FRACTURE CRITERIA DERIVED FROM ENERGY BALANCE CONCEPTS

R.N. Stevens[+] and F. Guiu[+]

Fracture energy balance concepts are reviewed. It is shown that when energy dissipating processes accompany crack growth the energy changes caused thereby should not be included in the calculation of a crack extension force. With this proviso it is shown that a fracture criterion based on a critical value of a crack-tip characterisation parameter such as K is identical to one based on a critical value of crack extension force, even when additional dissipative processes accompany crack growth. It is also shown that the work of plastic deformation does not appear in a fracture criterion as an effective surface energy.

The original Griffith theory of fracture [1] is an example of the theory of thermodynamic instability. This theory is based on the second law of thermodynamics and states that a system becomes unstable when its thermodynamic potential reaches a maximum value. Elastic bodies containing cracks and undergoing other dissipating processes are not purely mechanical (i.e., entropy-less) systems and it is essential to use the second law to discuss both the stable and unstable equilibrium of such systems.

For a system which is in the form of a rod of length L and cross-sectional area A with a tensile force f along its length, the Gibbs free energy, G, is the potential determining equilibrium when the system is held at constant T and L. Elastic strain energy is a Helmholtz free energy and has a significant entropy component. Surface energy is also a Helmholtz free energy. For the simple system being considered here G and F are related by

$$G = F - fL \tag{1}$$

+ Department of Materials, Queen Mary & Westfield College, University of London, Mile End Road, London, E1 4NS

The total Gibbs free energy change, $(dG_t)_{T,f}$ in a purely elastic system when a crack increases in area by da at constant T and f is therefore

$$(dG_t)_{T,f} = [dF_c - f\, dL_c]_{T,f} + 2\gamma\,(da)_{T,f} \qquad (2)$$

where the subscript c refers to the changes arising from changes in compliance of the system and γ is the surface energy. When the system is at unstable equilibrium dG_t is zero. When it is not the thermodynamic force motivating crack growth is defined by $-(\partial G_t/\partial a)_{T,f}$. This force can be divided into a driving term $\zeta = -(\partial G_c/\partial a)_{T,f}$ (where $dG_c = dF_c - f\, dL_c$), and a resistive term 2γ. The energy balance is satisfied when $\zeta = 2\gamma$ since this gives $dG_t = 0$. No real process will proceed with zero net driving force and thus, in a practical situation, $\zeta > 2\gamma$ and the Griffith energy balance will never be satisfied. However, the theory of thermodynamic instability does set a lower bound on the conditions for crack extension and, more importantly allows the identification of the driving force, ζ which can be expected to determine crack behaviour.

When $\zeta > 2\gamma$ crack growth is a dissipative process, even in the purely elastic case. In real materials crack growth is almost always accompanied by other energy dissipating processes, for example plastic deformation. Experience suggests that this makes crack growth more difficult. A common characteristic of these dissipative processes is that they result from the production and/or movement of elastic singularities such as dislocations in the case of plastic deformation. It is far from obvious how these additional dissipative processes should affect the energy balance and the calculation of the crack driving or resistive forces. It is possible to distinguish three different approaches. Two of these, although widely used, are fallacious, and will be considered first.

The overall energy balance. In this case a crack driving force is defined by adding all the free energy changes occasioned by every dissipative process occurring with crack extension and dividing this by the increment of crack area. The problem is that dissipative processes will decrease the Gibbs free energy and will appear to give rise to driving rather than resistive forces. Plastic deformation is taken as a typical dissipative process. The crack is assumed to remain sharp and have dislocations distributed around the tip and flanks. When it extends these move and more are generated, elongating the body by dL_p and changing the strain energy by dF_p as elastic strain is replaced by plastic strain. There will also be an increase dF_d due to strain energy stored in the strain fields of the extra dislocations produced. Hence

$$(dG_t)_{T,f} = [dF_c - f\, dL_c]_{T,f} + [dF_p - f\, dL_p]_{T,f} + (dF_d)_{T,f} + 2\gamma(da)_{T,f} \qquad (3)$$

If an overall instability condition exists, $(dG_t)_{T,f} = 0$ and

$$-[dF_c - f\, dL_c]_{T,f} - [dF_p - f\, dL_p]_{T,f} = (dF_d)_{T,f} - 2\gamma\,(da)_{T,f} \qquad (4)$$

