), \quad (3)$$

where L_n and L'_n are the width and height of the body of the n th level asperity. If $(V_0/Ka^3) > 1$, and $L'_n/L_n \sim 1$ for all n , asperities of all orders will be multistable, implying the occurrence of large static friction. Substituting for V_0 using $V_0 \approx Pa^3/c$, we find that this condition is equivalent to $P > cK$. If the condition given earlier for strong pinning at the zeroth order asperity interface, namely $P > c_a^{1/2} c_0 c_1 \cdots c_n K = (c/c_a^{1/2})K$ is satisfied, the condition for multistability on all levels, $P > cK$ is certainly satisfied.

Let us now make numerical estimates of whether the lowest order asperity is likely to be monostable, implying low static friction. To do this, we will apply the present model to a single micron-scale asperity in contact with the substrate. Typical values of K are of the order of $10^{11} N/m^2$. In Ref. [3], P at the micron-scale asperity level was estimated from the Greenwood-Williamson model [2] to be about $10^9 N/m^2$ (which is the light load limit, which is also the minimum value, of the load per unit contact area in the Greenwood-Williamson model [2]), and hence $V_0/Ka^3 \approx 0.01/c$, since $V_0 \approx Pa^3/c$. Thus from Eq. (1), if (L'_0/L_0) is of order 1, the zeroth order asperity will be monostable if $c = c_a c_0 c_1 \cdots c_n > 0.01$. This makes it likely that all order asperities and, hence, the interface will be monostable as well. It should be pointed out here that the mechanism for weak pinning discussed here is different from the mechanism discussed in Ref. [5] and in Appendix I of this paper in that it does not result from interactions between asperities (i.e., the collective pinning mechanism). It produces the same result as we found earlier assuming stiff asperities, however, namely that the friction between two asperities at a given length scale is reduced by a factor of the square root of the number of asperities at the next lower length scale order present on its sur-

face. This means that the friction coefficient is reduced from the strong pinning value (which is not much less than one) by a factor of $\prod_{n'=0}^n L_{n'-1}/L_{n'}$, where $L_{-1}=a$ and the product includes only values of n' corresponding to orders of asperities which are monostable. When the zeroth order asperity interface is nonostable, we saw that all higher order asperities will be monostable as well. In this case, the above product reduces to a/L_n for the factor by which the friction coefficient is reduced below one. As L_n in our numerical example is $\approx 1 \mu\text{m}$, this leads to a reduction factor of the order of 10^{-4} . We saw that the condition needed for the interface to be monostable and, hence, exhibit ultralow friction is that the ratio P/c be less than K . The quantity c , as it was defined earlier, can be thought of approximately as the ratio of the area of contact to the apparent area of the interface (although, as pointed out earlier, this is not precise). Since all calculations of this quantity for rough interfaces [2,17] give a constant value of this ratio, at least for small values of P , whether the surface is in the strong or weak pinning regime in the small P limit, it will remain in that regime as P increases, implying that the criterion for weak pinning depends only on the value of K and the degree of roughness of the surface as evidenced by the value of c . Since for larger values (for which $c \approx 0.05$), c was found in the finite elements calculation of Ref. [17] to be a sublinear function of P , it is possible that the interface will switch from weak to strong pinning for sufficiently large P . The results presented here only require that the restoring forces for horizontal distortions of the asperities be elastic. Therefore, they might not be invalidated even if compressions of the asperities normal to the interface, resulting from the load, are plastic [2,17]. At the numerical values of P for which we estimate that we will be in the low friction regime, however, the assumption of a purely elastic solid is quite likely to be valid.

