LETTER TO THE EDITOR

On the thermodynamic definition of surface stress

E M Gutman

Department of Materials Engineering, Ben-Gurion University of the Negev, PO Box 653, Beer-Sheva 84105, Israel

Abstract. Surface energy and surface stress have been thoroughly discussed in the literature. However, the widely accepted definition of surface stress as the strain derivative of the *total* free surface energy divided by the surface area (Shuttleworth, Herring, etc) is inconsistent with the correct definition of a stress tensor (given in continuum elasticity theory) as the strain derivative of the *specific* free energy. We analyse this anomaly in detail in order to prevent the erroneous use of the surface stress definition.

There exist two principal approaches to surface stress analysis: the surface layer of finite thickness method [1] and Gibbs' geometrical method [2]. The latter involves a certain dividing surface, and the excess of energy, entropy and other extensive properties is conventionally referred to it. The advantage of the method involving the surface layer lies in the fact that the surface layer of finite thickness actually exists. Its excessive free energy (with respect to the bulk phase) is due to a physical surface of discontinuity disturbing the mechanical equilibrium, inherent to the bulk phase, and leading to the layer deformation. Although the inhomogeneity distribution in the surface layer is, in general, of an asymptotic character, we can speak about its effective thickness. For instance, even in liquid CCl_4 at 293 K a deviation of 1% from Pascal's law can be achieved only at a distance of forty molecular spacings, whereas one of 0.1% can be achieved at the distance of eighty-six molecular spacings from the free surface [3]. The specific free energy of the surface may then be considered as Helmholtz's free energy calculated for an element of surface layer resting upon a unit surface area.

Surface stress in a solid was determined [4–10] as the derivative of the *total* surface free energy with respect to strain divided by the surface area. This is inconsistent with the definition given in the continuum elasticity theory [11]. Such a representation, in particular, results in the following: as in a liquid, in a globally relaxed state the stress in an anisotropic surface is given by the specific surface energy of the unstrained surface [5] or by that of the surfaces under 'equilibrium' strain, as corrected later [6]. This leads to the solid-surface representation using the model of an extensible liquid film, where the surface area grows at the expense of the material transfer from the bulk to the surface. As was noted by Gibbs [2], this is apparently possible only in certain particular cases, where a solid can be considered as a highly viscous fluid.

Definition of surface stress in terms of the strain derivative of the total surface free energy [5–10] leads to an expression for the surface stress in the form of a sum of two terms:

$$\sigma_{ij} = A^{-1}\frac{\partial(AF^s)}{\partial\epsilon_{ij}} = F^s\delta_{ij} + \frac{\partial F^s}{\delta\epsilon_{ij}} \tag{1}$$

that was proposed earlier by Shuttleworth [4] in the form

$$\gamma_0 = \frac{d(AF)}{dA} = F^s + A\frac{dF^s}{dA} \tag{2}$$

and became fundamental for all subsequent papers.

Here γ_0 is the normal component of the surface stress equal to the surface tension [4], F^s is the Helmholtz free surface energy per unit area of original (undeformed) surface area (Wolf [5], when citing Shuttleworth's work [4], points out that Shuttleworth derived equation (2) for an unstrained surface), A is the surface area, δ_{ij} is the Kronecker delta-function, σ_{ij} and ϵ_{ij} are the surface stress and surface strain tensors, respectively.

However, one can easily see that equations (1) and (2) are basically incorrect, since in the elasticity theory, stress tensor components are defined by the strain derivative of the *specific (not total)* Helmholtz free energy representing the total free energy referred to a unit of original volume or surface area. Therefore, the surface stress tensor, consisting, as usual, of a mean stress tensor and a deviator stress tensor [11, 12], should be described by

$$\sigma_{ij} = \left(\frac{\partial F^s}{\partial \epsilon_{ij}}\right)_{T,V} = \frac{1}{2}\left(\frac{\partial F^s}{\partial \epsilon_{kk}}\right)_{T,V}\delta_{ij} + \left(\frac{\partial F^s}{\partial \epsilon'_{ij}}\right)_{T,V} \tag{3}$$

where ϵ'_{ij} is a deviator strain tensor, instead of the equations (1) and (2).

