Possible microscopic explanation of the virtually universal occurrence of static friction

J. B. Sokoloff

Physics Department and Center for Interdisciplinary Research on complex Systems, Northeastern University, Boston, Massachusetts 02115

(Received 18 July 2001; revised manuscript received 6 December 2001; published 7 March 2002)

Perturbation theory, simulations, and scaling arguments predict that there should be no static friction for two weakly interacting, flat, atomically smooth, clean solid surfaces. The absence of static friction results from the fact that the atomic-level interfacial potential energy is much weaker than the elastic potential energy, which prevents the atoms from sinking to their interfacial potential minima. Consequently, we have essentially two rigid solids, for which the forces at randomly distributed ''pinning sites'' cancel. It is shown here that even fluctuations in the concentration of atomic-level defects at the interface do not account for static friction. The sliding of contacting asperities, which occurs when the problem is studied at the multi-micrometer length scale, relative to each other, however, involves the shearing of planes of atoms at the interface between a pair of asperities from the two surfaces in contact. Since this results in a force for the interaction of two in contact asperities which varies over sliding distances of the order of an atomic spacing, the contacting asperities at the surface are able to sink to their interfacial potential minima, with negligible cost in elastic potential energy. This results in static friction.

DOI: 10.1103/PhysRevB.65.115415 PACS number(s): 81.40.Pq, 46.55.+d, 62.20.Qp

I. INTRODUCTION

It is well known to every student in an elementary physics class that kinetic friction has very little velocity dependence in the slow sliding speed limit (often called ''dry friction''). Yet many atomic-level treatments of friction fail to give this behavior. For example, molecular dynamics simulations and analytic calculations[1–5] show that while commensurate interfaces are pinned for applied forces below a critical value (i.e., exhibit static friction), incommensurate surfaces are not pinned and exhibit viscous friction (i.e., friction proportional to the sliding velocity) for sufficiently weak interfacial forces. Perturbation theory calculations done for a nonmetallic monolayer film with underdamped phonons sliding on a nonmetallic substrate with some disorder, however, give nearly velocity independent sliding friction[4] and exhibit a divergence in the mean square atomic displacement in the limit of zero sliding velocity. The latter behavior signifies that the film will be pinned below a critical applied force. This behavior has been confirmed by recent molecular dynamics calculations on such a system.[5] Perturbation theory calculations done for a three-dimensional film sliding on a substrate, however, give viscous friction. This result is consistent with the notion that without multistability, there cannot be ''dry friction'' due to vibrational excitations in an elastic solid.[6,7] Dry friction is possible for the monolayer film, as mentioned above, however, because the two-dimensional phonon density of states of the film does not drop to zero as the frequency goes to zero, as it does for a three-dimensional solid.[4] As the sliding velocity drops to zero, the ''washboard frequency'' (the sliding velocity divided by a lattice constant) drops to zero. Since the phonon density of states does not drop to zero, there are phonons present at arbitrarily low frequency, which can be excited by the substrate potential. Since the density of states does fall to zero as the frequency falls to zero in three dimensions, however, the force of friction falls to zero as the velocity does. In models used for charge density waves (CDW's), in which the CDW is modeled as an elastic medium moving through a solid containing impurities distributed randomly throughout it, there is no pinning in four or more dimensions.[8] In contrast, in fewer than four dimensions, there is pinning. For a model used for friction,[9] consisting of a three dimensional elastic medium moving over a surface containing a random array of point defects, the critical dimension is 3.[9] As a consequence, although if the defect potential is sufficiently large, there will be static friction and ''dry friction,'' for weak defect potentials, there will be no static friction and the kinetic friction will be viscous (i.e., linear in the sliding velocity). The nonperiodic ''defect potential'' acting across the interface could be due to disorder on any length scale in the problem. For example, it can be due to atomic level point defects, such as vacancies and substitutional impurities at the interface, as has been assumed in Refs. 4 and 5, but it can also be due to the fact that the surfaces of the sliding solids are only in contact at micron scale randomly located protrusions, commonly known as asperities. On the atomic scale, it can also be due to adsorbed film molecules.[2]

In contrast to atomic level point defects, however, asperities and adsorbed molecules possess internal structure and as a consequence if they are sufficiently flexible, they can exhibit multistability (i.e., the existence of more than one stable configuration needed for the Tomlinson model[6] to apply). Caroli and Noziere[7] proposed an explanation for ''dry friction'' based on the Tomlinson model.[6] In the Tomlinson model the two bodies which are sliding relative to each other at relatively slow speeds remain stuck together locally until their centers of mass have slid a small distance relative to each other, at which point the stuck configuration of the two surfaces becomes unstable and the two surfaces locally slip rapidly with respect to each other until they become stuck again, and the process repeats itself. The slipping motion that takes place can either be local or can involve motion of the body as a whole. Then the actual friction acting locally at the