IV. CONCLUSIONS

Thus, the conclusion of this treatment of static friction is that atomically flat surfaces whose atoms interact primarily by hard core interactions would typically be in a weak pinning regime. This means that they would exhibit exceedingly low static friction. The virtually universal presence of multi-length-scale roughness, which is expected to exist for typical solid surfaces, however, will likely put the surface in the strong pinning regime, in which the surfaces can exhibit relatively large static friction. From the arguments presented here, we see that increasing the shear elastic constant by constructing the surfaces from a material of large shear elastic constant (such as saturated diamond or amorphous carbon films [8]) of sufficient thickness so that the film can be treated as a three-dimensional solid (i.e., thicker than typical asperity heights) will make it more likely that the surfaces are in the low friction weak pinning regime. In this regime, the friction can be several orders of magnitude smaller than in the strong pinning regime.

ACKNOWLEDGMENT

I wish to thank the Department of Energy (Grant No. DE-FG02-96ER45585).

APPENDIX A: A SUMMARY OF WEAK PINNING THEORY

For completeness the transition from strong to weak pinning for two three-dimensional elastic solids in contact at a disordered interface using scaling arguments, which was first presented in Ref. [3], will be discussed here. We expect that qualitatively correct results for this problem can be obtained by studying the simpler problem of a three-dimensional elastic solid in contact with a rigid disordered substrate. This problem was also studied in Ref. [12] in the context of a macroscopic solid in contact with a substrate at randomly distributed asperities using perturbation theory. In the present work, we will apply this model to a pair of asperities from two surfaces in contact. Here the disorder occurs over the area of contact of the two asperities. In collective pinning theory [9–11], there is a competition between a disordered potential and an elastic medium which interacts with this potential. In the strong pinning limit, the elastic solid is able to distort enough to essentially minimize its interaction with the disordered potential. In the weak pinning limit, the solid has little distortion over a volume, known as a Larkin domain, whose linear size, the Larkin length, is determined by minimizing the sum of the interaction with the disordered potential and the elastic distortion energy, provided the system's dimension is below the critical dimension for the problem. The three-dimensional solid interacting with a two-dimensional disordered substrate, however, is at its critical dimension [3].

Following a generalization of the discussion in Ref. [3], let us consider a homogeneous elastic solid interacting with a two-dimensional rigid disordered substrate, located at $z=0$. The energy of this system can be written as [15]

$$E = \int d^3r \left[K \sum_{\alpha,\beta} \left(\frac{\partial u_\alpha}{\partial x_\beta} \right)^2 + K' \left(\sum_\alpha \frac{\partial u_\alpha}{\partial x_\alpha} \right)^2 + V(\mathbf{r} + \mathbf{u}(\mathbf{r})\delta(z)) \right], \quad (\text{A1})$$

where α and β run over the components x , y , and z , $u_\alpha(\mathbf{r})$ denotes the α th component of the displacement field at the point \mathbf{r} in the elastic medium, K and K' are the elastic moduli (i.e., the Lamé coefficients [15]) and $V(\mathbf{r})$ denotes the substrate potential per unit area. We look for an approximate solution of the form $u_\alpha = u_\alpha(x/L, y/L, z/L')$, where u varies by an amount of the order of the range of a potential well of the substrate potential when x and y vary over a distance of L or z varies over a distance of order L' . These are the Larkin lengths along and perpendicular to the surface. We substitute this expression for u_α in Eq. (A1), and approximate the integral of the first two terms in the integrand of Eq. (A1) over a single Larkin domain, by the product of the average over a Larkin domain of first two terms in the integrand of Eq. (A1) and the volume of a Larkin domain, L^2L' , and then multiply by the number of domains, A/L^2 , where A is the area of the interface. Minimizing with respect to L' , we obtain

$$L' = L \left(\frac{\Gamma'}{\Gamma} \right)^{1/2}, \quad (\text{A2})$$

where $\Gamma = \sum_{\alpha,\beta=x,y} K \langle (\partial u_\alpha / \partial x_\beta)^2 \rangle + K' \sum_{\alpha=x,y} \langle (\partial u_\alpha / \partial x_\alpha)^2 \rangle$ and $\Gamma' = K \langle (\partial u_z / \partial z)^2 \rangle + K' \sum_\alpha \langle (\partial u_\alpha / \partial z)^2 \rangle$, where (x', y', z')

