

Journal of Research of the National Bureau of Standards

Volume 89

Number 6

November-December

Indentation Fractography: A Measure of Brittleness <i>B. R. Lawn and D. B. Marshall</i>	435
Controlled Indentation Flaws for the Construction of Toughness and Fatigue Master Maps <i>R. F. Cook and B. R. Lawn</i>	453
The Interactions of Composition and Stress in Crystalline Solids <i>F. C. Larché and J. W. Cahn</i>	467
Published Papers of the National Bureau of Standards	501
Index for Volume 89, January-December 1984	533

ISSN 0160-1741

Library of Congress Catalog Card No.: 63-37059

The Journal of Research of the National Bureau of Standards features advances in measurement methodology and analyses consistent with the NBS responsibility as the nation's measurement science laboratory. It includes reports on instrumentation for making accurate and precise measurements in fields of physical science and engineering, as well as the mathematical models of phenomena which enable the predictive determination of information in regions where measurements may be absent. Papers on critical data, calibration techniques, quality assurance programs, and well characterized reference materials reflect NBS programs in these areas. Special issues of the Journal are devoted to invited papers in a particular field of measurement science. Survey articles appear periodically on topics related to the Bureau's technical and scientific programs. As a special service to subscribers each issue contains complete citations to all recent NBS publications in NBS and non-NBS media.

David T. Goldman, Editor

Executive Editors
Donald R. Johnson
(Natl. Measurement Lab.)
John W. Lyons
(Natl. Engineering Lab.)

Board of Editors

John W. Cooper (Physics)	Howard J. M. Hanley (Boulder Laboratory)
Sharon G. Lias (Chemistry)	John W. Cahn (Materials)
Donald G. Eitzen (Engineering)	

Issued six times a year. Annual subscriptions: domestic \$17.00; foreign \$21.25. Single copy, \$3.00 domestic; \$3.75 foreign.

United States Government Printing Office, Washington: 1984

**Order all publications from the Superintendent of Documents
U.S. Government Printing Office, Washington, DC 20402**

The Secretary of Commerce has determined that the publication of the periodical is necessary in the transaction of the public business required by law of this Department. Use of funds for printing this periodical has been approved by the Director of the Office of Management and Budget through April 1, 1985.

Indentation Fractography: A Measure of Brittleness

B. R. Lawn

National Bureau of Standards, Gaithersburg, MD 20899

and

D. B. Marshall

Rockwell International Science Center, Thousand Oaks, CA 91360

Accepted: August 30, 1984

Indentation constitutes one of the most powerful test techniques for the systematic investigation of deformation and fracture responses in brittle materials. Indentations can be used to evaluate critical mechanical parameters (toughness, hardness, elastic modulus) with great simplicity and high accuracy.

Key words: crack initiation; crack propagation; erosion and wear; fractography; glass; indentation; radial cracks; strength.

1. Introduction

Indentation constitutes one of the most powerful test techniques for the systematic investigation of deformation and fracture responses in brittle materials. Indentations can be used to evaluate critical mechanical parameters (toughness, hardness, elastic modulus) with great simplicity and high accuracy. They can be used to introduce controlled cracks into strength-test specimens, and thence to obtain physical insight into failure mechanisms. They can be taken as a base for simulating "natural" surface damage processes such as particle impact, abrasive wear and machining. In short, indentations represent a model flaw system for quantifying a wide range of mechanical properties, and as such deserve detailed study.

Recourse to some of the review articles on the subject

About the Authors, Paper: B. R. Lawn, who is with NBS' Center for Materials Science, and D. B. Marshall, are physicists. Their work was sponsored in part by the U.S. Office of Naval Research, Metallurgy and Ceramics Program, and their paper is also appearing as a chapter in the *Fractography of Glass*, R. C. Bradt and R. E. Tressler, editors, Plenum Press, New York (in press).

[1-8]¹ reveals many facets of indentation analysis. For a start, contacts may be considered either "blunt" or "sharp," according to whether the local deformation prior to fracture is elastic or elastic-plastic. The latter, if relatively complex in its stress field characterization, presents us with some of the more interesting new phenomena in brittle fracture. Second, indentation events can occur under either equilibrium or kinetic conditions of deformation and fracture. Of these, the first lends itself more readily to detailed fracture mechanics formulation, but the second takes us closer to engineering design problems associated with "fatigue" (delayed failure) behavior. Again, distinction may be made between initiation and propagation stages in the contact fracture evolution. Propagating cracks are better understood because they develop in the contact far field, where high stress gradients smooth out. The ultimate crack configuration may nevertheless depend to a large extent on exactly where in the near field the initiation occurs, which in turn raises the question of availability of suitable starting nuclei (e.g., whether such nuclei are present or have to be created by the contact process itself). We can devise many more categories for the general indentation phenomenology, e.g., in accordance

¹Numbers in brackets indicate literature references at the end of this paper.

with loading type (normal vs tangential) or loading rate (static vs dynamic), attesting to a wide diversity in underlying micromechanical processes.

Here we consider these facets in relation to the fractography of glass. We begin with surveys of blunt and sharp contact patterns, describing basic features of the respective morphologies and outlining the essential fracture mechanics procedures for quantifying these features. If we pay more attention to the sharp contact configuration, this is because of the relatively dominant place it has occupied in indentation testing over the past decade. We then consider the ways that contact-induced cracks evolve when subjected to an ensuing tensile stress, with their attendant implications on strength and flaws. Finally, we look briefly at how indentation experiments can be used to provide a base for modelling surface damage processes related to wear, machining, etc. In some instances we shall draw from studies on materials other than glass, both to add to our insight into certain fracture mechanisms and to help place the broad topic of indentation fractography into a wider perspective.

2. Blunt Indenters

If contact conditions remain entirely elastic up to the onset of fracture the indenter is deemed "blunt" [1]. The classical example is the Hertzian cone fracture produced by indentation of a flat surface with a relatively hard sphere [9]. A detailed description of the evolution of Hertzian fractures was first given by Frank and Lawn [10]. Initiation occurs from pre-existing surface flaws in the region of high tensile stress just outside the circle of contact; the ensuing crack encircles the contact and subsequently propagates downward and outward into

its fully developed (truncated) cone configuration. As alluded to earlier, the second, propagation stage is much easier to understand, and so we shall deal with it first.

2.1 Crack Propagation

Under normal loading in an isotropic material like glass the Hertzian configuration assumes near-axisymmetry [9-12]. Figure 1 shows top and section views of such a crack formed by a steel ball of radius 12.7 mm on soda-lime glass [13]. It is apparent that the configuration can be closely represented as the frustum of a cone. Once formed, the cone crack remains stable, although some further, subcritical extension can occur under sustained loading if moisture is present in the environment.

It is this stability of the fully propagating cone crack which makes for simplicity in the fracture mechanics analysis. Further increases in the indenter load over and above the critical value for "pop in" simply cause the cone to expand its circular base in a controlled manner [11]; i.e., the configuration satisfies the growth conditions for simple penny-like equilibrium cracks, for which there is a standard solution [14]

$$P/c^{3/2} = A_2 K_c \quad (1)$$

where P is the load, c is the characteristic crack size, K_c is the critical stress intensity factor for equilibrium extension (the "toughness") and A_2 is a dimensionless constant. Extensive data confirming the interrelation between P and c for glass are contained in References [1] and [14]. It will be noted that eq (1) is independent of any term relating to the initiation conditions; the propagation mechanics are determined exclusively by the properties of the far field.

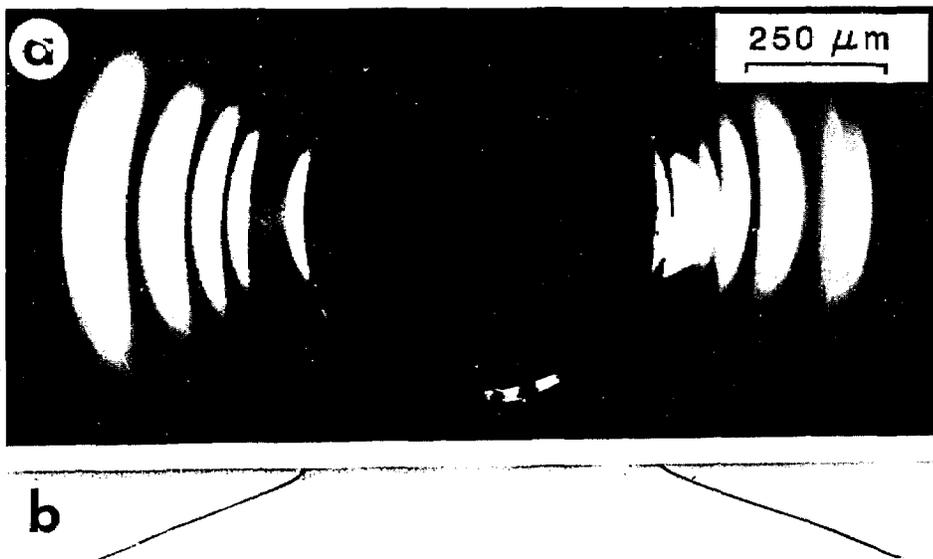


Figure 1—Hertzian cone crack in soda-lime glass: (a) view from beneath fully loaded specimen (light directed for specular reflection); (b) view in profile, after section-and-etch of unloaded specimen. After reference [13].

2.2 Crack Initiation

The precursor initiation micromechanics are more subtle because of the extremely high tensile stress gradients that exist about the contact circle [1,10,15]. It is not enough to say that initiation will occur when the surface tensile stresses reach the strength of the material; indeed, such a criterion leads to a totally incorrect relation between the critical contact load, P_c , and sphere radius, r . In terms of the modern fracture mechanics view, the crack evolves first as a shallow surface ring around the contact circle, and then extends downward, as the load increases, to a critical depth, at which point the system becomes unstable [10]. Implicit in this description is that there is a sufficient density of pre-existing surface flaws to guarantee the first stage of initiation, for the elastic contact conditions provide no means for the self-generation of suitable crack nuclei in the optimal tensile regions. (Contrast the sharp-indenter case, later.) On the other hand, the actual size of the starting flaw should not be a critical factor in the initiation condition, because of the stabilizing effect of the inhomogeneous stress field.

Detailed fracture mechanics calculations confirm the above description, and lead to the following expression for the critical load to cone initiation [1,10,15,16],

$$P_c = A_1 r K_c^2 / E \quad (2)$$

where E is Young's modulus and A_1 is another dimensionless constant. As expected, eq (2) is not dependent on the starting flaw size. This flaw independence has been verified by tests on glass surfaces with different

abrasion treatments [15]. The theory also predicts a linear relation between P_c and r , first observed empirically by Auerbach as long ago as 1891 [17]. "Auerbach's law," so called, had aroused much interest because, in combination with the Hertzian stress relation $\sigma \propto P/r^2$ [1,9], it implies a size effect in the critical stress level, $\sigma_c \propto 1/r$. (Reversing the argument, if fracture were to occur at a critical stress $\sigma_c = \text{const.}$ we would expect $P_c \propto r^2$, in clear violation of eq (2).) Note that the size effect is such that the necessary stress level increases as the sphere radius is reduced, suggesting the likelihood of some precursor "yield" as the indenter becomes "sharper."

Again, our formulation here is based on the assumption of ideal equilibrium conditions in the fracture process. When moisture is present subcritical crack growth can occur at $K < K_c$, and cone pop-in thereby enhanced, in a rate-dependent way, at loads $P < P_c$. Experimental studies of this environmental effect have been made [18,19].

2.3 Some Variants From Ideal Cone Fracture Configuration

There are several variants in the blunt-indenter crack patterns that can be produced. We consider just two examples here.

First, if the test material is not isotropic, homogeneous, or free of pre-existing residual stresses, departures can occur from the ideal cone geometry, even under axial loading. This is demonstrated in figure 2, for sphere indentations on single crystal silicon surfaces [20]. The broader feature of the surface ring crack is still apparent, but now the influence of crystallographic an-

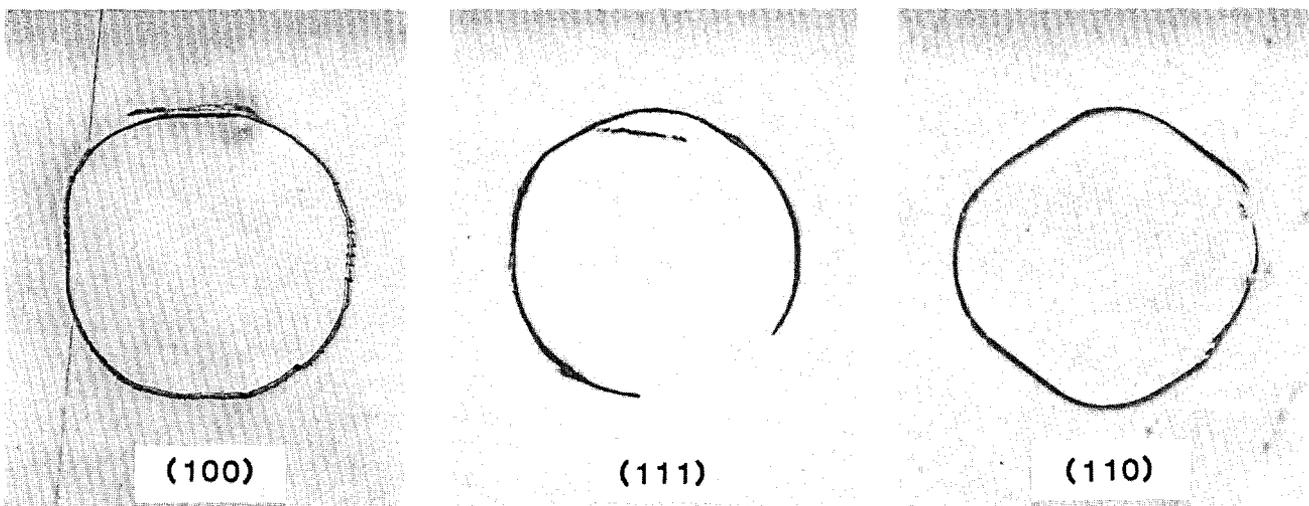


Figure 2—Hertzian fracture traces on silicon single crystals, for three surface orientations. Surfaces were etched after indentation. After reference [20].

isotropy has imposed itself on the patterns. In this case the crack path reflects the modifying effect of preferred $\{111\}$ -plane cleavage on the tendency for the fracture to follow tensile stress trajectories. Such effects may be considered secondary insofar as eqs (1) and (2) are concerned, necessitating only a re-evaluation of the constants A_1 and A_2 [20].

In the second example, we consider the changes in the pattern that ensue when the indenting sphere is made to slide across the test surface with frictional tractions. Figure 3 shows a sliding-indenter track on soda-lime glass [21]. Now not one, but several, intermittent (partial) cone cracks are generated in the wake of the advan-

where f is the coefficient of sliding friction. We may note the minor role of friction in this expression for crack size; for a fixed normal load P , a superposed sliding force at $f=1$ increases c by only some 25%. The initiation stage, however, is extremely sensitive to the friction-induced stresses. In this case the critical load equation, eq (2), is not so easily modified; solutions to the base fracture mechanics integration formulas have to be obtained numerically, and these solutions vary dramatically according to specific starting assumptions [23-25]. Suffice it to say that very small frictional forces can produce very large reductions in P_c .

3. Sharp Indenters

We alluded in section 2.2 to the increased prospect of some pre-fracture yield at the contact zone as the indenter tip becomes "sharper." In the extreme of zero tip curvature we may expect the immediate contact stress field to be determined by the plastic rather than the elastic properties of the test material [1,26]. This transition in contact response is seen in the micrographs of figure 4, for indentations on arsenic trisulphide glass using (a) a steel sphere of radius 200 μm and (b) a standard Vickers diamond pyramid. It is clear that the crack patterns are of a type totally different from that described in the previous section. In particular, the surface traces of the cracks are radially, rather than circumferentially, directed with respect to the contact area. The irreversible deformation has therefore significantly altered the nature of the tensile stress field. Moreover, it is possible to generate these same radial crack patterns on surfaces with the highest perfection, viz. pristine optical fibers [27], so the deformation process not only drives, but creates, the starting flaws. We consider these aspects of the fracture process below, following a similar course to that in our discussion of the blunt indenter case.

3.1 Crack Propagation

Let us now take a closer look at the crack pattern generated in sharp contact. If we take top *and* side views, as shown in figure 5 for a Vickers indentation on soda-lime glass, we find there are *two* operative crack systems [1]. Radial/median cracks (hereafter referred to simply as radial cracks) extend on median planes containing the load axis and an impression diagonal (where stress concentrations are highest). Lateral cracks extend from near the base of the subsurface deformation zone in a saucer-like manner, roughly parallel to the specimen surface. Under normal loading conditions both crack

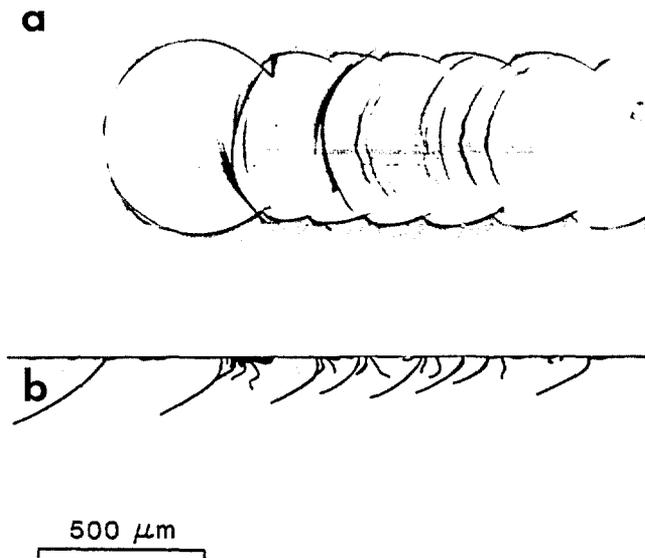
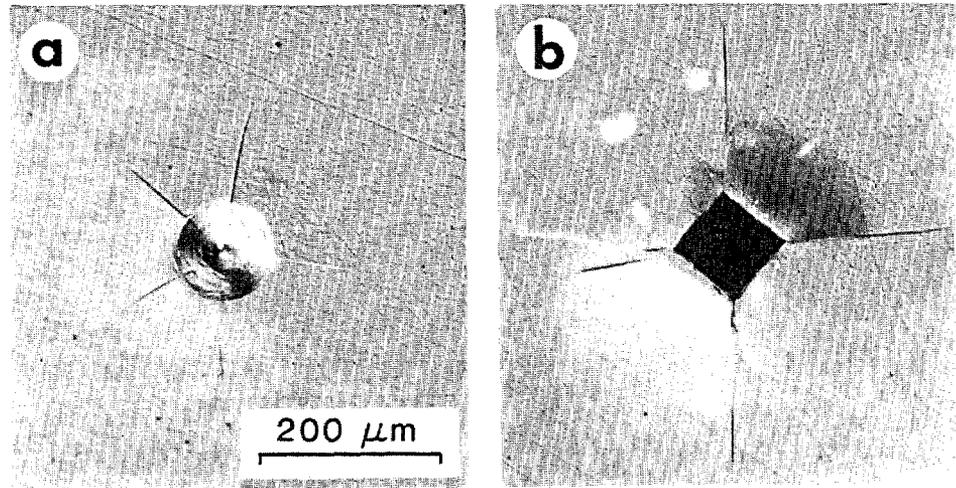


Figure 3—Partial cone crack damage track produced on soda-lime glass by sliding steel sphere (left to right), friction coefficient $f=0.1$: (a) surface view; (b) profile view, after section-and-etch. After reference [21].

cing indenter. The superposition of a tangential loading force, quite apart from altering the stress trajectory configurations, profoundly enhances the level of tensile stress (principally at the trailing edge of the contact area) [22]. These changes are not felt so strongly in the propagation stage of fracture, since the far field remains relatively insensitive to near-contact boundary conditions. Indeed, the geometry of individual cracks in the sequence may be simply regarded in terms of a simple tilting of the normal cone, where the load axis is rotated vectorially by the superposition of the friction force onto the axial load component. Using this simplistic argument it can be shown that eq (1) transforms to

$$P/c^{3/2} = A_2 K_c / (1 + f^2)^{1/2} \quad (3)$$

Figure 4—Radial crack patterns in arsenic trisulphide glass produced by (a) steel sphere (radius 200 μm) and (b) Vickers diamond pyramid.



types attain a penny-like configuration, in that their fronts are near-circular (or semi-circular, in the case of radial cracks). Thus on outward appearances we might be led to conclude that the mechanics of crack propagation in the contact far field are really no different for sharp indenters than for blunt indenters.

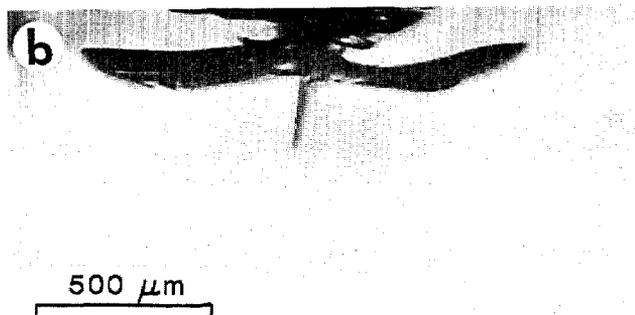
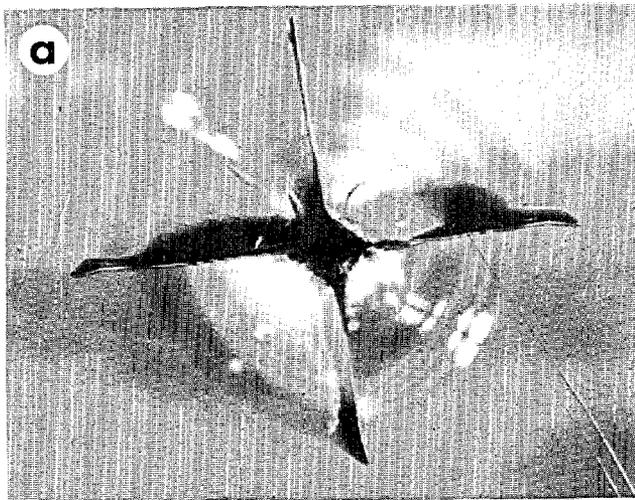


Figure 5—Vickers indentation in soda-lime glass: (a) surface view; (b) side view. Radial and lateral crack systems are evident. Courtesy B. J. Hockey.

However, if we take the trouble to follow the crack evolution *during* the actual contact cycle we discover some major differences between the two indenter types. The sequence of views in figure 6, taken in polarized light from beneath a Vickers indentation in soda-lime glass, shows some of these differences [28]. It is apparent that the radial and lateral crack systems both develop primarily during the *unloading* half-cycle, and that this development takes place more or less continuously. (As the peak indentation load is reduced this tendency to development in the later stages of the cycle is enhanced, until at low loads, toward the threshold for initiation, the entire evolution is confined to a narrow load interval immediately prior to final withdrawal [29,30].) It is also noted that the stress birefringence persists strongly in the last frame of the sequence. The sequence in figure 6 was shot in an inert (dry nitrogen gas) environment; on admitting laboratory air to the cracks at completion of the cycle the radial arms showed rapid and substantial post-contact growth, figure 7. These observations indicate that it is now the inelastic rather than the elastic component that dominates the driving forces on the cracks.

Further clues to the evolutionary process can be gained by inspecting the radial crack surface in side view, e.g., after breaking an indented specimen [28]. Fortuitous crack-front markers, presumably due to perturbations in the indentation loading system, are evident in the example shown in figure 8. By correlating these markers with the radial traces in figure 6, it can be deduced that crack extension during the loading half-cycle proceeds only in the downward direction, the surface component being almost completely suppressed. The explanation for this non-symmetrical growth is that the elastic component of the field, which is at full intensity at peak load, is essentially tensile *below* the surface but compressive *at* the surface. Unloading the in-

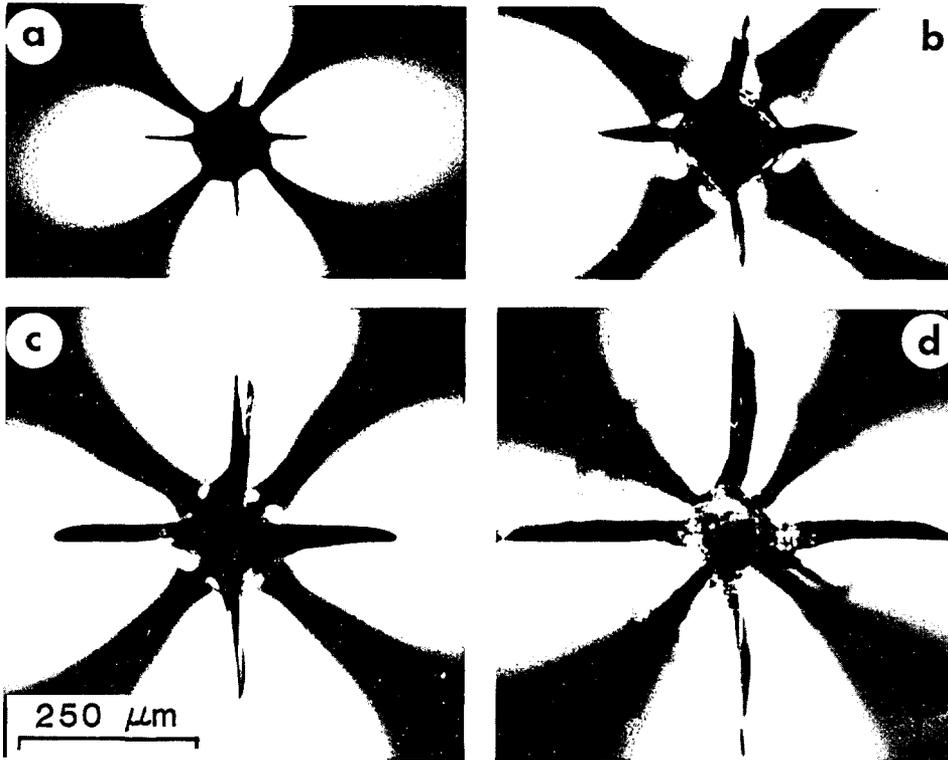


Figure 6—Vickers indentation in soda-lime glass, viewed from beneath indenter during loading cycle (polarized light). Sequence represents stages (a) half loaded, (b) full loaded, (c) half unloaded, (d) full unloaded. Lateral crack faintly visible in (d). After reference [28].

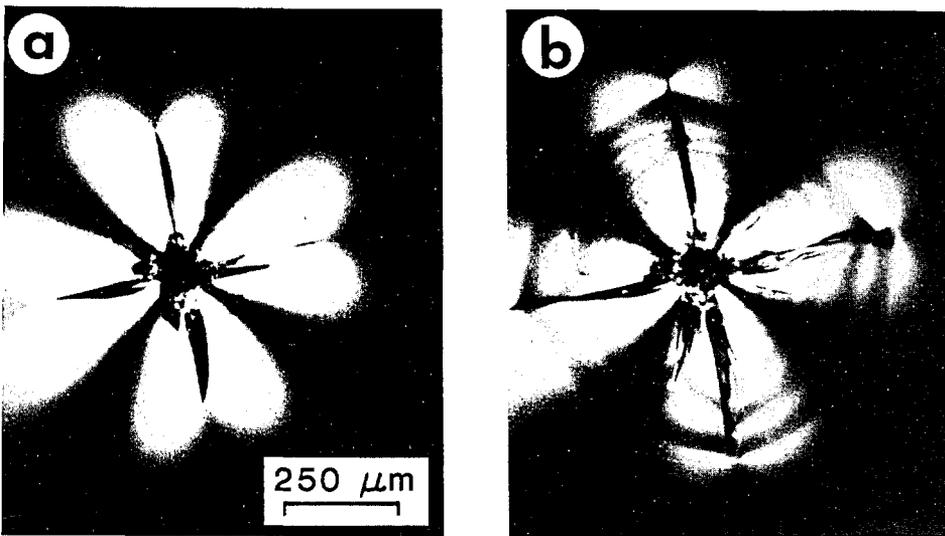


Figure 7—Same indentation pattern as in figure 6, (a) immediately after completion of contact cycle, and (b) one hour later on exposure to air.

denter is then equivalent to removing the elastic constraining stresses, leaving the crack system under the exclusive influence of the inelastic component of the field. This inelastic component acts to drive the cracks radially outward, accounting for the expansion into the ultimate symmetrical geometries on completion of the cycle.

Once this final state is achieved the cracks do indeed take on the universal propagation laws for penny-like configurations. Now, however, these laws must accom-

modate the residual-stress component as an essential element of the description. This can be done by regarding the contact deformation zone as a central loading force on the crack systems, for which a scaling analysis yields the stress intensity factor

$$K_r = \chi P / c^{3/2} \quad (4)$$

where χ is dimensionless constant. It will be noted that we have not yet specified *how* this residual driving force

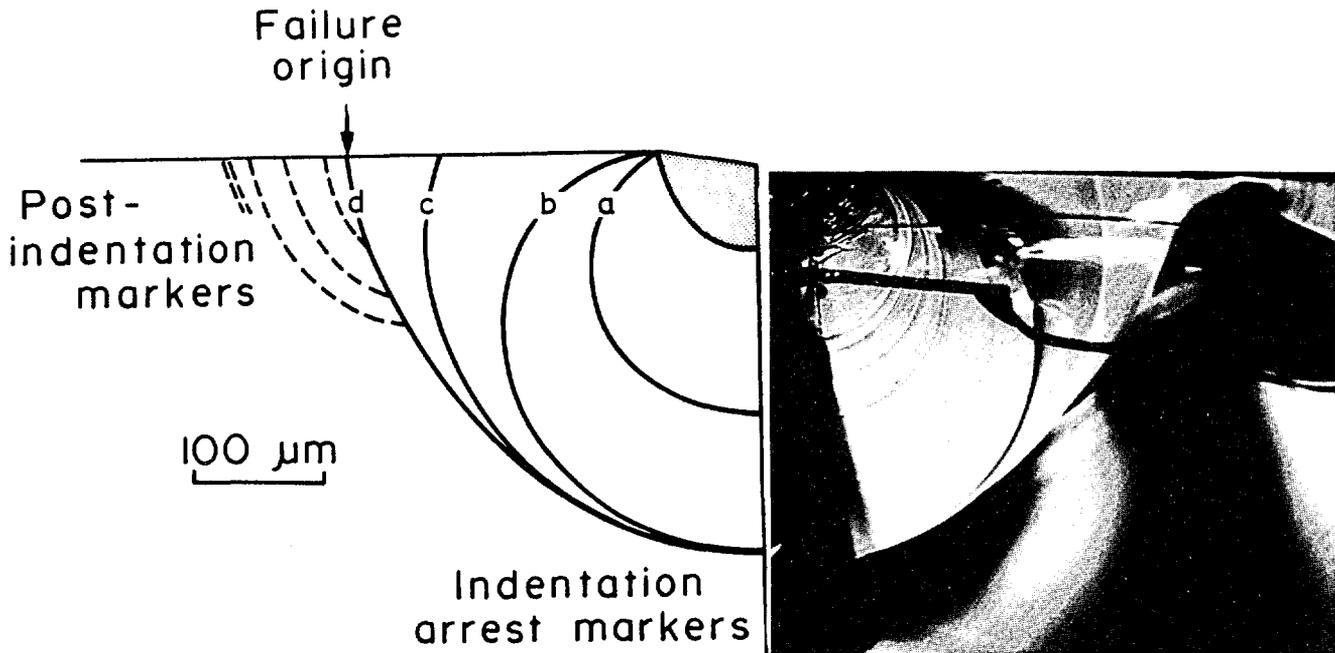


Figure 8—Same indentation as in figure 6, showing view of median crack plane after breaking specimen. Arrest positions marked correspond to stages (a)–(d) in figure 6. After reference [28].

develops; the underlying physical processes lie “buried” in the scaling term χ . Now, we have seen that the radial and lateral crack systems persist in a state of (stable) equilibrium throughout the unloading half-cycle, so we may set $K_r = K_c$ to obtain

$$P/c^{3/2} = B_2 K_c \quad (5)$$

for the immediate post-contact configurations, where we have put $B_2 = 1/\chi$ to produce a formulation in direct analogy to eq (1) for cone cracks. Data on radial cracks in soda-lime glass and several other ceramics [31] show that this relation is well satisfied over a wide range of crack sizes, even in the suspect region of small cracks where the size is not much greater than that of the hardness impression (i.e., where the central point loading assumption might be expected to break down). Data on lateral cracks are not so extensive [32,33], and some doubt remains as to how well eq (5) represents the mechanics of this system.

We have acknowledged that, in adopting a scaling argument to derive eq (4), we necessarily sacrifice all insight into the physical meaning of the χ term. To obtain some of this insight, we may usefully observe the shape of the residual hardness impressions. The vital clue here is that these impressions show evidence of strong “elastic recovery” effects. With Vickers indentations [34] this recovery is manifest as a “springback” in the contact depth, or as a “pincushion” appearance of

the surface impression. With Knoop indenters [35] it is apparent as a diminished minor axis relative to the major axis. The extent of the recovery correlates with the ratio of hardness to modulus, H/E . Figure 9 illustrates the point with examples of Knoop indentations in soda-lime glass and zinc sulphide [35], materials at the high and low end of this ratio spectrum, respectively; the impressions have the same nominal length in the two cases, but the recovery of the minor axis is clearly stronger in the glass. (In a material which shows no inelastic defor-

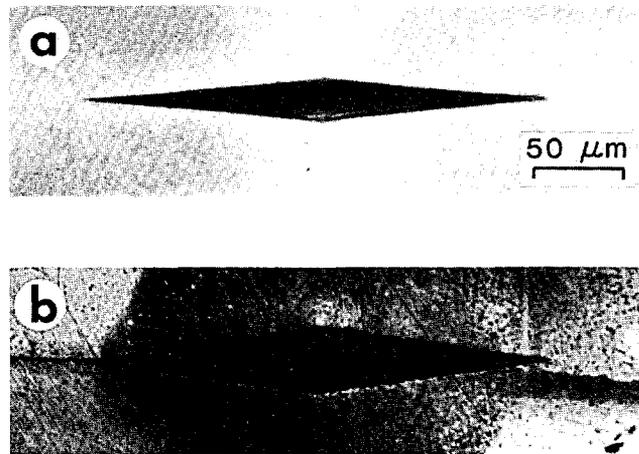


Figure 9—Knoop indentations in (a) soda-lime glass and (b) polycrystalline zinc sulphide. The glass shows stronger elastic recovery along the minor diagonal. (In unrecovered indentation, ratio of minor to major diagonal is 1:7.) After reference [35].

mation, such as rubber, this recovery would, of course, be complete, and no impression would be left at all.) The importance of these recovery phenomena in the context of fracture is that they relate to the intensity of the residual strain field about the contact area, and thence to the χ term in eq (4). In particular, they identify the important material parameter, H/E , which is expected to enter the formulation of this term. Unfortunately, quantitative theories of the contact problem incorporating this parameter remain in their infancy, owing to the general intractability of elastic-plastic stress analyses. The only available solution thus far is for isotropic, homogeneous materials which deform by shear-activated, volume-conserving radial flow, a description fitting soda-lime and other “normal” glasses [36]; this solution is of the form $\chi \propto (E/H)^{1/2} \propto 1/B_2$, [37] so the crack dimensions in eq (5) are no longer determined uniquely by the material toughness.