0163-1829/2002/65(11)/115415(9)/$20.00

interface could be viscous, but the rapid motion that takes place, even at slow sliding speeds, could still result in a sizable amount of friction, even in the limit of vanishing average sliding velocity. In the Caroli-Noziere model[7] interface contact only occurred at a very dilute concentration of interlocking asperities. It is the rapid stick-slip motion of these asperities, which gives rise to dry friction on the average in this model, if we assume that all of the kinetic energy released in the slip is dissipated, which is probably true in the zero velocity limit. This mechanism would seem to imply that the occurrence of dry friction depends on the existence of multistability; in situations in which the asperities are not multistable, there will be neither ‘‘dry friction’’ nor static friction. It is argued[7] that in the absence of plasticity in the model, the maximum force of dynamic friction, in the velocity approaches zero limit, must be equal to the force of static friction. It was pointed out in recent work by these workers, however, that the asperities are generally too stiff to undergo the Tomlinson model-like instabilities because of their shape.[7] Therefore, they proposed alternative mechanisms. In one mechanism, it is assumed that there exists a glassy film at the interface in the experimental systems that they are trying to describe. Glassy materials possess metastable atomic configurations (the equivalent of ‘‘two-level systems’’ which are believed to contribute to the specific heat) which can exhibit Tomlinson model-like instabilities during sliding, similar to those found by Falk and Langer in their study of the shearing of glassy materials.[10] This mechanism will, however, only be the correct explanation of ‘‘dry friction’’ for glassy interfaces. It is not clear, however, that all interfaces are glassy. He *et al.*[2] proposed that the occurrence of static friction between elastic solids requires the existence of adsorbed mobile molecules at the interface. It is important to know if the occurrence of a glassy interface or adsorbed mobile molecules is a requirement for the occurrence of static friction. If it is, it would imply that clean interfaces between crystalline solids would not exhibit static friction. Caroli and Noziere[7] also proposed that adhesive forces could provide the required multistability (because of the so called ‘‘jump to contact’’ instability) to give friction at slow speed. It is not clear that this will be significant for asperities even under light loads, however. In the absence of multistability, there is good reason to believe that there will be neither static friction nor dry friction, at least for light enough loads to put us in the ‘‘weak pinning limit’’ in the language of the charge density wave and vortex problems first studied by Larkin and Ovchinikov[11] and Fukuyama, Lee, and Rice.[12] The existence of multistability has also been shown to be a condition for pinning of CDW’s.[13] (One way of understanding this is that if there is static friction, the sliding velocity of the solid will only be nonzero if a force above the force of static friction is applied to the body. Alternatively, if we plot the force as a function of velocity, we can view this as implying that the force of friction approaches a nonzero value as the center of mass velocity approaches zero. The arguments in this reference and Ref. 7 tell us that there must be multistability for this to occur.)

It must be stressed that the present work, as all previous work on atomic level theories of friction quoted above, deals with the very light load limit, in which the hard cores of the atoms of the two surfaces which are in contact are not being pressed together with a lot of force. These treatments, however, can be thought of as one step closer to describing friction in every day applications, in that the loads are higher, than the extreme limit of contactless friction.[14] It is felt that it is important to study such a limiting case, as a first step towards understanding the higher load cases that occur in engineering applications. No claim is being made here that the present discussion will apply to much higher loads, which could be dominated by gross plastic deformations and wear. It is hoped, however, that the present discussion will stimulate more precise experiments in the light load limit, in order to understand the fundamental mechanisms for friction.

In Sec. II, a discussion is given of the scaling theoretic treatment of static friction. This is an outgrowth of a similar treatment by Fisher of the pinning of charge density waves. It is found that, at least for surfaces with defects that produce a relatively weak potential, the Larkin domains (i.e., the regions over which the solid distorts to accommodate defects at the interface) are as large as the interface, as was found by Persson and Tosatti using perturbation[9] theory. As a consequence, the forces from the various interfacial defects tend to cancel each other, resulting zero for the average static friction. Since all that remains are fluctuations, which are proportional to the square root of the number of defects at the interface, this implies that the static friction per unit interface area decreases as the inverse of the square root of the interface area, as was found for perfect crystalline interfaces by Muser *et al.*[2]

In Sec. III, it is shown that when one takes into account the distribution of contacting asperities at the interface that occurs when the problem is studied at the multimicrometer scale, one finds that the asperities are in the ‘‘strong pinning limit,’’ implying that there is static friction. This is shown to be a consequence of the fact that the shear force between two contacting asperities varies by a large fraction of its magnitude as the asperities are slipped relative to each other over slip distances of the order of atomic spacings. This must be true for a wide range of interfaces, both crystalline and disordered. (The only assumption that needs to be made is that the interface is made of planes of atoms in contact.) As a consequence, essentially all asperities can sink to their contact potential minimum by moving a distance parallel to the interface of the order of a lattice spacing (which is much smaller than typical interasperity spacing) with negligible cost in elastic energy. Thus, when an external stress is applied, the friction forces from all of the asperities will act coherently, resulting in a net force of static friction. Furthermore, the force constant for the contact potential is much larger than that due to the elastic force constant of the asperities. As a consequence, the asperities satisfy the criterion for the occurrence of multistability, shown in Ref. 7 to be a requirement for the occurrence of static friction. This treatment is in contrast to the treatment in Ref. 9, which argues that in the weak coupling limit the Larkin length for the asperities will be much larger than the width and length of the interface. It does not, however, tell us anything about whether the criterion for static friction that I use is satisfied, namely that the all of the asperities can sink to a contact potential minimum with negligible cost in elastic energy of the sliding solid. To put it another way, my argument for the lack of static friction due to atomic level defects in the weak pinning limit, given in Sec. II, is based on the scaling arguments that I apply in that section to Eq. (3), which show that for a three-dimensional solid sliding over another solid (or substrate) there exist only two regimes. There is a weak pinning regime, in which the elastic forces of the solid dominate over the interfacial forces (and the Larkin length is effectively infinite) and a strong pinning regime in which the interfacial forces dominate (in which the larkin length is effectively very small). In the former regime, there is no static friction for a macroscopic solid and in the latter regime, regime there is static friction. I find that for weak defect potentials, atomically flat surfaces can be in the weak pinning regime, and exhibit no static friction, but, in contrast, I find that when the problem is studied on the multiasperity length scale for surfaces that are only in contact at micron scale asperities, the interfacial forces dominate over the elastic forces, implying that we are in the strong pinning regime, and there is static friction.

Volmer and Natterman[15] have developed a theory to attempt to explain Amonton's law on a microscopic level. Their treatment of Amonton's law is qualitatively similar to that of Greenwood and Williamson.[16] Both treatments determine the area of contact as a function of load, and then determine the force of static friction by multiplying it by an estimate of the shear strength for the contacting asperities. In the estimate of the static friction in Ref. 15, no account is taken of the possibility that forces acting on the various contacting asperities can act in arbitrary directions and could, in principle, in the weak load limit cancel out. Their treatment for sliding surfaces is only able to give "dry friction" if the surface height correlators have a cusp in their position dependence. Analogies to charge density wave dynamics are used to argue that such behavior is expected to occur but will disappear on much smaller length scales (presumably comparable to atomic dimensions). This return to analytic behavior of the correlator on smaller length scales again leads to dynamic friction which approaches zero as the velocity approaches zero (i.e., essentially viscous friction).

II. SCALING TREATMENT AT THE ATOMIC LEVEL OF STATIC FRICTION

In this section, we will treat the problem of static friction due to disorder which results from atomic level defects, such as vacancies or substitutional impurities using scaling arguments. In the next section, we will consider random contacting asperities, which occur when the surface is viewed on the micron length scale.