$= (x/L, y/L, z/L')$ and $\langle \dots \rangle$ signifies an average over a Larkin domain. Since the derivatives of u are all of the order of atomic distances, $L' \approx L$. Assuming that $V(\mathbf{r})$ is completely random and $\mathbf{u}(\mathbf{r})$ varies by a negligible amount as \mathbf{r} runs over a domain of volume L^2L' , the integral of $V(\mathbf{r})$ over this domain is of the form $V_0 c^{1/2} L/a$, where V_0 is the root mean square (rms) value of the potential of interaction between a surface atom and the substrate and c is the fraction of the surface atoms which are in contact with the substrate. In arriving at this result, we assumed that the variation of $\mathbf{u}(\mathbf{r})$, when \mathbf{r} varies over a distance small compared to L , is negligibly small compared to the length scales on the substrate. Since the substrate is random, the integral over $V(\mathbf{r})$ is proportional to the square root of the number of surface atoms in contact with the substrate, which is of the order of $c(L/a)^2$ where a is a mean atomic spacing or potential well size. Since the number of surface atoms in contact with atoms from the second surface is of the order of cA/a^2 the mean force between two atoms in contact from each of the two surfaces is given by PA divided by this quantity or Pa^2/c . Hence, since $V(\mathbf{r})$ varies on a length scale a , $V_0 \approx Pa^3/c$. Then substituting Eq. (A2) in Eq. (A1), we obtain

$$E = [2(\Gamma\Gamma')^{1/2} - Pa^2/c^{1/2}]A/L \quad (\text{A3})$$

for the energy, which is minimized for infinite L if $2(\Gamma\Gamma')^{1/2} > Pa^2/c^{1/2}$ and for $L=0$ (which in practice means that L is as small as the smallest length scale in the problem rather than zero) if $Pa^2/c^{1/2} > 2(\Gamma\Gamma')^{1/2} \approx K$. Thus, it is clear that as c decreases, the interface can switch from weak pinning (if it was already in the weak pinning regime) to strong pinning. In the latter regime, by the arguments given in the last paragraph, the surfaces will be pinned together, i.e., there will be static friction. Because the interface area between two asperities in contact is only of micron size, there will be a transition from low to high, rather than from zero to nonzero static friction (as would occur for an infinite interface).

This problem can also be considered using perturbation theory in the weak pinning limit [12]. To do this, following Ref. [12], one calculates $\mathbf{u}(\mathbf{r})$ which results from the random forces found from $V(\mathbf{r})$ and from it calculates $\langle |\mathbf{u}(\mathbf{R}) - \mathbf{u}(\mathbf{0})|^2 \rangle$ using the standard expression for the elasticity Green's function [15]. Here, $\langle \dots \rangle$ signifies an average over the random substrate forces. R is considered to be equal to the Larkin length when this quantity is comparable to the square of the range of a substrate potential well, as this represents the distance over which the surface of the solid can be considered as rigid from the point of view of the random substrate potential. Following arguments similar to those in Ref. [12], we find a Larkin length that is an exponential function of the ratio of Young's modulus divided by V_0 , which can easily be quite large compared to any reasonable size solid interface when this ratio is reasonably large, as it is in the weak pinning regime. Thus, even though the variational method that I used above gives an infinite Larkin length in the weak pinning limit, whereas the perturbation theory method of Ref. [12] gives a finite Larkin length, since the Larkin length found in Ref. [12] is extremely large (i.e.,

an exponential function of a fairly large number) in the weak pinning limit, the two methods can be considered to give qualitatively the same result.