3.2 Crack Initiation

Below some threshold in the loading, radial and lateral cracks become suppressed. In glasses, the size of the hardness impression at which this threshold occurs is typically tens of micrometers, so it becomes difficult to observe the actual crack initiation processes by conventional optical microscopy. For this reason, we begin our discussion in this subsection with observations on another material, single crystal magnesium oxide, for which the critical events preceding fracture operate on a much larger scale.

Accordingly, figure 10 shows Vickers indentations on the (001) surface of magnesium oxide [38]; the views in (a) and (b) differ only in that the axis of the indenting pyramid has been rotated through 45°. Radial cracks are

evident as before, but note that these cracks do not extend beyond the well-defined slip-plane traces on the specimen surface. Now magnesium oxide is a relatively “soft” material (i.e., low H/E), so the elements of flow (in this case, dislocations) which accommodate the penetrating indenter extend well beyond the immediate hardness impression. The cracks we see are produced by intersection of the slip planes, in much the way as envisaged in the classical Cottrell “pile-up” model [39,40]. In other words, we are witnessing the initiation, and not the propagation, stage of fracture. We may reinforce this conclusion by noting that the cracks extend in identical directions for the two indenter orientations in figure 10; it is the slip-trace configuration, and not the alignment of the stress-concentrating contact diagonals, which dictates the fracture geometry. Moreover, the cracks lie on {110} planes, whereas the normal cleavage plane for fully propagating cracks in the rocksalt structure is {100}. Under such conditions our formulation for well-developed patterns in section 3.1 no longer has any basis and we must turn our attention to the mechanics of the precursor deformation process.

It is interesting to investigate how far, if at all, we may carry this description over to glass. We have mentioned the experimental difficulties associated with the relatively small scale of the hardness impressions in the initiation region. These difficulties are compounded by the relatively high value of H/E for glass, corresponding to a much greater confinement of the deformation elements within the immediate contact zone [37]; our observational techniques must therefore be capable of providing information on subsurface events. Some progress has nevertheless been made in this direction, notably by the examination of sectioned indentations using high resolution microscopy [30,41-43]. Examples

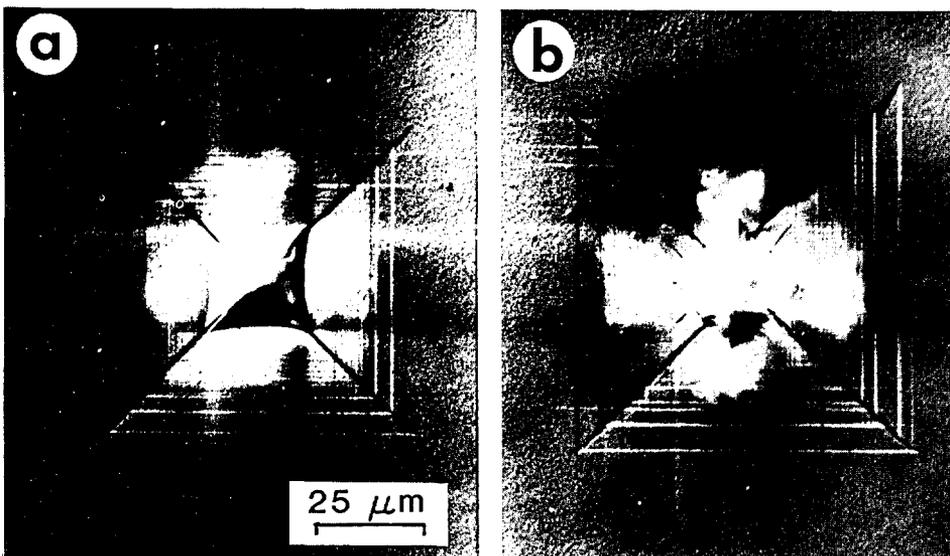


Figure 10—Vickers indentations on (001) surface of single crystal magnesium oxide. Edges of micrographs are parallel to $\langle 100 \rangle$ directions. Courtesy B. J. Hockey.

of the pertinent detail that can be revealed in this way are shown in figure 11(a) and (b) for soda-lime and arsenic trisulphide glass, respectively. There are traces of “shear faults” on the specimen surfaces, somewhat analogous to the slip lines in figure 10, but now without crystallographic features; the slip follows trajectories of maximum deviatoric stress in the near-contact field [41]. The fact that we are looking at an amorphous material here tells us that the fundamental faulting process is radically different from that generally described in terms of dislocation glide; not only does the classical concept of a dislocation Burgers vector break down in the non-crystalline structure but the operative shear stresses (as inferred from the hardness value) are close to the theoretical cohesive limit.

Whatever the true nature of the shear faulting mechanism, it is apparent that the process is severely disruptive, and therefore capable of generating centers of intense stress concentration for crack nucleation and growth. Indeed, close scrutiny of micrographs of the kind shown in figure 11 shows that one can invariably associate the initiation point of the radial and lateral cracks with one or more (usually intersecting) faults. Arguing along these lines, one may construct a simplistic, two-step model for the initiation [30]: in the first step, the stress concentration at the edge of shear fault reaches a sufficient intensity to nucleate an embryo microcrack; in step two, the microcrack grows to some critical instability point, whence pop-in ensues. For equilibrium fracture conditions the initiation relation may be determined once more by scaling principles [44],

$$P_c = B_1 K_c^4 / H^3 \quad (6)$$

with B_1 another dimensionless term. It is interesting to compare this result with its counterpart for cone cracks, eq (2); note that H replaces E as the controlling material contact parameter, reflecting the change from elastic to inelastic near-field conditions (notwithstanding the fact that B_1 , for the same reasons as discussed in relation to B_2 in section 3.1, will depend on H/E).

As with cone cracks (Section 2.2), radial and lateral cracks are susceptible to enhanced initiation at $P < P_c$ when moisture is present in the environment. Indeed, this enhancement can be substantial, with reductions of over two orders of magnitude in the threshold load when the indentation test is run in water [45]. Moreover, the initiation may now occur *after*, rather than *during*, the contact cycle (once more attesting to the vital role of the residual stress component), indicating that the effect of the hydrolytic interaction is rate dependent. The theoretical description is complex, for it can involve either of the two steps referred to above in connection with eq (6), and these steps are not easily differentiated in the kinetic data [30].

3.3 Some Variants From the Ideal Sharp Indenter Pattern

Let us look at some of the ways in which the somewhat simplistic descriptions in the previous subsection require modification when dealing with non-ideal contact systems.

Perhaps the most important of these non-ideal systems is that involving “anomalous” glasses, of which fused silica is the archetype [36]. These are glasses which deform by densification rather than by volume-conserving

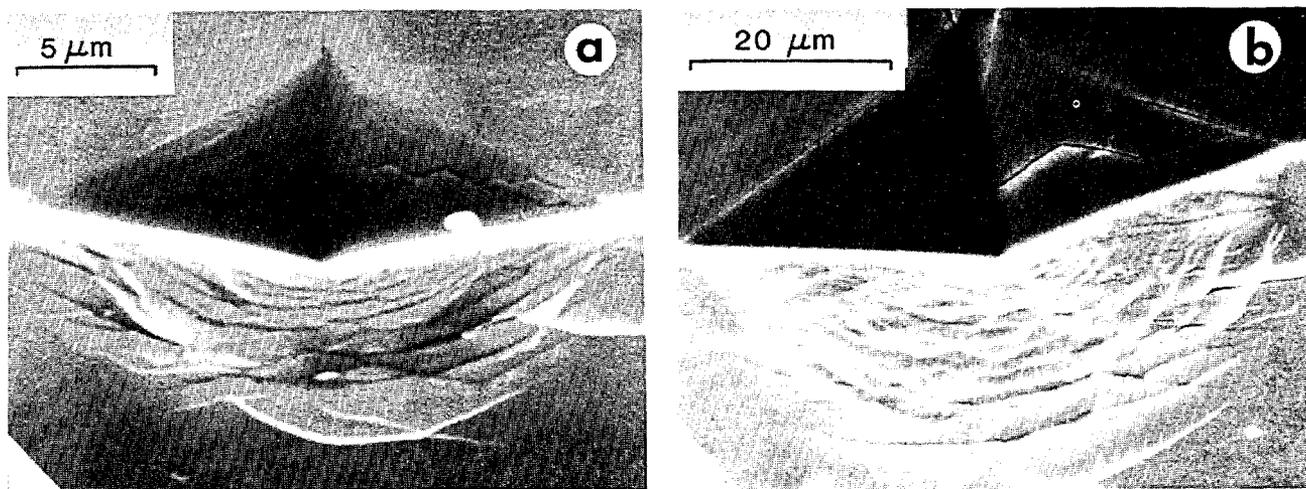


Figure 11 – Scanning electron micrographs of near-threshold Vickers indentations in (a) soda-lime and (b) arsenic trisulphide glasses. Specimens prepared by indenting across preexisting hairline crack in glass surfaces, then breaking in order to obtain half-surface and section views. After references [8] and [43].

flow. Figure 12 compares Vickers indentations in soda-lime and fused silica glasses. The patterns for the two glasses show clear differences, particularly in the near-contact region. It appears that in fused silica the intense local displacements induced by the penetrating indenter generate the same kind of shear offsets at the contact surface, but that these do not extend downward into the material as they do in soda-lime glass [45]; anomalous glasses do not contain the component of ionic modifiers thought necessary to produce “easy paths” for slip, along connected regions of terminal bonds [46]. Hence the silica structure accommodates the indentation volume by “collapsing” its structure. This densification mode is less disruptive, because it does not involve rupture, but rather realignment, of the network bonds [46]. Accordingly, the intensity of the residual contact field is much reduced in the anomalous glasses (a fact readily evident as a diminished birefringence in views of the kind shown in figure 6 [36]). In terms of eqs (5) and (6), this means one must completely re-evaluate the way the coefficients B_1 and B_2 relate to H/E , although the forms of the load/crack-size dependence remain unaffected. Another manifestation of the reduced driving force for radial and lateral cracking in anomalous glasses is the tendency for some cone cracking to occur, as in figure 12(b), suggesting that the contact field in this case may be more properly viewed as a hybrid of the blunt and sharp extremes.

As a second example, we consider how the crack patterns may be affected by inhomogeneities in the material system. This could be an important consideration in the development of new glasses with porosity or included phases. The potential for significant crack-

microstructure interactions can be most readily demonstrated by reference to figure 13, which shows Vickers indentations in alumina of three different grain sizes [47]; the sequence represents systems in which the radial cracks are (a) much larger than, (b) comparable with, and (c) much smaller than the scale of the microstructure. In cases (a) and (c) the crack patterns are well defined, as though the material were effectively homogeneous, but in (b) it is difficult to distinguish any true radial cracks at all. The implication here is that any microstructural elements which exist on a scale comparable with that of the contact itself can have a profound influence on the qualitative, as well as the quantitative, interpretation of the indentation crack patterns.

Finally, brief mention may be made of the effect of translating a sharp indenter across the test surface, effectively transforming the contact configuration from “point” to “line” loading [48]. Such a transformation changes the distribution, but not the nature, of the residual driving force on the crack. We show in figure 14 an end view, in polarized light, of a line flaw in soda-lime glass to re-emphasize that this residual driving force is by no means a minor factor. An analysis of the linear flaw geometry leads to the result [48,49]

$$P/c = B_2 K_c^2 \quad (7)$$

as the analogue of the point-flaw relation, eq (5).

4. Indentations as Controlled Flaws in Strength Analysis

Perhaps the most powerful of all applications of indentation fracture mechanics is in the analysis of

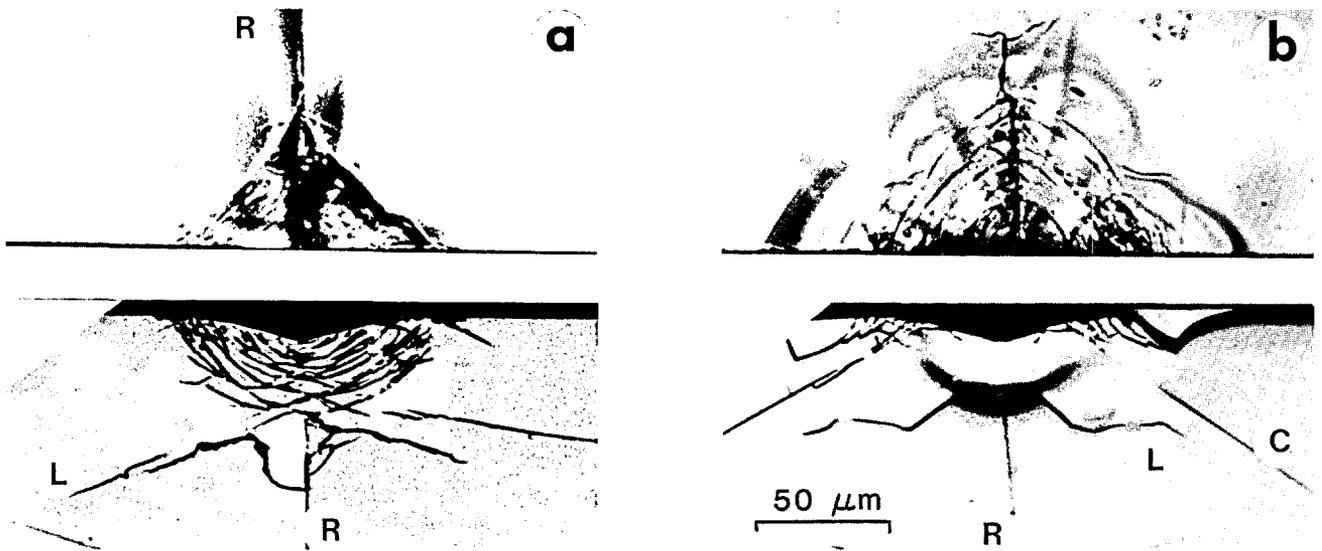


Figure 12 — Vickers indentations in (a) soda-lime and (b) fused silica glasses in half-surface (top) and section (bottom) views. (Specimens prepared in same way as those in Fig. 11.) Crack components indicated: R, radial; L, lateral; C, cone. After reference [36].

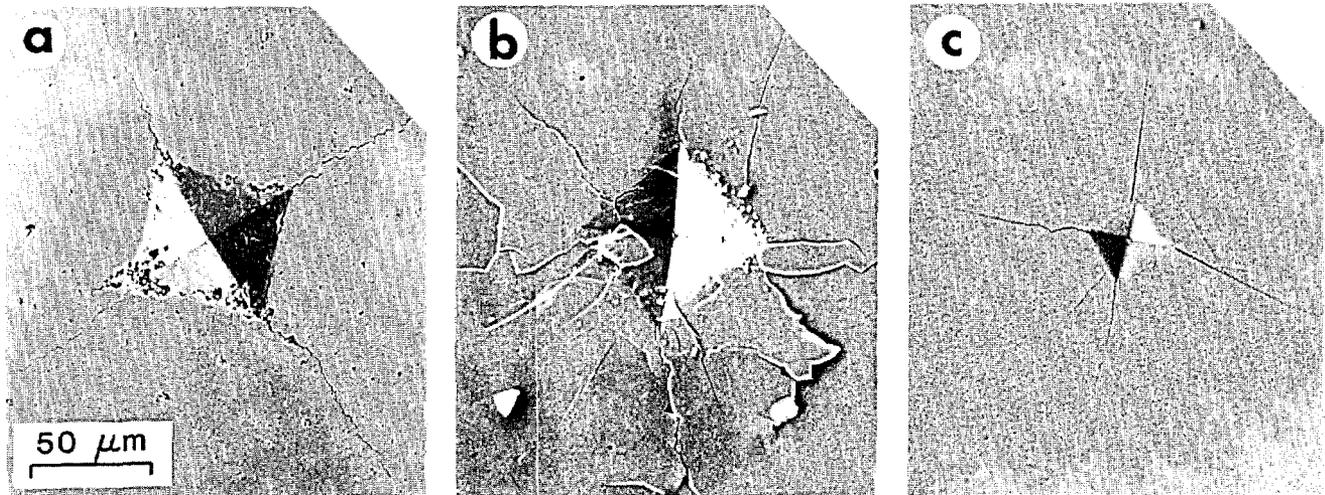


Figure 13—Scanning electron micrographs of Vickers indentations in three aluminas, (a) grain size 3 μm , (b) grain size 20 μm , and (c) single crystal (sapphire). After reference [47].

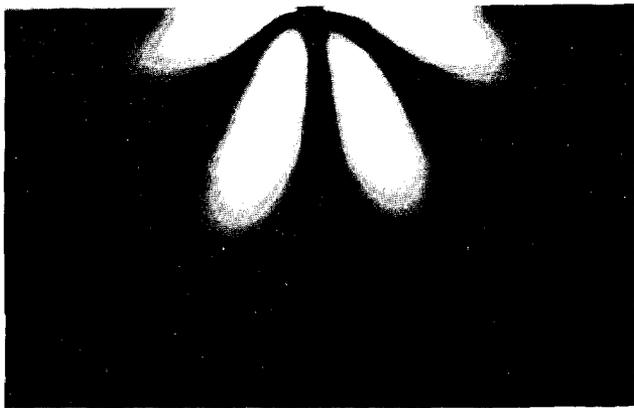


Figure 14—Line flaw in soda-lime glass, produced by tungsten carbide glass-cutting wheel. End view, in polarized light, showing residual stresses about track. After reference [48].

strength properties of brittle materials. For those concerned with materials evaluation, indentation provides a means of determining intrinsic fracture parameters with simplicity and accuracy. We shall make only brief reference to this aspect of the work in the sections below. Indentations can also be used to simulate the way in which naturally occurring surface flaws respond to applied stresses. It is here that most of our attention will be directed. We shall find that the results of controlled-flaw studies contain some surprises, particularly in relation to the time-honored Griffith concepts of strength.

4.1 Crack Morphologies in Failure

Let us begin with a survey of the fracture morphologies that result when the different indentation crack types are taken to failure in applied tensile loading. Our aim in this subsection is to make the reader

aware of some of the complications that may need to be considered when interpreting strength data. In this endeavor, we confine ourselves strictly to qualitative details.

Our first example pertains to the failure from cone-type cracks. Figure 15 shows such a failure, in this case from a sliding sphere track in soda-lime glass [21]. The test was carried out in an inert environment, so that moisture-assisted subcritical growth did not occur. Failure initiated from the base of one of the deeper cone cracks, and was spontaneous at the critical stress. In the end view of figure 15 we see large steps flanking the initiation point and curving around the cone base toward the specimen surface. The resulting intersection at the surface does not have a linear trace, but tails into a cusp which points along the original direction of sliding. It is clear that the fracture surface is far from planar, indicating the complexity in growth evolution that a curvilinear flaw will generally have to undergo as it attempts to align its plane normal to the tensile axis. In other words, we are dealing with a “mixed mode” fracture configuration, where shear components contribute to the crack driving force.

With radial cracks this issue of mixed mode loading can be avoided by taking care to orient the indenter so that principal median plane (i.e., the plane containing the long diagonal in the case of Knoop indenters, or either of the diagonals in the case of Vickers) coincides with that of the maximum tensile stress in the subsequent failure test. Now, however, a new and more far-reaching complication becomes evident. We see this in figure 16, which shows micrographs of Knoop indentations in silicon nitride *before* and *during* application of the stress (in this case, flexural) which leads to failure

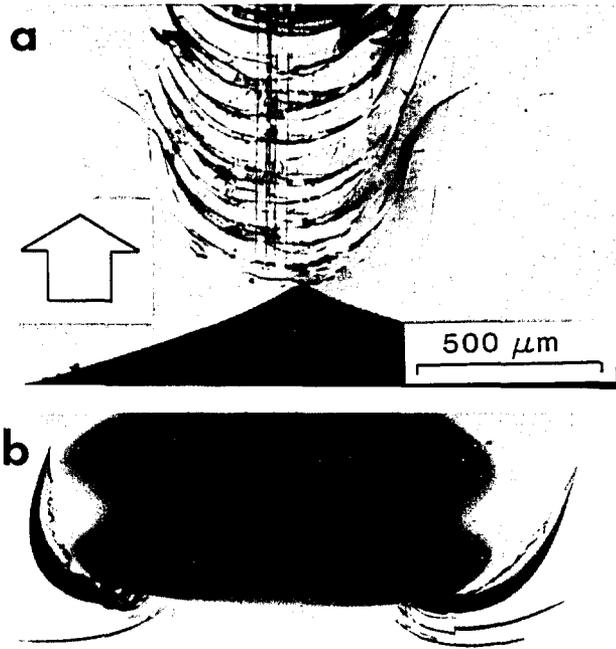


Figure 15—Failure from sliding-sphere track in soda-lime glass (friction coefficient 0.5): (a) surface view; (b) end view of fracture surface. In this case failure initiated from the first partial cone in the track. After reference [21].

[50]. The radial crack arms have clearly extended prior to failure, contrary to the conventional notion that the failure should occur spontaneously from an otherwise invariant flaw configuration. Lest it be argued that such precursor extensions could be due to moisture-assisted slow crack growth effects (and we shall indeed demonstrate below that slow crack growth *can* have a profound influence on the failure conditions), it is pointed out that the silicon nitride used in this work was chosen precisely because it is immune to such effects [50]. The same kind of response is obtained for silicate glasses tested in inert environments (vacuum, dry nitrogen) [51]. In view of our observations earlier concerning the evolution of radial cracks during the *unloading* stage of the indentation cycle (Sec. 3.1), we are led to conclude that the prior growth stage is a manifestation of the stabilizing influence of the residual contact stresses.

If the radial crack is *not* oriented normal to the tensile axis, or if the failure test is *not* conducted under equilibrium conditions, the added presence of mixed-mode and slow crack growth effects makes interpretation of the residual stress contribution somewhat less straightforward. These two complicating factors are apparent in figures 17 and 18, for Vickers indentations on soda-lime glass. In figure 17, the radial cracks were oriented to lie at 45° to the ensuing stress, which was applied to

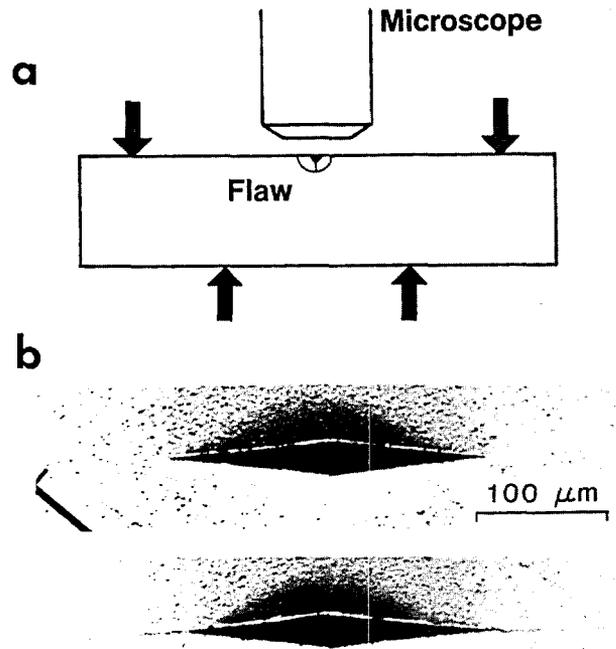


Figure 16—Growth of radial cracks from Knoop indentations in hot-pressed silicon nitride: (a) strength test arrangement; (b) crack pattern prior to and during stress application. After reference [50].

a level a little below that needed to cause failure [52]. The characteristic precursor growth is again in evidence, but note that the tendency for alignment of the cracks onto a plane of maximum tension is realized gradually during this growth stage. All theories of crack extension from inclined flaws based on the notion of spontaneous failure predict that this reorientation event should occur abruptly, producing a distinct kink in the surface crack path [52]. In figure 18, the radial cracks were oriented normally, but were taken to failure in water at a prescribed stressing rate; the two cases shown represent specimens (a) annealed before strength testing, to *remove* the residual stresses, and (b) tested in the as-indented state [53]. Although the pre-failure extension is well in excess of that observed in inert-strength tests, demonstrating that slow crack growth is no longer a minor factor, it is far more pronounced in the specimen where residual stresses remain operative.

All this is to emphasize the intractability of an exact strength formulation for materials which fail from *natural* flaws. Traditional theories of strength are based on highly simplistic descriptions of such flaws: given that the flaws do indeed have the nature of true microcracks (and even this may be open to question, particularly as strength levels approach the theoretical limit imposed

by cohesive forces, as they do in optical fibers), it is generally taken that the mode of failure is strictly tensile and, more significantly, that the sole driving force acting is that due to the applied loading system. It needs to be made clear that these descriptions are not based on any direct experimental evidence; natural flaws are

small, rarely larger than 100 μm in maximum dimension, and the location of the critical member in a large population is virtually impossible to predetermine. Fractography in such cases is restricted to post-failure analysis, which is severely limited in the information it can provide on flaw history.

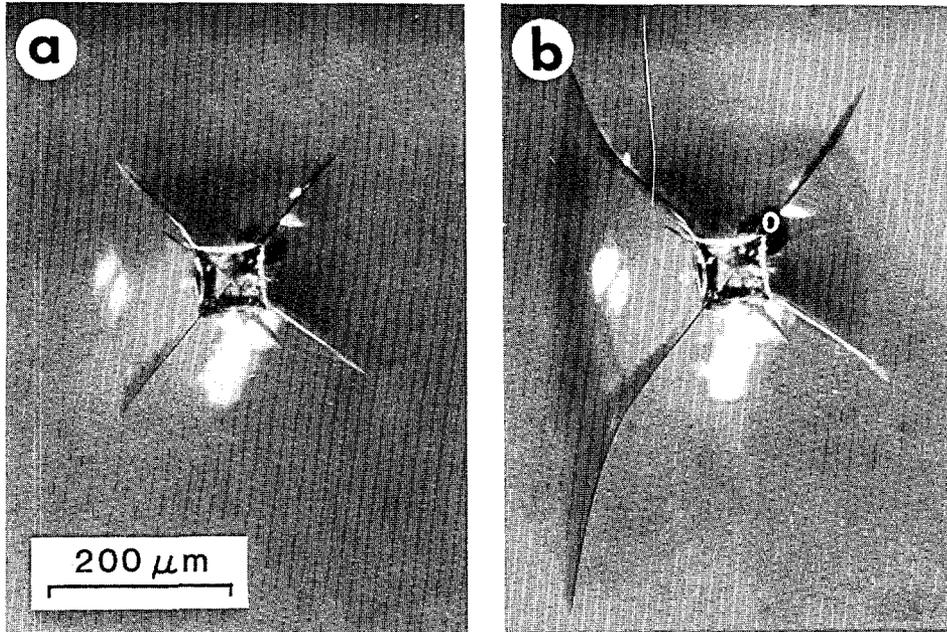


Figure 17—Growth of radial cracks from Vickers indentations in soda-lime glass during strength test: (a) prior to stressing; (b) during stressing. Indentation in this case was oriented with radial arms at 45° to prospective tensile axis (horizontally directed in this diagram) so as to produce mixed-mode loading.

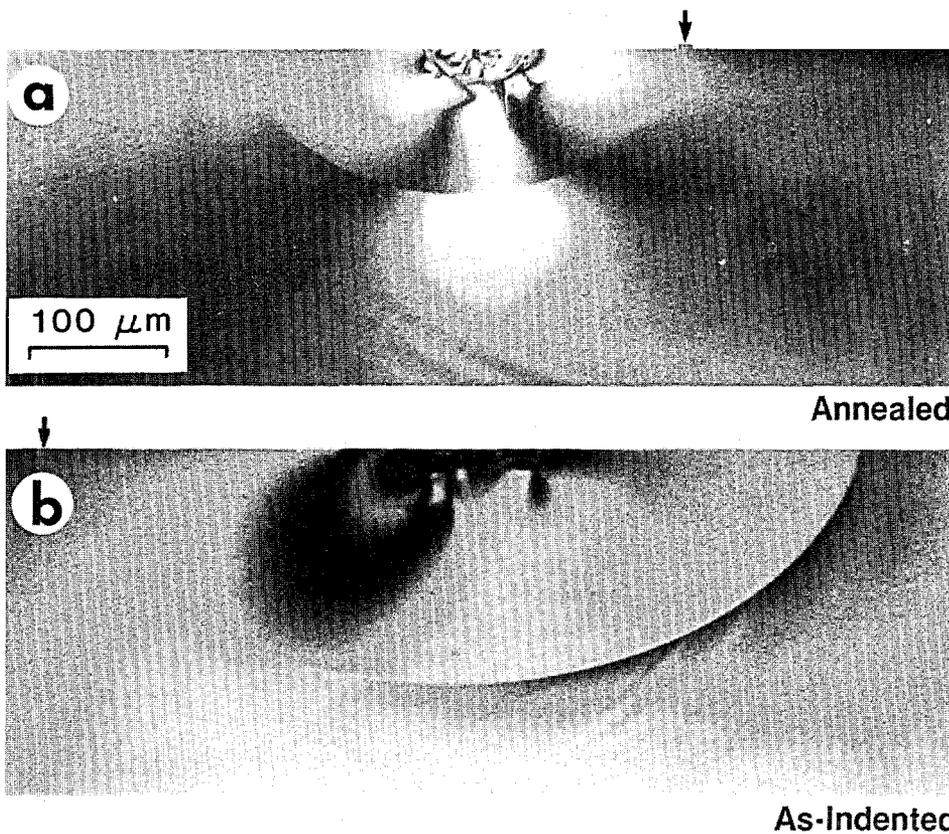


Figure 18—Fracture surfaces of Vickers-indented glass broken in water at fixed stressing rate. Contact load used to produce the indentations in the two cases illustrated was the same. Arrows designate points at which failure originated in the strength tests. After reference [53].

4.2 Strength Formulations

We have implied that the growth of flaws to failure is generally more complex than conventional strength theories would have us believe. In this subsection, we examine how the complicating factors, particularly those associated with residual contact stresses, may be incorporated into a broader fracture mechanics analysis.

The generalization may be achieved by writing the stress intensity factor for indentation cracks in the form [5,51,54]

$$K = \psi \sigma_a c^{1/2} + \chi P / c^{3/2} \quad (8)$$

where σ_a is the applied stress and ψ is a dimensionless, crack geometry term; the second term on the right of eq (8) is the residual component defined previously in eq (4). The condition for failure under equilibrium conditions is that $K = K_c$. In the traditional view the residual term is ignored, whence instability occurs at the initial crack size $c = c_0$ when $\sigma_a = \sigma_0$, say,

$$\sigma_0 = K_c / \psi c_0^{1/2}. \quad (9)$$

This result is applicable to cone cracks, or to radial cracks after annealing, with due allowance for mixed-mode effects in the ψ term [21,52]. However, for non-zero residual stresses instability does not occur until the crack has first extended from $c = c_0$ to a critical configuration $c = c_m$, at which $\sigma_a = \sigma_m$; this instability may be determined from the equilibrium form of eq (8) by putting $d\sigma_a/dc = 0$, thus

$$\sigma_m = 3K_c / 4\psi c_m^{1/2} \quad (10a)$$

$$c_m = (4\chi P / K_c)^{2/3}. \quad (10b)$$

Herein lies the formal description of the precursor stable growth stage referred to earlier in relation to virgin (non-annealed) radial crack systems, e.g., figure 16. It may be pointed out that eqs (9) and (10a) are of near-identical form (notwithstanding the factor 3/4) insofar as the relation between strength and critical flaw size is concerned [55], so the intricacies of pre-fracture flaw response may well pass unnoticed in the usual kind of fractographic observations. Some of the newer flaw detection techniques which are capable of characterizing flaw configurations *during* testing (e.g., acoustic scattering [54]) seem to suggest that the second category above is far more widespread than previously suspected, especially for surfaces with a contact-related history (machining, abrasion damage, etc.).

This influence of residual contact stresses can show up in subtle ways, as in the aging of glass. In a classical early study of this phenomenon, Mould showed that the strength of freshly abraded glass surfaces tended to increase gradually, typically by 30-40%, on prolonged exposure to water prior to breaking [56]. The prevailing view of this increase is that the crack tips are somehow "blunted" by the environmental interaction. However, if these studies are repeated under the same aging conditions, but with Vickers flaws instead, we find that strengthening occurs only while the cracks are actually extending [53,57], e.g., as in figure 7. This means, of course, that the aging process *cannot* be associated with any explanation in which crack *rounding* dominates crack *lengthening*. Now while the growth of the *radial* crack is not expected to affect the strength (since σ_m in eq (10a) depends only on the *critical* crack size c_m and not on the *initial* crack size c_0), the growth of the lateral crack can play an important role by relaxing the residual driving force on the system, i.e., diminishing χ in eq (10b). This interpretation is supported by the absence of any analogous strengthening on abraded or indented surfaces which have been annealed prior to breaking [56,57]. The results here bear on the fundamental nature of crack tips, implying that the intrinsic "sharpness" of flaws is not easily negated, even under adverse corrosion conditions.

The above fracture mechanics analysis can also be usefully adopted as a basis for design, in applications where brittle components are exposed to severe in-service contact conditions [2,4,51]. Thus for sharp, normal contacts eqs (10a) and (10b) may be combined to yield

$$\sigma_m = (3/4^{4/3} \psi \chi^{1/3}) K_c^{4/3} / P^{1/3} \quad (11)$$

where P is now to be interpreted as the maximum load that is likely to be encountered during the lifetime of the exposed surface. Clearly, if σ_m falls below the "laboratory strength" of the material, the potential exists for strength degradation. Fortunately, the dependence on P in eq (11) is weak, so the need for detailed information on prospective contact conditions is not great. From the material standpoint, toughness K_c is identified as the parameter to be optimized. It may be noted that an equation of the same form as eq (11) (differing only in the coefficient within the brackets) obtains for the counterpart blunt contact case, as can be seen by combining eq (9) with eq (1). Indeed, the formulation is readily extended to sliding contacts [21,48], or to normal contacts in impact loading [4,58].