The erroneous nature of equations (1) and (2) results from the confusing of different thermodynamic systems. The first of the terms is caused by forming a new surface with the same properties (the specific surface energy remaining unchanged). The other describes a changing specific surface energy due to the strain of the existing surface. However, these two processes (at least, in the limit) are relevant to completely different materials. The first represents the appearance of surface tension in a fluid with movable molecules, while the second represents the strain of a solid, where the total number of atoms in the surface layer remains constant, while their spacing changes. Therefore, it is not legitimate to sum these two terms to describe the surface stress in a solid.

The reason for this misuse is the incorrect application of the thermodynamic definition of stress which should be defined as the strain derivative of the specific free energy, i.e., of the energy referred to a unit volume or surface [11]. It is especially emphasized [11] that all thermodynamic values refer to the amount of matter confined in an unstrained volume or surface, the mentioned amount remaining unchanged in the process of strain of a solid, i.e., to a unit volume or surface of an unstrained solid. The total energy can always be obtained by integrating over the volume or surface of the unstrained solid [11].

It is only during the solid surface layer deformation that a reversible surface growth goes on with an unchanged amount of the material on the surface. The case of liquid film extension is quite different. Here the formation of the new surface proceeds not at the expense of elastic (reversible) strain, but at the expense of adding (irreversible transport phenomena) a certain amount of matter from the bulk material to the surface layer. In this case, the total free surface energy grows proportionally to the surface increment, the specific free energy remaining practically unchanged. Thus, the differentiation of the *total* surface energy with respect to the strain, taking into account that the area allowed within the differential sign represents a function of strain [4], has no physical meaning.

Indeed, equation (2) was obtained [4] using formal differentiation of the total Helmholtz free surface energy $F = AF^s$:

$$dF = d(AF^s) = \left(F^s + A\frac{dF^s}{dA}\right)dA. \tag{4}$$

It has been generally accepted practice to regard the expression in brackets in equation (4) as an ambiguous one, dependent on the way in which the surface changes [2]. The surface area may be increased in two different ways: (i) the first consists of forming a new surface identical to the original one in nature, at the expense of increasing the number of atoms on the surface coming from the bulk material; (ii) the second consists of stretching the *existing* surface, which alters its states of strain, whilst the number of surface atoms remains unchanged.

For case (i), the specific surface energy is independent of A, and we obtain the following for dF:

$$d_{(i)}F = F^s \, dA. \tag{5}$$

For case (ii), Shuttleworth [1] wrote the following

$$d_{(ii)}F = \gamma_0 \, dA \tag{6}$$

for an isotropic solid deformed by an equal stretch in all directions. He further combines equations (6) and (4), i.e. $dF = d_{(ii)}F$, and comes to equation (2) for γ_0. However, equating equation (6) to equation (4) is incorrect, since equation (4) formally describes both cases of surface area increase, while equation (6) only describes case (ii). Such equating carries us to absurdity, because these equations are relevant to *different* substances. In fact, one could as legitimately apply case (i) and formally equate equations (5) and (4), i.e. $dF = d_{(i)}F$, which would lead to

$$F^s = F^s + A\frac{dF^s}{dA}. \tag{7}$$

Obviously, such an operation could have been accomplished only under the condition $dF^s/dA = 0$, which only represents the condition of the realization of case (i).

Consequently, a respective limiting condition should exist for case (ii), as well. In fact, taking into account that $dA/A = d\epsilon_{kk}$, equation (6) may be rewritten as

$$d_{ii}F = A \, dF^s \tag{8}$$

since the increment of the specific surface energy due to deformation $dF^s = \gamma_0 \, d\epsilon_{kk}$ [11].