Let us now consider the transition from strong to weak pinning for a relatively soft three-dimensional solid coated with a relatively hard film, sliding over a rigid substrate. The three-dimensional solid will be assumed to have elastic constants that are sufficiently small so that if the solid were placed directly in contact with the substrate (i.e., there was no coating), the film would be in the strong pinning limit. Let us now assume that the surface of the solid in contact with the substrate is coated with a film of thickness L_f and Lamé coefficients K_f and K'_f , assumed to be larger than the corresponding coefficients for the three-dimensional solid, K and K' , in Eq. (A1). Then, the expression for the energy in Eq. (A1) gets replaced by

$$E = \int_{solid} d^3r \left[K \sum_{\alpha,\beta} \left(\frac{\partial u_\alpha}{\partial x_\beta} \right)^2 + K' \left(\sum_\alpha \frac{\partial u_\alpha}{\partial x_\alpha} \right)^2 \right] + \int_{film} d^3r \left[K_f \sum_{\alpha,\beta} \left(\frac{\partial u_\alpha}{\partial x_\beta} \right)^2 + K'_f \left(\sum_\alpha \frac{\partial u_\alpha}{\partial x_\alpha} \right)^2 - V(\mathbf{r} + \mathbf{u}(\mathbf{r}))\delta(z) \right], \quad (\text{A4})$$

where, as in Eq. (A1), α and β run over the components x , y , and z , $u_\alpha(\mathbf{r})$ and $V(\mathbf{r})$ denotes the substrate potential per unit area. We look for an approximate solution of the form $u_\alpha = u_\alpha(x/L, y/L, z/L')$ in the film and $u_\alpha = u_\alpha(x/L, y/L, z/L'')$, where u varies by an amount of the order of the range of a potential well of the substrate potential when x and y vary over a distance of L or z varies over a distance of order L' or L'' . We substitute these expressions in Eq. (A4), and approximate the integral of the first two terms in the integrand of Eq. (A4) over a single Larkin domain, by the product of the average over a Larkin domain of first two terms in the integrand of Eq. (A4) and the volume of a Larkin domain, $L^2(L'' - L_f)$, and then multiply by the number of domains, A/L^2 , where A is the area of the interface. Similarly, we approximate the integral of the second two terms in the integrand over a single Larkin domain, by the product of the average over a Larkin domain of first two terms in the integrand of Eq. (A4) and the volume of a Larkin domain in the film, which is given by L^2L_f , and then multiply by the number of domains, A/L^2 , where A is the area of the interface. If L_f is much larger than the width of a typical asperity, the hard film coating can be treated as a three-dimensional solid and, hence, using the arguments given in the last paragraph, the coated surface considered to be in the weak pinning regime for which the value of L which minimizes the energy (i.e., the Larkin) length is equal to the width of an asperity (i.e., the largest value that it can take). When L_f is small compared to the width of an asperity, we must treat the film as a two-dimensional hard solid attached to a three-dimensional soft solid and, hence, minimize the expression for E in Eq. (A4). When E is minimized with respect to L'' and L' , we find that L' (meaning that the distortions do not vary as we move across the thickness of the film) is infinite

and $L'' = (\Gamma'_1/\Gamma_1)^{1/2}L$, where Γ_1 and Γ'_1 are given by

$$\Gamma_1 = K \sum_{\beta=x,y} \left(\frac{\partial u'_\alpha}{\partial x'_\beta} \right)^2 - K' \left(\sum_{\alpha=x,y} \frac{\partial u'_\alpha}{\partial x'_\alpha} \right)^2,$$

$$\Gamma'_1 = (K + K') \left(\frac{\partial u'_\alpha}{\partial z'} \right)^2.$$

Substituting these results in Eq. (A4), and minimizing with respect to L , we obtain

$$L = \frac{A}{V_0 - B} L_f, \quad (\text{A5})$$

where

$$A = 2 \left[K_f \sum_{\alpha=x,y,z} \sum_{\beta=x,y} \left(\frac{\partial u''_\alpha}{\partial x'_\beta} \right)^2 + K'_f \left(\sum_{\alpha=x,y} \frac{\partial u''_\alpha}{\partial x'_\alpha} \right)^2 \right]$$