Following Fisher's treatment of the charge density wave (CDW) problem,[8] it is clear that we can also use a scaling argument for the friction problem in order to determine whether the pinning potential becomes irrelevant as the length scale becomes large. In order to accomplish this, let us formulate this problem in a way similar to the way that Fisher does, by considering the crystal lattice to be sliding over a disordered substrate potential under the influence of a force F, which is applied to each atom in the crystal. Then we can write the equation of motion as

$$m\ddot{\mathbf{u}}+m\gamma\dot{\mathbf{u}}_j=\sum_{j'}\mathbf{D}(\mathbf{R}_j-\mathbf{R}_{j'})\mathbf{u}_{j'}+\mathbf{f}(\mathbf{R}_j)+\mathbf{F}, \qquad (1)$$

where $\mathbf{D}(\mathbf{R}_j-\mathbf{R}_{j'})$ is the force constant matrix for the lattice, $\mathbf{f}(\mathbf{R}_j)$ is the force due to the substrate on the jth atom, and $\mathbf{R}_j$ is the location of the jth atom in the lattice. As a result of slow speed sliding of the lattice over the disordered substrate potential, low frequency acoustic phonons are excited. Since these modes have wave vector $\mathbf{k}$, small compared to the Brillouin zone radius $\mathbf{u}_j$, the displacement of the jth atom is a slowly varying function of $\mathbf{R}_j$. Then, following the discussion in Ref. 17, we can write the first term on the right hand side of Eq. (1) as

$$\mathbf{D}'(i^{-1}\nabla_j)\mathbf{u}_j, \qquad (2)$$

where $\mathbf{D}'(\mathbf{k})$ is the Fourier transform of $\mathbf{D}(\mathbf{R}_j-\mathbf{R}_{j'})$ and $\nabla_j=(\partial/\partial X_j,\partial/\partial Y_j,\partial/\partial Z_j)$, where $\mathbf{R}_j=(X_j,Y_j,Z_j)$ to a good approximation. Furthermore, to a good approximation we can expand $\mathbf{D}'$ to second order in ∇_j. Equation (1) then becomes

$$m\ddot{\mathbf{u}}+m\gamma\dot{\mathbf{u}}_j=-vE'\nabla_j^2\mathbf{u}_{j'}+\mathbf{f}(\mathbf{R}_j)+\mathbf{F}, \qquad (3)$$

where E' is an effective Young's modulus and v is the unit cell volume. We can then apply Fisher's scaling argument[8] to the resulting equation. This is accomplished by dividing the solid into blocks of of length L lattice sites parallel to the interface and L' lattice sites normal to the interface, assuming that these dimensions are chosen so that $\mathbf{u}_j$ varies slowly over each such a block. Then integrating Eq. (3) over a block which lies at the interface, we obtain

$$L^2L'[m\ddot{\mathbf{u}}_{j'}+m\gamma\dot{\mathbf{u}}_{j'}]=-(1/2)L'L^2(E'/L^2)\nabla_j'^2\mathbf{u}_{j'} - L\mathbf{f}'(\mathbf{R}_j')+L^2L'\mathbf{F}, \qquad (4)$$

where $\nabla_j'^2$ denotes the Laplacian in the rescaled coordinates $[(X_{j'}',Y_{j'}',Z_{j'}')=(X_j/L,Y_j/L,Z_j/L')]$, which are the coordinates of the centers of the boxes. Here we made use of the fact that $\mathbf{u}_j$ varies on length scales L and L', when we transform to the block coordinates. Hence, we may replace ∇_j^2 by $L^{-2}\nabla_j'^2$ and $\mathbf{f}(\mathbf{R}_j)$ by $L\mathbf{f}'(\mathbf{R}_{j'})$ The substrate force is only multiplied by L because the interaction of the defects with a single block at the surface is proportional to the square root of the surface area of the block. For a thick solid (i.e., one whose lateral and transverse dimensions are comparable), L and L' are always of comparable magnitude. Then, we conclude that no matter how large we make L and L', the ratio of the elastic force (the first term on the right hand side) and the substrate force (the second term on the right hand side) will remain the same. This implies that we are at the critical dimension for this problem since when the length scales L and L' are increased, neither the elasticity nor the substrate force becomes irrelevant. Whichever one dominates at one length scale will dominate at all scales. Equation (4) implies that the force of static friction per unit area acting at the interface is inversely proportional to the square root of the interface area. Instead of using the equations of motion, as is done here, we could alternatively formulate these scaling arguments by minimizing the energy of the system.[18]

Next, let us consider the effect of fluctuations in the defect concentration, for a thick solid. To do this, we again divide the solid into boxes of length L and examine what percentage of the blocks at the interface contain a large enough concentration of defects to put these blocks in the "strong pinning" regime (i.e., the regime in which the substrate forces dominate over the elastic forces between the blocks). To examine whether this is possible, let us define a parameter $\lambda = V_1/E'b^3$, where b is a lattice spacing and V_1 is the strength of the potential due to a defect acting on an atom. Let $n_c = c'L^2$ be the number of defects within a particular strongly pinned block and c' is the defect concentration large enough for it to be considered a strongly pinned block, i.e., a block whose interaction with the substrate is much larger than the elastic interaction between such blocks. (This concentration is necessarily noticeably larger than the mean defect concentration on the interface.) Then the interaction of a typical strong block with the substrate defects is $\lambda(c'L^2)^{1/2}$. The interface area surrounding each strong block is the total interface area A divided by the number of strong blocks at the interface, which is equal to $PA/(bL)^2$, where P is the probability of a particular block being a strong one $[A/(bL)^2$ is clearly equal to the total number of blocks at the interface, both strong and weak$]$. Then we obtain $(bL)^2/P$ for the area surrounding a strongly pinned block. Then the typical length for the elastic energy acting between two strong blocks is $L'b$, where $L' = L/P^{1/2}$. The ratio of the total elastic energy associated with each strong block to $E'b^3$ is the product of the volume surrounding a strongly pinned block$=(L')^3$ with $(L')^{-2}$, since $\nabla^2 u \propto L'^{-2}$ or L'. Then the criterion for a block to be a strong block is $\lambda(c'L^2)^{1/2} \gg L'$ or $\lambda \gg (c'P)^{-1/2}$. Since $c'P < 1$, this violates our previous assumption that $\lambda \ll 1$. Thus, we conclude that such fluctuations in the defect concentration will not lead to static friction.