If we now formally equate equations (8) and (4), i.e. $dF = d_{(ii)}F$, we obtain

$$d(AF^s) = A \, dF^s \tag{9}$$

which requires the limiting condition $A = $ constant. This means that the change in the surface area at the expense of surface stress appearance should be so small as to allow one to neglect the additional term $F^s \, dA$ in the increment of the total surface energy (9). This corresponds to the assertion of the elasticity theory [11] saying that all thermodynamic values are referred to the original (i.e., before the deformation) volume or surface area of the solid.

Hence, the equating of equations (5) and (6) taken separately from equation (4) is, on the whole, invalid in any case, because they are relevant to different substances (a liquid and a solid). For extreme cases, where either case (i) or case (ii) is realized, one should separately equate equations (5) and (6) to the respective terms in the right-hand part of equation (4).

For a certain hypothetical intermediate case (for instance, that of an isotropic amorphous body considered as a highly viscous supercooled liquid [2]), one should sum equations (5) and (6) as independent contributions of both cases and consider equation (4) as a complete differential with respect to the variables A and F^s. However, here it is also illegitimate to equate this equation to $A\sigma_{ij}\delta\epsilon_{ij}$ (i.e., to $d_{(ii)}F$), as Herring [8] does, because in the present intermediate case $dF \neq d_{(ii)}F$.

Thus, in the extreme case of a liquid, equation (6) denotes the work done against the surface stress to create a new surface. It is equivalent to the increase in the total surface energy according to equation (5), i.e., from the comparison of equations (5) and (6), $\gamma_0 = F^s$ is obtained, as usual. However, in the extreme case of a solid, the energy contribution due to a new surface formation according to equation (5) is negligible in comparison with the contribution of the surface layer deformation arising due to asymmetric force fields of surface atoms (in comparison with bulk atoms). Therefore, γ_0 and σ_{ij} are defined only by the value of the strain derivative of the *specific* (not total) free surface energy arising due to this strain.

There is a further point to be made. We can also show some elementary incorrectness in the formulation of principal equations. For instance, the last term in the right-hand part of equation (1.9) in [9] does not agree in dimension with other summands due to the factor A. Let us clear up whether the elimination of the factor A (if it was a typographical misprint) in equation (1.9) will change, and save, the situation. Equation (1.9) is written in [9] as follows:

$$dU = \left.\frac{\partial U}{\partial S}\right|_{V,N,A} dS + \left.\frac{\partial U}{\partial V}\right|_{S,N,A} dV + \left.\frac{\partial U}{\partial N}\right|_{S,V,A} dN + A\sum_{i,j}\left.\frac{\partial U}{\partial \epsilon_{ij}}\right|_{S,V,N} d\epsilon_{ij}$$

$$dU = T\,dS - P\,dV + \mu\,dN + A\sum_{i,j}\sigma_{ij}\,d\epsilon_{ij} \tag{10}$$

where U is the internal energy of a system, S is the entropy, V is the volume, N is the particle number, T is the temperature, P is the pressure and μ is the chemical potential. Let us first demonstrate that, even if we eliminate the factor A to save the dimensions in the last term of the right-hand part of the upper equation (10), the lower equation does not follow from the upper one, because the inequality

$$\sum_{i,j}\left.\frac{\partial U}{\partial \epsilon_{ij}}\right|_{S,V,N} d\epsilon_{ij} \equiv dU|_{S,V,N} \equiv \sum_{i,j}\left.\frac{\partial (U_0 A)}{\partial \epsilon_{ij}}\right|_{S,V,N} d\epsilon_{ij} \neq A\sum_{i,j}\sigma_{ij}\,d\epsilon_{ij} \tag{11}$$

is valid if A is *not constant* (the constant variables are S, V and N only). Here U_0 is the specific surface energy equal to γ in equation (1.5) of [9]:

$$U = TS - PV + \mu N + \gamma A. \tag{12}$$

An equality can take the place of this inequality (11) only under the condition A = constant. However, this contradicts the entire analysis in [9], A being a thermodynamic variable. Thus, there is the formal equality