The first two terms on the left-hand side of this equation are of the same sign and this implies that plastic deformation provides an additional driving force for crack growth, contrary to experience. The quantity $[dF_p - f\, dL_p]_{T,f}$ is the work of plastic deformation and Orowan [2] and Irwin [3] independently suggested that the

1004

work of plastic deformation, p , per unit area of crack extension should appear with a positive sign as a resistive not a driving force. According to this the instability condition is given by

$$- [dF_c - f\,dL_c]_{T,f} = 2(\gamma + p)\,(da)_{T,f} \qquad (5)$$

or $\qquad \zeta = 2(\gamma + p) \qquad\qquad\qquad\qquad (6)$

Clearly Equations (3) and (4) do not take into account correctly the effect of plastic deformation on crack growth. On the other hand it does not appear possible to fit the hypothesis of Orowan and Irwin into the framework of the theory of thermodynamic instability since it incorporates the work of plastic deformation with the wrong sign.

An energy balance according to the first law of thermodynamics. This balance is quite different from that of Griffith although this is not generally recognised. An early expression of it is by Gurney and Hunt [4] and it has been given more recently by Williams [5]. The balance states that the energy input in the form of work must equal the energy stored in the body less the energy lost as heat. This is the first law of thermodynamics and energy is therefore internal energy, U. Thus we have

$$-f\,dL = dU_e + 2u_s da - dQ \qquad (7)$$

where dQ is the heat absorbed by the system and dU , the internal energy change, has been separated into a component dU_e due to the change in elastic strain of the system and a component due to the increase in surface area, $2u_s da$ with u_s the internal energy per unit area of surface. It is important to note that U_e is not strain energy and u_s is not surface energy.

The physical content of this energy balance is quite different from that of Griffith. It is simply an energy conservation balance and is always obeyed whatever the crack length. The first law of thermodynamics can give no information about the equilibrium of a system hence no criterion for fracture can be derived from it nor can the driving forces for crack growth be identified from it. Nevertheless Equation (7) is often developed by assuming (wrongly) that when plastic deformation takes place $-dQ$ (= $p\,da$) is the work of plastic deformation dissipated as heat and therefore

$$-\frac{dU_e - f\,dL}{da} = 2(u_s + p) \qquad (8)$$

The left-hand side of (8) is then said to be the driving force for crack growth and the Orowan-Irwin hypothesis appears to be derived. This is quite false. Equation (7) is true at the instant of crack propagation only in the trivial sense that it must always be true whatever the crack length, and cannot give a fracture criterion or a driving force.

In spite of their radically different appearance Equations (3) and (7) are consistent, each being derivable from the other. With T and f held constant the change in length dL in (7) can be separated in components dL_c and dL_p while dU_e can be split into dU_c due to the change in compliance, dU_p due to the exchange of elastic

for plastic strain and dU_d due to the storage of internal energy in the strain fields of the extra dislocations produced. Thus (7) becomes

$$- [dU_c - f \, dL_c]_{T,f} - [dU_p - f \, dL_p]_{T,f} = 2u_s \, (da)_{T,f} + (dU_d)_{T,f} - dQ \qquad (9)$$

The second law of thermodynamics is now introduced. The difference in entropy of the system with crack of area $a + da$ and with the crack of area a is dS. If the change had taken place reversibly the heat absorbed would have been $T \, dS$. The actual heat absorbed, dQ, is given by

$$dQ = T \, dS + dQ' \qquad (10)$$

where dQ' is the heat irreversibly dissipated. The entropy change dS can also be split into components: dS_c due to the change in compliance, dS_p from the exchange of elastic for plastic strain, dS_d the change of strain entropy in the strain fields of the dislocations produced and finally that due to the change in the crack surface area, $2s_s da$ where s_s is the surface entropy per unit area. This enables Equation (10) to be written

$$dQ = T \, (dS_c + dS_p + dS_d + 2s_s da) + dQ' \qquad (11)$$

When (11) is put into (9) and all the dU_x terms collected with the corresponding TdS_x terms to form the Helmholtz free energies $dU_x - TdS_x$ we obtain

$$- [dF_c - f \, dL_c]_{T,f} - [dF_p - f \, dL_p]_{T,f} = (dF_d)_{T,f} - 2\gamma(da)_{T,f} - dQ' \qquad (12)$$

The true driving forces appear in (12). This is only possible because of the use of the second law. If equilibrium is assumed then $dQ' = 0$ and (4) is recovered.