If we assume that the sliding solid and the substrate surfaces at the interface are incommensurate and that the defects are either vacancies or substitutional impurities, which are centered around particular lattice sites in the substrate, there is another type of concentration fluctuation. For a uniform random distribution of defects over the substrate lattice sites, surface atoms of a completely rigid sliding solid will be found at all possible position within the various defect potential wells, which results in the net force on the solid due to defects being zero on the average. Let us again divide the solid into blocks of length L, but now the concentration of defects in each block will be taken to be equal to the mean defect concentration c. We will look for blocks in which the defects are distributed such that there is a sizable concentration of atoms located in that region of defect potentials, for which the force on the block is opposite the direction in which we are attempting to slide the block, due to the defects. Then we can divide each block into regions of equal size. If a defect lies in one region, the atoms which interact with the defect will have a force exerted on them opposing the attempt to slide the solid, and if it lies in the other region, the force on the atoms that it interacts with will be in the opposite direction. These two regions are, of course, fragmented. Then the net force on a block of length L at the interface due to the substrate is proportional to $L^2 \delta c P$, where δc is the mean difference in defect concentrations between the two regions in the block defined above (such that the mean defect concentration over the whole block is c) and P is the probability of having a concentration difference δc between these two regions. The un-normalized probability of having a concentration difference δc between the two regions in a block is given by

$$\frac{N_1!}{n_{c1}!\left(\frac{1}{2}N_1 - n_{c1}\right)!(cN_1 - n_{c1})!\left(\frac{1}{2}N_1 - cN_1 + n_{c1}\right)!}, \tag{5}$$

where the number of atoms in a block $N_1 = L^2$ and the number of atoms in the region in which the net force is against the direction in which we are attempting to slide the solid $n_{c1} = (1/2)[c + (1/2)\delta c]N_1$, whose normalized small δc approximation is

$$P \approx \exp\left(-\frac{3N_1}{2c(1-c)}\delta c^2\right). \tag{6}$$

Since in the large N_1 limit P decreases exponentially with increasing N_1, unless $\delta c \approx N_1^{-1/2}$, we conclude that the substrate defect force on the block will not increase with increasing N_1 and hence in the large N_1 limit, the elastic force on the block, which we showed above is proportional to $L = N_1^{1/2}$, will dominate. This implies that this type of concentration fluctuation will not lead to static friction for a macroscopic size block.

In the discussion under Eq. (4), it was argued that for an interface between two three-dimensional solids, which are sliding relative to each other, if the interfacial force dominates over the elastic forces at one length scale, it dominates at all length scales, implying that there is static friction for macroscopic solids. On the other hand, if the elastic forces dominate at one scale, they will dominate at all length scales, implying that there is no static friction for macroscopic solids. Let us now examine whether or not the interfacial dominates for an interface between two flat crystalline solids with defects in contact. The energy of the interface consists of two parts. One part is the single defect energy, which consists of the interaction energy of a defect with the substrate plus the elastic energy cost necessary for each defect to seek its minimum energy, neglecting its elastic interaction with other defects, which is independent of the defect density. It should be noted that there is a restoring force when the defect is displaced relative to the center of mass, even if the defect-defect interaction is neglected.[7,19] The second part is the elastic interaction between defects within the same solid, which depends on the defect density. In order to determine the effect of these energies, let us for simplicity model the interaction of the lth defect with the lattice by a spherically symmetric harmonic potential of force constant α_l. Assume that in the absence of distortion of the solid, the lth defect lies a distance $\mathbf{\Delta}_l$ from the minimum of its potential well. Let $\mathbf{u}_l$ be the displacement of the lth defect from its initial position. We use the usual elastic Green's function tensor of the medium at a distance r from the point at which a force is applied at the interface, but for simplicity, we approximate it by the simplified form $G(r)=(E'r)^{-1}$, where E' is Young's modulus.[19] Then the equilibrium conditions on the u's are

$$\mathbf{u}_l=(E'a)^{-1}\alpha_l(\mathbf{\Delta}_l-\mathbf{u}_l)+\sum_j (E'R_{l,j})^{-1}\alpha_j(\mathbf{\Delta}_j-\mathbf{u}_j), \tag{7}$$

where a is a parameter of the order of the size of the defect and $R_{l,j}$ is the distance between the lth and jth defects. This equilibrium condition is discussed in more detail in the Appendix. To lowest order in the interdefect interaction, the approximate solution for $\mathbf{u}_l$ is

$$\mathbf{u}_l=\mathbf{u}_l^0+[1+(E'a)^{-1}\alpha_l]^{-1}\sum_j (E'R_{l,j})^{-1}\alpha_j(\mathbf{\Delta}_j-\mathbf{u}_j^0), \tag{8a}$$

where

$$\mathbf{u}_l^0=\frac{\alpha_l}{E'a+\alpha_l}\Delta_l \tag{8b}$$

is the zeroth order approximation [i.e., the solution to Eq. (7) neglecting the second term on the right hand side of the equation]. Since the defects are randomly distributed over the interface, we can estimate the second term (i.e., the summation over j) on the right-hand side of Eq. (8a) by its root mean square (r.m.s.) average which is estimated by integrating the square of the summand over the position of the jth defect which is in contact with the substrate and multiplying by the density of asperities in contact with the substrate ρ, and then taking the square root. Since the angular integrals only give a factor of order unity, we need only consider the integral over the magnitude of $R_{l,j}$, giving an r.m.s. value of the sum over R^{-1} of order $[\rho \ln(W/a)]^{1/2}$ where W is the width of the interface and a is the mean defect size. For $W \approx 1$ cm and $a \approx 10^{-8}$ cm, $[\ln(W/a)]^{1/2}$ is of order unity. For a defect potential of strength V_1, $\alpha_l \approx V_1/b^2$, where b is of the order of a lattice constant. If $V_1 \approx 1$ eV and $b \approx 3 \times 10^{-8}$ cm, $\alpha \approx 2\times 10^3$ dyn/cm^2. For $E' \approx 10^{12}$ dyn/cm^2, $u_l \approx (\alpha_l/E'b)\Delta_l \approx 0.06\Delta_l$. This implies that the elasticity of the solid prevents the solid from distorting to any significant degree to accommodate the defects at the interface, which implies that we are in the weak pinning limit. From the scaling arguments of this section, we conclude that there will be no static friction in the macroscopic interface limit. For stronger defect potentials and/or smaller values of E', however, it is clear that we could also be in the strong pinning limit, then there is static friction. For almost any surface, contact only takes place at random asperities of mean size and spacing of the order of microns. In the next section it will be argued when one treats the interface on the multimicrometer scale as a collection of contacting asperities, one finds in contrast that it is almost certainly in the strong pinning limit.