$$\sum_{i,j}\left.\frac{\partial U}{\partial \epsilon_{ij}}\right|_{S,V,N} d\epsilon_{ij} \equiv dU|_{S,V,N} \equiv \sum_{ij}\left.\frac{\partial (U_0 A)}{\partial \epsilon_{ij}}\right|_{S,V,N} d\epsilon_{ij} = \sum_{i,j}\left(A\sigma_{ij}\,d\epsilon_{ij} + U_0\frac{\partial A}{\partial \epsilon_{ij}}\,d\epsilon_{ij}\right)$$

$$\equiv A\sum_{i,j}(\sigma_{ij} + \gamma\delta_{ij})\,d\epsilon_{ij} \tag{13}$$

as used in Shuttleworth's formalism [4] (of course, in this expression $\sigma_{ij} = \partial U_0/\partial \epsilon_{ij}$ is satisfied [11]). But even this formal equality (13), transforming the lower equation (10) into

$$dU = T\,dS - P\,dV + \mu\,dN + A\sum_{i,j}(\sigma_{ij} + \gamma\delta_{ij})\,d\epsilon_{ij} = T\,dS - P\,dV$$

$$+ \mu\,dN + \gamma\,dA + A\sum_{i,j}\sigma_{ij}\,d\epsilon_{ij}, \tag{14}$$

cannot help the author of [9], since in this case joint use of his equations (1.9) and (1.5) *does not lead* to his equation (1.10), represented here by equation (1), etc.

However, what is much more important (than the dimension problem) is that it is basically *impossible* to use equations (1.9) and (1.5) jointly in order to obtain equation (1.10) in [9]. In fact, equations (1.9) and (1.5) refer to *different* thermodynamic systems. Equation (1.5) describes the system with a new surface formed at the expense of '...a cleavage process', while equation (1.9) describes the system where '...variations in the area of the system...are realized...by stretching'. In the first case γ = constant, while in the second the specific surface energy changes due to deformation. Therefore, equations (1.5) and (1.9) involve different thermodynamic variables, and it seems absurd to apply them jointly to obtain equation (1.10), etc. Due to this absurdity, equation (1.10) that the author of [9] calls the Gibbs–Duhem equation cannot, in fact, bear this name, since it contains differentials of the *extensive* parameters A and ϵ_{ij}. This is inconsistent with classical thermodynamics stating that the Gibbs–Duhem equation presents the relationship between the *intensive* parameters in the differential form.

Thus, this simple analysis shows that Shuttleworth's definition of surface stress should be substituted with the definition of the stress tensor generally accepted in the elasticity theory as the strain derivative of the specific surface free energy.

References

[1] Guggenheim E A 1940 *Trans. Faraday Soc.* **36** 397
[2] Gibbs J W 1948 *Collected Works* vol 1 (New Haven, CT: Yale University Press) p 219
[3] Rusanov A I 1967 *Phase Equilibrium and Surface Phenomena* (St Petersburg: Khimia) (in Russian)
[4] Shuttleworth R 1950 *Proc. Phys. Soc.* A **63** 449
[5] Wolf D 1993 *Phys. Rev. Lett.* **70** 627
[6] Chou M Y, Siqing Wei and Vanderbilt D 1993 *Phys. Rev. Lett.* **71** 461
[7] Needs R J, Godfrey M J and Mansfield M 1991 *Surf. Sci.* **242** 215
[8] Herring C 1953 *Structure and Properties of Solid Surfaces* ed R Gomer and C S Smith (Chicago, IL: University of Chicago Press) p 14
[9] Zangwill A 1988 *Physics at Surfaces* (Cambridge: Cambridge University Press) p 9
[10] Cammarata R C 1992 *Surf. Sci.* **279** 341
[11] Landau L D and Lifshitz E M 1981 *Theory of Elasticity* (Oxford: Pergamon) p 8
[12] Dieter G E 1988 *Mechanical Metallurgy* (London: McGraw-Hill) p 43