Finally, controlled flaws can also be used to measure material parameters to high accuracy, with all the atten-

dant advantages of simplicity and economy characteristic of indentation methods. For example, by inverting eq (11) one can evaluate K_c in terms of strength and indentation load (given, of course, an appropriate "calibration" of the coefficient in this equation), without any recourse to the measurement of crack dimensions [59]. Similarly, one can combine the K relation in eq (8) with an appropriate crack velocity function to obtain material-environment fatigue parameters (e.g., exponent n in power-law velocity function) [53,60-62]. In all these analyses a proper quantitative evaluation requires a full accounting of the residual stress term. The controlled flaw methodology can be extended into the sub-threshold region, where important changes in the strength behavior become apparent [27,63-65], although the detailed micromechanical formalisms (the equivalents of eq (8) for crack *initiation*) have not yet been fully documented.

5. Erosive Wear and Machining

The cumulative effect of a large number of surface contacts with small particles can lead to significant amounts of material removal. Of the crack systems discussed in sections 2 and 3 it is the lateral which constitutes the most effective chipping mode [1] (although the cone crack can also be effective in this regard, especially when overlap between adjacent contact sites is frequent). These removal processes can be either highly deleterious or highly beneficial, depending on whether one is seeking to minimize or to maximize the removal process; that is, whether one is concerned with erosive wear [66-69] or with machining [70] properties.

Examples of surface removal damage are shown for soda-lime glass surfaces impacted with sharp silicon carbide particles in figure 19 [69]. The lateral cracking mode is clearly in evidence in the micrographs. In terms of the characteristic surface crack radius c and hardness impression diagonal a (recall that the lateral crack initiates from near the base of the deformation zone, the depth of which scales with a), we may estimate the potential chip volume for the i 'th normal contact event as

$$V_i = \pi c_i^2 a_i \quad (12)$$

This is the entry point for our indentation formalism. The fracture relation eq (5) may then be invoked to eliminate c_i , and the conventional hardness relation $P/\alpha a^2 = H$ (α a geometrical constant) likewise to eliminate a_i , from eq (12). Accordingly, we obtain a volume removal equation of the form

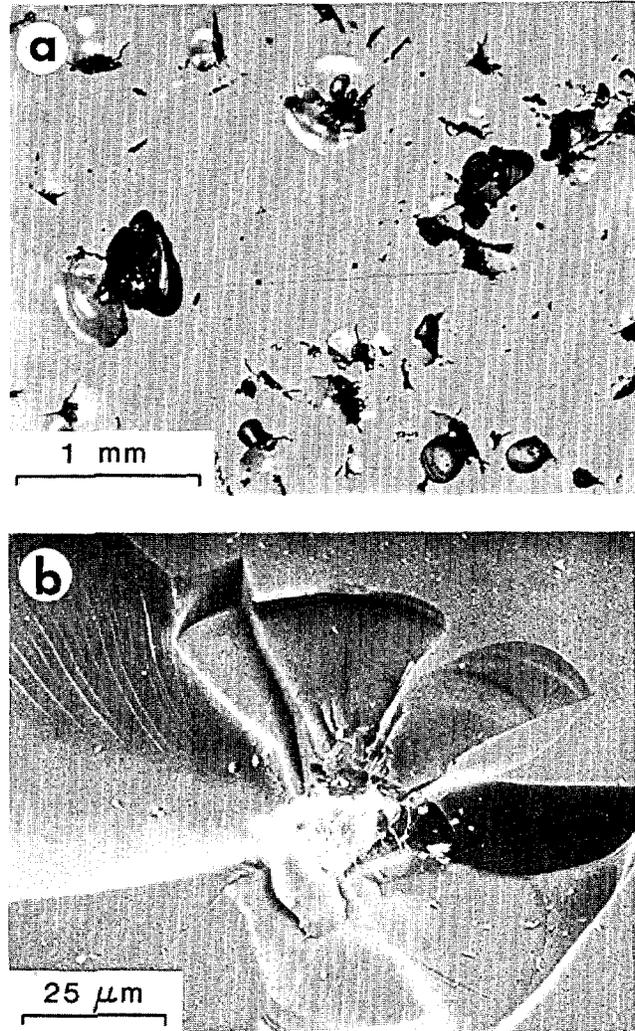


Figure 19—Erosion damage in soda-lime glass impacted with 150 μm silicon carbide particles at normally incident velocity 90 m s^{-1} : (a) optical micrograph, showing multiple impact sites; (b) scanning electron micrograph, showing single impact site. Lateral cracking is clearly the dominant surface removal mode. After reference [69].

$$V_i = (\omega/K_c^{4/3} H^{1/2}) P_i^{11/6} \quad (13)$$

where $\omega = \omega(H/E)$ (recall the dependence of indentation constants on hardness to modulus, Sect. 3.1) is a "wear coefficient." The simplest theories then proceed on the tacit assumption that all such individual contacts are of the same severity and are non-interacting, so that the total volume removed is simply $V = NV_i$, where N is the number of events. For contacts in impact loading, P_i can be eliminated in favor of incident kinetic energy U_i (via an appropriate contact equation) [58] to obtain an erosion equation. The most important predictions of this model, concerning the manner in which the removal rate increases with contact load or energy and decreases

with material toughness, have been confirmed in extensive experimental studies on erosive wear in brittle glasses and ceramics [66–69].

The procedure for constructing indentation-based models for machining wear rates is basically the same, with due allowance for a tendency to linear rather than point contact geometry. The true nature of machining damage, in terms of the essential interrelation between the near-surface deformation processes and the ensuing cracks, has only recently become clear, primarily as a result of strength studies of the kind referred to in section 4 [54,70]. Development of a *detailed* theory of machining, with proper account of such important extraneous influences as near-contact lubricants, tool geometry, etc., awaits the next generation of indentation analysts.

References

- [1] Lawn, B. R., and T. R. Wilshaw, Indentation Fracture: Principles and Applications, *J. Mater. Sci.* **10**: 1049 (1975).
- [2] Lawn, B. R., and D. B. Marshall, Indentation Fracture and Strength Degradation in Ceramics, in "Fracture Mechanics of Ceramics," R. C. Bradt, D. P. H. Hasselman and F. F. Lange, eds., Plenum Press, New York (1978), Vol. 3, p. 205.
- [3] Lawn, B. R., and D. B. Marshall, Mechanisms of Micro-Contact Damage in Brittle Solids, in "Lithic Use-Wear," B. Hayden, ed., Academic Press, New York (1979), p. 63.
- [4] Lawn, B. R.; D. B. Marshall, P. Chantikul, and G. R. Anstis, Indentation Fracture: Applications in the Assessment of Strength of Ceramics, *J. Aust. Ceram. Soc.* **16**: 4 (1980).
- [5] Lawn, B. R., The Indentation Crack as a Model Indentation Flaw, in "Fracture Mechanics of Ceramics," R. C. Bradt, A. G. Evans, D. P. H. Hasselman, and F. F. Lange, eds., Plenum Press, New York (1983), Vol. 5, p. 1
- [6] Lawn, B. R., and S. M. Wiederhorn, Contact Fracture in Brittle Materials, in "Contact Mechanics and Wear of Rail/Wheel Systems," J. Kalousek, R. V. Dukkipati, and G. M. Gladwell, eds., University of Waterloo Press, Vancouver (1983), p. 133.
- [7] Lawn, B. R.; B. J. Hockey and H. Richter, Indentation Analysis: Applications in the Strength and Wear of Brittle Materials, *J. Microscopy*, **130**:295 (1983).
- [8] Lawn, B. R., Indentation: Deformation and Fracture Processes, in "Strength of Glass," C. R. Kurkjian, ed., Plenum Press, New York, in press.
- [9] Hertz, H. H. *J. Reine Angew. Math.* **92**:156 (1881); *Verhandlungen des Vereins zur Beforderung des Gewerbe Fleisses* **61**:449 (1882). Reprinted, in English in "Hertz's Miscellaneous Papers," MacMillan, London (1896), Chs. 5,6.
- [10] Frank, F. C., and B. R. Lawn, On the Theory of Hertzian Fracture, *Proc. Roy. Soc. Lond.* **A299**:291 (1967).
- [11] Roesler, F. C., Brittle Fractures Near Equilibrium, *Proc. Phys. Soc. Lond.* **69**:981 (1956).
- [12] Culf, C. J., "Fracture of Glass Under Various Liquids and Gases," *J. Soc. Glass Technol.* **41**:157 (1957).
- [13] Williams, J. S.; M. V. Swain and B. R. Lawn, Cone Crack Closure in Brittle Solids, *Phys. Stat. Sol. (A)* **3**:951 (1970).
- [14] Lawn, B. R., and E. R. Fuller, Equilibrium Penny-Like Cracks in Indentation Fracture, *J. Mater. Sci.* **10**:2016 (1975).
- [15] Langitan, F. B., and B. R. Lawn, Hertzian Fracture Experiments on Abraded Glass Surfaces as Definitive Evidence for an Energy Balance Explanation of Auerbach's Law, *J. Appl. Phys.* **40**:4009 (1969).
- [16] Warren, R., Measurement of the Fracture Properties of Brittle Solids by Hertzian Indentation, *Acta Met.* **26**:1759 (1978).
- [17] Auerbach, F., Measurement of Hardness, *Ann. Phys. Chem.* **43**:61 (1891).
- [18] Langitan, F. B., and B. R. Lawn, Effect of a Reactive Environment on the Hertzian Strength of Brittle Solids, *J. Appl. Phys.* **41**:3357 (1970).
- [19] Mikosza, A. G., and B. R. Lawn, A Section-and-Etch Study of Hertzian Fracture Mechanics, *J. Appl. Phys.* **42**:5540 (1971).
- [20] Lawn, B. R., Hertzian Fracture in Single Crystals with the Diamond Structure, *J. Appl. Phys.* **39**:4828 (1968).
- [21] Lawn, B. R.; S. M. Wiederhorn and D. E. Roberts, Effect of Sliding Friction Forces on the Strength of Brittle Solids, *J. Mater. Sci.*, in press.
- [22] Hamilton, G. M., and L. E. Goodman, The Stress Field Created by a Circular Sliding Contact, *J. Appl. Mech.* **33**:371 (1966).
- [23] Lawn, B. R., Partial Cone Crack Formation in a Brittle Material Loaded with a Sliding Spherical Indenter, *Proc. Roy. Soc. Lond.* **A299**:307 (1967).
- [24] Chiang, S., and A. G. Evans, Influence of a Tangential Force on the Fracture of Two Contacting Elastic Bodies, *J. Amer. Ceram. Soc.* **66**:4 (1983).
- [25] Enomoto, Y., Sliding Fracture of Soda-Lime Glass in Liquid Environments, *J. Mater. Sci.* **16**:3365 (1981).
- [26] Lawn, B. R., and M. V. Swain, Microfracture Beneath Point Indentations in Brittle Solids, *J. Mater. Sci.* **10**:113 (1975).
- [27] Dabbs, T. P.; D. B. Marshall and B. R. Lawn, Flaw Generation by Indentation in Glass Fibers, *J. Amer. Ceram. Soc.* **63**:224 (1980).
- [28] Marshall, D. B., and B. R. Lawn, Residual Stress Effects in Sharp-Contact Cracking: I. Indentation Fracture Mechanics, *J. Mater. Sci.* **14**:2001 (1979).
- [29] Ishikawa, H., and N. Shinkai, Critical Load for Median Crack Indentation in Vickers Indentation of Glasses, *J. Amer. Ceram. Soc.* **65**:C-124 (1982).
- [30] Lawn, B. R.; T. P. Dabbs and C. J. Fairbanks, Kinetics of Shear-Activated Indentation Crack Initiation in Soda-Lime Glass, *J. Mater. Sci.* **18**:2785 (1983).
- [31] Marshall, D. B., and A. G. Evans, Reply to Comment on Elastic/Plastic Indentation Damage in Ceramics: The Median/Radial Crack System, *J. Amer. Ceram. Soc.* **64**:C-182 (1981).
- [32] Evans, A. G., and T. R. Wilshaw, Quasi-Static Solid Particle Damage in Brittle Solids, *Acta Met.* **24**:939 (1976).
- [33] Marshall, D. B.; B. R. Lawn and A. G. Evans, Elastic/Plastic Indentation in Ceramics: The Lateral Crack System, *J. Amer. Ceram. Soc.* **65**:561 (1982).
- [34] Lawn, B. R., and V. R. Howes, Elastic Recovery of Hardness Indentations, *J. Mater. Sci.* **16**:2475 (1981).
- [35] Marshall, D. B.; T. Noma and A. G. Evans, A Simple Method for Determining Elastic-Modulus to Hardness Ratios Using Knoop Indentation Measurements, *J. Amer. Ceram. Soc.* **65**:C-175 (1982).
- [36] Arora, A.; D. B. Marshall, B. R. Lawn, and M. V. Swain, Indentation Deformation/Fracture of Normal and Anomalous Glasses, *J. Non-Cryst. Solids* **31**:915 (1979).
- [37] Lawn, B. R.; A. G. Evans and D. B. Marshall, Elastic/Plastic Indentation Damage in Ceramics: The Median/Radial

- Crack System, *J. Amer. Ceram. Soc.* **63**:574 (1980).
- [38] Hockey, B. J., unpublished work.
- [39] Cottrell, A. H., Theory of Brittle Fracture in Steel and Similar Metals, *Trans. Met. Soc. A.I.M.E.* **212**:192 (1958).
- [40] Lawn, B. R., and T. R. Wilshaw, "Fracture of Brittle Solids," Cambridge University Press, London (1975), Ch. 2.
- [41] Hagan, J. T., and M. V. Swain, The Origin of Median and Lateral Cracks at Plastic Indents in Brittle Materials, *J. Phys. D: Appl. Phys.* **11**:2091 (1978).
- [42] Hagan, J. T., Shear Deformation Under Pyramidal Indentations in Soda-Lime Glass, *J. Mater. Sci.* **15**:1417 (1980).
- [43] Dabbs, T. P.; C. J. Fairbanks and B. R. Lawn, Subthreshold Indentation Flaws in the Study of Fatigue Properties of Ultra-High Strength Glass, in "Methods for Assessing the Structural Reliability of Brittle Materials," S. W. Freiman, ed., ASTM Special Technical Publication, in press.
- [44] Lawn, B. R., and A. G. Evans, A Model for Crack Initiation in Elastic/Plastic Indentation Fields, *J. Mater. Sci.* **12**:2195 (1977).
- [45] Multhopp, H.; B. R. Lawn and T. P. Dabbs, Deformation-Induced Crack Initiation by Indentation of Silicate Materials, in Plastic Deformation of Ceramic Materials, R. E. Tressler and R. C. Bradt, eds., Plenum Press, New York, in press.
- [46] Ernsberger, F. M., Mechanical Properties of Glass, *J. Non-Cryst. Solids* **25**:293 (1977).
- [47] Anstis, G. R.; P. Chantikul, D. B. Marshall, and B. R. Lawn, A Critical Evaluation of Indentation Techniques for Measuring Fracture Toughness: I. Direct Crack Measurements, *J. Amer. Ceram. Soc.* **64**:533 (1981).
- [48] Symonds, B. L.; R. F. Cook and B. R. Lawn, Dynamic Fatigue of Brittle Materials Containing Indentation Line Flaws, *J. Mater. Sci.* **18**:1306 (1983).
- [49] Swain, M. V., Median Crack Initiation and Propagation Beneath a Disc Glass Cutter, *Glass Technol.* **22**:222 (1981).
- [50] Marshall, D. B., Controlled Flaws in Ceramics: A Comparison of Knoop and Vickers Indentation, *J. Amer. Ceram. Soc.* **66**:127 (1983).
- [51] Marshall, D. B.; B. R. Lawn and P. Chantikul, Residual Stress Effects in Sharp-Contact Cracking: II. Strength Degradation, *J. Mater. Sci.* **14**:2225 (1979).
- [52] Marshall, D. B., Mechanisms of Failure From Surface Flaws in Mixed Mode Loading, *J. Amer. Ceram. Soc.*, in press.
- [53] Marshall, D. B., and B. R. Lawn, Flaw Characteristics in Dynamic Fatigue: The Influence of Residual Contact Stresses, *J. Amer. Ceram. Soc.* **63**:532 (1980).
- [54] Marshall, D. B., Failure From Surface Flaws, in "Methods for Assessing the Structural Reliability of Brittle Materials," S. W. Frieman, ed., ASTM Special Technical Publication, in press.
- [55] Marshall, D. B.; B. R. Lawn and J. J. Mecholsky, Effect of Residual Contact Stresses on Mirror/Flaw-Size Relations," *J. Amer. Ceram. Soc.* **63**:358 (1980).
- [56] Mould, R. E., Strength and Static Fatigue of Abraded Glass Under Controlled Ambient Conditions: III. Aging of Fresh Abrasions," *J. Amer. Ceram. Soc.* **43**:160 (1960).
- [57] Marshall, D. B., and B. R. Lawn, Surface Flaws in Glass, in "Strength of Glass," C. R. Kurkjian, ed., Plenum Press, New York, in press.
- [58] Wiederhorn, S. M., and B. R. Lawn, Strength Degradation of Glass Impacted With Sharp Particles: I. Annealed Surfaces, *J. Amer. Ceram. Soc.* **62**:66 (1979).
- [59] Chantikul, P.; G. R. Anstis, B. R. Lawn, and D. B. Marshall, A Critical Evaluation of Indentation Techniques for Measuring Fracture Toughness: II. Strength Method, *J. Amer. Ceram. Soc.* **64**:539 (1981).
- [60] Chantikul, P.; B. R. Lawn and D. B. Marshall, Micromechanics of Flaw Growth in Static Fatigue: Influence of Residual Contact Stresses, *J. Amer. Ceram. Soc.* **64**:322 (1981).
- [61] Lawn, B. R.; D. B. Marshall, G. R. Anstis, and T. P. Dabbs, Fatigue Analysis of Brittle Materials Using Indentation Flaws: I. General Theory, *J. Mater. Sci.* **16**:2846 (1981).
- [62] Cook, R. F.; B. R. Lawn and G. R. Anstis, Fatigue Analysis of Brittle Materials Using Indentation Flaws: II. Case Study on a Glass Ceramic, *J. Mater. Sci.* **17**:1108 (1982).
- [63] Dabbs, T. P.; B. R. Lawn and P. L. Kelly, A Dynamic Fatigue Study of Soda-Lime and Borosilicate Glasses Using Small-Scale Indentation Flaws, *Phys. Chem. Glasses* **23**:58 (1982).
- [64] Dabbs, T. P., and B. R. Lawn, Fatigue of High-Strength Soda-Lime Glass: A Constant Stressing Rate Study Using Subthreshold Flaws, *Phys. Chem. Glasses* **23**:93 (1982).
- [65] Dabbs, T. P., and B. R. Lawn, Acid-Enhanced Crack Initiation in Glass, *J. Amer. Ceram. Soc.* **65**:C-37 (1982).
- [66] Evans, A. G.; M. E. Gulden and M. E. Rosenblatt, Impact Damage in Brittle Materials in the Elastic-Plastic Response Regime, *Proc. Roy. Soc. Lond.* **A361**:343 (1978).
- [67] Hockey, B. J.; S. M. Wiederhorn and H. Johnson, Erosion of Brittle Materials by Solid Particle Impact, in "Fracture Mechanics of Ceramics," R. C. Bradt, D. P. H. Hasselman, and F. F. Lange, eds., Plenum Press, New York (1978), Vol. 3, p. 379.
- [68] Gulden, M. E., Correlation of Experimental Erosion Data With Elastic-Plastic Impact Models, *J. Amer. Ceram. Soc.* **64**:C-59 (1981).
- [69] Wiederhorn, S. M., and B. J. Hockey, Effect of Material Parameters on the Erosion Resistance of Brittle Materials, *J. Mater. Sci.* **18**:766 (1983).
- [70] Marshall, D. B.; A. G. Evans, B. T. Khuri-Yakub, J. W. Tien, and G. S. Kino, The Nature of Machining Damage in Brittle Materials, *Proc. Roy. Soc. Lond.* **A385**:461 (1983).

Controlled Indentation Flaws for the Construction of Toughness and Fatigue Master Maps

R. F. Cook

University of New South Wales, Kensington, N. S. W. 2033, Australia

and

B. R. Lawn

National Bureau of Standards, Gaithersburg, MD 20899

Accepted: August 30, 1984

A simple and economical procedure for accurate determinations of toughness and lifetime parameters of ceramics is described. Indentation flaws are introduced into strength test pieces, which are then taken to failure under specified stressing and environmental conditions. By controlling the size of the critical flaw, via the contact load, material characteristics can be represented universally on "master maps" without the need for statistical considerations.

This paper surveys both the theoretical background and the experimental methodology associated with the scheme. The theory is developed for "point" flaws for dynamic and static fatigue, incorporating load explicitly into the analysis. A vital element of the fracture mechanics is the role played by residual contact stresses in driving the cracks to failure. Experimental data on a range of Vickers-indented glasses and ceramics are included to illustrate the power of the method as a means of graphic materials evaluation. It is demonstrated that basic fracture mechanics parameters can be measured directly from the slopes, intercepts and plateaus on the master maps, and that these parameters are consistent, within experimental error, with macroscopic crack growth laws.

Key words: ceramics; fatigue; indentation flaw; lifetime prediction; master maps; materials evaluation; strength; toughness; universal curves.

Introduction

The increasing use of glasses and ceramics as structural materials has prompted the development of new and accurate techniques for evaluating intrinsic fracture

About the Authors, Paper: A graduate student in the Department of Applied Physics in Australia's University of New South Wales, R. F. Cook is currently working under B. R. Lawn, a physicist in the NBS Center for Materials Science. Their paper will also appear in a forthcoming ASTM Special Technical Publication, *Methods for Assessing the Structural Reliability of Brittle Materials*, S. W. Freiman, editor. Funding for the work was provided by the Australian Research Grants Committee and the U.S. Office of Naval Research (Metallurgy and Ceramics Program).

parameters. Chief among these parameters are the fracture toughness, K_{Ic} , and the crack velocity exponent, n , which respectively characterize the equilibrium and kinetic crack growth responses. In the context of brittle design it is essential to achieve an adequate level of precision in such parameter evaluations. This is particularly so in the consideration of component integrity under sustained stresses and chemical environments, where apparently minor uncertainties can translate into order-of-magnitude discrepancies in lifetime predictions.

A standard method of determining basic fracture parameters for design is to measure the strengths of representative test specimens in flexure. However, for specimens with typically as-received or as-prepared surfaces, these strengths depend not only on the intrinsic material properties but on the flaw distributions as well. It is then not possible to investigate these two elements of the

problem in *any truly independent way*. Evaluation of material parameters becomes a mere exercise in statistical data manipulation, with little or no physical insight into the nature of the critical flaws responsible for failure [1-2].¹ This probabilistic approach makes it difficult to assess the relative merits of different materials from the standpoint of intrinsic properties alone.

A controlled-flaw technique which effectively eliminates the statistical component from strength testing has been described in a series of recent articles [3-12]. A single dominant flaw of predetermined size and geometry is introduced into the prospective tensile surface of each specimen using a standard diamond indenter. The specimens are then stressed to failure in the usual way. With the indentation and flexure testing conditions held fixed, any variations in the strength behavior can be taken as direct reflections of the intrinsic material response. The only need for statistical treatments then resides in the trivial accountability of random scatter in the data. Quite apart from the ensuing improvements in data reproducibility, the indentation procedure confers several advantages in strength analysis: (i) greater specimen economy, (ii) because the location of the critical flaw is predetermined, closer observation of the fracture mechanics to failure, and (iii) a reasonable simulation of the damage processes that are responsible for a great many brittle failures [13-15]. One apparent complication attending the technique is the existence of a strong residual contact field about the elastic/plastic deformation zone, necessitating the incorporation of additional terms in the governing stress intensity factor. However, closed-form solutions of the fracture mechanics formulations are now available for both equilibrium [4] and kinetic [16] conditions of failure; analytical determinations of toughness and fatigue parameters from the strength data may accordingly be made in as straightforward a manner as for "Griffith" flaws without the residual stress term.

The capacity to control the scale of the critical flaw via the indentation load is a potent tool in the investigation of material fracture properties. The load actually replaces initial crack size as a variable in the fracture equations, thereby eliminating the need for onerous measurements of crack dimensions (although some observations of crack growth are useful for confirming the validity of the theory) [15]. Size effects in the micro-mechanics may then be studied systematically: important changes in the nature of low-load contact flaws have been thus revealed on reducing the crack size to

the scale of the deformation zone [17] or of the microstructure [18]. Systematic variations in the load dependence of indentation-strength characteristics can also be used to evaluate pre-existing stress states in brittle materials, e.g., in tempered glass [19]. Again, some materials may produce ill-defined indentation patterns outside certain ranges of flaw size, or be restricted in specimen dimensions, in which case the geometrical requirements of standard strength-testing procedures may make it impossible to operate at a single contact load. The theoretical analysis allows one to compensate for any such changes in the working contact conditions, effectively reducing all data to an "equivalent" load.

This paper illustrates a procedure for representing the intrinsic strength properties of brittle materials on an indentation "master map." A suitable "normalization" scheme incorporating indentation load into the plotting coordinates allows for the reduction of all inert and fatigue strength data on to "universal" curves for the various test materials. In this sense the scheme is reminiscent of that developed earlier by Mould and Southwick [20], except that their use of relatively ill-defined abrasion flaws necessitated a totally empirical approach in the data reduction. On our master map the position of a given curve may be taken as a graphic indicator of the intrinsic toughness and fatigue susceptibility. Quantitative determinations may accordingly be made of K_c and n without recourse to statistically based theories of strength.

Background Theory

Stress Intensity Factor for Indentation Cracks

The starting point in the analysis is the stress intensity factor for an indentation crack of characteristic dimension c produced at peak contact load P and subjected to subsequent applied tensile stress σ_a . For "point" flaws produced by axially loaded indenters the general form of this stress intensity factor is [4]

$$K = \chi P/c^{3/2} + \psi \sigma_a c^{1/2} \quad (1)$$

where χ and ψ are dimensionless parameters. The second term in eq (1) is the familiar contribution from the applied field; ψ depends only on crack geometry, here assumed to be essentially "penny-like" [21]. The first term is the contribution from the residual contact field; for materials which deform irreversibly by a constant volume process

$$\chi = \xi(E/H)^{1/2} \quad (2)$$

¹Figures in brackets indicate literature references at the end of this paper.

approximately [22], where E is Young's modulus, H is hardness and ξ is a numerical constant.

In the event of any pre-existent stress acting on the crack a third term would have to be included in eq (1) [4,9]. Other than to note that this potential complication needs to be heeded when preparing the surfaces of test specimens we shall consider it no further in our mathematical derivations.

Equilibrium Solutions: Inert Strengths

Equilibrium conditions of crack growth are closely realized experimentally by testing in an inert environment. In terms of fracture mechanics notation the criterion for equilibrium is that $K = K_c$; if $dK/dc < 0$, the equilibrium is stable, if $dK/dc > 0$, it is unstable. Now it is evident from eq (1) that K for given values of P and σ_a passes through a minimum in its functional dependence on c ; thus at subcritical configurations $K(\min) < K_c$ there is a stable and an unstable equilibrium, to the left and to the right of the minimum, respectively [16]. In an inert strength test, σ_a is increased steadily until these two equilibria merge at $dK/dc = 0$, which defines the critical variables

$$\sigma_m = 3K_c/4\psi c_m^{1/2} \quad (3a)$$

$$c_m = (4\chi P/K_c)^{2/3} \quad (3b)$$

at which crack growth proceeds without limit. We may note that any relaxation of the residual stress field, as reflected in a reduction in χ (or, more specifically, in ξ in eq (2)), will cause σ_m to expand and c_m thence to contract.

It can be shown that the ideal indentation crack is in a state of equilibrium immediately after completion of the contact cycle [22]. The size of this crack is found by setting $\sigma_a = 0$, $K = K_c$ in eq (1);

$$c_0 = (\chi P/K_c)^{2/3}. \quad (4)$$

From eq (3b) we have $c_0 \approx 0.40c_m$. On subsequently applying the tensile stress, the crack extends stably from c_0 to c_m , whence spontaneous failure ensues at $\sigma_a = \sigma_m$ [4]. In reality, deviations from this ideal behavior are observed; relaxation effects can cause c_m to contract, as already mentioned, and subcritical, moisture-assisted crack extension within the residual contact field can cause c_0 to expand, to c'_0 say. Nevertheless, unless the condition $c'_0 < c_m$ is violated, some precursor crack growth will still precede failure, in which case σ_m remains a measure of the inert strength.

Equation (3) may then be conveniently rearranged to

eliminate all terms in crack size, and then combined with eq (2) to yield

$$\sigma_m P^{1/3} = (3/4\psi)(1/4\xi)^{1/3} [(H/E)^{1/8} K_c]^{4/3}. \quad (5)$$

This expression conveniently relates the test variables on the left side to the material properties, primarily the toughness, on the right side. We emphasize once more that this formulation is contingent on the absence of all spurious pre-present stresses.

Kinetic Solutions: Dynamic Fatigue

When cracks are exposed to moisture or other inter-active environmental species, extension can occur in the subcritical region, $K < K_c$. The major characteristic of this kind of extension is its rate dependence, which is in turn highly sensitive to the crack driving force. The basic equation of kinetic fracture accordingly takes the form of a crack velocity $v(K)$. In the interest of obtaining closed-form solutions to the ensuing fracture mechanics relations we choose the empirical power-law function [23]

$$v = v_0(K/K_c)^n \quad (6)$$

where v_0 and n are material/environment parameters. Materials with lower values of n are said to be more "susceptible" to kinetic crack growth effects.

The most practical loading arrangement for the systematic study of rate effects in strength properties is that of "dynamic fatigue," in which the time differential of stress is held fixed up to the point of failure, i.e., $\dot{\sigma}_a = \sigma_a/t = \text{const}$. We may thus combine eqs (1) and (6) to obtain a differential equation for this stressing configuration,

$$dc/dt = v_0 [\chi P/K_c c^{3/2} + \psi \dot{\sigma}_a c^{1/2} t/K_c]^n. \quad (7)$$

This equation has to be solved at given P and $\dot{\sigma}_a$ for the time to take the crack from its initial configuration, $K = K(c'_0)$, to its final configuration, $K = K_c$, at which point the stress level defines the dynamic fatigue strength, $\sigma_a = \sigma_f$ [16];

$$\sigma_f = (\lambda' \dot{\sigma}_a)^{1/(n'+1)} \quad (8)$$

where

$$n' = 3n/4 + 1/2 \quad (9a)$$

$$\lambda' = (2\pi n')^{1/2} \sigma_m^{n'} c_m / v_0. \quad (9b)$$

The solution in eq (8) is identical in form to that for "Griffith" flaws ($\chi=0$) [23]. However, the slopes and intercepts from a linear plot of $\log \sigma_f$ against $\log \dot{\sigma}_a$ are very different in the two instances. In the present case ($\chi \neq 0$), n' and λ' may be regarded as "apparent" fatigue parameters, in the sense that transformation equations are required to convert these to "true" crack velocity exponent and coefficient terms. Thus, eq (9a) may be inverted to obtain n directly from n' , and eq (9b) similarly (in conjunction with measured values of σ_m and c_m) to obtain v_0 from λ' . It is again seen that initial crack size does not enter the results, as long as the condition $c'_0 \leq c_m$ remains operative [9].

Implicit in the derivation of eq (8) is the usual assumption that the prospective test surfaces are free of spurious stresses. The introduction of such stresses leads to nonlinearities in the dynamic fatigue plotting scheme, thereby destroying the basis for the above analysis [9,10].

It is convenient at this point to incorporate the indentation load as a working test variable into the dynamic fatigue relations. Whereas n' in eq (9a) is independent of all test variables, λ' in eq (9b) can be expressed as an explicit function of P via the quantities σ_m and c_m in eq (3). In this way we may write

$$\lambda' = \lambda'_p / P^{(n'-2)/3} \quad (10)$$

where λ'_p is a *modified* intercept term, totally independent of P , given by

$$\lambda'_p = (2\pi n')^{1/2} (3K_c/4\psi)^{n'} (K_c/4\chi)^{(n'-2)/3} / v_0 \quad (11)$$

Equation (10) tells us that fatigue data obtained on one material but using different indentation loads will fall on different straight lines, mutually translated but without change of slope. Now by inserting eq (10) into eq (8) we may appropriately modify the dynamic fatigue relation, thus

$$\sigma_f P^{1/3} = (\lambda'_p \dot{\sigma}_a P)^{1/(n'+1)}, \quad (12)$$

such that by plotting $\log (\sigma_f P^{1/3})$ against $\log (\dot{\sigma}_a P)$ all data should fall on to a universal fatigue curve. This plot would, of course, cut off at a limiting level on the ordinate corresponding to the inert strength plateau defined in eq (5). The procedure for evaluating crack velocity parameters from the slopes and intercepts of such representations is the same as before, but with eq (10) serving as an intermediary to eq (9).

Kinetic Solutions: Static Fatigue

Of more practical interest from a design standpoint is the issue of component lifetime under fixed stress rather than stress *rate*. Ideally, it would seem desirable to formulate a universal static fatigue relation in direct analogy to eq (12) retaining, as far as possible, the same adjustable parameters. Lifetime predictions could then be made from dynamic fatigue data alone, without having to resort to delayed failure experiments. This formulation may be achieved in two steps. First, eliminate stressing rate in favor of time to failure, $\dot{\sigma}_a = \sigma_f / t_f$. This step introduces the lifetime concept without yet altering the status of eq (12) as a dynamic fatigue relation. Then, convert to equivalent static fatigue variables by replacing σ_f with σ_A , i.e., the level of the invariant applied stress, and t_f with $(n' + 1)t_f$ [16]. The resulting static fatigue relation is

$$t_f / P^{2/3} = \lambda'_p / (n' + 1) (\sigma_A P^{1/3})^{n'} \quad (13)$$

We reiterate here, at the risk of laboring the point, that the variables P , σ_A and t_f in eq (13) relate to prospective static fatigue conditions, whereas the parameters n' and λ'_p are adjustables, as defined by eqs (9) and (10), to be determined from dynamic fatigue data.

Experimental

Materials Selection and Preparation

The materials in this study were chosen in accordance with two major criteria: first, they should cover a range of toughness and crack velocity characteristics, as determined by independent fracture techniques; second, they should be of some technical importance. Table 1 lists these materials and their pertinent properties.

All specimens were prepared in the usual manner for strength testing. However, particular attention was paid to surface preparation, bearing in mind our repeated assertion that pre-existing stress states can greatly influence the interpretation of strength data. The glass specimens were therefore annealed [19] and the ceramics surface-polished to a mirror finish with diamond paste [10] to ensure removal of any such stresses.

Indentation and Strength Testing Procedure

All specimens were routinely indented centrally along their length using a Vickers diamond pyramid indenter to produce dominant flaws for the subsequent

Table 1. Materials used in this study.