III. STATIC FRICTION DUE TO DISORDER ON THE MICRON LENGTH SCALE

The arguments in the last section seem to imply that weakly interacting disordered surfaces cannot exhibit static friction. We shall see, however, that unlike weak atomic scale defects, for which the elastic interaction between them can dominate over their interaction with the second surface, for contacting asperities that occur when the problem is studied on the multimicrometer scale, the interaction of two contacting asperities from the two different surfaces dominates over the elastic interaction between two asperities in the same surface. It is suggested here that this could be responsible for the virtual universal occurrence of static friction. Roughness due to asperities is well described by the Greenwood-Williamson (GW) model,[16] in which there are assumed to be elastic spherically shaped asperities on a surface with an exponential or Gaussian height distribution in contact with a rigid flat substrate, especially for relatively light loads. As mentioned in the introduction, Volmer and Nattermann's approximate way of accounting for static friction[17] is not qualitatively different from that of Ref. 16. In the GW model, the total contact area is given by

$$A_c=2\pi\sigma N R_c\int_h^\infty ds\,\phi(s)(s-h), \tag{9}$$

where $\phi(s)$ is the distribution of asperity heights s, in units of a length scale σ associated with the height distribution, R_c is the radius of curvature of a typical asperity, and h is the distance of the bulk part of the sliding solid from the flat surface in which it is in contact, measured in units of σ. Since the force of static friction exerted on a single asperity is expected to be equal to the product of the contact area and a shear strength for the interface, it is proportional to this quantity.

The number of contacting asperities per unit surface area is given by

$$\rho(h)=(N/A)\int_h^\infty ds\,\phi(s), \tag{10}$$

where A is the total surface area and N is the total number of asperities whether in contact with the substrate or not. The load is given in this model by

$$F_L=(4/3)E'N(R_c/2)^{1/2}\sigma^{3/2}\int_h^\infty ds\,\phi(s)(s-h)^{3/2}. \tag{11}$$

A Gaussian distribution is assumed here for $\phi(s)$ [i.e., $\phi(s)=(2\pi)^{-1/2}e^{-s^2/2}$].

Let us now apply the equilibrium conditions expressed in Eqs. (7) and (8) (used in the last section to treat atomic level defects), to the asperities.[18] It is argued in the Appendix that these equations should give a correct description of contacting asperities. The shearing of the junction at the area of contact of two asperities involves the motion of two atomic planes relative to each other, and hence the sliding distance over which the contact potential varies must be of the order of atomic distances. Then, if we denote the width of the asperity contact potential well (i.e., the length scale over which the contact potential varies) by b, of the order of atomic spacings, we must choose a typical value for α such that αb is of the order of the shear rupture strength of the asperity contact junction ($\approx E'\pi a^2$). Thus, $\alpha/E'a \approx a/b = 10^4$. Then, applying Eq. (11b) to the contacting asperities, we find that $\mathbf{u}_l^{(0)} \approx \mathbf{\Delta}_l$, i.e., the contacting asperities lie at the minima of the contact potential. This is very easy to understand. Since the contact potential varies over distances of the order of an atomic spacing, the asperities can all sink very close to their contact potential minima by moving a distance of the order of an atomic spacing, with negligible cost in elastic potential energy. This is what distinguishes the present treatment from those of Refs. 7 and 9. In those references, it is assumed that continuum mechanics accurately describes the asperity-asperity contact. In contrast, I propose that it is essential to take into account the fact that the shearing stress of such a contact must vary on an atomic length scale. In this section, a is taken to be a parameter of the order of the size of the asperity. Since the contacting asperities are randomly distributed over the interface, we can again estimate the second term (i.e., the summation over j) on the right hand side of Eq. (7) by its root mean square (r.m.s.) average which is estimated by integrating the square of the summand over the position of the jth asperity, which is in contact with the substrate, over its position and multiplying by the density of asperities in contact with the substrate ρ, giving an r.m.s. value of the sum over R^{-1} of order $[\rho \ln(W/a)]^{1/2}$ where W is the length of the interface and a is the asperity size. For $W \approx 1$ cm and $a \approx 10^{-4}$ cm, $[\ln(W/a)]^{1/2}$ is of order unity.

Here, we have considered only micron and atomic length scales to be important. There can, of course, also be roughness on length scales between the latter two. As the load is increased, asperities on smaller than microm length scales could potentially also become important. As long as typical asperity spacings are still much larger than atomic spacings, however, the arguments given in this section should still apply to them.