$$- 2 \left[K \sum_{\alpha=x,y,z} \sum_{\beta=x,y} \left(\frac{\partial u'_\alpha}{\partial x'_\beta} \right)^2 + K' \left(\sum_{\alpha=x,y} \frac{\partial u'_\alpha}{\partial x'_\alpha} \right)^2 \right], \quad (\text{A6})$$

and

$$B = (\Gamma'_1/\Gamma_1)^{1/2} K \sum_{\alpha=x,y,z} \left[\sum_{\beta=x,y} \left(\frac{\partial u'_\alpha}{\partial x'_\beta} \right)^2 + \frac{\Gamma_1}{\Gamma'_1} \left(\frac{\partial u'_\alpha}{\partial z'} \right)^2 \right], \quad (\text{A7})$$

where $u'_\alpha(x', y', z') = u_\alpha(x/L, y/L, z/L')$ and $u''_\alpha = u_\alpha(x/L, y/L, z/L'')$. The value of L given in Eq. (A5) minimizes E if V_0 is sufficiently large so that the denominator is positive. For smaller values for which it becomes negative, the value of L which minimizes E is $L = \infty$.

For the situation in which the film behaves like a two-dimensional solid, the Larkin length is finite and given by Eq. (A5). The interaction with the substrate and, hence, the force of friction is reduced by a factor $N_c^{-1/2}$, where N_c is the number of atoms in contact with the substrate in one Larkin domain, and $N_c \approx cL^2/a^2$, where c is the fraction of surface atoms which are in contact with the substrate. For a stiff film for which the elastic energy per surface atom $K_f a^3$, where a is the mean atomic spacing, is large compared to the interaction with the substrate $V_0 a^2$, $N_c^{-1/2} \approx c^{-1/2} (V_0/K_f L_f)$. If the soft solid is assumed to be in the strong pinning regime, V_0 is greater than K_a . If they are of comparable magnitude, our expression for $N_c^{-1/2}$ reduces to $c^{-1/2} (K/K_f) (a/L_f)$, which can lead to a reasonable reduction of the friction since the shear modulus of diamond is about six times that of most other materials.

APPENDIX B: MULTISTABILITY OF ASPERITIES

Since the occurrence of friction in the slow sliding speed limit requires that there exist times of local rapid motion of asperities, it was argued in Ref. [14] that each pair of asperities in contact must possess more than one equilibrium solution. As the surfaces slide relative to each other, each pair of contacting asperities makes transitions from a higher energy

equilibrium configuration to a lower energy one, resulting in the production of kinetic energy, which gets quickly dissipated among the various excitations of the system (i.e., lattice vibrations and electronic excitations). This is argued to be the source of energy dissipation due to kinetic friction in the slow sliding speed limit. This mechanism has traditionally been known as the Tomlinson model [16]. It is also argued in Ref. [14] that the existence of multistability is a requirement for the existence of static friction as well. The argument is basically as follows: Consider a single asperity contact in two dimensions. Reference [18] shows that there will be multistability in two dimensions as long as the asperities are anisotropic. Let $V(\mathbf{r}+\mathbf{s})$ represent the interface potential of interaction between two asperities in contact, where \mathbf{r} represents the displacement of the tip of the asperity with respect to the bulk solid to which it is attached, as a result of a shear distortion of the asperity, and \mathbf{s} is the position of the base of the asperity on the bulk solid. Then the energy of the asperity under consideration can be written as

$$E = V(\mathbf{r} + \mathbf{s}) + (1/2) \mathbf{r} \cdot \lambda \cdot \mathbf{r}, \quad (\text{B1})$$

where the second term is the elastic energy and λ is the elasticity tensor. The equilibrium configuration of the asperity for a given value of s is found from

$$\frac{\partial E}{\partial x} = \frac{\partial V}{\partial x} + \hat{x} \cdot \lambda \cdot \mathbf{r} = 0, \quad (\text{B2a})$$

and

$$\frac{\partial E}{\partial y} = \frac{\partial V}{\partial y} + \hat{y} \cdot \lambda \cdot \mathbf{r} = 0, \quad (\text{B2b})$$

which is equivalent to

$$\frac{\partial V}{\partial s_x} + \hat{x} \cdot \lambda \cdot \mathbf{r} = 0 \quad (\text{B3a})$$

and

$$\frac{\partial V}{\partial s_y} + \hat{y} \cdot \lambda \cdot \mathbf{r} = 0. \quad (\text{B3b})$$