Materials (sources footnoted)	Independent parameters				Indentation parameters		
	E GPa	H GPa	K_c MPa m ^{1/2}	n	K_c MPa m ^{1/2}	n	log v_0 ms ⁻¹
Soda-Lime Glass ^a	70	6.6	0.74*	16–19*	0.97	18	–1.6
Borosilicate Glass ^b	89	6.5	0.77*	31–37*	1.2	36	1.6
Fused Silica ^c	72	7.6	0.81*	38*	1.2	44	2.2
Synroc ^d	190	10.3	1.9	—	1.8	35	0.2
P.Z.T. ^e	88	3.1	0.87	—	1.0	43	–0.5
Alumina ^f	400	16	4.4	46*	3.8	59	1.7
Silicon Carbide ^g	435	24	4.1*	118*	3.7	222	8.4
Glass Ceramic ^h	108	8.4	2.5*	63,* 84*	2.2	117	5.0

*Determinations by other workers. (See references, below).

^aSchott-Ruhrglas GMBH [11,31,32].

^bSchott-Ruhrglas GMBH [11,31,32].

^cSchott-Ruhrglas GMBH [31,33].

^dSynroc B, Australian Atomic Energy Research Establishment [34].

^eLead Zircon Titanate, Plessey Australia.

^fF99, Friedrichsfeld GMBH [35].

^gNC203, Norton Co. [7,36].

^hPyroceram C9606, Corning Glass Co. [7,10,37,38].

failure tests. The Vickers geometry was chosen both for its proven capacity to produce well-defined radial crack patterns and for its general availability in hardness testing facilities. The glasses were indented at several loads, ranging from 0.05 to 100 N, whereas the ceramics were each indented at single loads, 10, 20, or 100 N. In all cases the radial cracks extended well beyond the central hardness impression, but never to a length in excess of one tenth the specimen thickness.

The indented specimens were then broken in four-point flexure [24] in a universal testing machine at constant crosshead speed. Care was taken to center the indentation on the tension side, with one set of radial cracks aligned normal to the long axis. The breaking loads were recorded using conventional strain gage and piezoelectric load cells [10], and the corresponding rupture stresses thence evaluated from simple beam theory. Inert strengths, σ_m , were measured in dry nitrogen or argon or silicone oil environments, with the crosshead running at its maximum speed. Dynamic fatigue strengths, σ_f , were measured in distilled water over the allowable range of crosshead speeds. At least six specimens were broken in each strength evaluation, from which means and standard deviations were computed.

Measurement of Critical Crack Dimensions

For the purpose of confirming the necessary condition that the initial crack size c'_0 should never exceed the instability value c_m for equilibrium failure, and for verifying certain aspects of the fatigue solutions presented earlier, an optical examination of representative critical

indentations is recommended. The technique used here was to place three indentations instead of one on a given test surface, and then take the specimen to failure under inert conditions [10]. On the understanding that all three indentations must have had nearly identical growth histories, the procedure leaves two “dummies” in the broken test piece from which to measure the required crack dimensions. The Vickers geometry proves particularly useful in this technique, for while the set of radial cracks perpendicular to the tensile direction provides a measure of c_m , the set parallel to this same direction remains free of external stress and hence provides a measure of c'_0 .

In all materials studied in this work some precursor crack growth was indeed found to occur prior to failure.

Results

Inert Strengths and Toughness

In this section we begin by examining the dependence of inert strength on indentation load for the three glasses studied. With this dependence established, we then investigate how the inert strength data may be reduced to a composite toughness parameter for all of the test materials.

Figure 1 accordingly shows σ_m as a function of P for the glasses. The straight lines are best fits of slope $-1/3$ in logarithmic coordinates, as per eq (5). This same dependence has been confirmed elsewhere for several other brittle materials [7,18,25,26].

Values of the composite parameter $\sigma_m P^{1/3}$ are thus

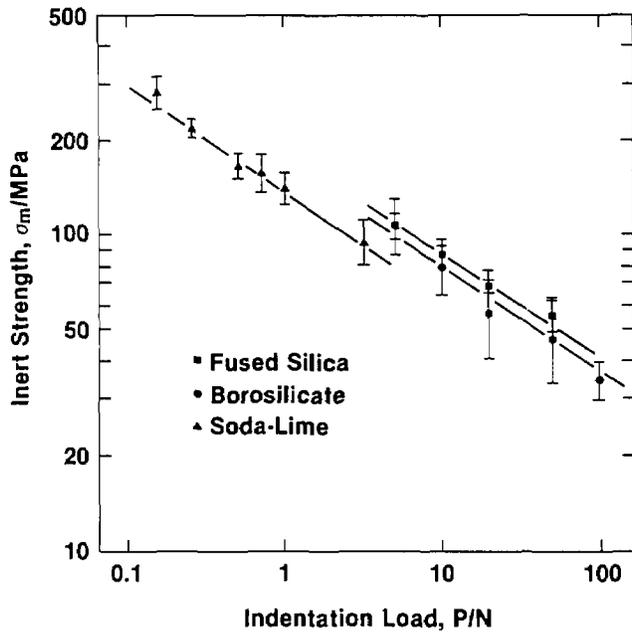


Figure 1—Inert strength as function of indentation load for the silicate glasses. (Data courtesy of T. P. Dabbs).

evaluated for each of the glasses and ceramics, and are plotted as a function of $(H/E)^{1/8}K_c$ (from table 1) in figure 2. The straight line is a fit of logarithmic slope 4/3 in accordance with eq (5), using a “calibration” value $(3/4\psi)(1/4\xi)^{1/3} = 2.02$ from an earlier, more comprehensive study [7]. The trends in figure 2 appear to be in reasonable accord with prediction, although some deviations are evident, particularly for the fused silica and borosilicate glasses. Estimates of the “indentation

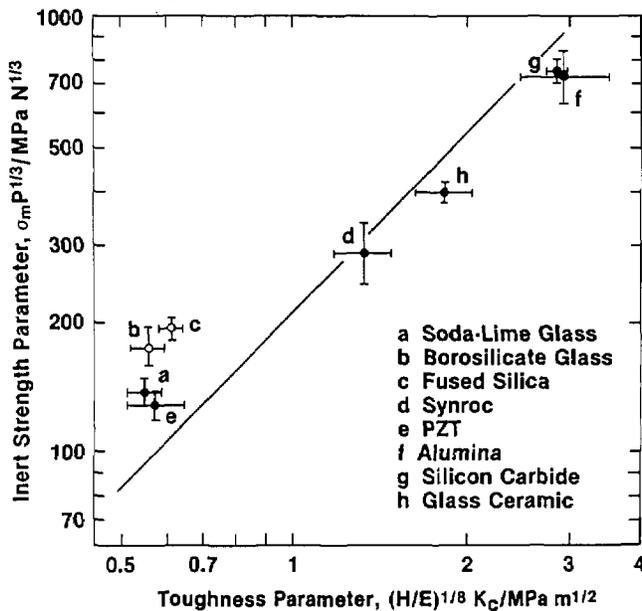


Figure 2—Inert strength parameter $\sigma_m P^{1/3}$ as function of toughness parameter $(H/E)^{1/8}K_c$ for the glasses and ceramics.

toughness” obtained directly from $\sigma_m P^{1/3}$ by inverting eq (5) are included in table 1 for comparison with the independently determined values.

Dynamic Fatigue and Crack Velocity Parameters

We consider now the dynamic fatigue responses, again beginning with the glasses to examine the functional influence of contact load, and outline the procedure for determining the exponent and coefficient in the crack velocity function.

Figure 3 shows these responses for the glass compositions in water. The straight lines drawn through individual sets of data at fixed P are best fits to eq (8), regressed for each glass on all the data consistent with the intercept relation eq (10). Thus we obtain families of lines of constant slope, with systematic displacements to lower strength levels with increasing load. Analogous plots are shown in figure 4 for the five ceramics in the same water environment, but now for a single load in each case. The inert strength limits are included in all plots as a reference baseline for assessing the degrees of fatigue.

From the regressed slopes and intercepts we obtain values of the apparent fatigue parameters n' and λ' in eq (8). Inversion of eq (9) (together with the inert strength data) then allows us to evaluate the true crack velocity parameters, n and v_0 . These evaluations are summarized in table 1; comparisons may be made in this tabulation with independent measurements of the crack velocity exponent.

Master Maps

We have set the base for determining universal fracture curves for the materials studied, and thence to construct master maps. We do this for dynamic and static fatigue conditions in turn.

The presentation of the dynamic fatigue results on a single master map requires conversion of all data to appropriate load-adjusted variables $\sigma_i P^{1/3}$ and $\dot{\sigma}_a P$ in eq (12). Figure 5, an appropriate composite of all data thus converted from figures 3 and 4 (but with error bars omitted for clarity), is such a map. Each material is now conveniently represented by a universal curve, independent of the contact loads used to obtain the data. The curves plotted in this diagram represent numerical solutions of the basic fatigue differential equation, eq (7), obtained for the ranges of P and $\dot{\sigma}_a$ covered experimentally for each material, using the inert and kinetic parameters already determined along with the measured

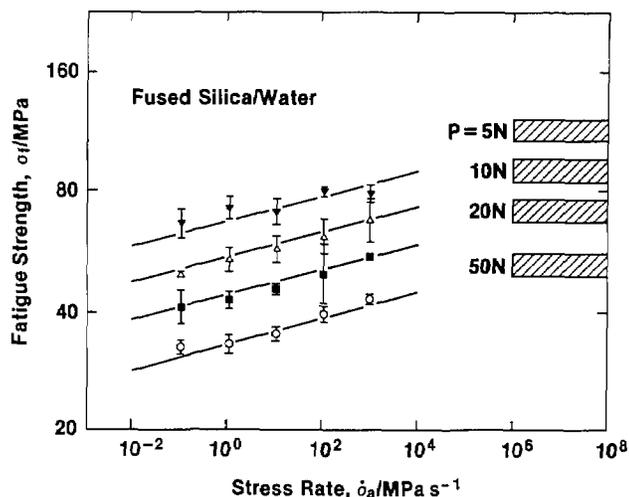
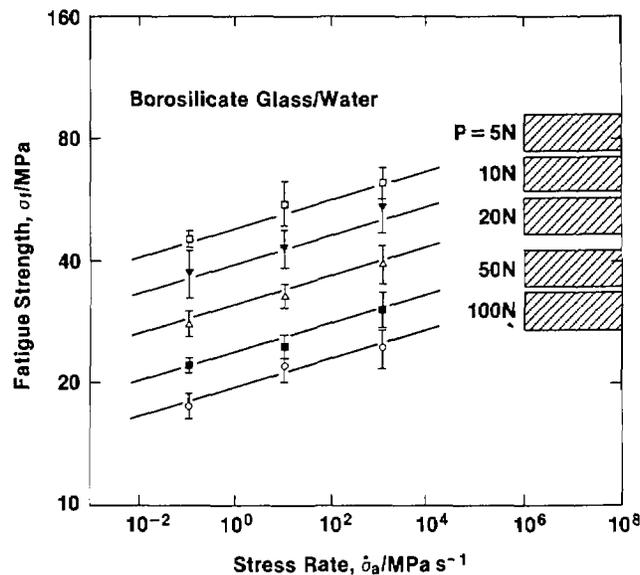
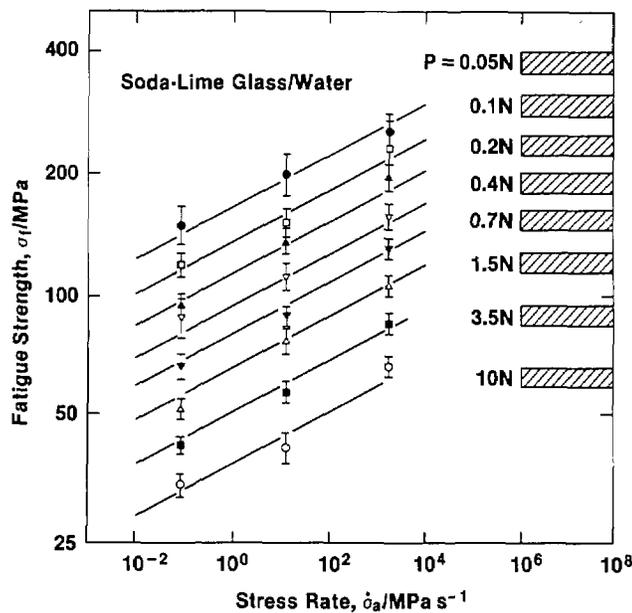


Figure 3—Dynamic fatigue responses of glasses indented at different loads. The hatched bands indicate inert strength levels. (Data courtesy of T. P. Dabbs).

initial crack sizes [10]. The fact that the curves regenerated in this way are effectively coincident with the data is, of course, no real surprise, since the regression analyses used in the parameter evaluations were performed in accordance with the solutions of the differential equation in the first place. An exercise of this kind nevertheless serves two useful purposes: (i) to confirm that the solutions referred to, which are of closed form, are indeed reasonably reliable, and (ii) to show how closely the curves remain linear in the fatigue region, and then plateau out at the inert strength levels, $\sigma_m P^{1/3}$ (fig. 2).

The equivalent construction for static fatigue is obtained from the constant stressing rate results using the rationale described earlier in the derivation of eq (13). Thus we generate the plots shown in figure 6 directly from the best-fit values of n' and λ' (or more strictly, via

eq (10), λ'_p) determined by the data regressions shown in figures 3 and 4. Cutoff levels on the abscissa again correspond to inert strength limits. Because the construction in figure 6 is not obtained this time from regenerated solutions of the basic differential equation, we are unable to plot the curved transition between the fatigue and inert regions; however, the abruptness of the corresponding crossover points in figure 5 suggests that we may reasonably ignore any such curvature in the life-time maps.

Discussion

Quantitative Evaluation of Fracture Parameters

The scheme presented here for reducing fatigue data to universal curves for any specified material/

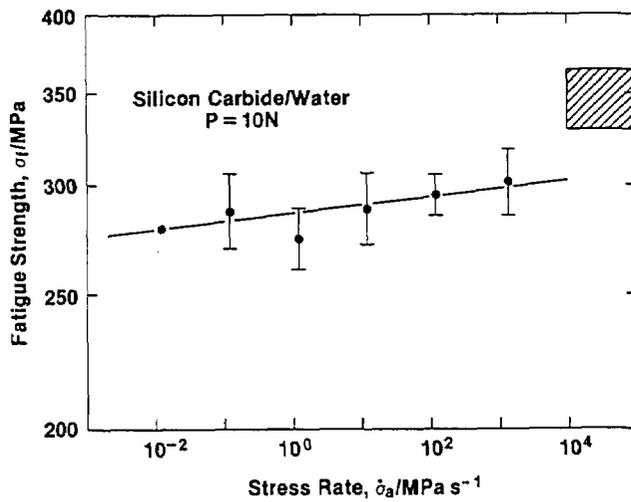
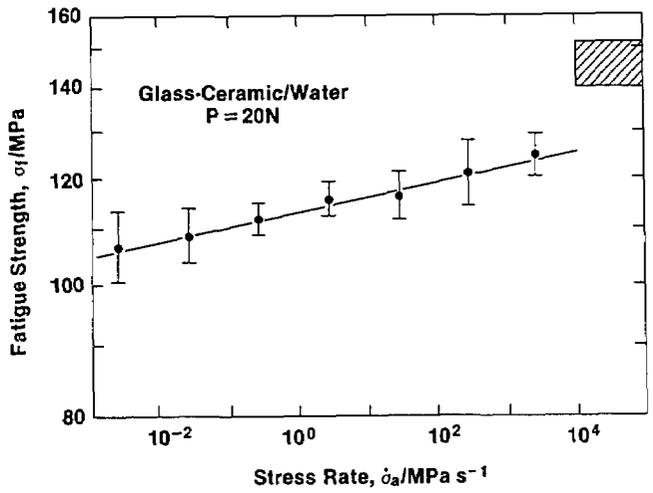
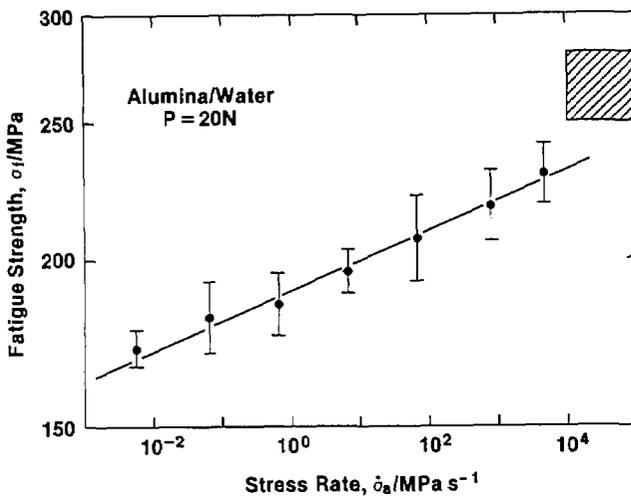
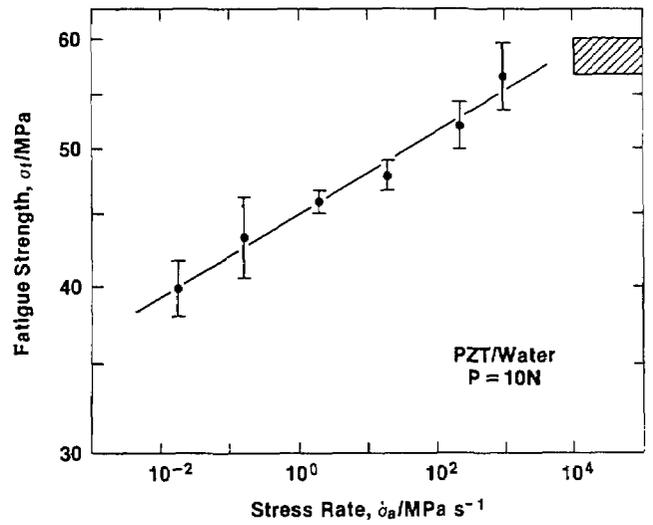
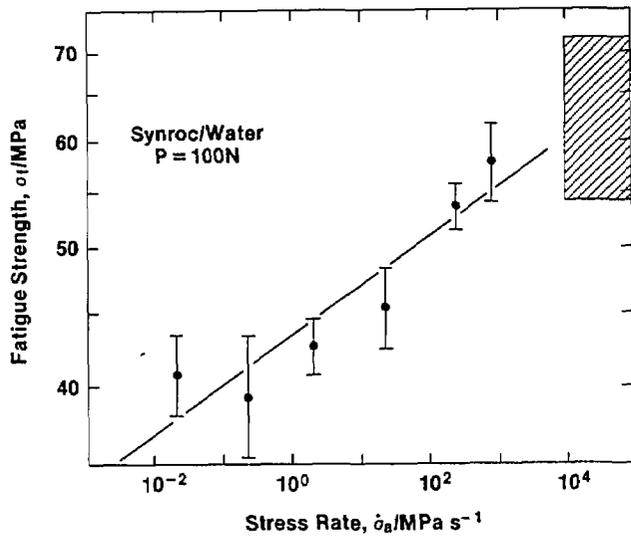


Figure 4—Dynamic fatigue responses of ceramics indented at single loads. Hatched bands indicate inert strength levels.

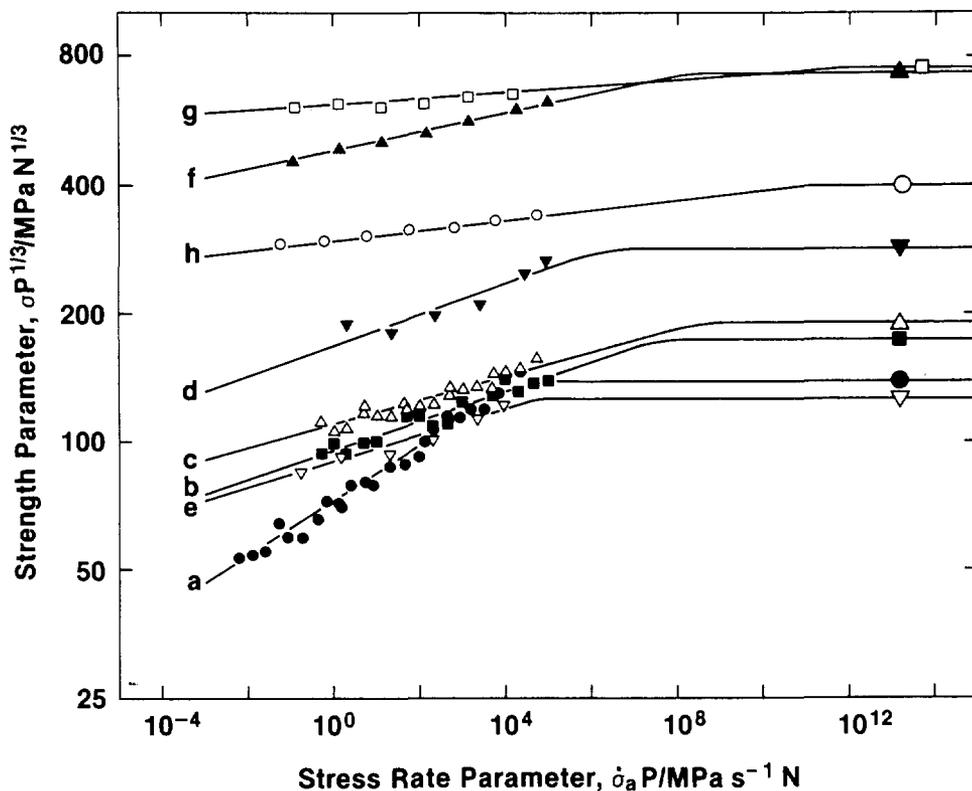


Figure 5—Dynamic fatigue master map. Error bars omitted from data points for clarity. See table 1 for key to materials.

environment system, and thence for constructing master maps to facilitate comparisons between these curves, provides an attractive route to simple, accurate, and economical evaluation of fracture parameters for design. In the following subsections we discuss how these constructions may be used as a quantitative tool for parameter determinations in different regions of the curves.

Inert Strength Levels. The position of the inert strength cutoff level, $\sigma_m P^{1/3}$, may be taken as an indicator of material toughness, K_c . Intrinsically tougher materials will therefore exhibit cutoffs further toward the top of a dynamic fatigue map (fig. 5) and toward the right of a static fatigue map (Fig. 6).

It should be emphasized that the correspondence implied here is not exact. To clarify this point we may invert eq (5) to obtain an explicit expression for toughness,

$$K_c = (256\psi^3\xi/27)^{1/4} (E/H)^{1/8} (\sigma_m P^{1/3})^{3/4}. \quad (14)$$

Thus, K_c depends on the elastic/plastic term E/H as well as on $\sigma_m P^{1/3}$. On the other hand, since E/H varies only between 10 and 25 over the range of materials listed in table 1, the use of an invariant, representative mean value $\langle (E/H)^{1/8} \rangle = 1.50$ in eq (5) would lead to errors of no more than 10%. Another potential source of dis-

crepancy lies in the implicit assumption that geometrical similarity is preserved in the indentation pattern from material to material, as reflected in the constancy of the parameters ξ and ψ . We have already pointed out that relaxation effects in the residual contact field can lead to reductions in the ξ term. Systematically low values of ξ will also be manifest in materials which deform by other than a constant-volume process or exhibit plastic pile up at the impression edges [22]. Fused silica and borosilicate glass, which tend to deform by densification [27], fall into this category, thereby explaining the tendency for the data points representing these two materials to lie above the general trend in figure 2. Finally, it has been taken as given that the radial crack patterns are always well defined, and in the materials used here this generally has been found to be so. But in materials where the microstructure is comparable in scale with the indentation event, the symmetry of the crack pattern can become severely disrupted [8,28], with consequent variations in both ξ and ψ .

It may be argued that the "effective" toughness reckoned from the cutoff position on a master map, while perhaps not an accurate measure of its macroscopically determined counterpart, nevertheless may more closely characterize the response of "natural" flaws. This is certainly likely to be so where the strength-controlling flaw in a component is created by a surface contact

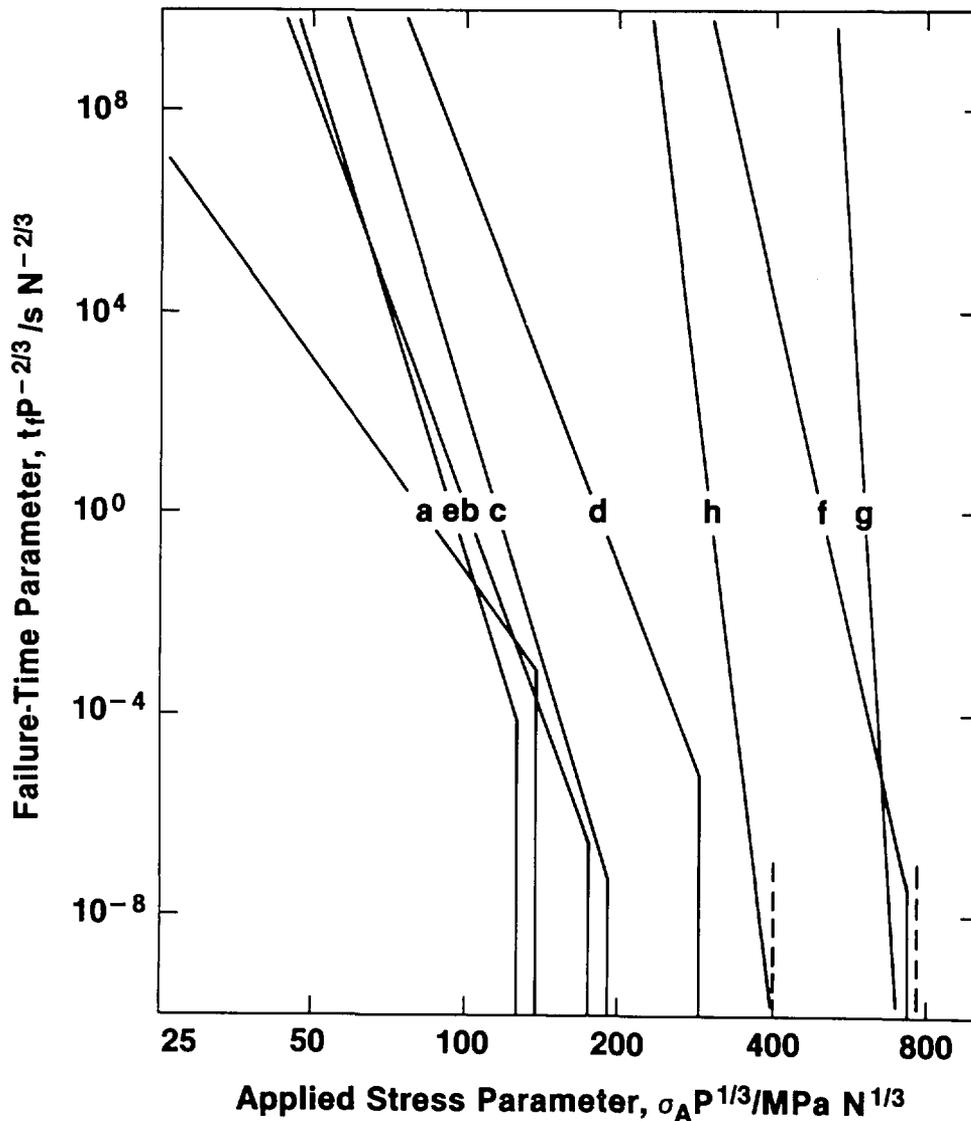


Figure 6—Static fatigue master map, generated from figure 5. See table 1 for key to materials.

event, as in sharp-particle impact or in a machining operation. In this sense the master map approach might well be expected to serve more appropriately as a source of design parameters than the more conventional methods involving large-scale fracture specimens.

Fatigue Curve Slopes. We have noted from eqs (12) and (13) that the slope of a universal fatigue curve is a measure of the intrinsic susceptibility to slow crack growth. Thus materials with lower values of the crack velocity exponent n , and hence of n' , eq (9a), will have greater slopes on dynamic fatigue master maps and, conversely, lower (negative) slopes on static fatigue maps.

As with the toughness, certain caution needs to be exercised when using master map data to determine n values. This is because in applying the inverted form of eq (9a),

$$n = 4n'/3 - 2/3, \quad (15)$$

it is implicit that certain necessary conditions are met. The most important of these is the proviso $c'_0 < c_m$, which we have considered at some length in this work. It is interesting to note that if this proviso is satisfied even the “anomalous” glasses which deform by non-volume-conserving processes may be analyzed in terms of eq (15); the fatigue properties are not sensitive to the origin of the residual contact field, as long as this field is of sufficient intensity to generate some precursor crack growth [11]. If such a precursor stage were not to be evident in the failure mechanics the “apparent” term n' would tend closer in value to the true n [5,9,12,13]. A second condition that needs to be met is that the flaws should indeed be produced in axial loading; other indentation loading systems, e.g., linear translation, give rise

to flaws which are governed by a transformation equation with coefficients significantly different from those in eq (15) [12,16].

It is seen in table 1 that the exponents obtained from this study agree well with the independent determinations for the glasses, but not for the ceramics. The relatively good agreement in the case of the glasses is attributable in part to the "model" behavior of this class of materials: transparency, isotropy, absence of microstructural complication, and ease in specimen preparation are factors which contribute to this behavior. Also, the n values of the glasses are comparatively low, so fatigue effects show up more strongly. This last point, coupled with a growing realization that conventional testing techniques used to obtain macroscopic velocity data are themselves subject to uncertainty (particularly the double torsion specimen [29]), could account for the discrepancies evident in the data for the ceramics.

Fatigue Curve Intercepts. The intercept terms in the master map representations do not have such a simple interpretation in terms of basic fracture parameters. This is clear from eq (11); λ'_p is a function of several quantities. Given the fatigue slope and inert strength evaluations as described in the two previous subsections, along with a direct measurement of the critical flaw size c_m , λ'_p effectively determines the crack velocity coefficient v_0 . Due to the compounding of errors (particularly from the n' exponent), determinations of this kind are subject to gross uncertainty. There accordingly seems to be little value in trying to retain v_0 as a design parameter, particularly since the λ' terms, which can usually be determined to within 15% from dynamic fatigue data, may be used directly in lifetime formulae. In studies of the basic physics and chemistry of crack growth, of course, v_0 remains a useful coefficient for scientific analysis.

Practical Implications of Master Maps

The major appeal of the master map construction advocated here lies in the provision of a graphic indicator of the intrinsic toughness and fatigue properties of brittle materials. Each material is represented by a universal curve, the relative position of which determines the merit of that material for structural applications. The marked superiority of such materials as silicon carbide and alumina become vividly apparent in the maps of figures 5 and 6. Useful distinctions may also be made between materials which cross over within the data range, e.g., soda-lime glass and PZT. On the basis of straight inert-strength testing we might reckon the first of these as the stronger material, whereas for applica-

tions involving sustained stresses it is the second which would tend to the larger lifetimes. Such crossovers would not be so obvious from the raw fracture mechanics parameters. It will be appreciated that this kind of intercomparison is made on the basis of "equivalent" flaw sizes: in this respect the indentation method, through its control over the flaw severity via the contact load, is unique in its capacity for reducing strength data to a common denominator.

In arguing the merits of this approach we do not mean to imply that it is only the *intrinsic* fracture properties which play an important role in the determination of component strengths and lifetimes; the effective sizes of the naturally occurring flaws which ultimately cause failure must also be known. Our procedure, by introducing flaws greater in severity than any of these natural flaws, automatically excludes information concerning the latter from the data. What our scheme effectively allows us to do is to determine the intrinsic parameters in a truly independent manner. All necessary extrinsic flaw parameters should be obtainable from straightforward inert strength tests (run at a single stressing rate), in the form of the usual statistical distribution functions. Lifetime predictions for as-prepared components could then be made without ever having to accumulate vast quantities of fatigue data [2]. In adopting this strategy one needs to keep in mind the strong influence that any persisting residual stress concentrations associated with the original initiation processes (in our case the elastic/plastic deformation) might exert on the subsequent flaw evolution. In the absence of information as to this aspect of flaw characterization steps should be taken to design conservatively, on the basis of "worst-case" configurations wherever possible. This last point is dealt with in greater detail in reference [25].

It has been indicated at several points that the existence of any spurious stresses incurred during the mechanical, chemical, or thermal history of a material would necessitate a third contribution to the starting stress intensity factor in eq (1), with consequent deviations from the currently determined toughness and fatigue relations. The fact that such deviations were not observed in the materials studied here may be taken as evidence that this potential complication has been successfully avoided. Again, it may be well to emphasize that it may not be so simple to confirm the elimination of spurious stresses from surfaces whose strengths are controlled by natural flaws, particularly in materials with typically wide flaw distributions; nor, of course, may we wish to eliminate them, bearing in mind that these stresses are most often compressive.

Finally, a comment may be made concerning the convenience of indentation load as a variable for in-

investigating fundamental flaw size effects. By systematically reducing the load we can produce corresponding smaller flaws, thereby providing a link between macroscopic and microscopic crack behavior. Any change in the nature of the indentation flaw will then become evident as deviations from universal plots, much as just described in relation to the spurious stress influence. In this way it has been possible to demonstrate that indentation flaws in glasses undergo an abrupt transition in properties below a threshold load (corresponding to a flaw size $\approx \mu\text{m}$): above this threshold the macroscopically determined laws of crack growth remain perfectly valid, regardless of scale, provided that the residual contact term is duly accounted for [11]; below the threshold the universal curves no longer apply, and failure becomes dominated by initiation micro-mechanics [17,30]. The indentation technique should prove similarly useful for studying size effects in ceramics, particularly for polycrystalline materials with relatively coarse microstructures.

Conclusions

1) The indentation-flaw technique provides an attractive route to the evaluation of intrinsic fracture parameters. Coupled with independent determinations of natural flaw distributions, the approach offers the prospect of accurate lifetime predictions with optimum specimen economy.

2) The control over the nature, shape, and above all, the size (via the contact load) of the indentation flaw allows for the derivation of a universal fracture formulation. Each material is represented by a single curve which incorporates the toughness and fatigue properties. Composite plots of these curves produce master maps, affording a simple graphic format for materials comparisons.

3) The inert strength cutoff on such a master map is a measure of effective material toughness. For "well-behaved" materials this effective toughness is consistent with macroscopically measured K_c values. In cases where inconsistency is observed the toughness reckoned from indentation data may provide a more reliable indication of the response of the typical natural flaw.

4) The slope of the fatigue curve on a master map is a measure of the susceptibility of a material to subcritical crack growth. The crack velocity exponent determined from this slope is an apparent value, n' , which is converted to the true value, n , via a simple transformation equation.

5) Deviations from universality on a master map indicate an extraneous influence in the fracture mechanics, e.g., spurious stress states, microstructure/crack interactions, and threshold size effects.

The authors thank T. P. Dabbs for providing raw fracture data on the glasses, and L. Respass and S. J. Mann for their help with specimen preparation.