Let us now give sample numerical values for some of the quantities which occur in the application of the GW model to this problem. Following Ref. 16, we choose $\sigma = 2.4 \times 10^{-4}$ mm and $R_c = 6.6 \times 10^{-2}$ mm, and assume that there is a density ρ of 4.0×10^3 asperities/mm^2. Then by performing the integrals in Eqs. (9)–(11), we find that for $F_L/A = 3.98 \times 10^{-4} N/\text{mm}^2$, where A is the apparent area of the interface, the total contact area divided by A is 3.03×10^{-5}, and the contact area per asperity from the ratio of Eqs. (9) and (10) is 2.44×10^{-5} mm^2. Also, $\rho(h)^{1/2}$, which is equal to the square root of Eq. (10) is 1.11 mm^{-1}. Since in the present case $\alpha_l \gg E'a$, Eq. (8a) becomes

$$\mathbf{u}_l \approx \mathbf{\Delta}_l + (E'a/\alpha_l)\{-\mathbf{\Delta}_l + [\ln(W/a)]^{1/2}\rho^{1/2} a \mathbf{\Delta}_l'\}, \quad (12)$$

where $\mathbf{\Delta}_l'$ is a vector whose magnitude is of the same order as $\mathbf{u}_l$, but it can be in any direction. It is determined by the contribution to $\mathbf{\Delta}_l$ from neighboring asperities. In arriving at Eq. (12) we have replaced $R_{l,j}^{-1}$ by its root mean square value. Using the above numerical estimates of the parameters in this problem, we find that $\rho^{1/2} a \approx 10^{-3}$ and $E'a/\alpha_l \approx 10^{-4}$ from which we conclude that $\mathbf{u}_l \approx \mathbf{\Delta}_l$ to a good approximation. This implies that the asperities are always in the strong pinning regime, in which all asperities lie essentially in their contact potential minimum.

Although it has been argued here that the GW model predicts the occurrence of a sufficiently dilute concentration of asperities with strong enough forces acting on them due to the second solid to consider the asperities to be essentially uncorrelated, this still does not necessarily guarantee that there will be static friction, since it has been argued that even for uncorrelated asperities, static friction will only occur if the asperities exhibit multistability.[7,13] The condition for multistability to occur at an interface,[9] namely, that the force constant due to the asperity contact potential be larger than that due to the elasticity of the asperity ($\approx E'a$), however, is satisfied, as discussed under Eq. (12).

In conclusion, when one considers atomically smooth surfaces, the disorder at an interface between two weakly interacting nonmetallic elastic solids in contact will not result in static friction. When one considers the distribution of asperities that occur on multimicrometer length scales, however, one finds that the asperities are virtually always in the "strong pinning regime," in which asperities always lie close to a minimum of the potential due to their contact with a second solid. This accounts for the fact that there is almost always static friction. Muser and Robbins' idea,[2] however, is not invalidated by this argument. Their result will still apply for a smooth crystalline interface. It will also apply in the present context to the contact region between two asperities, implying that for a clean interface the shear force between contacting asperities is proportional to the square root of the contact area.[20] The GW model predicts for this case that the average force of friction is proportional to the 0.8 power of the load,[16] but this load dependence is not significantly different from when the asperity contact force is proportional to the contact area. This is illustrated in Fig. 1, where the both the integral over s in Eq. (9) and the integral

$$\int_h^\infty ds\, \phi(s)(s-h)^{1/2} \quad (13)$$

are plotted as a function of the dimensionless integral over s in Eq. (11) for the load. This quantity, and hence the static friction, are approximately proportional to the 0.8 power of the load. Furthermore, some simple arguments show that although the Muser-Robbins[2] picture, when the effects of asperities considered in the present work are taken into account, does not allow one to conclude that there will be no static friction for clean surfaces, it does predict that the static friction for clean surfaces is much smaller than what is normally observed. The argument is as follows: If the interface between two asperities is either in the strong pinning limit or using the Muser-Robbins[2] picture, it contains a submonolayer of mobile molecules, the force of static friction per asperity is given by

$$F_s/N = E_r \langle A_c \rangle, \quad (14)$$

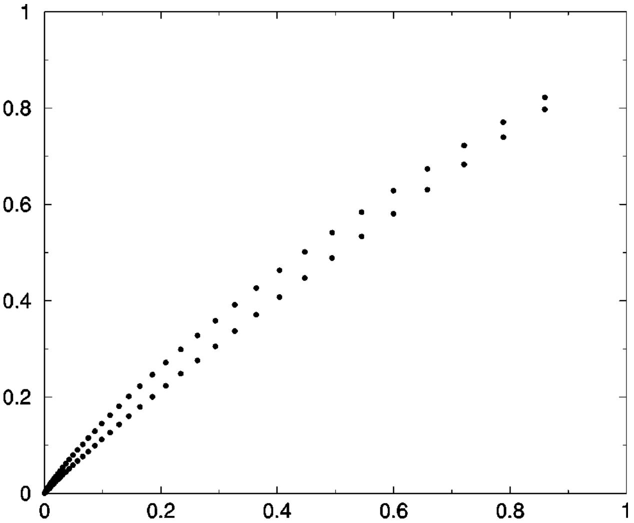

FIG. 1. The integral over s in Eq. (9) (the lower curve) and the integral over s in Eq. (13) (the upper curve) are plotted as a function of the integral over s in Eq. (11), which is proportional to the load. All quantities are dimensionless.

where one expects for the shear rupture strength at the asperity contact region E_r

$$E_r \approx cV_0/b^3, \tag{15}$$

where c is the concentration of defects at the interface, V_0 is the strength of the defect potential, and b is of the order of a lattice constant. Using the sample parameters given earlier in this section, we obtain as an estimate of the static friction coefficient

$$\mu_s = F_s/F_L \approx 0.1. \tag{16}$$

According to the Muser-Robbins argument, for clean surfaces, Eq. (14) is replaced by

$$F_s/N \approx E_r\langle A_c(b^2/cA_c)^{1/2}\rangle = E_r b\langle A_c^{1/2}\rangle/c^{1/2} \tag{17}$$

which when one again substitutes the sample parameters given earlier in this section gives

$$\mu_s \approx 10^{-5}. \tag{18}$$

On the basis of this argument, one concludes that the ideal static friction coefficient between clean, weakly interacting surfaces in the light load limit, is much smaller than what one typically observes.

IV. CONCLUSIONS

The main conclusions are as follows. There will always be static friction for any macroscopic size interface (i.e., an interface with lateral dimensions much larger than typical asperity spacing). The physical reason for this is that the potential acting between two asperities in contact varies on length scales of the order of typical atomic spacings. Since this distance is much smaller than typical asperity spacing, all asperties can sink to their potential minima with negligible cost in elastic potential energy of the contacting solids. As a consequence, when an external stress is applied, all asperities will be pushed out of their potential minima by almost the same amount, resulting in a net force of static friction. The mechanism of Ref. 2 can still be applied to the area of contact between two contacting asperities, however. It can result in very small (although still nonzero) values for the static friction between two macroscopic solids.