The empty nature of these two forms of the energy balance can be illustrated well as follows. If crack growth and plastic deformation both take place then the dissipated heat, dQ' can be divided into terms dQ'_c and dQ'_p due to crack growth and due to plastic deformation respectively. However, the heat dQ'_p is equal to the difference $- [dF_p - f \, dL_p]_{T,f} - (dF_d)_{T,f}$ between the plastic work and the energy stored as dislocations, and all terms relating to plastic deformation disappear from (12) leaving

$$- [dF_c - f \, dL_c]_{T,f} = 2\gamma(da)_{T,f} - dQ'_c \qquad (13)$$

If plastic deformation is the only dissipator of heat then $dQ'_c = 0$ and the Griffith equation is recovered from (13).

The local energy balance: While Equation (3) and all the equations derived from it are thermodynamically correct, they have paradoxical consequences. This is because implicit in our discussion is the assumption that there is an overall instability condition encompassing all processes accompanying crack growth. Clearly each elastic singularity involved in the dissipative processes will have its own thermodynamic driving force to which it will respond. Hence the energy increment used to determine the thermodynamic driving force on each entity should be the energy decrease caused by its displacement alone. Thus free energy changes caused

by plastic deformation should not be counted in the evaluation of the crack extension force. A crack extension force so determined is called a local crack extension force since the energy changes taking place away from the crack are not counted.

Based on this view, the local crack extension force can be calculated in principle as follows. A change in Gibbs free energy at constant temperature and tensile force caused by increasing the crack area by da is, from Equation (1).

$$(dG)_{T,f} = (dF_e)_{T,f} + 2\gamma(da)_{T,f} - f(dL)_{T,f} \tag{14}$$

where dF_e is the change in strain energy and γ is the surface free energy. This equation gives the difference in Gibbs free energy of the system with a crack of area $a + da$ and the system with a crack of area a.. As indicated the extension of the crack occurs <u>with all singularities (such as dislocations) apart from the crack fixed in state and in position</u>. The following process allows a reversible transition between these states. A cut is made from the edge of the crack increasing its area by da with tractions applied to the cut surfaces to prevent them from opening. The tractions are then changed reversibly until they are zero and the enlarged crack is fully open. Work $d\omega_{ce}$ is done by the system against the elastic tractions during this process as well as work $d\omega_{cc}$ against the atomic cohesive forces. Work is also done against the external tensile force, f. Since the Helmholtz free energy changes by $(dF)_T = -d\omega$ in a reversible isothermal process where $d\omega$ is the total work done by the system, then during the crack opening process

$$(dF)_{T,f} = -(d\omega_{cc})_{T,f} - (d\omega_{ce})_{T,f} + f(dL)_{T,f} \tag{15}$$

Combining Equations (14) and (15) and using the fact that $(-d\omega_{cc})_T = 2\gamma(da)_T$ we find

$$-(d\omega_{ce})_{T,f} = (dG_e)_{T,f} = (dF_e)_{T,f} - f(dL)_{T,f} \tag{16}$$

Thus the crack extension force is

$$\zeta = -\left(\frac{\partial G_e}{\partial a}\right)_{T,f} = +\left(\frac{d\omega_{ce}}{da}\right)_{T,f} \tag{17}$$

The work $d\omega_{ce}$ will depend on the tractions across the surface of area da before it is opened and on the elastic properties of the body. The tractions will include contributions from the external loading system together with contributions from the stress fields of all the singularities around the crack tip. The latter will cause toughening if they decrease the crack extension force below the value it would have if the external load were the only source of the crack tip stresses.

Although this notional process was carried out with all the singularities except the crack fixed in position and state, the value of $d\omega_{ce}$ due to infinitesimal displacement of the other singularities during crack opening would only be altered by a higher order term which will disappear when taking the limit. In such a case the total decrease in elastic Gibbs free energy will be greater than $d\omega_{ce}$ but $d\omega_{ce}$ still measures that part of the total due to crack extension alone. This provides the justification for

saying that the energy changes brought about by other processes should not be counted in the determination of crack extension force.

In order to relate the local energy balance and the global energy balance the energy changes in Equation (3) have to be partitioned in a different way. We first of all consider the energy change in a system <u>without dislocations</u> but otherwise identical and undergoing an identical change in crack area. The Gibbs free energy of this system changes by $[dF'_c - f\,dL'_c]_{T,f} + 2\gamma(da)_{T,f}$. Next we consider the energy change in a system <u>without a crack</u> but containing an identical distribution of dislocations to those in the system with a crack plus dislocations and which make identical displacements to those in this system when the crack increases in area by da. When this happens the energy change will be $[dF'_d - f\,dL'_p]_{T,f}$ where the first term is the energy stored in the stress fields of the extra dislocations produced and dL'_p is the plastic elongation of the system.