References

- [1] Wiederhorn, S. M.; E. R. Fuller, J. Mandel, and A. G. Evans, *J. Amer. Ceram. Soc.*, Vol. 59, 1976, pp. 404-411.
- [2] Ritter, J. E.; N. Bandyopadhyay and K. Jakus., *Ceram. Bull.*, Vol. 60, 1981, pp. 798-806.
- [3] Marshall, D. B., and B. R. Lawn, *J. Mater. Sci.*, Vol. 14, 1979, pp. 2001-2012.
- [4] Marshall, D. B.; B. R. Lawn and P. Chantikul, *ibid.*, Vol. 14, 1979, pp. 2225-2235.
- [5] Marshall, D. B. and B. R. Lawn, *J. Amer. Ceram. Soc.*, Vol. 63, 1980, pp. 532-536.
- [6] Chantikul, P.; B. R. Lawn and D. B. Marshall, *ibid.*, Vol. 64, 1981, 322-325.
- [7] Anstis, G. R.; P. Chantikul, B. R. Lawn, and D. B. Marshall, *ibid.*, Vol. 64, pp. 533-538.
- [8] Chantikul, P.; G. R. Anstis, B. R. Lawn, and D. B. Marshall, *ibid.*, Vol. 64, pp. 539-543.
- [9] Lawn, B. R.; D. B. Marshall, G. R. Anstis, and T. P. Dabbs, *J. Mater. Sci.*, Vol. 16, 1981, pp. 2846-2854.
- [10] Cook, R. F.; B. R. Lawn and G. R. Anstis., *ibid.*, Vol. 17, 1982, pp. 1108-1116.
- [11] Dabbs, T. P.; B. R. Lawn and P. L. Kelly, *Phys. Chem. Glasses*, Vol. 23, 1982, pp. 58-66.
- [12] Symonds, B. L.; R. F. Cook and B. R. Lawn, *J. Mater. Sci.*, in press.
- [13] Marshall, D. B. and B. R. Lawn, *J. Amer. Ceram. Soc.*, Vol. 64, 1981, pp. C6-C7.
- [14] Marshall, D. B.; A. G. Evans, B. T. Khuri-Yakub, J. W. Tien, and G. S. King, to be published.
- [15] Lawn, B. R., in *Fracture Mechanics of Ceramics*, R. C. Bradt, A. G. Evans, D. P. H. Hasselman and F. F. Lange, Eds., Plenum, New York, 1983, Vol. 5, pp. 1-25.
- [16] Fuller, E. R.; B. R. Lawn and R. F. Cook, *J. Amer. Ceram. Soc.*, Vol. 66, 1983, pp. 314-321.
- [17] Dabbs, T. P., and B. R. Lawn, *Phys. Chem. Glasses*, Vol. 23, 1982, pp. 93-97.
- [18] Cook, R. F., unpublished work.
- [19] Marshall, D. B. and B. R. Lawn, *J. Amer. Ceram. Soc.*, Vol. 61, 1978, pp. 21-27.
- [20] Mould, R. E., and R. D. Southwick, *ibid.*, Vol. 42, 1959, pp. 542-547, 582-592.
- [21] Lawn, B. R. and E. R. Fuller, *J. Mater. Sci.*, Vol. 10, 1975, pp. 2016-2024.
- [22] Lawn, B. R.; A. G. Evans and D. B. Marshall, *J. Amer. Ceram. Soc.*, Vol. 63, 1980, pp. 574-581.

- [23] Wiederhorn, S. M., in *Fracture Mechanics of Ceramics*, R. C. Bradt, D. P. H. Hasselman and F. F. Lange, Eds., Plenum, New York, 1974, vol. 2, pp. 613-646.
- [24] *Flexure Testing of Glass*, A.S.T.M. Annual Book of Standards, ASTM, Philadelphia, 1979, Part 17, C158-C172.
- [25] Gonzalez, A. C.; H. Multhopp, R. F. Cook, B. R. Lawn, and S. W. Frieman, a paper in *Methods for Assessing the Structural Reliability of Brittle Materials*, ASTM Special Technical Publication, Philadelphia (in press).
- [26] Lawn, B. R.; D. B. Marshall, P. Chantikul, and G. R. Anstis, *J. Austral. Ceram. Soc.*, Vol. 16, 1980, pp. 4-9.
- [27] Arora, A.; D. B. Marshall, B. R. Lawn, and M. V. Swain, *J. Non-Cryst. Solids*, Vol. 31, 1979, pp. 415-428.
- [28] Smith, S. S., and B. J. Pletka, in *Fracture Mechanics of Ceramics*, R. C. Bradt, A. G. Evans, D. P. H. Hasselman and F. F. Lange, Eds., Plenum, New York, 1983, Vol. 6, pp. 189-210.
- [29] Pletka, B. J.; E. R. Fuller and B. G. Koepke, in *Fracture Mechanics Applied to Brittle Materials*, S. W. Freiman, Ed., A.S.T.M. Special Technical Publication 678, ASTM., Philadelphia, 1978, pp. 19-37.
- [30] Dabbs, T. P.; C. J. Fairbanks and B. R. Lawn, a paper in *Methods for Assessing the Structural Reliability of Brittle Materials*, ASTM Special Technical Publication. Philadelphia (in press).
- [31] Wiederhorn, S. M, *J. Amer. Ceram. Soc.*, Vol. 52, 1969, pp. 99-105.
- [32] Wiederhorn, S. M., unpublished work.
- [33] Ritter, J. E., and C. L. Sherbourne, *J. Amer. Ceram. Soc.*, Vol. 54, 1971, pp. 601-605.
- [34] Cook, R. F.; B. R. Lawn, T. P. Dabbs, K. D. Reeve, E. J. Ramm, and J. L. Woolfrey, *J. Amer. Ceram. Soc.*, Vol. 65, 1982, pp. C172-C173.
- [35] Gonzalez, A. C., and S. W. Freiman, unpublished work.
- [36] McHenry, K. D.; T. Yonushonis and R. E. Tressler. *J. Amer. Ceram. Soc.*, Vol. 59, 1976, pp. 262-263.
- [37] Koepke, B. G. Unpublished work.
- [38] Pletka, B. J., and S. M. Wiederhorn. *J. Mater. Sc.*, Vol. 17, 1982, pp. 1247-1268.

The Interactions of Composition and Stress in Crystalline Solids

F. C. Larché

Université de Montpellier 2, 34060 Montpellier-Cedex, France

and

J. W. Cahn

National Bureau of Standards, Gaithersburg, MD 20899

Accepted: August 30, 1984

The thermodynamics of stressed crystals that can change phase and composition is examined with particular attention to hypotheses used and approximations made. Bulk and surface conditions are obtained and for each of them practical expressions are given in terms of experimentally measurable quantities. The concept of open-system elastic constants leads to the reformulation of internal elastochemical equilibrium problems into purely elastic problems, whose solutions are then used to compute the composition distribution. The atmosphere around a dislocation in a cubic crystal is one of several examples that are completely worked out. The effects of vacancies and their equilibrium within a solid and near surfaces are critically examined, and previous formulas are found to be first order approximations. Consequences of the boundary equations that govern phase changes are studied with several examples. Finally, problems connected with diffusional kinetics and diffusional creep are discussed.

Key words: creep; crystals; heterogeneous solids; multicomponent; phase equilibria; solid; stress; thermodynamics; vacancies.

Introduction

The literature of the thermodynamics of solids spans more than a century and has appeared in many fields. It has been marked by long controversies, some even regarding the very existence of equilibrium under conditions of nonhydrostatic stress. The resulting concepts and relations have been used in applications to global equilibrium problems, and as local equilibrium conditions in nonequilibrium problems of diffusion, creep,

electrochemistry, and phase changes. The formulations have been gradually generalized to include multicomponent anisotropic solids, containing vacancies and other defects, that are nonhydrostatically and nonuniformly stressed. Considerable attention has been given to multi-phase systems and to conditions of equilibrium at interfaces between phases that are in mechanical and thermal contact, that can exchange matter and under conditions of slip or no slip (incoherent and coherent, resp.). In view of the importance of the field, a clarification of the controversies seems in order.

Thermodynamics lends itself to many formulations based on different definitions, conventions and notations. When properly done, all these formulations should identify the same measurable quantities and give identical relationships among them. Discrepancies arise when the formulations differ in assumptions made about the behaviour of matter. There are also many simplifications that may not be valid or necessary. Invalid assumptions have been made about the laws of

About the Authors, Paper: F. C. Larché, a member of the faculty at the Université de Montpellier, and J. W. Cahn, who is with NBS' Center for Materials Science, are frequent collaborators in scientific papers. This paper, which is also appearing in *Acta Metallurgica*, is presented here through the courtesy of *Acta Metallurgica* and its publisher, Pergamon Press.

thermodynamics and about the conditions for equilibrium. We will examine the main formulations for their assumptions to find their range of validity. Whenever possible we will identify the most general formulation and show how the other formulations follow as special cases, compare predictions, and identify sources of discrepancies. But since general formulations are often more cumbersome to apply, we will examine a set of simple applications to display how one uses the main results in this field.

It may be worthwhile to categorize broadly the main controversies and to illustrate with one simple example how they arise. These center around: 1) the question of the existence of equilibrium if diffusion is permitted; 2) the various methods of distinguishing solids from fluids in a formulation, these involving models of solids and constraints on the variations that can occur in solids; 3) the definitions of chemical potential of species inside solids, since in some formulations one cannot arbitrarily add atoms to the interior of a crystal without removing other atoms or destroying vacancies; 4) how one formulates the conditions for equilibrium when the familiar minimum Gibbs-free energy which works only for constant hydrostatic pressure is inapplicable, and when so many different chemical potential conventions have been proposed; and 5) clear distinctions between the accretions that can occur at surfaces and at interior defects, such as climbing dislocations, and the addition of atoms to sites inside of crystals.

In addition, there are a variety of simplifications with obvious limitations on the applicability of the results. Among them is one, homogeneity, which has led to major misconceptions. Many situations will lead to homogeneous systems at equilibrium, but if one requires in tests for equilibrium that all variations keep the system homogeneous, one may constrain the system unnecessarily.

With these controversies in mind, let us examine the simple example of a solid cylinder containing one or more components and a straight axial dislocation. Let us first ignore surface effects and let the cylinder be infinite in all directions. Let there be no restriction on diffusion. If the solid is crystalline, an equilibrium will be reached with the dislocation retained in which the solid is heterogeneously and nonhydrostatically stressed. If the solid is multicomponent, it will also be compositionally heterogeneous. The system can reach an equilibrium which of course means that all diffusional flow has ceased, in spite of the shear stresses and the heterogeneity.

If the cylinder had been a highly viscous liquid in which the dislocation had been introduced by a cutting, displacing, and welding procedure, the dislocation

would disappear on annealing. Equilibrium would not be compatible with shear stress or heterogeneity. It is apparent that crystallinity imposes restrictions on the variations that lead to a different type of equilibrium.

Even in a one component solid, there will be a gradient in the Helmholtz-free energy density at equilibrium. Any definition of a chemical potential, which for a one-component system reduces to the local free energy per atom, cannot subsequently be used by asserting that such chemical potentials must be constant at equilibrium or, if not constant, will lead to diffusional fluxes. Care must be exercised in the definition of chemical potentials in one or multicomponent systems to ensure that they are useful.

The constraint which crystallinity imposes in this example is that some of the atoms cannot be moved at will without a counterflux of some other species, including vacancies, to take their place in the crystal structure. At the surface and at the core of dislocations capable of climbing, this constraint does not apply and atoms can be inserted or removed at will.

To illustrate the importance of separate equilibrium conditions at surfaces, let the cylinder in our example have a finite radius and permit surface rearrangement. An equilibrium shape could be reached where transfer of small amounts of any species of atoms from one surface location to another does not change the appropriate free energy. This would be a thermodynamically stationary state in which all fluxes would cease, but it would be metastable or possibly unstable equilibrium because moving the dislocation out of the cylinder would lead to a lowered energy.

2. What Is a Solid?

Formulations of thermodynamics differ considerably in how the essential aspects of solidity are represented mathematically. Many authors purporting to deal specifically with solids reach conclusions that are the same as for very viscous liquids that may take a long time to reach an equilibrium that does not support shears.

Various models, composed of springs and dashpots, have been proposed to represent the viscoelastic behavior of matter. Whereas the Maxwell model creeps continuously under load, the Meyer-Kelvin-Voigt [1]¹ solid reaches a mechanical equilibrium when the load is entirely carried by the spring. The elements of these solids do not dissolve or diffuse, and Gibbs [2] devised a model of a solid that did both.

Gibbs introduced the idea of a solid component which does not diffuse. Like Mayer-Kelvin-Voigt's solid, it

¹Figures in brackets indicate literature references at the end of this paper.

can deform elastically but it always retains its connectivity. In addition Gibbs considered surfaces, where he did permit the solid to grow by accretion or to shrink by vaporization, to melt or to dissolve into contacting fluids. He also incorporated the concept of a fluid component which can diffuse and distort the solid. He fully developed the thermodynamic properties of such a solid, including its equilibria, and revealed a variety of surprising properties. Since the solid component was not involved in any chemical variations except at the surface, there was no need to define a chemical potential in the solid. When the solid was equilibrated with a fluid, the chemical potential of this solid component in the fluid was readily calculated. One important result was that the chemical potential in the saturated fluids in contact with a homogeneously stressed solid depends on the orientation of the surface. There is thus not only no need to define a chemical potential of the solid component, but it does not seem to be definable. The fluid component on the other hand has a defined chemical potential that is constant at equilibrium throughout all phases even if they are heterogeneously stressed. Gibbs' solid is therefore quite active chemically and yet it is different from a fluid. The key was the solid component. Even though this component can dissolve, essential solid properties are obtained.

Gibbs was strongly influenced by the law of definite proportions and required his solid component to be a single element or a stoichiometric compound. If it was a compound, the chemical potential in the saturated fluids is calculated even if the compound dissociates or reacts with the solvent. Modern examples of Gibbs solids are polymer fibers which also can absorb solvent molecules, and silicate glasses in which the silicate network is the solid component while modifier ions can diffuse about. A very good example of the kind of equilibrium Gibbs was able to calculate is the bending of a damp wooden beam in which the water redistributes at equilibrium and affects the compliance. Li, Darken, and Oriani [3] pointed out that mobile interstitials in metals at temperatures where the substitutional atoms did not move was a valid metallurgical example of a Gibbs solid with a fluid component. An example of the equilibria of a dissolving Gibbs solid occurs in stressed electrodes. The equations predict the effect of elastic stress on the electrode potential [4].

Solid state diffusion of every component is counter to the strict definition of Gibbs' solid component. As a result most thermodynamic formulations that permit unrestricted diffusion to take place do not ascribe to the solid any property that differs from a viscous fluid. As the example in the introduction points out, unrestricted

diffusion consistent with our knowledge of the solid does permit new kinds of equilibria.

Gibbs' solid component, because it did not diffuse, served as network for defining displacement and hence strain, as well as the local composition of the fluid component. The local energy and entropy density were functions of the local strain and composition. What was needed was a network which continued to define unambiguously the same place in the solid even if all atoms were capable of diffusing. In crystalline structures, the lattice serves this function, and a thermodynamics has been developed. Robin [5] has simply let the lattice itself be the solid component, and has found that "component differences" become the exact analogues of Gibbs' fluid components. Instead of modifying Gibbs' concept we have defined a network solid as one in which there is an unambiguous method of locating the same place after diffusion, and where the thermodynamic properties are functions of the strain and local composition defined by this network [6]. Gibbs solid component is one example of such a network; the lattice is another example.

Most of our work has been with simple crystal structures in which there is one type of substitutional site and one type of interstitial site. Atoms of a given species are assumed to be either substitutional or interstitial. The substitutional sites served as a network. Bravais solids where lattice sites are occupied by substitutional atoms are an example. Recently attention has focused on species which could occupy both interstitial or substitutional sites [7], and this has led to the generalization of structures in which many different sites are occupied in a unit cell and where a particular species can occupy several sites. One can even include the case where no species occupies the origin in the unit cell which serves as network marker.

In crystal structures, the network imposes what we have called the network restriction. A site exists, regardless of the species that occupies it, or even if it is empty. Atoms exchange among sites

$$A_I + B_J = A_J + B_I \quad (2.1)$$

where I and J are different types of sites: Sites that are mostly filled are occupied by what are called substitutional atoms, while sites that are mostly vacant are occupied by what are called interstitial atoms.²

²The term interstitial compound is an unfortunate term in which the interstitials are merely small atoms fully occupying a site in the structure [8]. The usual definition of interstitials, that these are atoms occupying sites that are mostly empty, has important consequences in thermodynamic formulations. An empty substitutional site is called a vacancy, while empty interstitial sites are usually ignored, since their concentration or activity in, e.g., the law of mass action, hardly differs from unity.

Vacancies are capable of diffusing or reacting with atoms on other sites. Letting B be a vacancy, (2.1) becomes

$$V_I + A_J = A_I + V_J \quad (2.2)$$

where I and J are different sites. If I is an interstitial site, this can also be written

$$A_J = A_I + V_J. \quad (2.3)$$

One of the main results of the network restriction is that there is no need to define separate chemical potentials of individual network species. Within the crystal only their differences are ever needed.

The network is unambiguously defined only as long as the structure is not severely distorted. The network can be modified at surfaces and dislocations and these have led to special equilibrium conditions. Of particular interest is the fact that there are differences between solid-fluid interfaces and solid-solid interfaces regarding equilibrium conditions. Two types of solid-solid boundaries have been treated [10]: incoherent interfaces where there are two independent networks with no relationship between them and coherent interfaces where there is an exact correspondence between network sites in the two crystals, and a connectivity across the interface that survives the distortions of a phase change that transfers sites from one crystal to the other. Thus many restrictions in Gibbs' solid have been eliminated. Modern understanding of solid solutions, crystalline defects, and diffusion have been incorporated. In addition, solid-solid equilibria, interfaces, and phase changes have been considered.

3. Derivations of Usable Equilibrium Conditions

3.1 Thermodynamic Formulation

The basic two laws of thermodynamics are quite general and applicable not only to all equilibrium conditions but also in specifying what cannot happen in non-equilibrium conditions. They often are cumbersome to use, but from them special conditions have been derived (such as constant temperature at equilibrium) that are easier to apply. In addition, there are certain restrictions or constraints that occur commonly that permit even simpler specialized but rigorously applicable procedures to be developed. A good example is the Gibbs free energy. Under the special restriction that temperature, pressure, and the mass of various species be held constant, it can be shown that the laws of thermodynamics reduce to the simple condition that the Gibbs free en-

ergy monotonically decreases to a minimum. For these common restrictions, it is not longer necessary to start from the basic laws. For equilibrium, one begins with the minimization of Gibbs free energy knowing that this is fully equivalent to the basic laws. The procedure is a general one, subject only to the easily verifiable restrictions on temperature, pressure, and mass. The restrictions are important. When temperature decreases (as in an endothermic reaction held adiabatically), pressure increases or mass is added, the Gibbs free energy can increase and has lost its usefulness as a simple condition for equilibrium.

Whenever we encounter new restrictions or constraints, it is necessary to return to the two basic laws to find new conditions for equilibrium that are general, subject only to the restrictions or constraints. It is important that the restrictions or constraints are verifiable and that they be general enough to include many important situations, but not so general as to lead to cumbersome conditions. The procedures for finding simpler equilibrium conditions subject to new restrictions or constraints are straightforward and if done with mathematical rigor, need only be done once. Applications then follow from these derived conditions. The derivation often identifies the useful free energy. It is dangerous to assert conditions for equilibrium under new restrictions (some type of free energy to be minimized or some potential to be constant) without a derivation that begins with the basic two laws.

There are various derivations in the literature. They differ in the model of "what is a solid" expressed in terms of restrictions on possible variations. They also differ on whether or not they require homogeneity. They differ on whether they begin with the basic two laws, or with some derived law.

It is not difficult to start with the basic laws used by Gibbs: "For the equilibrium of any isolated system, it is necessary and sufficient that in all possible variations in the state of the system which do not alter its entropy, the variation of its energy shall either vanish or be positive" [9, p. 56]. It is quite straightforward to permit the system to be heterogeneous.

Since the general state of a solid is heterogeneous, the energy, entropy and mass of its various components will be integrals over the volume, and the minimization procedure is done by standard variational calculus. Such a formulation permits the solid to change its shape by elastic deformation or by a process of network modification which we will call either accretion, dissolution, or phase change.

These methods of variational calculus were used by Gibbs every time the system under consideration was not homogeneous; the influence of gravity [9, p. 144],

stressed solids [2], surfaces [9, p. 238], multiphase systems [9, p. 64], etc. A variational statement of the first and second laws of thermodynamics for the multi-component network solid has been carried out [6]. It very neatly produces all the conditions for equilibrium—mechanical, thermal, and chemical—in the bulk and at the interfaces. There is usually no need to assume linearity, ideality, or isotropy. The derived equations identify and define important functions and usually can be manipulated to suggest methods of measurement.

The imposed constraints are incorporated into the formulation as Lagrange multipliers and this introduces quantities which must be constant throughout the system at equilibrium. Since sites in a unit cell or a network exist whether occupied by atoms or not, vacancies appear as a conserved species within a network. We formulated three different rules for the transfer of material across an interface [10]. Network sites could be added or subtracted to the solid at solid-fluid and at incoherent solid-solid interfaces. At a coherent solid-solid interface, a single network describes both solids, and during phase changes, sites are transferred but do not change their relative locations.

3.2 State Variables and Notations

The procedure outlined can be followed once the state variables have been identified. With network solids, a strain can be defined. The energy density is assumed to be a function of that strain (either the usual small strain, or the deformation gradient to include the cases of large strains), of the entropy density, and of the density of the various atomic or molecular species.

The choice of the strain or deformation gradient as a state variable that describes the mechanical state of the solid by no means exhausts the possible choices. Continuum mechanicians and others [11-14] have described much more complex solids, where higher gradients of displacement or composition come in the picture. We feel that our choice is sufficient to describe many metallurgical materials. In any case, thermodynamics uses as input data the results of measurements of mechanical and thermal properties, and inadequate specification of state variables would become apparent.

Only small strain theory will be explicitly used here. The relations that are valid without this approximation have been derived [10, 15], and effects that might modify the small strain results will be mentioned and discussed in the course of this article.

The reference state for strain in the solid is quite arbitrary. It can be at zero stress, or under hydrostatic pressure, and at any arbitrary constant composition. It merely serves to identify the same point \underline{x}' in a solid

after composition change and strain. For many elastic energy equations, a convenient reference state is zero stress. There are also useful standard states for thermodynamic quantities. These are often at hydrostatic stress that is not zero and at definite compositions. As a result there are advantages to be flexible about the reference state for strain. We will try to point out in each application which reference state we have used.

When the point \underline{x}' of a solid is displaced by \underline{u} , the small strain is defined by³

$$E_{ij} = 1/2(u_{i,j} + u_{j,i}). \quad (3.1)$$

A change of reference state from \underline{x}' to \underline{x}'' (\underline{x}' where $\underline{x}'' - \underline{x}' = \underline{v}$ leads in the small strain approximation to a new strain E''_{ij} given by

$$E''_{ij} = E_{ij} + 1/2(v_{i,j} + v_{j,i}). \quad (3.2)$$

The density of energy, entropy and component I are respectively denoted by ϵ , s , and ρ_I . Because the elementary volume of solid is affected by its state of strain, densities per unit volume in the deformed state always contain a strain effect. As such they are not very convenient to use. Much better variables are the densities per unit volume in the reference state. These will be noted by primed symbols. The relations between primed and unprimed densities are

$$\epsilon'/\epsilon = s'/s = \rho'_I/\rho_I = \rho'_0/\rho_0 = \dots \quad (3.3)$$

$$= V_0/V'_0 = 1 + E_{kk} \quad (3.4)$$

where ρ_0 is the molar density of lattice sites, and its inverse V_0 is the molar volume of lattice sites.

All of our chemical densities ρ_I and ρ'_I will be atomic or molar densities (moles/volume). This is especially preferred to mass densities when we consider vacancies as a species. It is useful to introduce dimensionless composition variables

$$c_I = \rho'_I/\rho'_0 = \rho_I/\rho_0.$$

This is the classical mole fraction for single-site substitutional alloys. For an interstitial alloy with no vacancies on the substitutional sites, c_I given above is the

³All vectors and tensors are expressed in terms of components with respect to an orthonormal axis system. Small subscripts like i and j are understood to have value 1, 2, or 3. Repeated indices are understood to be summed (Einstein convention) and subscripts preceded by a comma are derivatives, e.g.

$$E_{ii} = E_{11} + E_{22} + E_{33}$$

$$u_{i,j} = \partial u_i / \partial x_j$$

molal composition. The mole fraction \bar{c}_I is then

$$\bar{c}_I = \rho_I / (\rho_0 + \rho_I) = c_I / (1 + c_I).$$

which reduces to c_I at small concentration. We will drop the distinction between c_I and \bar{c}_I .

3.3 Lagrange Multipliers

From the entropy constraint comes the standard condition that the temperature is everywhere equal to a Lagrange multiplier, and is therefore constant. It allows us to define a Helmholtz free energy density by a Legendre transform

$$f' = \epsilon' - \theta s' \quad (3.5)$$

which we subsequently use because it is more convenient in many practical applications.

From the conservation of mass conditions come Lagrange multipliers that differ substantially from standard fluid equilibrium, a direct consequence of the network constraint. As with fluids, conservation of N chemical components lead to N Lagrange multipliers that are constants at equilibrium. Whereas for fluids they can be identified with N chemical potentials, for a system consisting of a network solid containing N substitutional species only $N-1$ quantities can be identified with physical processes replacing one specie with another on a site. The quantities thus identified with Lagrange multiplier differences we have called diffusion potentials. The notation is M_{IK} , where K is the dependent species. Vacancies are considered a species that can be ignored in some applications. Because of their definition as Lagrange multipliers, the M_{IK} , like the temperature are constants, and take on a precise local meaning everywhere within the system

$M_{IK} = \text{constant}$ everywhere within the system.

$$= (1/\rho_0') (\partial f' / \partial c_{IK})_{\theta, E_{IJ}}. \quad (3.6)$$

Since the c_I are not independent, we have introduced the differential operator

$$(\partial / \partial c_{IK}) = (\partial / \partial c_I)_{c_{J \neq I, K}} \quad (3.7)$$

for a unit composition increase of species I , an equal decrease in species K , holding the composition of all other substitutional species on that site fixed. For binaries we drop the subscripts and adopt the convention $c = c_1$ and $(\partial / \partial c_{12}) = (\partial / \partial c)$.

From this definition we have

$$M_{IJ} + M_{JK} + M_{KI} = 0 \quad (3.8)$$

$$M_{IJ} = -M_{JI}; M_{II} = 0. \quad (3.9)$$

In the case of equilibrium with a fluid, M_{IK} is equal to the difference in chemical potential of I and K in the fluid

$$M_{IK} = \mu_I^f - \mu_K^f \quad (3.10)$$

If the vacancy is chosen as K , we have

$$M_{Iv} = \mu_I^f. \quad (3.11)$$

It might seem natural to use the M_{Iv} , and keep the formalism of hydrostatic thermodynamics. This has been done in a number of formulations [7]. However, it has practical drawbacks (see sect. 5.5), and we have found it preferable to keep the flexibility of choice for the dependent species K .

The N th Lagrange multiplier which we will call μ_K cannot be identified in many problems. It is eliminated from all equilibrium calculations for internal equilibrium of a crystal away from surfaces and dislocations that can climb. It also is eliminated from all equilibrium calculations at coherent boundaries. Only in fluids, at incoherent boundaries and climbable dislocations can we identify μ_K with the chemical potential of the K specie.

The chemical potentials of interstitials are constant and equal to the chemical potentials of the corresponding species in the other phases,

$$M_I = \mu_I^f. \quad (3.12)$$

We shall see in section 5, where multisite solids are considered, that there is no need to differentiate between substitutional and interstitial sites. An increase of composition of the interstitial species I , holding the composition of all other interstitial species fixed, results in an equivalent decrease of vacancies on interstitial sites. But unlike vacancies on substitutional sites, vacancies on interstitial sites always have a concentration close to the total number of possible sites and can be dropped from consideration. In order to standardize and simplify the notation, we also call these chemical potentials diffusion potentials, and in order to simplify the notation in the various expressions, M_{IK} is understood to represent all diffusion potentials.

The restriction in the number of potentials necessary to calculate an equilibrium is a direct consequence of the crystalline nature of the solid and therefore should apply to the same solid under hydrostatic stress. In this case it can be shown (Appendix 1) that the previous equations, together with the boundary conditions to be discussed thereafter, are strictly equivalent to the standard conditions for equilibrium between fluids.

3.4 Mechanical Equilibrium

The variational calculus gives us [6,10] the very standard form of the mechanical equilibrium equation. It states that the divergence of the stress tensor is zero

$$T_{ij,j}=0. \quad (3.13)$$

This equation is also true for the large strain case, but the derivative is with respect to variables \underline{x} rather than \underline{x}' , a distinction that is not made in the small strain approximation. Large strain forms involving \underline{x}' have been obtained [15].

3.5 Interface Conditions

Along each interface, there are conditions for mechanical equilibrium, and a condition for phase change equilibrium. They both depend on the nature of this interface.

3.5.1 Solid-Fluid Interfaces

For solid-fluid interfaces, the mechanical equations state that the normal is a principal direction of stress. The principal value associated with it is equal in magnitude to the pressure in the liquid and opposite in sign. The pressure is here the classical thermodynamic pressure, which is positive in fluids, and the convention for stress is such that the stress corresponding to a tension is positive.

The phase change equation can be written

$$f - \sum \mu_i^L \rho_i = -P \quad (3.14)$$

where μ_i^L are the chemical potentials in the fluid, while the ρ_i and f pertain to the solid. Because of the $(N-1)$ equalities (3.10)

$$f - \sum_{I \neq K} M_{IK} \rho_I - \mu_K^L \rho_0 = -P. \quad (3.15)$$

Because $M_{KK}=0$ the summation over all species is the same as the summation over all species but K. We can therefore drop the restriction and adopt the notation that Σ without any qualification means summation over all species I. To simplify notation it is convenient to define the ω function as

$$\omega \equiv f - \sum M_{IK} \rho_I - \mu_K \rho_0 \quad (3.16)$$

where μ_K is the Lagrange multiplier associated with the K^{th} species. At this stage neither ω nor μ_K has physical

meaning. Once all the equilibrium equations are written they will have a specific meaning, or are eliminated. In a fluid ω is equal to minus the pressure, and thus because $\mu_K = \mu_K^L$ eq (3.15) could be rewritten

$$\omega^s = \omega^L. \quad (3.17)$$

We should emphasize that these equations are between unprimed quantities, that are usually not convenient to use for solids. The conversion follows eq (3.4) and gives

$$\omega^{ts} = -P(1 + E_{kk}). \quad (3.18)$$

3.5.2 Incoherent Interfaces

Along an incoherent solid-solid boundary, the equilibrium equations are

$$T_{ij}^\alpha n_j^\alpha = \omega^\alpha n_i^\alpha \quad (3.19)$$

$$T_{ij}^\beta n_j^\beta = \omega^\beta n_i^\beta \quad (3.20)$$

$$\omega^\alpha = \omega^\beta \quad (3.21)$$

where n_i^α (resp. n_i^β) are the components of the normal to the interface oriented from α to β (resp. β to α). They all contain ω and hence the Lagrange multiplier μ_K .

Equations (3.19) and (3.20) imply that the normal is a principal stress axis and that in this case ω is the value of that principal stress. Multiplication of (3.19) by n_i^α and summation over i gives

$$\omega^\alpha = T_{ij}^\alpha n_i^\alpha n_j^\alpha. \quad (3.22)$$

From (3.20) we can obtain a similar expression for ω^β . Therefore ω^α and ω^β are identified for this problem.

Using the definition of ω we obtain

$$\mu_K = V_0^\beta (f^\beta - \sum M_{IK} \rho_I^\beta - T_{ij}^\beta n_i^\beta n_j^\beta). \quad (3.23)$$

Substituting this value of μ_K in (3.21) and (3.19) gives the equivalent system of equations

$$\begin{aligned} \mu_K &= V_0^\alpha (f^\alpha - \sum M_{IK} \rho_I^\alpha - T_{ij}^\alpha n_i^\alpha n_j^\alpha) \\ &= V_0^\beta (f^\beta - \sum M_{IK} \rho_I^\beta - T_{ij}^\beta n_i^\beta n_j^\beta) \end{aligned} \quad (3.24)$$

$$T_{ij}^\alpha n_j^\alpha = -T_{ij}^\beta n_j^\beta = T_{kj}^\alpha n_k^\alpha n_i^\alpha n_j^\alpha \quad (3.25)$$

Equations (3.24) and (3.25) contain only known quantities and are the usable ones. Equation (3.23) can be interpreted as a definition for the chemical potential of the K species and this potential is constant along the interface. Along an incoherent interface we can then calculate a chemical potential for every specie, something which is not possible at any other location within

the bulk of the α and β phase. Let us note that each side of eq (3.24) depends on what specie is chosen for K. Because the expression

$$\Sigma M_{IK} \left(\frac{\rho_1^\alpha}{\rho_0^\alpha} - \frac{\rho_1^\beta}{\rho_0^\beta} \right) = \Sigma M_{IK} (c_1^\alpha - c_1^\beta)$$

is independent of K, the equation itself is independent of this choice. A comparison of (3.23) and (3.15) shows the similarities between solid-fluid and incoherent solid-solid equilibria.

3.5.3 Coherent Solid Interfaces

In a coherent solid-solid equilibrium, the mechanical boundary conditions

$$T_{ij}^\alpha n_j^\alpha = -T_{ij}^\beta n_j^\beta \quad (3.26)$$

indicate that the tractions (but not necessarily the stress tensor) are continuous across the interface. If the same reference state for strain is chosen for α and β the phase change equation (Appendix 2) reads

$$\begin{aligned} V_0' f^\alpha - \Sigma M_{IK} c_1^\alpha + V_0' (-T_{ij}^\alpha n_i^\alpha n_j^\alpha + 2\Omega_{ij}^\alpha T_{jk}^\alpha n_i^\alpha n_k^\alpha) \\ = V_0' f^\beta - \Sigma M_{IK} c_1^\beta + V_0' (-T_{ij}^\beta n_i^\beta n_j^\beta \\ + 2\Omega_{ij}^\beta T_{jk}^\beta n_i^\beta n_k^\beta) \end{aligned} \quad (3.27)$$

where Ω_{ij} is the small rotation tensor

$$\Omega_{ij} = \frac{1}{2}(u_{i,j} - u_{j,i}). \quad (3.28)$$

For this type of interface equilibrium, the Lagrange multiplier μ_K has disappeared from the equations. In contrast to the two cases treated before, no definition of individual chemical potentials for each species arises, even at the interface. As we will see none are needed to solve problems. This is a direct consequence of the restrictions in a fully coherent phase change, where no network site is created or destroyed.