ACKNOWLEDGMENT

I wish to thank the Department of Energy (Grant No. DE-FG02-96ER45585) for their financial support. I also wish to thank R. Markiewicz for useful discussions.

APPENDIX: EQUILIBRIUM CONDITION FOR PINNING CENTERS AT AN INTERFACE

The interaction forces between two solids in contact act at various points along the interface for atomically flat surfaces are expected to act at the points of contact of the atoms of both solids at the interface. Since it has been established, however, that weakly interacting perfect incommensurate surfaces exhibit no static friction,[1–3] we expect that any static friction that occurs is due to defects. Therefore, let us consider the interaction of point defects, randomly distributed over the interface. We must consider both the potential of interaction of a defect of one surface with an atom on the second surface and the elastic energy cost that one must pay when the solid distorts in order to minimize the interfacial and elastic potential energies. If $\mathbf{f}_j$ represents the force acting on an atom at site j in the solid due to its interaction with a second solid with which it is in contact, the displacement of the atom at the lth lattice site $\mathbf{u}_l$ is given by

$$\mathbf{u}_l = \mathbf{G}_{l,j}\cdot\mathbf{f}_j, \tag{A1}$$

where

$$\mathbf{G}_{l,j} = \mathbf{D}_{l,j}^{-1}, \tag{A2}$$

where $\mathbf{D}_{l,j}$ is the dynamical matrix.[17] Then Eq. (A1) can be written as

$$\mathbf{u}_l = \mathbf{G}_{l,l}\cdot\mathbf{f}_l + \sum_{j\neq l}\mathbf{G}_{l,j}\cdot\mathbf{f}_j. \tag{A3}$$

Following the discussion in Ref. 14, we find that

$$\mathbf{G}_{l,j} = v(2\pi)^{-3}\sum_\gamma \int d^3k e^{i\mathbf{k}(\mathbf{R}_l-\mathbf{R}_j)}\frac{\hat{\boldsymbol{\epsilon}}_{\mathbf{k}}^{\gamma}\hat{\boldsymbol{\epsilon}}_{\mathbf{k}}^{\gamma}}{m\omega_\gamma^2(\mathbf{k})}, \tag{A4}$$

where m is the mass of an atom in the solid, $\hat{\boldsymbol{\epsilon}}_{\mathbf{k}}^{\gamma}$ is the unit vector which gives the polarization of the γth phonon mode of wave vector $\mathbf{k}$, $\mathbf{R}_j$ is the location of the jth atom, and v is the unit cell volume. In order to simplify the problem, let us replace the tensor $\hat{\boldsymbol{\epsilon}}_{\mathbf{k}}^{\gamma}\hat{\boldsymbol{\epsilon}}_{\mathbf{k}}^{\gamma}$, by the unit tensor, which should give results of the correct order of magnitude. Then Eq. (A4) becomes when the integral over k is done in the Debye approximation

$$\mathbf{G}_{l,l} \approx \frac{9}{mc^2k_D^3}\int_0^{k_D} dk \frac{\sin(kR)}{k}, \tag{A5}$$

where $R=|\mathbf{R}_l-\mathbf{R}_j|$, c is the mean sound velocity and we have used the fact that the Debye wave vector k_D is related to v by

$$v(2\pi)^{-3}4\pi k_D^3/3=1. \tag{A6}$$

In Ref. 17, it is shown that the elastic constants are given by

$$-(1/2v)\sum_{\mathbf{R}} \mathbf{R}\cdot\mathbf{D}(\mathbf{R})\mathbf{R}. \tag{A7}$$

The magnitude of a typical value of an elastic constant E is given by

$$E=(1/2)(2\pi)^{-3}b^2\sum_{\gamma}\int d^3km\omega_{\gamma}^2(\mathbf{k}). \tag{A8}$$

When Eq. (A8) is evaluated in the Debye approximation, we obtain

$$E=(3/10\pi^2)mc^2k_D^5b^2. \tag{A9}$$

For $k_DR \gg 1$, we find using Eqs. (A5) and (A9), taking $k_Db \approx \pi$

$$\mathbf{G}_{l,j}=(E'R)^{-1}, \tag{A10}$$

where $E'=(40/9)E$. For $k_DR \ll 1$,

$$G \approx (E'b)^{-1}. \tag{A11}$$

For simplicity, we assume that $\mathbf{f}_j$ has the form

$$\mathbf{f}_j=\alpha_j(\mathbf{\Delta}_j-\mathbf{u}_j). \tag{A12}$$

Then, from Eqs. (A1), (A3), (A10), (A11), and (A12), we obtain Eq. (7).

The equilibrium condition expressed in Eq. (10) can also be applied to an interface for which the contact takes place only at a dilute concentration of randomly placed contacting asperities, giving for the displacement at a point on the lth asperity

$$\mathbf{u}_l=\sum_j \int d^2r_j'(E'|\mathbf{r}_l-\mathbf{r}_j-\mathbf{r}_j'|)^{-1}\mathbf{p}(\mathbf{r}_j'), \tag{A13}$$

where $\mathbf{r}_j$ is the location of a central point in the contact area of the jth asperity, $\mathbf{r}_j'$ gives the location of an arbitrary point on this asperity relative to $\mathbf{r}_j$, and $\mathbf{p}(\mathbf{r}_j')$ is the shear stress at the point $\mathbf{r}_j'$. We have replaced the summation over atomic positions in Eq. (A1) by the integral over $\mathbf{r}_j'$ over the contact area of the jth asperity. In the dilute asperity limit, in which $|\mathbf{r}_l-\mathbf{r}_j| \gg r_j'$, Eq. (A13) is to a good approximation

$$\mathbf{u}_l=\sum_{j\neq l}(E'|\mathbf{r}_l-\mathbf{r}_j|)^{-1}\mathbf{f}_j+\int d^2r_l'(E'|\mathbf{r}_l+\mathbf{r}_l'|)^{-1}\mathbf{p}(\mathbf{r}_l'), \tag{A14}$$

where $\mathbf{f}_j=\int d^2r_j'\mathbf{p}(\mathbf{r}_j')$, where the range of integration is over the contact area of the jth asperity. For simplicity, we may replace $\mathbf{p}(\mathbf{r}_l')$ in the integral over r_l' by its average value, denoted by $\mathbf{f}_l/(\pi a^2)$. Then we need to estimate the integral