When the energy changes in the crack-free and the dislocation-free systems are added, the sum is <u>not</u> the energy change in the system with both a crack and dislocations. The difference is called the interaction energy, $(dG_I)_{T,f}$. It arises because of mutual cancellation or enhancement of the stress fields of the crack and the other singularities. Putting the change in the elastic self-energy of the crack, $[dF'_c - f\,dL'_c]_{T,f}$ equal to $(dG'_c)_{T,f}$ we have for the total Gibbs free energy change, $(dG_t)_{T,f}$

$$(dG_t)_{T,f} = (dG'_c)_{T,f} + (dG_I)_{T,f} + (dF'_d)_{T,f} - (f\,dL_p)_{T,f} + 2\gamma(da)_{T,f} \qquad (18)$$

If crack extension takes place without movement or creation of dislocations the true crack extension force can be determined from the energy change. The only non-zero terms on the right-hand side of Equation (18) in this case are the first two and the last. Thus we have

$$\left(\frac{\partial G_t}{\partial a}\right)_{T,f} = \left(\frac{\partial G'_c}{\partial a}\right)_{T,f} + \left(\frac{\partial G_I}{\partial a}\right)_{T,f} + 2\gamma \qquad (19)$$

It is clear that $(dG'_c)_{T,f} + (\partial G_I)_{T,f}$ is equal to $(dG_e)_{T,f}$ as defined in Equation (16) and hence from Equation (17)

$$\zeta = \left(\frac{d\omega_{ce}}{da}\right)_{T,f} = -\left(\frac{\partial G'_c}{\partial a}\right)_{T,f} - \left(\frac{\partial G_I}{\partial a}\right)_{T,f} \qquad (20)$$

The first of the two right-hand terms represents the crack extension force arising from the external load and the second the contribution from the internal stresses of the singularities around the crack tip.

It will be recognised that what is being described here is usually referred to as crack shielding in studies of fracture toughening. As is well known, if the elastic part of the response of the material to a deforming force is linear these calculations can be very conveniently carried out in terms of the stress intensity factor, K. The crack tip shielding approach to fracture toughening is now widely used although its relation to the local energy balance is not generally recognized. It is worth noting that only the elastic part of the system's response to deforming forces is involved in such an

1008

energy balance and therefore non-linearity due to plastic dissipation and hysteretic effects is not directly relevant to toughening and only affects crack propagation through its influence on the elastic stress in the crack tip region.

The simplest assumption upon which a fracture criterion can be based is that crack propagation takes place when the local crack extension force (or equivalently, the local stress intensity factor) reaches a critical value. Fracture toughening is then a result of the reduction of the crack extension force by internal stress relaxation resulting from energy dissipating processes. Such an assumption, like any other criterion for fracture, can only be validated by comparison with experiment.

Conclusions: It is clear that an overall energy balance based on the first law of thermodynamics is empty of meaning as far as a criterion for crack instability is concerned. The second law of thermodynamics must be used to discuss the unstable equilibrium of cracked bodies. However crack instability criteria based on a second law overall energy balance have been shown to be equally without value. When such an overall energy balance is done properly it yields paradoxical results.

The energy dissipated per unit area of crack extension by processes accompanying crack growth does not appear as an effective surface energy or as a crack resistive term. This work should not be included in the determination of a crack extension force. The value of the true crack extension force must be obtained from a local energy balance and only in this way can the original Griffith theory be extended to the more general case. The local energy balance is identical to crack tip shielding calculations.

It is of considerable interest to note that a crack tip field characterisation approach to the problem of crack propagation is identical to an energy balance, not only in the purely elastic case, as is generally acknowledge, but also when energy dissipating processes accompany crack extension.

REFERENCES

(1) Griffith, A.A., Phil. Trans. Roy. Soc. (London), Vol. A221, 1920, pp.163-197

(2) Orowan, E. "Fatigue and fracture of metals" (W.M. Murray, Ed.) John Wiley, New York, 1952, pp.139-167

(3) Irwin, G.R., "Fracturing of metals", ASM Cleveland, Ohio, 1948, pp 147-166

(4) Gurney, C., & Hunt, J., Proc. Roy. Soc. (London) Vol. A229, 1967, pp.508-524

(5) Williams, J.G., "Fracture mechanics of polymers" Horwood-Wiley, New York, 1984