The force exerted by the bulk solid (i.e., the force of friction) is given by

$$F_x = - \frac{\partial V}{\partial s_x} \quad (\text{B4a})$$

and

$$F_y = - \frac{\partial V}{\partial s_y}. \quad (\text{B4b})$$

Let us now slide the surface by an amount s_0 in an arbitrary direction, which we will take to be along the x axis. This is accounted for by adding s_0 to s_x in Eq. (B1). Let us consider an asperity which is monostable. Taking the total derivative of E with respect to s_0 , we obtain

$$\frac{dE}{ds_0} = \frac{\partial V}{\partial s_x} \left(\frac{\partial x}{\partial s_x} - 1 \right) + \frac{\partial V}{\partial s_y} \frac{\partial y}{\partial s_x} + \frac{\partial \mathbf{r}}{\partial s_x} \cdot \lambda \cdot \mathbf{r}. \quad (\text{B5})$$

From Eq. (B3), we find that

$$F_x = -\frac{\partial V}{\partial s_x} = -\frac{dE}{ds_0}. \quad (\text{B6})$$

If there is a uniform distribution of asperities centered at locations denoted by the variable s , we find that the average contribution to the force of static friction due to an asperity is given by

$$\langle F_x \rangle = \ell^{-2} \int_{-\ell/2}^{+\ell/2} ds_x ds_y \left[-\frac{dE}{ds} \right] = -\frac{[E(\ell/2) - E(-\ell/2)]}{\ell}, \quad (\text{B7})$$

where ℓ is the length of the solid. Thus in the large ℓ limit there is no static friction, but for a finite length interface, as at the region of contact of two asperities, there is a net force of friction proportional to the length of the region of contact, as has been assumed in Sec. III. This argument will not work for multistable asperities because if any of the asperities are multistable, E cannot be an exact differential because it has

multiple branches. As a consequence, the value of E for a given configuration of the asperities is a function of how that configuration was established (i.e., we must specify which branch of its potential energy versus s curve each asperity is on). Clearly, if there were a net force in the x direction, as there is likely to be if the asperities are multistable, the integral in Eq. (B7) would be proportional to ℓ^2 . While this argument for the smallness of static friction in the absence of multistability is only strictly correct for a perfectly uniform distribution of asperities over the interface, the zero velocity limit of the kinetic friction can certainly be shown to be small in the absence of multistability for all asperity distributions [19].

Thus, we conclude that the contribution to the static friction from each asperity of n th order can only be proportional to the number of $n-1$ order asperities on its surface if those asperities are multistable. If this is not the case, the contribution to the static friction will be only proportional the square root of the number of $n-1$ order asperities in contact with the substrate.