4. The Data Base

We have identified a number of important thermodynamic quantities that determine the state of a system, and a number of functions of these state variables that enter into the equations of equilibrium. We now examine how one might determine these quantities from the usual quantities that are measured and available in compilations. They turn out to be identical to those used in ordinary solution thermodynamics and elasticity.

4.1 Geometric Variables

The lattice constants are readily-determined non-linear functions of composition, temperature, and stress. From the lattice constants in the reference state we can compute ρ_0' . From a comparison of the lattice shape in the actual state and the reference state, we can compute the strain, or, if the strain is large, the deformation gradient. Since the actual state and the reference state are usually chosen to be at the same temperature but not necessarily at the same composition, the strain E_{ij} is a sum of a contribution due to composition change with no change in stress, E_{ij}^c , and one due to stress. The general case when neither contribution is isotropic has been treated [15]. The tensor E_{ij}^c is subject to the same crystal symmetry restrictions as the thermal expansion tensor [17]. For the present we will concentrate mostly on the isotropic case. Defining k such that

$$E_{ij}^c = k \delta_{ij} \quad (4.1)$$

and assuming Hooke's law of linear elasticity we can write

$$E_{ij} = \left(k - \frac{\nu}{E} T_{kk} \right) \delta_{ij} + \frac{1+\nu}{E} T_{ij}. \quad (4.2)$$

The dilatation E_{kk} is given by

$$E_{kk} = \frac{1-2\nu}{E} T_{kk} + 3k. \quad (4.3)$$

In cubic crystals, E_{ij}^c is also isotropic, so that the formula in eq (4.1) is still valid.

The constant ρ_0' appears repeatedly in various formulas because elastic energy naturally appears as energy per unit volume, whereas other energies will be per mole. ρ_0' is the conversion factor that transforms one into the other. Its inverse V_0' is the molar volume of the lattice sites. Combining (3.3), (3.4) and (4.3) we have for isotropic solids

$$V_0'/V_0' = \rho_0'/\rho_0 = 1 + \frac{1-2\nu}{E} T_{kk} + 3k. \quad (4.4)$$

The derivative of E_{ij}^c with respect to composition in binary allows also occurs commonly

$$\eta_{ij} = dE_{ij}^c/dc. \quad (4.5)$$

For systems with orthogonal axes

$$\eta_{ii} = (\partial \ln a_i / \partial c) \text{ (no summation)} \quad (4.6)$$

where the a_i are the lattice parameters. When E_{ij}^c is isotropic

$$\eta_{ij} = (dk/dc)\delta_{ij} = \eta\delta_{ij}. \quad (4.7)$$

In binary isotropic and cubic systems η is also related to the partial molar volumes

$$\eta = (\bar{V}_1 - \bar{V}_2)/3V'_0. \quad (4.8)$$

If η is constant

$$k = (c - c_0)(\bar{V}_1 - \bar{V}_2)/3V'_0 \quad (4.9)$$

where c_0 is the composition of the reference state chosen to measure the strain. It is to be emphasized that the anisotropic and nonlinear versions of these equations are readily available [15].

4.2 Thermochemical Quantities

The two important quantities to be determined are f' and M_{IK} . There are several convenient paths of integration from a hydrostatic state, where these quantities can be determined with standard thermodynamic methods, to the actual stressed state. We begin with the differential of f'

$$df' = T_{ij}dE_{ij} - s'd\theta + \rho'_0 \sum M_{IK}dc_i. \quad (4.10)$$

The function ϕ' , defined by a Legendre transform

$$\phi' = f' - T_{ij}E_{ij} \quad (4.11)$$

proves to be useful. Its differential

$$d\phi' = -E_{ij}dT_{ij} - s'd\theta + \rho'_0 \sum M_{IK}dc_i \quad (4.12)$$

permits us to deduce the following Maxwell relation

$$-\rho'_0(\partial M_{IK}/\partial T_{ij})_{c_j} = (\partial E_{ij}/\partial c_{IK})_{T_{kl}}. \quad (4.13)$$

Hooke's law at constant composition is

$$T_{ij} = C_{ijkl}(E_{kl} - E_{kl}^c) \quad (4.14)$$

or

$$E_{ij} = E_{ij}^c + S_{ijkl}T_{kl} \quad (4.15)$$

where the C_{ijkl} are moduli of elasticity, and the S_{ijkl} compliances. Both are composition and temperature dependent. From (4.15) we deduce

$$\left(\frac{\partial E_{ij}}{\partial c_{IK}}\right)_{T_{kl}} = \left(\frac{\partial E_{ij}^c}{\partial c_{IK}}\right) + \left(\frac{\partial S_{ijkl}}{\partial c_{IK}}\right)T_{kl}. \quad (4.16)$$

Chemical potentials are assumed known at a hydrostatic pressure P , and composition c_1, c_2, \dots

$$M_{IK}(P, c_1, c_2, \dots) = \mu_I(P, c_1, c_2, \dots) - \mu_K(P, c_1, c_2, \dots). \quad (4.17)$$

It is customary to define standard chemical potentials μ_i^0 and activity coefficients such that

$$\mu_i(P, c) = \mu_i^0(P) + R\theta \ln \gamma_i c_i \quad (4.18)$$

where γ_i is chosen for convenience. Depending on the problem, it is chosen to approach 1 either for dilute or concentrated solution. Vacancy potentials also are fit to this convention. Since $\mu_v(P, \bar{c}_v) = 0$, where \bar{c}_v is the equilibrium vacancy concentration at P ,

$$\mu_v^0(P) = -R\theta \ln \gamma_v \bar{c}_v \quad (4.19)$$

where γ_v is the vacancy activity coefficient. If it is constant, the chemical potential of vacancies under pressure P can also be written

$$\mu_v(P, c_v) = R\theta \ln(c_v/\bar{c}_v) \quad (4.20)$$

The expressions for the chemical potentials are introduced into eq (4.13) and the resulting expression integrated along a constant composition path to the stress T_{ij} . For a binary solution

$$M_{12}(T_{ij}, c) = \mu_1^0(P) - \mu_2^0(P) + R\theta \ln \frac{\gamma_1 c}{\gamma_2(1-c)} - V'_0 \eta_{ij} T_{ij} - \frac{V'_0}{2} \frac{dS_{ijkl}}{dc} T_{ij} T_{kl} + \frac{V'_0}{2} \frac{dS_{jjkk}}{dc} P^2 - V'_0 \eta_{kk} P. \quad (4.21)$$

If the solid is isotropic, this expression becomes

$$M_{12}(T_{ij}, c) = \mu_1^0(P) - \mu_2^0(P) + R\theta \ln \frac{\gamma_1 c}{\gamma_2(1-c)} - V'_0 \eta T_{kk} + \frac{V'_0}{2} \frac{d}{dc} \left(\frac{\nu}{E}\right) (T_{kk})^2 - \frac{V'_0}{2} \frac{d}{dc} \left(\frac{1+\nu}{E}\right) T_{ij} T_{ij} - 3V'_0 \eta P + \frac{3V'_0}{2} \frac{d}{dc} \left(\frac{1-2\nu}{E}\right) P^2. \quad (4.22)$$

These expressions contain terms both linear and quadratic in stress. They simplify considerably when the elastic coefficients are not composition dependent. Equation (4.22) for instance becomes

$$M_{12}(T_{ij}, c) = \mu_1^0(P) - \mu_2^0(P) + R\theta \ln \frac{\gamma_1 c}{\gamma_2(1-c)} - V'_0 \eta (T_{kk} + 3P). \quad (4.23)$$

To obtain f' we calculate ϕ' with eq (4.12). It is first integrated along a path of constant composition, from pressure P to stress T_{ij} . Using Hooke's law (4.15), this gives

$$\begin{aligned} \phi'(T_{ij},c) - \phi'(P,c) = & -\frac{1}{2}S_{ijkl}T_{ij}T_{kl} - E_{kk}^c T_{kl} \\ & + \frac{1}{2}S_{jjkk}P^2 - E_{kk}^c P \end{aligned} \quad (4.24)$$

and using (4.11)

$$f'(T_{ij},c) - f'(P,c) = \frac{1}{2}S_{ijkl}T_{ij}T_{kl} - \frac{1}{2}S_{jjkk}P^2 \quad (4.25)$$

Since under hydrostatic stress, the familiar liquid thermodynamics is valid, the Helmholtz free energy $f'(P,c)$ is known. It may be obtained from the more commonly tabulated molar Gibbs free energy G_m by subtracting PV_0 and dividing by V_0' . This gives

$$f'(P,c) = \rho_0' G_m - P \rho_0' / \rho_0. \quad (4.26)$$

Since

$$\rho_0' / \rho_0 = 1 + E_{kk} = 1 + E_{kk}^c - S_{jjkk}P \quad (4.27)$$

one obtains, after replacement of G_m by its value as a function of composition

$$\begin{aligned} f'(P,c) = & \rho_0' \{ c [\mu_1^0(P) + R\theta \ln \gamma_1 c] + (1-c) [\mu_2^0(P) \\ & + R\theta \ln \gamma_2 (1-c)] \} - P(1 + E_{kk}^c) + S_{jjkk}P^2. \end{aligned} \quad (4.28)$$

Combination of (4.25) and (4.28) gives the final result

$$\begin{aligned} f'(T_{ij},c) = & \rho_0' \{ c [\mu_1^0(P) + R\theta \ln \gamma_1 c] + (1-c) [\mu_2^0(P) \\ & + R\theta \ln \gamma_2 (1-c)] \} - P(1 + E_{kk}^c) \\ & + \frac{1}{2} S_{ijkl} T_{ij} T_{kl} - \frac{1}{2} S_{jjkk} P^2. \end{aligned} \quad (4.29)$$

For an isotropic solid, this relation becomes

$$\begin{aligned} f'(T_{ij},c) = & \rho_0' \{ c [\mu_1^0(P) + R\theta \ln \gamma_1 c] + (1-c) [\mu_2^0(P) \\ & + R\theta \ln \gamma_2 (1-c)] \} - P(1 + 3k) - \frac{\nu}{2E} (T_{kk})^2 \\ & + \frac{1+\nu}{2E} T_{ij} T_{ij} - \frac{3(1-2\nu)}{2E} P^2. \end{aligned} \quad (4.30)$$

Because it always appears in the boundary conditions, the expression for the quantity $V_0' f' - M_{12} c$ is useful. Combining (4.22) and (4.30) we get, in the isotropic case

$$\begin{aligned} V_0' f' - M_{12} c = & \mu_2^0(P) + R\theta \ln \gamma_2 (1-c) + V_0' \left[-P(1 + 3k) \right. \\ & \left. - \frac{\nu}{2E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} + c \eta (T_{kk} + 3P) - \frac{3(1-2\nu)}{2E} P^2 \right] \end{aligned}$$

$$\begin{aligned} -\frac{c}{2} \frac{d}{dc} \left(\frac{\nu}{E} \right) (T_{kk})^2 + \frac{c}{2} \frac{d}{dc} \left(\frac{1+\nu}{E} \right) T_{ij} T_{ij} \\ - \frac{3c}{2} \frac{d}{dc} \left(\frac{1-2\nu}{E} \right) P^2 \end{aligned} \quad (4.31)$$

When the elastic coefficients are not composition dependent, this becomes

$$\begin{aligned} V_0' f' - M_{12} c = & \mu_2^0(P) + R\theta \ln \gamma_2 (1-c) + V_0' \left[-P(1 + 3k) \right. \\ & \left. - \frac{\nu}{2E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} - \frac{3(1-2\nu)}{2E} P^2 \right. \\ & \left. + c \eta (T_{kk} + 3P) \right]. \end{aligned} \quad (4.32)$$

In a crystal of arbitrary symmetry, this expression is

$$\begin{aligned} V_0' f' - M_{12} c = & \mu_2^0(P) + R\theta \ln \gamma_2 (1-c) + V_0' \left[-P(1 + E_{kk}^c) \right. \\ & \left. + \frac{1}{2} S_{ijkl} T_{ij} T_{kl} - \frac{1}{2} S_{jjkk} P^2 + c \eta_{ij} T_{ij} \right. \\ & \left. + \frac{c}{2} \frac{dS_{ijkl}}{dc} T_{ij} T_{kl} + c \eta_{kk} P - \frac{c}{2} \frac{dS_{jjkk}}{dc} P^2 \right]. \end{aligned} \quad (4.33)$$

Expressions (4.21) to (4.23) apply to substitutional binary solutions. For interstitial binary solutions the integration along a constant composition path from the hydrostatic stress to the stress T_{ij} using (4.13) gives the elastic terms identical to those in (4.21) to (4.23). Because there is no network constraint or interstitial concentration we use (3.12) for M_1 and obtain for dilute interstitial solutions

$$\begin{aligned} M_1(T_{ij}/c) = & \mu_0'(P) + R\theta \ln \gamma_1 c - V_0' \eta_{ij} T_{ij} \\ & - \frac{V_0'}{2} \frac{dS_{ijkl}}{dc} T_{ij} T_{kl} - V_0' \eta_{kk} P + \frac{V_0'}{2} \frac{dS_{jjkk}}{dc} P^2. \end{aligned} \quad (4.34)$$

Equations for the special cases of isotropy and constant elastic coefficients are like (4.34) except that the elastic terms take the forms they have in (4.22) and (4.23). We will see in section 5.7 that there is no need to distinguish between interstitial and substitutional solutions. Had we chosen the vacancy on the interstitial site as component 2 we could have obtained (4.34) directly from (4.21) by noting that $\mu_2^0 = 0$ for the vacancy.

5. Internal Equilibrium

The study of internal equilibrium requires the simultaneous solution of the equations of elasticity and those of chemical equilibrium. The method we have found useful

recognizes that the strain is a function of stress and composition. But the composition at equilibrium with a given diffusion potential is determined by the local stress alone. Thus the strain at a given diffusion potential is a function of stress alone. If we obtain this stress-strain function, we can solve these problems as if they were ordinary elastic problems, without any further regard to chemical problems whose effects are now implicitly accounted for.

There are several derivations. The simplest and most easily generalized for large strains and nonlinear effects parallels in its first steps the thermodynamic methods used to derive the relationships between isotropic (adiabatic) and isothermal elasticity. In the first section we review the main results and then apply them to various problems.

5.1 Open-System Elastic Constants

After a straightforward manipulation of partial derivatives, the following expression, valid for a two-component solid, is obtained (Appendix 3)

$$\left(\frac{\partial E_{ij}}{\partial T_{kl}}\right)_{M_{12}} = \left(\frac{\partial E_{ij}}{\partial T_{kl}}\right)_c + V'_0 \eta_{ij} \eta_{kl} / \left(\frac{\partial M_{12}}{\partial c}\right)_{T_{mn}} \quad (5.1)$$

Making the usual small strain approximations, and an expansion of the strain around $T_{ij}=0$ produces the constant chemical-potential form of Hooke's law

$$E_{ij} = S_{ijkl}^* T_{kl} \quad (5.2)$$

The coefficients of the stress have been called open-system compliances, S^* and are related to the constant composition compliances S by

$$S_{ijkl}^* = S_{ijkl} + V'_0 \eta_{ij} \eta_{kl} / \left(\frac{\partial M_{12}}{\partial c}\right)_{T_{mn}} \quad (5.3)$$

where $(\partial M_{12}/\partial c)_{T_{mn}}$ is evaluated at $T_{mn}=0$ and where all the quantities except V'_0 are functions of c . The second order terms that are neglected in this expansion have been discussed [15]. Introducing the notation

$$1/\chi = \rho'_0 \left(\frac{\partial M_{12}}{\partial c}\right)_{T_{ij}=0} \quad (5.4)$$

i.e.,

$$1/\chi = \frac{\rho'_0 R \theta}{c} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c}\right) \quad (5.5)$$

for interstitial solutions, and

$$1/\chi = \frac{\rho'_0 R \theta}{c(1-c)} \left(1 + \frac{\partial \ln \gamma_1}{\partial \ln c}\right)$$

for substitutional binary solutions, the open system compliances, for isotropic solids are given by

$$\begin{aligned} E^* &= E/(1 + \chi \eta^2 E) \\ \nu^* &= (\nu - \chi \eta^2 E)/(1 + \chi \eta^2 E) \\ (K^{-1})^* &= 3(1 - 2\nu^*)/E^* = K^{-1} + 9\chi \eta^2 \end{aligned} \quad (5.6)$$

$$G^* = G$$

where K is the bulk modulus and G the shear modulus.

Far away from spinodals and critical points, the expression (5.3) is not very sensitive to the composition. It is then appropriate to use the values of the open-system constants, at a composition near the average composition of the specimen. The elastic coefficients become constants, and the elastic part of the problem is now independent of the compositional part. For a closed system, the obvious choice is the average composition. For a system that is in contact with a chemical reservoir, the composition at equilibrium under zero stress is usually a good choice. In the case of a very high average stress, the equilibrium composition at some high pressure may be more appropriate. With such replacement of the composition in (5.3) or (5.4) to (5.6), all the solutions of ordinary linear elasticity become directly applicable to elasto-chemical problems.

5.2 Finding the Composition Field

Finally, even though we have eliminated the composition to solve the elastochemical problem, the composition field is easily obtained from the solution. At constant diffusion potential, composition is uniquely determined by the local stress. For a binary for example (4.21) can be solved for the composition

$$\frac{\gamma_1 c}{\gamma_2 (1-c)} = \text{constant} \times \exp[\text{elastic terms}/R\theta] \quad (5.7)$$

$$\text{where constant} = \exp[\{M_{12} - (\mu_1^0 - \mu_2^0)\}/R\theta]. \quad (5.8)$$

A useful linearized version of eq (5.7) is obtained by linearizing the elastic terms of that equation or of (4.21) to (4.23) and differentiating at constant M_{12} , P , and θ . Using (5.5) this gives

$$dc/\chi = \eta_{ij} dT_{ij} \quad (5.9)$$

or

$$c = c_0 + \chi \eta_{ij} T_{ij} \quad (5.10)$$

where c_0 is a constant of integration and is the com-

position that an element of unstressed solid would have if it were in equilibrium with the system.

For the isotropic case this becomes

$$c - c_0 = \chi \eta T_{kk} \quad (5.11)$$

Had we linearized about a hydrostatic pressure P the result would have been

$$c - c(P) = \chi \eta (T_{kk} + 3P) \quad (5.12)$$

There are several ways of evaluating the constants in (5.8) or (5.10), but basically they are all methods of evaluating M_{12} at equilibrium. If the system is in contact with a materials reservoir with specified M_{12} the answer is straightforward. If it is equilibrated with a fluid phase, eq (3.10) applies. If the composition and stress are specified at some point in the system, eq (4.21) can be used. This occurs in some problems where almost all of the solid acts as a reservoir in the sense that most of it is homogeneous in composition and stress, and that transfer of components to small inhomogeneously stressed parts of the system hardly affects the composition of the homogeneous part.

For the typical case of a closed heterogeneous system the overall composition is specified. At equilibrium the diffusion potentials become a constant whose value must be determined as part of the solution. This is a standard procedure in the method of Lagrange multipliers. Equation (5.7) is a one-parameter family of composition profiles. For each assumed value of the parameter M_{12} , we can determine the overall composition by integration. The one that satisfies the specified composition is the solution and this fixes M_{12} .

This procedure is simplified if linearization of (5.7) to give (5.10) is valid. Using this to obtain c_0 from which we can obtain M_{12} . We use the conservation of mass for the entire solid of total volume Ω' in the reference state and average composition \bar{c}

$$\int_{\Omega'} c dV' = \Omega' \bar{c}. \quad (5.13)$$

Substituting (5.10) we obtain

$$c_0 = \bar{c} - \frac{\chi \eta_{ij}}{\Omega'} \int_{\Omega'} T_{ij} dV' \quad (5.14)$$

which can be substituted into (4.21) to (4.23) to obtain M_{12} . Once c_0 is known we have the composition profile of the inhomogeneously stressed system

$$c - \bar{c} = \chi \eta_{ij} (T_{ij} - \frac{1}{\Omega'} \int_{\Omega'} T_{ij} dV') \quad (5.15)$$

or

$$c - \bar{c} = \chi \eta_{ij} (T_{ij} - \bar{T}_{ij})$$

where \bar{T}_{ij} is a component of the volume averaged stress, and χ and η_{ij} are evaluated at c . This is the linearized equation for composition in a closed system.

5.3 Internal Equilibrium of Vacancies

We consider a single component solid with vacancies as the second component. If, as is often assumed [18], there is no relaxation around a single vacancy at any level of applied stress and the elastic constants do not depend on vacancy concentrations, the diffusion potential M_{v1} , given by eq (4.23), is a function of composition only. Therefore a constant diffusion potential would imply a vacancy composition field that is constant regardless of the stress distribution. Even with these assumptions we will later see (sec. 6.2) that the local equilibrium vacancy concentration at the interface does depend on stress at the interface.

A more realistic model assumes relaxation. Let the partial molar volume of vacancies differ from the molar volume of the species. If the elastic constants do not depend on vacancy concentration, eq (4.23) yields with $P=0$

$$M_{v1} = M_{v1}^0 + R\theta \ln \frac{c_v}{1-c_v} - (\bar{V}_v - \bar{V}_1) T_{kk} / 3 \quad (5.16)$$

At equilibrium, this is constant, leading to a vacancy concentration field given by (with $c_v \ll 1$)

$$c_v = \bar{c}_v \exp\left(\frac{\bar{V}_v - \bar{V}_1}{3R\theta} T_{kk}\right) \quad (5.17)$$

where \bar{c}_v is the equilibrium concentration of vacancies at $P=0$.

5.4 Dislocation Atmospheres

5.4.1 Atmosphere Around a Dislocation in an Isotropic Solid

Let us consider a substitutional two-component infinite isotropic solid, with a negligible concentration of vacancies. A straight edge dislocation with a Burgers vector of magnitude b is located in the solid along the z axis. If the sizes of components 1 and 2 are different, there will be a segregation around the dislocation. This problem has been solved, considering one of the atoms as a defect [19]. This means that its concentration has to be relatively small. Indeed in many cases only vacancies or interstitials are considered. These are unnecessary restrictions as we shall see.

Far from the dislocation, the solid is at composition c_0 , and is stress-free. Therefore we can think of this far-away solid as a chemical reservoir. The solid with the dislocation and its atmosphere has the same diffusion potential as the stress-free solid at c_0 . For convenience, we choose the solid at c_0 as the reference for strain. Since we have shown that under small strain approximation, the elastic part of the problem is equivalent to a constant composition problem with the open-system elastic coefficients, eq (5.6), the stress field, with the atmosphere present, is given by

$$\begin{aligned} T_{rr} = T_{\phi\phi} &= \frac{-Gb\sin\phi}{2\pi(1-\nu^*)r} \\ T_{r\phi} &= \frac{Gb\cos\phi}{2\pi(1-\nu^*)r} \\ T_{zz} &= \frac{-Gb\nu^*\sin\phi}{\pi(1-\nu^*)r} \end{aligned} \quad (5.18)$$

and the composition field is, to a first approximation, using eqs (5.11) and (5.18)

$$\Delta c = -\chi\eta \frac{(1+\nu^*)Gb\sin\phi}{(1-\nu^*)\pi r}. \quad (5.19)$$

(These equations correct an algebraic error in reference [6].) Replacing the open systems constant by their values, we finally obtain

$$\begin{aligned} T_{rr} = T_{\phi\phi} &= \frac{-Gb(1+\chi\eta^2 E)\sin\phi}{2\pi(1-\nu+2\chi\eta^2 E)r} \\ T_{r\phi} &= \frac{Gb(1+\chi\eta^2 E)\cos\phi}{2\pi(1-\nu+2\chi\eta^2 E)r} \\ T_{zz} &= \frac{-Gb(\nu-\chi\eta^2 E)\sin\phi}{\pi(1-\nu+2\chi\eta^2 E)r} \\ \Delta c &= \frac{-\chi\eta(1+\nu)Gb\sin\phi}{\pi(1-\nu+2\chi\eta^2 E)r} \end{aligned} \quad (5.20)$$

where the subscript 0 has been dropped from all the variables since all of them have to be evaluated at composition c_0 , including the Burgers vector magnitude. In our case (substitutional solution), χ is given by eq (5.5) and η by (4.7) and (4.8).

We first note that, since χ is positive for a stable solid solution, the stresses are decreased, by a fraction of the order of $\chi\eta^2 E$. This factor tends to zero for highly dilute solutions. But for a concentrated solution, it can be significant. Taking an ideal solution, $c_0=0.5$, $\rho_0=10^5 \text{ mol m}^{-3}$, $R\theta=10^4 \text{ J mol}^{-1}$, $E=10^{11} \text{ Nm}^{-2}$, and $\eta=0.1$ gives a value of 0.25 for $\chi\eta^2 E$. This change in the stress field,

which is readily obtained here, has, to our knowledge, not been calculated within the framework of the defects model.

At low concentration, the following approximation holds

$$\chi \approx c_0 V_0' / R\theta$$

and

$$\chi\eta = \frac{c_0(\bar{V}_1 - \bar{V}_2)}{3R\theta}$$

and we can neglect $2\chi\eta^2 E$ in comparison to $(1-\nu)$ obtaining thereby the classical point-defect solution

$$\Delta c \approx \frac{-c_0(\bar{V}_1 - \bar{V}_2)(1+\nu)Gb\sin\phi}{3\pi R\theta(1-\nu)r}.$$

But it is to be emphasized that the composition eq (5.7) can be solved exactly by numerical methods. Our result is more general in that it includes in a self-consistent way all the interactions that may be present, specifically in concentrated solutions, between the defects themselves and the defects and the matrix. In particular, it takes into account the nonideality of the solid solutions in a phenomenological way that is model independent. If no measured value is available for the activity coefficient function γ_1 , specific statistical mechanical models [20-22] can of course be used and the result directly introduced in the value of χ .

5.4.2 Dislocation Atmosphere in a Cubic Crystal

Analytic expressions are rarely known for the elastic fields caused by point-forces in a medium of arbitrary symmetry [23]. Hence the usual integral methods for calculating atmospheres cannot be used. On the other hand the introduction of open system compliances is not restricted to isotropic solids, and formulas have been developed for the most general elastic solids [15]. Because the elastic field has been found for several cases of dislocations in these non-isotropic single-component crystals, the concept is most valuable.

By a simple substitution of the open-system elastic coefficients, the same elastic calculations are valid for solid solutions equilibrated to constant diffusion potentials. The composition fields are given to first order by eq (5.10) or more exactly from the solution of eq (5.7). We shall treat the case of a [111] screw dislocation in a cubic crystal. The x_3 axis is along the dislocation, the x_2 axis is along [110] and x_1 along [112]. The stress field has been given by Steeds [24]. Because the equations are rather long, we shall derive only the composition field.

In cubic crystals, the change in composition with stress is given to first order by

$$\Delta c = \chi \eta T_{kk} \quad (5.21)$$

as for the isotropic case. At constant composition, T_{kk} has the value

$$T_{kk} = \frac{Gb \delta s_{44} \sin 3\phi}{4\sqrt{2}\pi r (1 - \delta \cos^3 3\phi)(1 - \delta)^{1/2} S (3s_{11} - 2S)} \quad (5.22)$$

with

$$S = s_{11} - s_{12} - s_{44}/2$$

a factor which is zero for isotropic crystals,

$$\delta = \frac{2S^2}{9(s_{11} + s_{44} + 5S/6)}$$

and the s_{ij} are the standard two indices compliances, referred to the cube axis. For cubic crystals, the open system compliances are

$$s_{ij}^* = s_{ij} + \chi \eta^2 \quad i \text{ and } j < 3 \quad (5.23)$$

$$s_{ij}^* = s_{ij} \quad i \text{ and } j > 3$$

therefore

$$S^* = S$$

and

$$\delta^* = \frac{2S^2}{9(s_{11}^* + s_{44} + 5S/6)}$$

Combining (5.21), (5.22), and (5.23), we obtain the composition field

$$\Delta c = \frac{9\chi\eta b \delta^* s_{44} \sin 3\phi}{4\sqrt{2}\pi r (1 - \delta^* \cos^3 3\phi)(1 - \delta^*)^{1/2} S (3s_{11}^* - 2S)} \quad (5.24)$$

where all the constants that depend on the material have to be taken at c_0 , the composition far away from the dislocation. This result, obtained by a simple algebraic manipulation, has, to our knowledge, never been obtained by other methods.

5.4.3 Dislocation Atmospheres: Nonlinear Effects

At constant diffusion potentials, when the composition changes from the unstressed to the stressed state are small, we have shown that the strain is linearly

related to the stress, as in the usual theory of elasticity. But this law has a smaller range of applicability than in the constant composition case. The thermodynamics of solutions introduce nonlinear terms in the stress-strain law. When the strain is expanded as a function of stress, we have identified four second-order effects [15]: (a) non-linear stress-strain laws at constant composition, due, for instance, to rearrangement of interstitial atoms into sites that become nonequivalent under stress; (b) change of compliances with composition; (c) deviation from Vegard's law; and (d) non-linearity of the solution thermodynamics. The first two effects have been considered within the framework of defects theories. It does not seem that the two others have been treated [25]. Since solutions of non-linear elastic problems have been found [26], they can be used, with the second-order open-system compliances, to find second-order effects on dislocation atmospheres.

5.5 Internal Equilibrium of a Binary Substitutional Solid With Vacancies

We have seen in section 4 that, for a binary substitutional solid with vacancies, in equilibrium with a fluid, the following is true

$$M_{1v} = \mu_1^f \quad (5.25)$$

$$M_{2v} = \mu_2^f \quad (5.26)$$

where μ_1^f and μ_2^f are the chemical potentials of species 1 and 2 in the fluid. It seemed therefore rather natural to use these equations, which have the same form as those for fluid equilibrium, rather than the mathematically equivalent

$$M_{12} = \mu_1^f - \mu_2^f \quad (5.27)$$

$$M_{v2} = -\mu_2^f. \quad (5.28)$$

From a theoretical point of view, there is no difference. Although these equations are valid for nonlinear inhomogeneous and anisotropic solids, we give as an example expressions for constant elastic coefficients and isotropy

$$M_{1v} = M_{1v}^0 + R\theta \ln \frac{\gamma_1 c_1}{\gamma_v c_v} - \frac{\bar{V}_1 - \bar{V}_v}{3V_0'} T_{kk} \quad (5.29)$$

$$M_{12} = M_{12}^0 + R\theta \ln \frac{\gamma_1 c_1}{\gamma_2 c_2} - \frac{\bar{V}_1 - \bar{V}_2}{3V_0'} T_{kk}. \quad (5.30)$$

The concentration of vacancies is small compared to c_1 and c_2 . Measurement of c_v , γ_v , and \bar{V}_v are therefore sub

ject to potentially large errors. These affect eqs (5.25), (5.26), and (5.28) but not (5.27). For computational purposes, it is then better to use the second formulation. Besides, if we are only interested in the composition c_1 and c_2 , we can neglect the vacancies and use only eq (5.30) for equilibrium calculations. By keeping the flexibility of choice for the dependent substitutional species, we can eliminate species whose concentration has been found to have a negligible effect on the chemical behavior of the solid solutions, including vacancies, even if they are essential to the mechanisms by which chemical equilibrium is attained.

5.6 Multisite Solids

Up to this point, we have focused our attention on crystalline solids that are most common in the metallurgical world, where there is only one substitutional site, that is highly occupied, and an interstitial site that is lightly occupied. But in many instances crystals have several non-equivalent sites, occupied by mixed species of atoms or molecules or vacancies. The fraction of empty sites can vary for each type of site from 0 to 1. In the description we can of course eliminate sites that are and remain empty. They don't contribute to the energy or entropy of the system. For all other sites, we can describe their status by the densities of the atoms and the densities of vacancies on each of them. As for the substitutional site with which we have been dealing in the preceding section, there will be a constraint condition: the total density of atoms and vacancies is constant for each site. Using the method described in section 4, it can be shown that at equilibrium, the diffusion potentials are constant, equal on all sites, and equal to the corresponding difference in chemical potentials when equilibrated with a fluid

$$M_{IK}^1 = M_{IK}^2 \dots = M_{IK}^v = \mu_I^L - \mu_K^L \quad (5.31)$$

where the superscripts label the different sites. There are cases where there is no species K that is present on all sites, or where it is not convenient to use the same K-species for all sites. The formulas can easily be transformed, using eqs (3.8) and (3.9)

$$M_{IK} + M_{KJ} = M_{IJ} \quad (5.32)$$

If a species is not present on one site, it cannot be used as the dependent species on that site, and its diffusion potential equation drops from the set of eqs (5.31). The vacancies are to be considered as a species, since an exchange of an i -site vacancy for a j -site vacancy pro-

duces no change of state, exactly as the exchange of a K atom on an i -site with a K atom on a j -site.

Equations (5.31) govern the equilibrium partitioning of I atoms on the different sites. If only the total density is of interest, one can interpret eqs (5.31) differently. They state that along an equilibrium path, the Helmholtz free energy density is only a function of the total density of the $(N-1)$ independent species.⁴ Calling M_{IK} the common value of the diffusion potential for each site, we have

$$df'' = s' d\theta + \sum M_{IK} dp'_i \quad (5.33)$$

Equation (5.33) shows that the formulas developed in the preceding section can also be applied, with the total density of each species as composition variables (or the ratio ρ'_i/ρ'_0 , ρ'_0 being a chosen total density, like the total density of sites, or the density of sites I, $(I=1, \dots, v)$ whatever is most useful).

In the equations used in section 5, the interstitial site was sparsely occupied, and we used eq (4.34) for the diffusion potential of this species. But rigorously its diffusion potential is M_{1v} , where v represents the vacancies on interstitial sites

$$M_{1v} = M_{1v}^0 + R\theta \ln \frac{\gamma_1 c_1}{\gamma_v c_v} + \text{elastic terms} \quad (5.34)$$

If there are v interstitial sites per substitutional site, $\gamma_v c_v$ tends to one as c_v tends to v . Therefore, in dilute interstitial solutions

$$M_{1v} \approx \mu_1^0 + R\theta \ln \gamma_1 c_1 + \text{elastic terms} \quad (5.35)$$

which is the expression we have used. In almost all cases, site occupancy is either high or low. Phase transformations occur before intermediate occupancy is reached. But hydrogen in metals is an important case where the occupancy can span all the possible composition field without a phase change. In such cases, the rigorous diffusion potential has to be used. Equations for the internal equilibrium between sites have been given, with the preceding approximation by Li et al. [27]. It is clear that there is no need to make the distinction between interstitial and substitutional atoms. A single formalism with multisite occupation is possible and avoids the confusion that can arise if a specie occupies both substitutional and interstitial sites [7]. For most metal-

⁴When a function $F(x_1, x_2, \dots, x_n)$ is such that, for all values of the x_i ,

$$\partial F / \partial x_1 = \partial F / \partial x_2 = \dots = \partial F / \partial x_n$$

then F is a function only of the sum $(x_1 + x_2 + \dots + x_n)$.

lurgical examples, species do seem to occupy only one site.