$$\int d^2r'|\mathbf{r}+\mathbf{r}'|^{-1}, \tag{A15}$$

where $\mathbf{r}$ denotes a point on the lth asperity and the integral runs over the contact area of this asperity. Taking the contact area to be a circle of radius a, this integral can easily be shown to be equal to

$$4\int_0^{\infty} dr'r'(r+r')^{-1}K\left(\frac{4rr'}{r+r'}\right), \tag{A16}$$

where $K(k)$ is the complete elliptic function. It has a logarithmic singularity at $r'=r$, which is integrable, and is of order 1 away from the singularity. Consequently, the integral is of order a and we obtain a contribution of order $(E'a)^{-1}\mathbf{f}_l$ for the last term in Eq. (A14). Hence Eq. (A14) becomes

$$\mathbf{u}_l=(E'a)^{-1}\mathbf{f}_l+\sum_{j\neq l}(E'|\mathbf{r}_l-\mathbf{r}_j|)^{-1}\mathbf{f}_j. \tag{A17}$$

If for simplicity, we replace $\mathbf{f}_j$ by $\alpha_j(\mathbf{\Delta}_j-\mathbf{u}_j)$, as was done in Sec. II, and we obtain the equilibrium condition for the defects used in that section. Since $\mathbf{u}_l$ was estimated in that section to be only $0.06\mathbf{\Delta}_l$, the use of the harmonic approximation was not really justified in Sec. II. It was only used there for simplicity to illustrate the fact the system is in the weak pinning limit. In contrast, in the application of Eq. (A14) to the multiasperity problem in Sec. III, we shall see in the next paragraph that this approximation is justified.

In zeroth order in the interasperity interaction, we have

$$\mathbf{u}_l^{(0)} \approx (E'a)^{-1}\mathbf{f}_l=(E'a)^{-1}\mathbf{f}[(\mathbf{\Delta}_l-\mathbf{u}_l^{(0)})/b], \tag{A18}$$

where b represents the length scale over which $\mathbf{f}$ varies with $\mathbf{u}_l^{(0)}$, which was argued in Sec. III to be of the order of an atomic spacing. Making the substitution $(\mathbf{\Delta}_l-\mathbf{u}_l^{(0)})=b\mathbf{v}$, Eq. (A18) becomes

$$E'a(\mathbf{\Delta}_l-b\mathbf{v})=\mathbf{f}(\mathbf{v}). \tag{A19}$$

Since $\mathbf{f}$ oscillates on a length scale b, it will almost certainly have a zero in the vicinity of $\mathbf{v}=\mathbf{\Delta}/b$. Then since $E'ab \ll |\partial\mathbf{f}/\partial v_x|, |\partial\mathbf{f}/\partial v_y|$, there will be solutions to Eq. (A18) with $|\mathbf{f}|$ much less than its maximum value, because if we plot the left hand side of Eq. (A19) versus v for $\mathbf{v}$ along $\mathbf{\Delta}$, the slope of the straight line on the left hand side is much less than the slope of the f versus v curve on the right hand side. Hence, the point of intersection of these two curves is at a point which is much below the maximum value of f. Thus, we are justified in expanding $\mathbf{f}$ in a Taylor series in $\mathbf{v}$ around the nearest zero of $\mathbf{f}$ to $\mathbf{v}=\mathbf{\Delta}/b$. If the potential that $\mathbf{f}$ is derived from is chosen for simplicity to be a spherically symmetric function of $\mathbf{v}$, we may write $\mathbf{f}(\mathbf{v}) \approx \alpha\mathbf{v}$, where α is a constant.

[1] J. E. Sacco and J. B. Sokoloff, Phys. Rev. B **18**, 6549 (1978).

[2] 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).

[3] S. Aubry, in *Solitons and Condensed Matter*, edited by A. R. Bishop and T. Schneider (Springer, New York, 1978), p. 264.

[4] J. B. Sokoloff, Phys. Rev. B **51**, 15 573 (1995); J. B. Sokoloff and M. S. Tomassone, *ibid.* **57**, 4888 (1998).

[5] M. S. Tomassone and J. B. Sokoloff, Phys. Rev. B **60**, 4005 (1999).

[6] G. A. Tomlinson, Philos. Mag. **7**, 905 (1929).

[7] C. Caroli and Ph. Nozieres, Eur. Phys. J. B **4**, 233 (1998); *Physics of Sliding Friction*, edited by B. N. J. Persson and E. Tosatti, *NATO ASI Series E: Applied Sciences*, Vol. 311 (Kluwer, Dordrecht, 1996).

[8] D. S. Fisher, Phys. Rev. B **31**, 1396 (1985); Phys. Rev. Lett. **50**, 1486 (1983).

[9] 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 Publishers, Boston, 1995), p. 179; V. L. Popov, Phys. Rev. Lett. **83**, 1632 (1999).

[10] M. L. Falk and J. S. Langer, Phys. Rev. E **57**, 7192 (1998).

[11] A. I. Larkin and Yu. N. Ovchinikov, J. Low Temp. Phys. **34**, 409 (1979).

[12] H. Fukuyama and P. A. Lee, Phys. Rev. B **17**, 535 (1977); P. A. Lee and T. M. Rice, *ibid.* **19**, 3970 (1979).

[13] J. B. Sokoloff, Phys. Rev. B **23**, 1992 (1981).

[14] B. C. Stipe, H. J. Mamin, T. D. Stowe, T. W. Kenny, and D. Rugar, Phys. Rev. Lett. **87**, 096801 (2001).

[15] A. Volmer and T. Nattermann, Z. Phys. B: Condens. Matter **104**, 363 (1997).

[16] J. A. Greenwood and J. B. P. Williamson, Proc. R. Soc. London, Ser. A **295**, 3000 (1966); J. I. McCool, Wear **107**, 37 (1986).

[17] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976), pp. 443–444.

[18] J. B. Sokoloff, Phys. Rev. Lett. **86**, 3312 (2001).

[19] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Pergamon Press, New York, 1970), p. 30.

[20] L. Wenning and M. H. Muser, Europhys. Lett. **54**, 693 (2001).