-
- [1] L. Kogut and I. Etsion, *J. Colloid Interface Sci.* **261**, 372 (2003).
- [2] J. A. Greenwood and J. B. P. Williamson, *Proc. R. Soc. London, Ser. A* **295**, 300 (1996).
- [3] A. Volmer and T. Natterman, *Z. Phys. B: Condens. Matter* **104**, 363 (1997); J. B. Sokoloff, *Phys. Rev. Lett.* **86**, 3312 (2001); *Phys. Rev. B* **65**, 115415 (2002).
- [4] M. Muser, in *Dynamics of Small Confining Systems*, Materials Research Society Symposium Proceedings, Boston, edited by J. M. Drake, J. Klafter, P. Levitz, R. M. Overney, and M. Urbakh (Materials Research Society, Warrendale, PA, 2001) Vol. 651. G. He, M. H. Muser, and M. O. Robbins, *Science* **284**, 1650 (1999); M. H. Muser and M. O. Robbins, *Phys. Rev. B* **61**, 2335 (2000); M. H. Muser, L. Wenning, and M. O. Robbins, *Phys. Rev. Lett.* **86**, 1295 (2001); G. He and M. Robbins, *Tribol. Lett.* **10**, 7 (2001); M. H. Muser, *Phys. Rev. Lett.* **89**, 224301 (2002).
- [5] J. B. Sokoloff, *Phys. Rev. E* **71**, 056107 (2005).
- [6] A. Majumdar and B. Bhushan, *J. Tribol.* **113**, 1 (1991); D. Avnir, D. Farin, and P. Pfeifer, *Nature (London)* **308**, 261 (1984); B. N. J. Persson, *Phys. Rev. Lett.* **87**, 116101 (2001); S. Roux, J. Schittbuhl, J.-P. Vilotte, and A. Hansen, *Europhys. Lett.* **23**, 277 (1993); J. F. Archard, *Proc. R. Soc. London, Ser. A* **243**, 190 (1957).
- [7] J. P. Gao, W. D. Luedtke, D. Gourdon, M. Ruths, J. N. Israelachvili, and U. Landman, *J. Phys. Chem. B* **108**, 3480 (2004).
- [8] J. A. Heimberg, K. J. Wahl, and I. L. Singer, *Appl. Phys. Lett.* **78**, 2449 (2001); G. T. Gao, P. T. Mikulski, and J. A. Harrison, *J. Am. Chem. Soc.* **124**, 7202 (2002); G. T. Gao, P. T. Mikulski, G. M. Chateauneuf, and J. A. Harrison, *J. Phys. Chem. B* **107**, 11082 (2003); N. Saka (unpublished).
- [9] A. I. Larkin and Yu. N. Ovchinnikov, *J. Low Temp. Phys.* **34**, 409 (1979).
- [10] H. Fukuyama and P. A. Lee, *Phys. Rev. B* **17**, 535 (1977); P. A. Lee and T. M. Rice, *ibid.* **19**, 3970 (1979).
- [11] D. S. Fisher, *Phys. Rev. B* **31**, 1396 (1985); *Phys. Rev. Lett.* **50**, 1486 (1983).
- [12] B. N. J. Persson and E. Tosatti, *Solid State Commun.* **109**, 739 (1999); in *Physics of Sliding Friction*, edited by B. N. J. Persson and E. Tosatti (Kluwer Academic Boston, 1995), p. 179; V. L. Popov, *Phys. Rev. Lett.* **83**, 1632 (1999).
- [13] B. N. J. Persson, *Phys. Rev. B* **44**, 3277 (1991); B. N. J. Persson and A. Volokitin, *J. Chem. Phys.* **103**, 8679 (1995); *J. Phys.: Condens. Matter* **9**, 2869 (1997).
- [14] C. Caroli and Ph. Nozieres, *Eur. Phys. J. B* **4**, 233 (1998); in *Physics of Sliding Friction*, edited by B. N. J. Persson and E. Tosatti, NATO ASI Series E: Applied Sciences (Kluwer Academic, Dordrecht, 1996), Vol. 311; M. Brillouin, *Notice sur les Travaux Scientifiques* (Gautier-Villars, Paris, 1909); G. A. Tomlinson, *Philos. Mag.* **7**, 905 (1929).
- [15] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 2nd ed. (Pergamon Press, New York 1970).
- [16] G. A. Tomlinson, *Philos. Mag.* **7**, 905 (1929).
- [17] S. Hyun, L. Pei, J. F. Molinari, and M. O. Robbins, *Phys. Rev. E* **70**, 026117 (2004); M. O. Robbins (unpublished).
- [18] A. Tanguy and Ph. Nozieres, *J. Phys. I* **6**, 1251 (1996).
- [19] T. Baumberger and C. Caroli, cond-mat/0506657 (unpublished).