We next turn to phase change equilibrium at solid-fluid interfaces. The case of a stoichiometric compound already illustrates the principal features. Let species A completely occupy a equivalent sites α per unit cell, species B b equivalent sites β , etc. Because there is only one species on each site we cannot define a diffusion potential. In the liquid each species has a well defined chemical potential. The equation for equilibrium is

$$f - (a\mu_A^L + b\mu_B^L + c\mu_C^L \dots)\rho_0 = -P \quad (5.36)$$

where ρ_0 is the total density of sites in a unit cell. This is a straightforward expression of chemical equilibrium for the dissolution of the compound $A_aB_bC_c \dots$, which continues to hold under stress. It is Gibbs' eq (393) [9] since he quite clearly considered solids to be compounds (CP) and defined a single chemical potential μ^{CP} for them in the fluid even if they dissociated

$$\mu^{CP} = a\mu_A + b\mu_B + c\mu_C + \dots \quad (5.37)$$

In defining μ^{CP} there is a rigid adherence to a law of definite proportions dictated by the numbers of equivalent sites fully occupied in the crystal structure.

If we now let the α sites be occupied by several species I, J, K including vacancies we obtain diffusion potentials. Choosing species K as the counterspecies the equilibrium equation is

$$f - \rho_0 \sum M_{IK} c_i^\alpha - \rho_0 (a\mu_K^L + b\mu_B^L \dots) = -P. \quad (5.38)$$

The term in the parenthesis is the chemical potential for the stoichiometric compound $K_aB_bC_c \dots$. There are obvious advantages to choosing K to be the major species on site α . If site α is a lightly occupied interstitial site the compound is $B_bC_c \dots$ and μ_K is set to zero.

If several sites are each occupied by more than one species the equations are not changed if a different species is chosen as counter species for each site. If the same species is chosen as counter species of several sites the terms combine. In particular if the same counter species K is used for all sites we obtain

$$f - \rho_0 \sum_I \sum_\alpha M_{IK} c_i^\alpha - (a+b+c+\dots)\mu_K \rho_0 = -P. \quad (5.39)$$

Summing over all sites we obtain

$$f - \rho_0 \sum M_{IK} c_i - (a+b+c+\dots)\mu_K \rho_0 = -P. \quad (5.40)$$

This is identical with eq (3.15) if we redefine ρ_0 in terms of atom site density instead of unit cell densities.

6. Interface Equilibria

In this section we illustrate various aspects of equilibria involving three kinds of interfaces that stressed solids can have but ignoring capillary effects. Most of our examples will be uniformly stressed, and have only as many components as are necessary to illustrate the points to be made. When the solid is multicomponent and nonuniformly stressed, the interior equilibria can be solved by the methods of the open-system elastic constants of the previous section. This converts a multicomponent elastic and thermochemical problem into an elastic problem alone, although possibly a nonlinear one.

6.1 Change of Solubility With Stress

Our first example is a Gibbs solid—a pure substance for instance—in equilibrium at pressure P with a fluid in which it can dissolve along a flat interface. Forces are applied to the solid so that its state of stress is now T_{ij} . To maintain mechanical equilibrium, one of the principal values of T_{ij} is $-P$, and the corresponding principal direction of stress is normal to the fluid-solid interface. What is the change in the chemical potential of the fluid necessary to keep the system in chemical equilibrium? The only equation, besides mechanical equation, is the boundary conditions, eq (3.18) which becomes for a one component linear elastic solid

$$f' - \mu^L \rho_0' = -P(1 + E_{kk}). \quad (6.1)$$

Following Gibbs [9, p. 196], we compare this equilibrium with that of the same solid phase equilibrated under hydrostatic stress with the same fluid. Using bars to indicate the values of the thermodynamic quantities in this equilibrium we write

$$\bar{f}' - \bar{\mu}^L \rho_0' = -P(1 + \bar{E}_{kk}). \quad (6.2)$$

Subtracting these two equations, we obtain

$$f' - \bar{f}' + P(E_{kk} - \bar{E}_{kk}) = \rho_0'(\mu^L - \bar{\mu}^L) \quad (6.3)$$

$(f' - \bar{f}')$ is the elastic energy stored in the solid on going from pressure P to stress state T_{ij} and $P(E_{kk} - \bar{E}_{kk})$ is the work done on the solid by the liquid. The l.h.s. of eq (6.3) is thus the work that has to be done to bring a hydrostatically stressed solid to the nonhydrostatic state while surrounded by the liquid. It is necessarily positive, and the fluid in equilibrium with a nonhydrostatically stressed solid is always supersaturated with respect to precipitating a hydrostatically stressed solid by the

amount given in (6.3). If we let c_L and \bar{c}_L be the concentration of the solid component in the fluid in equilibrium with respect to the nonhydrostatically and hydrostatically stressed solid, we can use eq (4.32) to obtain

$$\rho'_0 R \theta \ln(\gamma_L c_L / \bar{\gamma}_L \bar{c}_L) = -\frac{\nu}{2E}(T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} + \frac{3(1-2\nu)}{2E} P^2 + \frac{1-2\nu}{E} T_{kk} P. \quad (6.4)$$

Let t_1 , t_2 , and $-P$ be the principal values of stress. If the change in solubility is small, and the solution is dilute or ideal, we get

$$\frac{c_L - \bar{c}_L}{\bar{c}_L} = \frac{1}{2\rho'_0 R \theta E} \times [t_1^2 + t_2^2 - 2\nu t_1 t_2 + 2(1-\nu)(t_1 + t_2 + P)P]. \quad (6.5)$$

Because $-1 < \nu < 1/2$, the right hand side of eq (6.5) is positive, except of course when $t_1 = t_2 = -P$, where it is zero. The solubility of the solid in the liquid is always increased when a stress is applied to the solid. The solution is supersaturated with respect to a hydrostatically stressed solid at pressure P , a classical result that was derived by Gibbs.

We now turn to the case of a two-component solid in equilibrium with a melt. We have two conditions for equilibrium

$$f' - \mu_1^L \rho'_1 - \mu_2^L \rho'_2 = -P(1 + E_{kk}) \quad (6.6)$$

$$M_{12} = \mu_1^L - \mu_2^L. \quad (6.7)$$

We compare again to the equilibrium of the solid with the fluid under pressure P .

$$\bar{f} - \bar{\mu}_1^L \bar{\rho}'_1 - \bar{\mu}_2^L \bar{\rho}'_2 = -P(1 + \bar{E}_{kk}) \quad (6.8)$$

$$\bar{M}_{12} = \bar{\mu}_1^L - \bar{\mu}_2^L. \quad (6.9)$$

Subtraction of (6.8) from (6.6) and (6.9) from (6.7) gives two equations for the change of composition in the fluid and the solid to maintain equilibrium under stress.

Assuming for simplicity (i) $P=0$, (ii) terminal solutions (i.e., both solid and liquid are dilute solutions), (iii) no change in elastic coefficients with composition, we get

$$R \theta \ln\left(\frac{1-c}{1-\bar{c}}\right) + V'_0 \left[-\frac{1}{2} \frac{\nu}{E} (t_1 + t_2)^2 + \frac{1+\nu}{2E} (t_1^2 + t_2^2) + c \eta(t_1 + t_2) \right] = R \theta \ln\left(\frac{1-c_L}{1-\bar{c}_L}\right) \quad (6.10)$$

$$R \theta \ln\left(\frac{c(1-\bar{c})}{\bar{c}(1-c)}\right) - V'_0 \eta(t_1 + t_2) = R \theta \ln\left(\frac{c_L(1-\bar{c}_L)}{\bar{c}_L(1-c_L)}\right). \quad (6.11)$$

As usual, this system of equations can be solved numerically, or, if the changes are small, we can linearize the equations and solve with Cramer's rule.

6.2 Vacancies Equilibrium in a One-Component Solid

Consider a cylinder of isotropic hydrostatically stressed solid in contact with a fluid in which it cannot dissolve at pressure P , with an equilibrium concentration of vacancies \bar{c}_v . A load is applied that produces a stress whose components are T_{zz} , $T_{rr} = T_{\theta\theta}$. We want to calculate the equilibrium concentration of vacancies along the surfaces S_1 and S_2 . Since the components of the solid don't appear in the fluid, there is no equation like (3.12). But the phase change eq (3.15) applies, and in this case since μ_K is identified with $\mu_v^L = 0$, the equation becomes

$$V'_0 f' - (1-c_v) M_{1v} = -P V'_0 (1 + E_{kk}) \quad (6.12)$$

where $-P$ is the normal traction. Let us first adopt Herring's simplifying assumptions that (a) there is no volume relaxation around vacancies, (b) there is no change in elastic constants with vacancy concentration, and (c) the solid obeys the law of dilute solutions. Using (4.32) we get (i) under pressure \bar{P}

$$\mu_v^0(\bar{P}) + R \theta \ln \bar{c}_v = 0 \quad (6.13)$$

(ii) under stress, along S_2

$$\begin{aligned} \mu_v^0(\bar{P}) + R \theta \ln c_v^z + V'_0 \left[-\bar{P} - \frac{1}{2} \frac{\nu}{E} (2T_{rr} + T_{zz})^2 + \frac{1+\nu}{2E} (2T_{rr}^2 + T_{zz}^2) + \frac{3(1-2\nu)}{2E} \bar{P}^2 \right] \\ = V'_0 T_{zz} \left[1 + \frac{1-2\nu}{E} (2T_{rr} + T_{zz}) \right] \end{aligned} \quad (6.14)$$

(iii) under stress, along S_1

$$\begin{aligned} \mu_v^0(\bar{P}) + R \theta \ln c_v^r + V'_0 \left[-\bar{P} - \frac{1}{2} \frac{\nu}{E} (2T_{rr} + T_{zz})^2 + \frac{1+\nu}{2E} (2T_{rr}^2 + T_{zz}^2) + \frac{3(1-2\nu)}{2E} \bar{P}^2 \right] \\ = V'_0 T_{rr} \left[1 + \frac{1-2\nu}{E} (2T_{rr} + T_{zz}) \right]. \end{aligned} \quad (6.15)$$

It is quite clear that c'_v and c^z_v are different, unless $T_{rr} = T_{zz}$, i.e., when the system is under hydrostatic stress. Since we have assumed no relaxation around vacancies, $\eta = 0$, and therefore according to eq (4.23), M_{1v} is different on S_z and S_r . As a result, a vacancy flux will appear. This is further discussed in section 8.4.

Making the further assumption that $P = 0$, and neglecting quadratic terms in stress, subtraction of (6.13) from (6.14) and (6.15) gives

$$\ln(c'_v/\bar{c}_v) = V'_0 T_{rr} / R \theta \quad (6.16)$$

$$\ln(c^z_v/\bar{c}_v) = V'_0 T_{zz} / R \theta.$$

This is Herring's [18, 28] well-known formula: to first order in stress, only the normal pressure affects the equilibrium vacancy concentration at an interface. We will get the same results, whether this interface is a solid-fluid interface or an incoherent solid-solid interface.

The order of magnitude of the quadratic terms can be easily obtained by making $T_{rr} = 0$ so that linear terms disappear in (6.15). We obtain, along S_r

$$\ln(c'_v/\bar{c}_v) = V'_0 T_{zz}^2 / 2ER \theta. \quad (6.17)$$

Within the small strain approximation, this effect is less than 1% of the effect on S_z . But there are cases where it might be significant (cf. sec. 8.4).

Conditions (a), (b), and (c) can easily be removed through the use of the general formulas developed in section 4. As an example we treat the case where there is a volume relaxation around a vacancy. Using (4.32), assuming $P = 0$, and following the above procedure, we get, to first order in stress

$$\ln(c'_v/\bar{c}_v) = \frac{V'_0}{R \theta} [T_{rr} - \eta(2T_{rr} + T_{zz})] \quad (6.18)$$

$$\ln(c^z_v/\bar{c}_v) = \frac{V'_0}{R \theta} [T_{zz} - \eta(2T_{rr} + T_{zz})]. \quad (6.19)$$

The corrective term, proportional to η , contains the trace of the stress tensor. As such other components than the normal pressure influence the vacancy concentration at a particular interface, if elastic relaxation around vacancies are taken into account.

6.3 Using Open-System Elastic Constants for Multicomponent Phase Equilibrium

For the general multicomponent phase-equilibrium under stress, the fact that the M_{JK} are constant gives $(N-1)$ relationships between stress and composition. As shown earlier, it is possible to solve these equations for

composition as a function of stress and obtain the strain E_{ij}^s that results from composition changes. The result is a stress-strain relation at constant M_{JK} . This relationship was used to solve elastic problems within a single phase as if it were composed of a single component.

These same relationships apply to each individual phase in a multiphase equilibrium, but the phase change boundary conditions of section 3.5 contain a similar coupling between stress and composition. In the present section we shall demonstrate that by using open-system-elastic constants, the compositional part of these equations can also be eliminated. In fact this method allows us to treat multicomponent equilibrium as if each phase were a one-component purely elastic part of the system, and that for such a solid, the ω function is equal to the elastic energy apart from a constant (cf. eq (3.16)). Finally once the elastic problem has been solved, the composition field is obtained by the methods of section 5.2.

We will use as an example binary isotropic linear solids, although the proof can be made for a multicomponent anisotropic system. We shall further assume constant elastic coefficients, and that, at zero stress and potential M_{12} , the composition is c . Let Δc be the change of composition due to a change of stress. Expanding f' around the unstressed state we find using (3.6) and (5.4)

$$f'(T_{ij}, c + \Delta c) = f'(0, c) + \rho'_0 M_{12} \Delta c + (\Delta c)^2 / \chi - \frac{\nu}{2E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij}. \quad (6.20)$$

Let us consider the function

$$f'^* = f'(0, c) - \frac{\nu^*}{2E^*} (T_{kk})^2 + \frac{1+\nu^*}{2E^*} T_{ij} T_{ij} \quad (6.21)$$

where we have added to the free energy of the solid under zero stress and at potential M_{12} , an elastic energy computed with open-system elastic constants at M_{12} . Replacing these constants by their values (5.6) we obtain

$$f'^* = f'(0, c) + \frac{1+\nu}{2E} T_{ij} T_{ij} - \frac{\nu}{2E} (T_{kk})^2 + \frac{1}{2} \chi \eta^2 (T_{kk})^2. \quad (6.22)$$

But the change in composition Δc is given by (5.11) so that (6.22) can be written

$$f'^* = f'(0, c) + \frac{1+\nu}{2E} T_{ij} T_{ij} - \frac{\nu}{2E} (T_{kk})^2 + (\Delta c)^2 / 2\chi. \quad (6.23)$$

The function $[f' - \rho'_0(c + \Delta c)M_{12}]$ that appears repeatedly in the phase change boundary equations (cf. (3.24) and (3.27)) is thus obtained as

$$f' - \rho'_0(c + \Delta c)M_{12} = f'^* - \rho'_0 c M_{12} \quad (6.24)$$

Or, if we replace M_{12} and $f'(0,c)$

$$f' - \rho'_0(c + \Delta c)M_{12} = -\frac{\nu^*}{2E^*}(T_{kk})^2 + \frac{1 + \nu^*}{2E^*}T_{ij}T_{ij} - \rho'_0\mu_2(0,c). \quad (6.25)$$

Thus the various phase change boundary conditions are expressed in terms of an open-system Helmholtz free energy for each phase. This free energy has the same form as a Helmholtz free energy of a one-component phase. Its elastic constants are the open-system elastic constants of section 5.1. The reference state of each phase is the unstressed multicomponent phase with the same value of M . Its composition is c in (6.24) and (6.25), its lattice parameter is used to define strain, and its constant composition elastic constants are to be used in eqs (5.3) or (5.6) to calculate the open-system constants.

By examination of (6.25), we can see that the use of these open-system constants allows us to treat, as far as the stress is concerned, any multicomponent system just as if it were a one-component system. Thus elastic solutions developed for one component inclusions, for instance [23], can now be used for similar multicomponent inclusions.

After finding the stress field, the results of section 5.2 can be used to obtain the composition field.

An interesting consequence of the preceding results occurs in a binary system in which both phases have the same conventional elastic constants. In an infinite single component system the Bitter-Crum theorem [16] holds. There is no elastic interaction between particles. The system is degenerate with respect to particle shape and dispersion. In a binary system if the χ or η 's differ, the open system elastic constants would differ even if the conventional elastic constants did not. As a result there is now elastic interaction between particles that is entirely the result of the compliance due to composition changes.

7. Partial Equilibrium—Local Equilibrium

When the general conditions for equilibrium are not satisfied, the system will tend to equilibrium. The rates of various processes are usually so different that in the time scale of an experiment we may often assume that some processes have reached equilibrium while others have not occurred at all. In this section we briefly discuss these partial equilibria. When processes are too fast for thermal and chemical relaxation, we obtain the results of classical adiabatic elasticity. The relation between isothermal constant composition elastic coeffi-

cients S_{ijkl}^{θ} and adiabatic elastic coefficients S_{ijkl}^s is a well known thermodynamic result [17]

$$S_{ijkl}^{\theta} = S_{ijkl}^s + \alpha_{ij}\alpha_{kl}\left(\frac{\partial\theta}{\partial s}\right)_{T_{ij}} = S_{ijkl}^s + \alpha_{ij}\alpha_{kl}\theta/C^T \quad (7.1)$$

α_{ij} is the thermal expansion coefficient, and C^T the heat capacity, both at constant stress.

When thermal and elastic equilibration occur but without diffusion or interface motion, we have classical isothermal elasticity. Comparing eqs (5.3) and (7.1) we note that they are quite similar except that temperatures instead of compositional derivatives are used. Thus the relationship between adiabatic, isothermal, and open-system elastic constants is one of increasing equilibration first with thermal and then with materials reservoirs.

Diffusion of some species, e.g., interstitials, often is orders of magnitude faster than that of other species. Such a partial equilibrium, called paraequilibrium [29], is often reached in phase transformations of multicomponent alloys. Only hydrostatic cases seem to have been treated. When stresses are important the modification from corresponding binary interstitial alloy problems seems straightforward.

Interface processes, crystal growth or dissolution and grain growth all involve network modification processes that may be quite slow. Grain boundary sliding may not occur. For calculation of such partial equilibria, the corresponding equilibrium equations must be suppressed. Polycrystalline averages of the properties can be used to obtain corresponding averages for stress and composition fields.

The most common partial equilibrium occurs when all processes except diffusion have relaxed to equilibrium. The only suppressed condition is that M_{1K} need be constant, but M_{1K} remains continuous across all interfaces that have reached equilibrium. This partial equilibrium is called local equilibrium at interfaces.

Many experiments are done under conditions where partial equilibrium is maintained while some or all of the remaining variables are observed while they relax to equilibrium. The laws of most of the relaxation processes have been studied. Interface relaxation is complicated and often nonlinear. On the other hand, heat flow in response to thermal gradients is coupled with elasticity and constitutes the subject of thermoelasticity. Diffusion in response to nonuniformity of the M_{1K} is also well understood, regardless of whether the origin of the gradients in M_{1K} are from composition gradients, stress gradients or interface conditions. The next section ex

amines a set of problems involving diffusional equilibration under isothermal conditions with local equilibrium assumed.

8. Diffusional Kinetics and Creep

Many problems of diffusion involve stress. In diffusional creep the applied stress is the motivating force for the diffusion. Compositional heterogeneity results in a self-stress that affects diffusion in a way that is too often ignored in the diffusion calculation. As we have seen, stress affects the diffusion potential and interface equilibrium conditions. It has an effect both on the rate and direction of the diffusional flux within each grain and on the boundary conditions to the diffusion equations at each interface.

Often only some of the effects of stress have been considered, or approximations have been made that ignored effects of the same order or larger than the effects considered. In this section we will examine the effects of stress on diffusion and creep, inside the grains and at interfaces, and with both applied stresses and the self-stresses that arise from the compositional inhomogeneity.

We begin with a formulation for multicomponent diffusion that is consistent with our thermodynamic formulation and has the proper invariances with respect to arbitrary choices of the species K . We then examine problems of inhomogeneous stress when the network is unaltered. Much of this was the subject of a recent overview [30] in which a hierarchy of increasingly difficult problems was discussed. We next turn our attention to diffusional network alteration phenomena, such as creep and phase change, both under applied stress and self-stress. Because of the importance of vacancies in this problem, interesting phenomena occur even in one-component systems. We reformulate and simplify the general equations to examine a few problems of diffusional creep in a one-component system with vacancies.

8.1 Multicomponent Diffusion in Isothermal Network Solids

As shown in [31] the invariant formulation of substitutional multicomponent diffusion flux J_I in an isothermal isotropic or cubic network solid⁵ is given by

$$-J_I = \sum_{j=1}^N B_{IJ} \text{grad } M_{JK} \quad I=1, \dots, N. \quad (8.1)$$

⁵The reference geometry for diffusion is usually the unstressed state. With the notation we have used, the fluxes should be noted with a prime. Since there is no confusion possible, we shall drop it here.

B_{IJ} is a mobility, function of composition and stress at a given temperature. It has been shown that the B_{IJ} are independent of the choice of the species K . There are $(N-1)$ chemical species plus vacancies. There are $(2N-1)$ independent network restrictions on the B_{IJ}

$$\sum_I B_{IJ} = 0 \quad J=1, \dots, N \quad (8.2)$$

$$\sum_J B_{IJ} = 0 \quad I=1, \dots, N. \quad (8.3)$$

As a result there are $(N-1)^2$ independent coefficients which is the expected number of phenomenological coefficients for the diffusion of $(N-1)$ interacting species without a network constraint. It is also the number expected for $(N-1)$ interstitial species. For a one-component solid with vacancies there is only one term

$$J_1 = -J_v = B_{v1} \text{grad } M_{v1}. \quad (8.4)$$

Similarly for the diffusion of a single interstitial species there is one term

$$-J_1 = B_1 \text{grad } M_1. \quad (8.5)$$

For a two-component substitutional solution there are four independent B . With vacancies as the K species the M_{v} terms disappear and we have

$$-J_1 = B_{11} \text{grad } M_{1v} + B_{12} \text{grad } M_{2v} \quad (8.6)$$

$$-J_2 = B_{21} \text{grad } M_{1v} + B_{22} \text{grad } M_{2v}$$

$$-J_v = B_{v1} \text{grad } M_{1v} + B_{v2} \text{grad } M_{2v}$$

with the restrictions that

$$B_{11} + B_{21} + B_{v1} = 0$$

$$B_{12} + B_{22} + B_{v2} = 0.$$

Using species 2 as the K species we have the same coefficients in different combinations with the diffusion potential M

$$-J_1 = B_{11} \text{grad } M_{12} + B_{1v} \text{grad } M_{v2} \quad (8.7)$$

$$-J_2 = B_{21} \text{grad } M_{12} + B_{2v} \text{grad } M_{v2}$$

$$-J_v = B_{v1} \text{grad } M_{12} + B_{vv} \text{grad } M_{v2}.$$

The knowledge that B remains the same in various formulations should permit flexibility both in gathering data and in formulating and applications.

Stress affects both B and M in the flux equations. B is affected by the level of stress alone. We expand about a stress state which can be either zero

$$B_{JKij} = B_{JKij}^0(c, \theta) + B_{JKijkl}^1(c, \theta)T_{kl} + \dots \quad (8.8)$$

or some other convenient state T^0

$$B_{JKij}(c, \theta, T) = B_{JKij}^0(c, \theta, T^0) + B_{JKijkl}^1(c, \theta, T^0)(T_{kl} - T_{kl}^0). \quad (8.9)$$

The gradient of M depends on the stress and the stress gradient. From the Maxwell eq (4.13) the coefficient of the stress gradient is the strain produced by a unit composition change

$$\left(\frac{\partial M_{JK}}{\partial T_{ij}}\right) = -V_0' \left(\frac{\partial E_{ij}}{\partial c_{JK}}\right)_{T_{kl}} \quad (8.10)$$

which is precisely defined and readily estimated from lattice parameter-composition data. For cubic or isotropic cases

$$\partial M_{JK} / \partial T_{ij} = -V_0' \eta_{JK} \delta_{ij} \quad (8.11)$$

and

$$\nabla M_{JK} = R \theta [(\nabla c_J / c_J) - (\nabla c_K / c_K)] - V_0' \eta_{JK} \nabla (tr T). \quad (8.12)$$

Strictly this should be at the actual stress, but in most cases data for unstressed crystals should be adequate, and lead to a linear formulation. Combining (8.1) with (8.12) and retaining only terms linear in T we obtain for cubic or isotropic cases

$$-J_I = -A_I (\nabla tr T) + \rho_0' \sum_{J \neq K} D_{IJ(K)} \nabla c_J. \quad (8.13)$$

The factor ρ_0' needs to be introduced since the c are defined to be dimensionless rather than molar densities, where

$$A_I = V_0' \sum_J B_{IJ} \eta_{JK} \\ D_{IJ(K)} = V_0' R \theta B_{IJ} \left(\frac{1}{c_J} + \frac{1}{c_K}\right) \quad (8.14)$$

Because diffusion fluxes and gradients are independent of the choice of K , A_I and the B_{IJ} can be shown also to be independent of that choice, but to be consistent the $D_{IJ(K)}$ must depend on the choice in the way shown in (8.14). To avoid large uncertainties in the $D_{IJ(K)}$ it is again clearly advantageous to choose K to be the major species, rather than vacancies.

8.2 Diffusion Without Network Changes

Conservation of matter is expressed by the equation

$$\rho_0' \frac{\partial c_I}{\partial t} + \text{div} J_I = 0. \quad (8.15)$$

Compositional heterogeneity produces a long-range stress field and changing compositions change this field. Since stress and stress gradients affect B and M , the stress and diffusion equations have to be solved simultaneously. It has been common to ignore this mutual interaction and to study either the stress resulting from diffusion or the effect of stress on diffusion alone. When the ignored effects are small, this is valid, but for most cases it is not.

A straightforward technique for solving the stress and diffusion equations has been developed [30]. As in section 5 the relationship between elastic stress and an arbitrary composition field often remains solvable and can be used to eliminate stress from the diffusion equation. Plastic stress accommodation would render this technique invalid.

A hierarchy of increasingly complicated problems was examined for cases of diffusion in binary alloys in which there was no applied stress. All stress was due to compositional heterogeneity alone.

The mutual interaction in most cases is a major factor. In the case of spinodal decomposition, it can change the sign of the diffusional flux and is responsible for the metastability between the chemical and coherent spinodal [32]. The stress effect is so long ranged that compositional heterogeneity can affect diffusion elsewhere. Fick's law which states that the flux depends only on local gradients is often not valid. Because this stress effect is proportional to the local concentration it can be neglected in dilute solutions.

Interface boundary conditions for diffusion in interstitial solutions have been examined for cases in which the network is chemically inactive. The boundary condition is a simple continuity of M at a fixed location in the reference state. It depends on the level of stress at the boundary. For local equilibrium eq (5.7) is applicable.

8.3 Diffusion with Self-stress and Phase-change at the Boundary

In our previous work [30] on the effect of self-stress on diffusion the network was conserved at the boundary. There are many metallurgical problems, such as diffusion controlled phase growth, where the network is not conserved, but where equilibrium prevails at the interface. This equilibrium is governed by eq (5.7) and a phase-change equation that depends on the nature of the boundary.

Self-stress is what we call the stress that is the result of sample heterogeneity. Generally its value at a point is a function of the composition distribution everywhere. For special geometries its value becomes a simple expression involving principally the local composition,

and the effects of self-stress on the thermodynamic variables can be expressed in terms of the local composition only reducing self-stress problems to composition problems.

One such geometry is the semi-infinite solid with concentration fields that are functions only of the distance from the surface. We will consider the case of a semi-infinite couple, with diffusion in α and β , and an incoherent boundary. Under pressure P , the equilibrium compositions are \bar{c}^α and \bar{c}^β . When diffusion takes place, the compositions are c_0^α and c_0^β far away from the boundary, and \tilde{c}^α and \tilde{c}^β at the boundary (fig. 1). We shall further assume, for simplicity, that the pressure P is zero, and that the diffusing sample is under zero external pressure. This implies that the tractions are zero at the α - β boundary. We also assume no change of elastic constant with composition for either phase. Under these hypotheses, the mechanical equilibrium at the interface, eq (3.25), is always fulfilled. Equations (5.7) and (3.24) become, using (4.22) and (4.32)

$$\begin{aligned} & \mu_1^{0\alpha} - \mu_2^{0\alpha} + R\theta \ln \frac{\tilde{\gamma}_1^\alpha \bar{c}^\alpha}{\tilde{\gamma}_2^\alpha (1 - \bar{c}^\alpha)} - V_0'^\alpha \eta^\alpha T_{kk}^\alpha \\ & = \mu_1^{0\beta} - \mu_2^{0\beta} + R\theta \ln \frac{\tilde{\gamma}_1^\beta \bar{c}^\beta}{\tilde{\gamma}_2^\beta (1 - \bar{c}^\beta)} - V_0'^\beta \eta^\beta T_{kk}^\beta \end{aligned} \quad (8.16)$$

and

$$\mu_2^{0\alpha} + R\theta \ln [\tilde{\gamma}_2^\alpha (1 - \bar{c}^\alpha)] + V_0'^\alpha \left[-\frac{1}{2} \frac{\nu^\alpha}{E^\alpha} (T_{kk}^\alpha)^2 \right]$$

$$\begin{aligned} & + \frac{1 + \nu^\alpha}{2E^\alpha} T_{ij}^\alpha T_{ij}^\alpha + \tilde{c}^\alpha \eta^\alpha T_{kk}^\alpha \Big] \\ & = \mu_2^{0\beta} + R\theta \ln [\tilde{\gamma}_2^\beta (1 - \bar{c}^\beta)] \\ & + V_0'^\beta \left[-\frac{1}{2} \frac{\nu^\beta}{E^\beta} (T_{kk}^\beta)^2 + \frac{1 + \nu^\beta}{2E^\beta} T_{ij}^\beta T_{ij}^\beta + \tilde{c}^\beta \eta^\beta T_{kk}^\beta \right]. \end{aligned} \quad (8.17)$$

At equilibrium under zero pressure, these equations become

$$\begin{aligned} \mu_1^{0\alpha} - \mu_2^{0\alpha} + R\theta \ln \frac{\tilde{\gamma}_1^\alpha \bar{c}^\alpha}{\tilde{\gamma}_2^\alpha (1 - \bar{c}^\alpha)} & = \mu_1^{0\beta} - \mu_2^{0\beta} \\ & + R\theta \ln \frac{\tilde{\gamma}_1^\beta \bar{c}^\beta}{\tilde{\gamma}_2^\beta (1 - \bar{c}^\beta)} \end{aligned} \quad (8.18)$$

$$\mu_2^{0\alpha} + R\theta \ln [\tilde{\gamma}_2^\alpha (1 - \bar{c}^\alpha)] = \mu_2^{0\beta} + R\theta \ln [\tilde{\gamma}_2^\beta (1 - \bar{c}^\beta)]. \quad (8.19)$$

We first have to find the stress field. In a half-space specimen, we have found [30] that its trace depends only on the local composition

$$T_{kk}^\alpha = -2Y^\alpha \eta^\alpha (\bar{c}^\alpha - c_0^\alpha) \quad (8.20)$$

$$T_{kk}^\beta = -2Y^\beta \eta^\beta (\bar{c}^\beta - c_0^\beta). \quad (8.21)$$

Where $Y = E/(1 - \nu)$. Introducing these values in (8.16) and (8.17), and after subtraction of (8.18) from (8.16) and (8.19) from (8.17), we obtain the system of

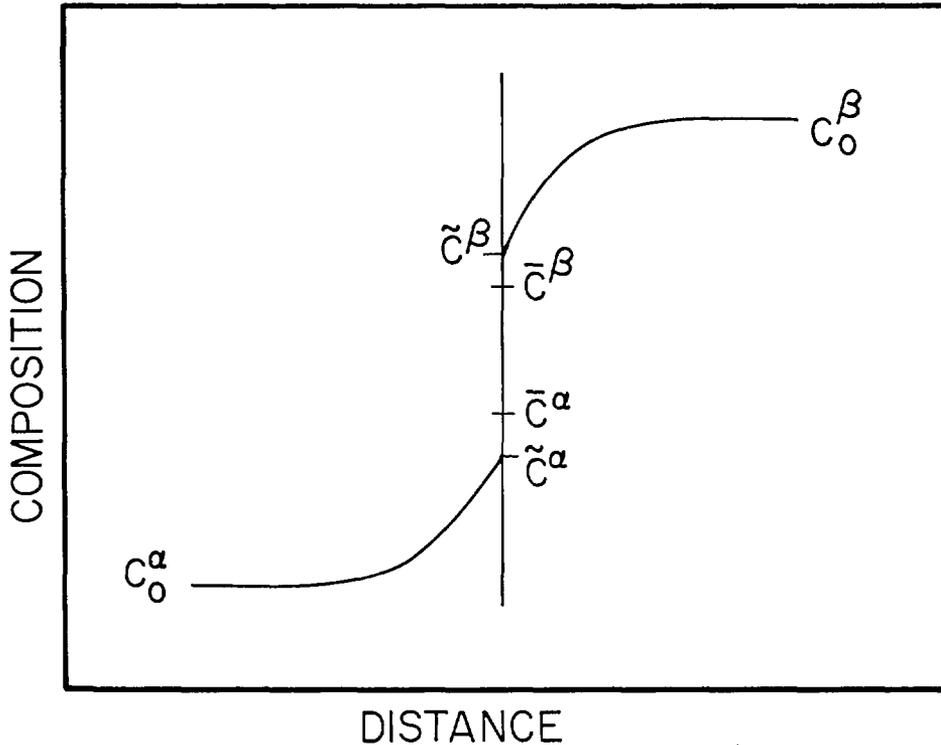


Figure 1—Compositions in a self-stressed diffusion couple with an incoherent interface. The compositions far away from the interface are c_0^α and c_0^β . The self-stress generated by the composition gradient has shifted the equilibrium composition at the boundary to \tilde{c}^α , \tilde{c}^β from their unstressed phase diagram values of \bar{c}^α , \bar{c}^β .

equations to solve for \bar{c}^α and \bar{c}^β . As we have seen before, it can be solved numerically or, if $(\bar{c}^\alpha - c_0^\alpha)$ and $(\bar{c}^\beta - c_0^\beta)$ are small, it can be linearized, and the resulting system of equations solved by Cramer's rule.

Under the assumption that there is no normal stress across the α - β interface, a common tangent construction is possible (see Appendix 4 for the demonstration). To the Helmholtz free energy per mole we have to add the elastic energy per mole, which is just a function of the local composition. Its value is

$$\hat{f}_{el} = \frac{V_0' E}{1-\nu} \eta^2 (\bar{c}^\alpha - c_0^\alpha)^2 \quad (8.22)$$

where V_0' is the molar volume at composition c_0 . The construction is shown in figure 2. This type of construction has been used by Hillert [33] for the case of massive transformation, in which it is proper to assume that the phase which is forming is homogeneous, and by Purdy et al. [34] for diffusion-induced grain boundary migration.

8.4 Effect of Vacancies: General Formulation

When vacancies, in addition to providing a mechanism for diffusion, also interact with the stress, and provide a means for creating or destroying network at

an interface, new phenomena appear, in particular diffusional creep. In this section, we consider only one-component systems, where these effects are not obscured by the phenomena previously described in this chapter. We first formulate the creep as a boundary value problem and then turn our attention to specific creep problems.

The Partial Differential Equation

The flux of vacancies \underline{J} is given by

$$-J_i = B_{ij}(M_{v1})_{,j} \quad (8.23)$$

where B_{ij} is a tensor function of the temperature θ , c , (the concentration of vacancies) and the stress. An expansion around $T=0$ gives

$$B_{ij} = B_{ij}^0(c, \theta) + B_{ijkl}^1(c, \theta) T_{kl} + \dots \quad (8.24)$$

The coefficient of order 0 is given by

$$B_{ij}^0 = D_{ij} c_v (1 - c_v) / R \theta \rho_0' \quad (8.25)$$

where D_{ij} is the self-diffusion matrix. Usually it is not very much dependent on the vacancy concentration.

The tensors B_{ij}^0 and B_{ij}^1 being properties of a crystalline material follow the rules of crystalline sym-

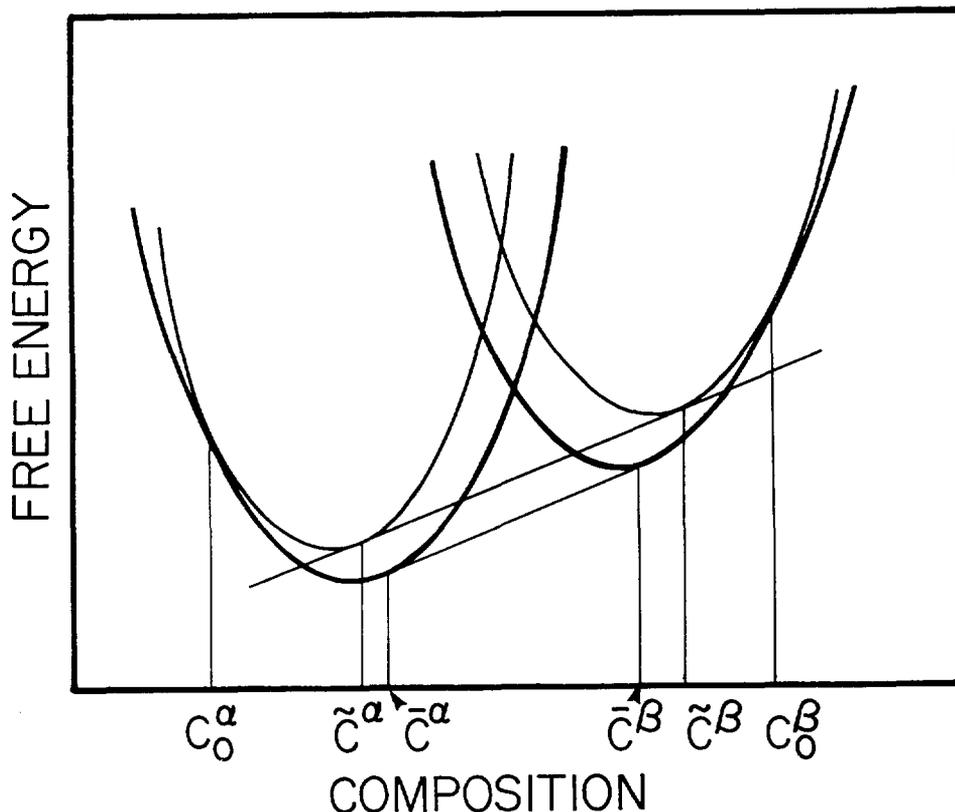


Figure 2—Common tangent construction that gives the composition of figure 1. The unstressed free energies (heavy lines) are shifted by an amount equal to the elastic energy $V_0' E \eta^2 (c - c_0)^2 / (1 - \nu)$ to give the light curves. The common tangent construction gives \bar{c}^α and \bar{c}^β .

metries. For isotropic materials

$$B_{ij}^0 = B^0 \delta_{ij} \quad (8.26)$$

and

$$B^0 = c_v(1 - c_v)D/R\theta\rho'_0 \quad (8.27)$$

The tensor B_{ijkl}^1 has the same form as an elastic tensor for an isotropic material

$$B_{ijkl}^1 T_{kl} = \beta T_{kk} \delta_{ij} + \gamma T_{ij} \quad (8.28)$$

where β and γ are two constants. This equation reveals that if the tensor B_{ij} is stress dependent, it introduces a stress-coupled anisotropy in an otherwise isotropic diffusion coefficient.

Neglecting second order effects in stress in M_{vi} , that is assuming that the elastic coefficients do not depend on vacancy concentration, the gradient of the vacancies diffusion potential can be written

$$(M_{vi})_{,j} = \frac{R\theta}{c_v(1-c_v)} \left[1 + \frac{\partial \ln \gamma_v}{\partial \ln c_v} \right] (c_v)_{,j} - V'_0 \eta_{kl} T_{kl,j} \quad (8.29)$$

If dilute solution laws apply, this equation simplifies into

$$(M_{vi})_{,j} = \frac{R\theta}{c_v} (c_v)_{,j} - V'_0 \eta_{kl} T_{kl,j} \quad (8.30)$$

which, for isotropic material becomes

$$\nabla M_{vi} = (R\theta/c_v) \nabla c_v - V'_0 \eta \nabla (tr T). \quad (8.31)$$

The conservation equation is expressed as usual

$$\rho'_0 \frac{\partial c_v}{\partial t} + J_{i,i} = s \rho'_0 \quad (8.32)$$

The source and sink terms, which is the number of vacancies created per unit volume, come, for instance, from the vacancy source at a moving dislocation. The complete diffusion equation for vacancies is obtained by combining [8.23] with [8.32]

$$\rho'_0 \frac{\partial c_v}{\partial t} = \rho'_0 s + [B_{ij}(M_{vi})_{,j}]_{,i} \quad (8.33)$$

In an isotropic solution, one gets

$$\frac{\partial c_v}{\partial t} = s + D \nabla_{c_v}^2 - \frac{DV'_0 \eta}{R\theta} \nabla_{c_v} \cdot \nabla T_{kk} - \frac{c_v DV'_0 \eta}{R\theta} \nabla^2 T_{kk} \quad (8.34)$$

where we have neglected the stress dependence of B_{ij} . When the relaxation of the lattice around a vacancy can be neglected, the last two terms of the r.h.s. disappear, and one obtains the simple equation

$$\frac{\partial c_v}{\partial t} = s + D \nabla^2 c_v \quad (8.35)$$

Initial Conditions. The initial conditions consist in a given vacancy concentration field. For steady state, these conditions are not needed. They are unimportant at long times, as long as a steady state can be reached.

Boundary Conditions. The boundary conditions depend of course on the problem that is treated. The most useful seems to be given by an equilibrium condition along all surfaces of the solid. Written for an isotropic solid, constant elastic coefficients, a reference pressure $P=0$ (with an equilibrium vacancy concentration c_v), dilute solution behavior, and a reference composition $c_v=0$ for strain, this reads (eqs (3.18) and (4.31))

$$\begin{aligned} \mu_v^0(o) + R\theta \ln c_v = -PV'_0 \left(1 + \frac{1-2\nu}{E} T_{kk} + 3c_v \eta_v \right) \\ - V'_0 \left[-\frac{1}{2} \frac{\nu}{E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} - (1-c_v) \eta_v T_{kk} \right] \end{aligned} \quad (8.36)$$

or

$$\begin{aligned} R\theta \ln(c_v/\bar{c}_v) = -PV'_0 \left(1 + \frac{1-2\nu}{E} T_{kk} + 3c_v \eta_v \right) \\ - V'_0 \left[-\frac{1}{2} \frac{\nu}{E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} - (1-c_v) \eta_v T_{kk} \right]. \end{aligned} \quad (8.37)$$

Since $c_v \ll 1$, these equations can be simplified into

$$\begin{aligned} \mu_v^0(o) + R\theta \ln c_v = -PV'_0 \left(1 + \frac{1-2\nu}{E} T_{kk} \right) \\ - V'_0 \left[-\frac{1}{2} \frac{\nu}{E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} - \eta_v T_{kk} \right]. \end{aligned} \quad (8.38)$$

Because it is the dominant term linear in stress, the r.h.s. is usually $-PV'_0$. Only this term was taken into account in Herring's theory of diffusional creep. We shall see in the next section cases where the quadratic terms are important for new effects.

Network modification along the surfaces due to the vacancy flux is simply given by

$$n'_i \left(\frac{\partial x'_i}{\partial t} + V_0 J_i \right) = 0 \quad (8.39)$$

where the x'_i are the coordinates of a point of the interface. This equation tells us that the shape of the specimen changes as diffusion takes place, due to the vacancy creation and annihilation at the surfaces.

Stress Equilibrium. Up to now we have been concerned

with the diffusion equation. Stress equilibrium in this quasi-static model obeys the partial differential equation (3.13)

$$T_{ij,j} = 0 \quad (8.40)$$

with proper boundary conditions. In most problems they will be given in terms of tractions along the surface. It is important to note that, because of the network modifications there, they are specified on a changing (and usually unknown) surface.

To specify the problem fully in term of stress, we need the Beltrami-Mitchell equations [11,30]. For isotropic materials, the expression is

$$(1 + \nu)T_{ij,kk} + T_{kk,ij} + E\eta \left[\frac{1+\nu}{1-\nu} \delta_{ij}(c_v)_{,kk} + (c_v)_{,ij} \right] = 0. \quad (8.41)$$

8.5 Some Creep Problems

8.5.1 Herring's Classical Problems: Diffusional Viscosity of a Polycrystalline Solid

Let us first show that with Herring's assumptions and approximations [18] the equations presented in section 8.4 become identical to his starting equations. Only steady state is considered. There is no volume change associated with a vacancy (i.e., the average volume of a vacancy is equal to the atomic volume). This implies $\eta = 0$; therefore the interactions between stress and composition appear only in the boundary condition pertaining to network modification. Furthermore all terms nonlinear in stress are neglected, and the reference pressure is zero. The solution of atoms and vacancies is ideal (i.e., there is no interaction with vacancies and their concentration is very small). Finally, there is no source term within a grain.

With these approximations, the diffusion eq (8.33) becomes

$$\nabla^2 M_{vi} = 0. \quad (8.42)$$

The expression for the diffusion potential is

$$M_{1v} = \mu_1^0(0) - \mu_v^0(0) + R\theta \ln[(1 - c_v)/c_v] \quad (8.43)$$

and the boundary condition (8.29) becomes

$$\mu_v^0(0) + R\theta \ln c_v = -PV'_0. \quad (8.44)$$

Subtracting (8.44) from (8.43), and neglecting $\ln(1 - c_v)$, one gets

$$M_{1v} = \mu_1^0 + PV'_0. \quad (8.45)$$

This is the boundary condition used by Herring (his eq (2)) for the partial differential eq (8.42) since our P equals his $-P_{zz}$. The stress equilibrium equation is the same, and he implicitly used condition (8.37) to get the rate of displacement of the interface (e.g., to go from (3) to (4) in his paper). Thus within the assumptions explicitly spelled out at the beginning of this section, we recover Herring's equations and boundary conditions.

His solutions combined a mean field (the average of the stress tensor within a grain is equal to the applied stress) and a perturbation analysis (the shape of the grain does not change as diffusion proceeds).

The formulation of the problem with fewer assumptions is possible using the equations of the previous section which contains important additional terms in the diffusion eq (8.33) and boundary conditions (8.29). We next explore a few problems chosen to illustrate the physical consequences of these additional terms.

8.5.2 Quadratic Effects

Usually the linear term of the r.h.s. of (8.36) is the dominant one, but, whenever the specimen surfaces are all immersed in a fluid of constant pressure, this term is constant and at steady state does not contribute any gradient. Under these conditions the higher order terms are the only ones present. We consider two examples in which we approximate condition for which P is constant over the surfaces of interest.

The first treated by Roitburd [35] is a pore in a specimen under uniaxial stress in which he examined the shape change by vacancy fluxes that redistributed material around the pore. Other vacancy sinks and sources were assumed so far away that fluxes between them and pores could be neglected. Because P in the pore is constant, the effects depend entirely on the quadratic terms. The result of the calculation is that a spherical pore will distort to an oblate spheroid with the minor axis along the stress axis. Because this conclusion arises from quadratic terms the same result is obtained regardless of whether the specimen is under tension or compression.

A closely related problem is a long single crystal rod of nonuniform cross section under a uniaxial load applied at the ends. If the characteristic length of the nonuniformities is short compared to the specimen length, we may examine the redistribution of material along the lateral surfaces by vacancy flux and ignore the fluxes between these surfaces and the specimen ends. Along the surface P is again constant. If we assume $\eta_v = 0$ and that the elastic constants are independent of c_v , (8.36) becomes

$$\mu_v^0(0) + R\theta \ln c_v = -V'_0 \left[-\frac{1}{2} \frac{\nu}{E} (T_{kk})^2 + \frac{1+\nu}{2E} T_{ij} T_{ij} \right]. \quad (8.46)$$

The r.h.s. is minus the elastic energy of the solid. Let us note that the rod is unstable to necking. A small indentation (or any change in cross section) will produce a higher stress at its root (or at the minimum cross section). Vacancy flux will remove material from the root (or at minimum cross section) and deposit it nearby at a place of lowered elastic energy. The rod is unstable to necking by diffusion creep regardless of whether it is under tension or compression. This is the same result as Roitburd's pore, which can be considered an internal notch.

This counterintuitive result is consistent with thermodynamics. Consider the work done by the loading system, applied force times distance moved. The compliance of a rod with nonuniform cross section increases if the rod necks down, and thus the load system does work on the specimen. Conversely if the rod were to become more uniform under load, its compliance would decrease and it would have to do work on the load system. This would be in violation of thermodynamic principles.

Another interesting result of eq (8.46) is the case of a uniform rod, in which we again can ignore the ends as vacancy sources or sinks. The equation states that for $\eta_v = 0$ and elastic constants independent of c_v , the equilibrium vacancy concentration is a maximum at zero stress, and is lowered equally by tensile and compressive stresses. This result is again understood if we realize that the cross-section will be reduced if vacancies leave the system, increasing the specimen's compliance. The result will be modified if we assume that the elastic constants are a function of c_v and if we let η_v differ from zero, but for small changes it will not affect the sign.

8.5.3 Balancing Quadratic and Linear Effects. The 2π Wedge Disclination

Linear effects do not automatically dominate quadratic effects. An interesting example where both are present and cancel identically is a hollow tube composed of a 2π wedge disclination in which there is a pressure difference between the inside and outside of the tube.

To form the 2π wedge disclination we take a rectangular sheet of a perfect single crystal, bend it into a tube and weld the seam to insure perfect matching of lattice planes (fig. 3).

At this stage there are tangential compressive stresses at the inner surface and tensile stresses at the outer surfaces. M_{1v} at the two surfaces is the same because the stresses at the two surfaces have the same magnitude. Because of this the system reaches a vacancy equilibrium in this heterogeneously stressed system in which vacancy gradients and stress gradients combine to give a constant M_{1v} throughout.

Now apply a pressure difference between the inside and outside and permit vacancy flow. It is readily shown that in spite of the pressure difference the value of M_{1v} at the inner surface equals that at the outer surface. In the presence of the higher pressure at the inside there is a change in elastic free energy density, a reduction at the inner surface and an increase at the outer surface, and vice versa if the sign of the pressure difference is changed. The elastic energy is quadratic in the stress, but the change in stress due to the imposed pressure difference is linear in ΔP . The result is that the linear terms in P in M_{1v} cancel identically the changes in the quadratic terms in the tangential stresses. The linear and quadratic terms balance identically to give the same M_{1v} at the two surfaces. Again an equilibrium is reached in which M_{1v} is constant throughout and vacancy concentration gradients compensate for stress gradients.

This surprising result that the 2π wedge disclination will not creep by vacancy flow even when there is a pressure difference can also be understood by considering the consequence of the transfer of an entire plane of atoms from the inside to the outside. If we start with either of the flat single crystal plates and create the disclination we see that the tube is the same whether the atom layer is transferred or not (fig. 3).

9. Summary and Conclusions

We have reviewed and applied the thermodynamics that has been developed for multicomponent multiphase stressed crystalline solids. We have found equilibria in which the solids were neither homogeneous in stress nor in composition. We have considered equilibria for three types of multiphase contact: solid-fluid, incoherent, and coherent solid-solid. We have also examined simple non-equilibrium cases where potential gradients determine diffusion. Diffusional creep in particular was used to illustrate the importance of a full thermodynamic treatment.

Crystalline solids differ fundamentally from liquids in that they possess long range three-dimensional translational order. This implies that we can define a lattice and site occupancy. The number density and type of sites is known, and a local change in composition can only be made by redistributing atoms and vacancies among these sites. This fundamental restriction in the interior of a crystalline solid introduces important differences between the thermodynamics of solids and those of liquids. Because these restrictions apply at coherent boundaries but not at other boundaries, we find different equilibrium conditions at the various boundaries.

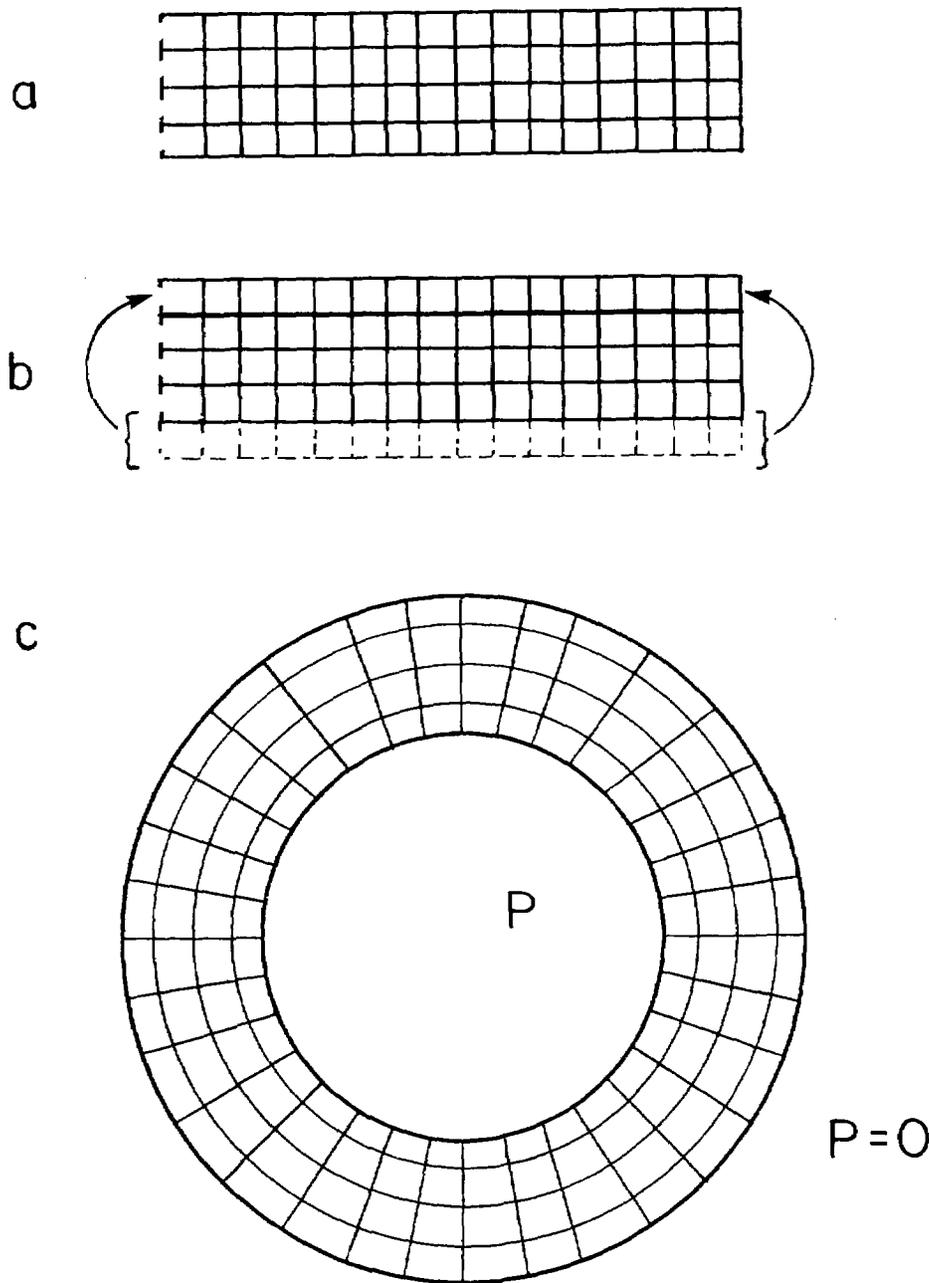


Figure 3—Radial vacancy fluxes that remove layers from the inner surfaces and deposit them on the outer surface of a 2π wedge disclination do not enlarge the disclination and therefore no work is done by any pressure difference. To see this, consider the cross section (c) of 2π wedge disclination made by elastically bending the perfect crystal (a) into a circular cylindrical shell and joining the ends. The 2π wedge disclination after radial diffusion is unchanged because it can be made from (b) which is identical to (a) except for translation of bottom layers to top. It will therefore reach the same equilibrium geometry in the presence of the pressure differences.

The equations that result from the thermodynamics consists of a set of coupled partial differential equations, algebraic equations and boundary conditions for stress and composition. For the kinetics, the diffusion equations are added. Although full nonlinear and large strain formulations exist, we have concentrated on examples where the essential features were displayed with small-strain approximations and linearized thermodynamics.

The thermodynamics has resulted in identifying and precisely defining the important phenomenological quantities needed for predictive calculation. The defini-

tions in particular are important and much of the controversy in the literature is judged to be the result of inadequate definitions of quantities. Furthermore the necessary data needed for evaluating the equations turn out to be computable from classically measured quantities, such as free energies of hydrostatically stressed solid solutions, elastic coefficients, and lattice parameters.

One important method for solving the equilibrium equations uses the notion of open-system elasticity. This method eliminates the composition variable from the

system of equations, and leaves a purely elastic problem to be solved. Central to the method are the open-system elastic constants, and in this paper we show that the same technique applies to multiphase solid equilibria. With this technique a large number of elastochemical problems are now solved because they become identical to solved problems of chemically homogeneous elastic solids. Once the stress field is known, only algebraic equations have to be solved to obtain the composition in the solid. As an example of the use of this concept, we have solved the dislocation atmosphere (stress field and composition field) in an isotropic and a cubic solid, automatically taking into account in a self-consistent way the thermodynamics of the solid solutions. Another example is the inclusion problem, although we have not found in the literature the shapes that satisfy the phase equilibrium boundary condition other than sphere, circular rod, and plate.

The question of the need for defining separate chemical potentials for each chemical species inside the solid has been a subject of controversy ever since Gibbs. We hope that we have shown that problems of equilibria can be solved without defining or using them. Gibb's famous example of a homogeneously stressed solid which gave three different chemical potentials when equilibrated with three fluids each at a pressure equal to minus a principal stress should alert everyone to the danger of attempting a definition. Of course our M_1 , could be construed to be a chemical potential of the i th specie, but we prefer for clarity to retain the vacancy as the counter specie.

Questions of species that occupy more than one site needed to be addressed. As our section 5.6 shows, the classical notion of chemical reactions among species on different sites very nicely resolves any confusion. Treating interstitials as atoms occupying sites that are mostly empty resulted in a unified treatment and clearly demonstrated the principle. From this more general treatment we showed it is possible to develop a treatment in which interstitials require a different and more convenient formulation.

We have supplemented an earlier overview on the effect of self-stress on diffusion by adding boundary conditions that permit phases to grow or shrink at the interface.

Diffusional creep is an important field in which the linearized and simplified treatment of Herring was an important first step. However Herring's definitions were not precise and this has led to much later confusion. We have presented a detailed derivation of a fuller treatment in which each term is fully defined and related to the data base. To emphasize the importance of the nonlinear terms, which Herring alluded to, but dis-

carded, we gave two examples each of which seems counterintuitive but thermodynamically correct: a long rod which in compression is unstable to necking by diffusional creep, and a tube composed of a perfect 2π wedge disclination which does not bulge by radial vacancy flux even when there is a pressure difference between the interior and exterior. The former is a case where Herring's linear term is zero and we must resort to the quadratic terms, and the latter is a case where the linear term identically cancels changes in the quadratic terms. The fuller equation contains several other terms usually ignored in creep theories that also can become important.

Capillary effects (surface strain and surface free energy) are not included. A formulation exists for some types of interfaces or specific geometries [36,37]. Theories of equilibrium of stressed solids with capillarity effects for the three types of interfaces considered here are being developed [38].

Although the elastic energy is usually small compared to the free energy change resulting from a composition change, there are domains where the interactions of composition and stresses are likely to be important. Self-stresses resulting from the presence of defects or heterogeneity of the material can have sizable consequences. The depression of the consolute critical point and the spinodal is a well known example. In systems without critical points coherent equilibrium is also strongly affected. Coherent phase diagram features have recently been found [39,40] that differ markedly from incoherent phase diagrams. The equations that could be used to calculate these phase diagrams have been obtained in sections 3 and 4.

Interesting consequences originate from the long range nature of the elastic forces. For instance this introduces non-local effects in the diffusion equation. Under hydrostatic pressure, a multi-phase incoherent dispersion at equilibrium is degenerate with respect to the shape of the phases, i.e., the equilibrium is independent of the shape of the precipitates. Under a more general state of stress (coherent precipitates, for instance), this simple result is no longer valid. The equilibrium equations have to be solved on an unknown boundary and the equilibrium shape is to be determined as part of the solution (a so-called free boundary problem). With the use of the open-system elastic constants such problems can be expressed as a purely elastic problem. The phase equilibrium boundary condition is the one that makes the problem different from classical elastic inclusion problems for which a shape is imposed. The solutions of the elastic equation of general shape will not be consistent with the phase equilibrium boundary condition. The catalog of the shapes that produce an elastic field

that in turn satisfies this condition has not yet been found. The introduction of capillarity would modify this condition. Work has been done on the subject [41].

We are grateful to W. C. Johnson and R. F. Sekerka for helpful discussions and criticism. We are especially grateful to M. Hillert for questioning the need to treat interstitials differently from other species. Out of our discussion with him the ideas of section 5.6 evolved. J. Hirth kindly called our attention to misprints in [37] which have been corrected in this article.

References

- [1] Truesdell, C., and R. A. Toupin, *The Classical Field Theories*, in *Encyclopedia of Physics*, S. Flugge, ed. **III/1**, Springer-Verlag (1960).
- [2] Gibbs, J. W., *Scientific Papers*, **1**, pp. 184-218, Longman (1906).
- [3] Li, J. C. M.; R. A. Oriani and L. S. Darken, *Z. Phys. Chem. Neue Folge* **49**, 271 (1966).
- [4] Yang, L.; G. T. Horne and G. M. Pound, *Proceedings of a Symposium on Physical Metallurgy of Stress Corrosion Cracking*, Pittsburgh, Interscience (1959), p. 29.
- [5] Robin, P.-Y. F., *American Mineralogist*, **59**, 1286 (1974).
- [6] Larché, F. C., and J. W. Cahn, *Acta Metall.* **21**, 1051 (1973).
- [7] Mullins, W. W., in *Proceeding of an International Conference on Solid-Solid Phase Transformations*, Pittsburgh, Met. Soc. AIME (1982), p. 49.
- [8] Bennett, L. H.; A. J. McAlister and R. E. Watson, *Physics Today*, **30**, 34 (1977).
- [9] Gibbs, J. W., *Scientific Papers*, **1**, Longman (1906).
- [10] Larché, F. C., and J. W. Cahn, *Acta Metall.* **26**, 1579 (1978).
- [11] Malvern, L. E., "Introduction to the Mechanics of a Continuous Medium," Prentice-Hall (1969).
- [12] Van der Waals, J. D., translated by J. S. Rowlinson, *J. Stat. Phys.* **20**, 197 (1979).
- [13] Cahn, J. W., and J. E. Hilliard, *J. Chem. Phys.* **28**, 258 (1958).
- [14] Hart, E. W., *Phys. Rev.*, **113**, 412 (1958).
- [15] Larché, F. C., and J. W. Cahn, *Acta Metall.* **26**, 53 (1978).
- [16] Bitter, F., *Phys. Rev.* **37**, 1527 (1931). Crum, M. M., as cited by F. R. N. Nabarro, *Proc. Roy. Soc.* **A175**, 519 (1940).
- [17] Nye, J. F., *Physical Properties of Crystals*, Clarendon Press, Oxford (1957).
- [18] Herring, C. J., *Appl. Phys.* **21**, 437 (1950).
- [19] Cottrell, A. H., "Report of a Conference on Strength of Solids," University of Bristol, The Physical Society, London (1948).
- [20] Louat, N., *Proc. Phys. Soc.* **B69**, 459 (1956).
- [21] Beshers, D. N., *Acta Metall.* **6**, 521 (1958).
- [22] Johnson, R. A., *Phys. Rev.* **B24**, 7383 (1981).
- [23] Eshelby, J. D., *Adv. Solid State Phys.* **3**, 79 (1956).
- [24] Steeds, J. W., *Introduction to Anisotropic Elasticity Theory of Dislocations*, Clarendon Press, Oxford (1973).
- [25] Balluffi, R. W., and A. V. Granato, in "Dislocations in Solids," F. R. N. Nabarro, ed., Vol. 4, p. 2, North-Holland, Amsterdam (1979).
- [26] Gairola, B. K. D., in "Dislocations in Solids," F. R. N. Nabarro, ed., Vol. 1, p. 223, North-Holland, Amsterdam (1979).
- [27] Li, J. C. M.; F. V. Nolfi and C. A. Johnson, *Acta Metall.* **19**, 749 (1971).
- [28] Herring, C., in "The Physics of Powder Metallurgy," W. E. Kingston, ed., McGraw-Hill (1951).
- [29] Hillert, M., in "Alloy Phase Diagrams," L. W. Bennett, T. B. Massalski, and B. C. Giesen, eds., North-Holland (1983).
- [30] Larché, F. C., and J. W. Cahn, *Acta Metall.* **30**, 1835 (1982).
- [31] Cahn, J. W., and F. C. Larché, *Scripta Met.* **17**, 927 (1983).
- [32] Cahn, J. W., *Acta Met.* **9**, 795 (1961).
- [33] Hillert, M., *Met. Trans.* **15A**, 411 (1984).
- [34] Tashiro, K., and G. R. Purdy, *Scripta Met.* **17**, 455 (1983).
- [35] Roitburd, A. L., *Sov. Phys. Solid State* **23**, 622 (1981).
- [36] Cahn, J. W., *Acta Metall.* **28**, 1333 (1980).
- [37] Cahn, J. W., and F. C. Larché, *Acta Metall.* **30**, 51 (1982).
- [38] Alexander, J. I., and W. C. Johnson, to be published.
- [39] Williams, R. O., *Met. Trans.*, **A**, **11**, 247 (1980).
- [40] Cahn, J. W., and F. C. Larché, *Acta Metall.* **32**, 1915 (1984).
- [41] Johnson, W. C., and J. W. Cahn, *Acta Metall.* **32**, 1925 (1984).

Appendix 1. Solid-Liquid Equilibrium Under Hydrostatic Stress

We consider the case of a substitutional binary solid. In equilibrium with a fluid under hydrostatic stress (for instance if it is entirely surrounded by the fluid), the mechanical equilibrium eqs (3.13) and (3.14) implies that the stress is equal to

$$T_{ij} = -P\delta_{ij} \quad (\text{A1.1})$$

where P is the pressure in the fluid. The stress being uniform, the constancy of the diffusion potential implies that the composition is uniform. Therefore the solid is uniform. The boundary condition

$$f - \mu_1^L \rho_1 - \mu_2^L \rho_2 = -P \quad (\text{A1.2})$$

can be combined with the equation for the diffusion potential

$$M_{12} = \mu_1^L - \mu_2^L \quad (\text{A1.3})$$

to give

$$\mu_1^L = (f + P + \rho_2 M_{12}) V_0 \quad (\text{A1.4})$$

$$\mu_2^L = (f + P - \rho_1 M_{12}) V_0 \quad (\text{A1.5})$$

Because the *solid is uniform*, these expressions are valid everywhere. The quantities on the right hand side of (A1.4) and (A1.5) depend only on the value of the state variables. Let us call them μ_1^s and μ_2^s .

$$\mu_1^s \equiv (f + P + \rho_2 M_{12}) V_0 \quad (\text{A1.6})$$

$$\mu_2^s \equiv (f + P - \rho_1 M_{12}) V_0 \quad (\text{A1.7})$$

Elimination of M_{12} between these two equations gives

$$f = -P + \rho_1 \mu_1^s + \rho_2 \mu_2^s$$

and, because of the uniformity, we can multiply by V_0 to get the total Helmholtz free energy

$$F = -P V_0 + N_1 \mu_1^s + N_2 \mu_2^s$$

where N_1 and N_2 are the total number of moles of components 1 and 2 respectively. The differential of f' is

$$df' = T_{ij} dE_{ij} + M_{12} d\rho_1$$

M_{12} is replaced by its value obtained from (A1.6) and (A1.7). Using the definition of ρ_1' , and after multiplication by V_0' , one obtains

$$dF = -P dV_0 + \mu_1^s dN_1 + \mu_2^s dN_2$$

Therefore

$$\mu_1^s \equiv \left(\frac{\partial F}{\partial N_1} \right)_{V, N_2}$$

$$\mu_2^s \equiv \left(\frac{\partial F}{\partial N_2} \right)_{V, N_1}$$

We have recovered all the classical formula for fluid-fluid equilibrium. Despite network constraints, a solid under hydrostatic stress behaves as if it were a fluid.

Appendix 2. The Boundary Conditions for Coherent Phase Change: Small Strain Approximation

The full large strain boundary condition for coherent phase change is [15]

$$\omega'^\alpha - \omega'^\beta - \mathbf{n}'^\alpha \bullet (\mathbf{F}^{\alpha T} - \mathbf{F}^{\beta T}) \bullet (\partial f'^\alpha / \partial \mathbf{F}) \bullet \mathbf{n}'^\alpha = 0 \quad (\text{A2.1})$$

where the same reference state is chosen for both phases. The superscript T stands for transpose and \mathbf{F} is the deformation gradient. $(\partial f' / \partial \mathbf{F})$ is the first Piola-Kirchhoff tensor \mathbf{T}_R . It is related to the Cauchy stress tensor \mathbf{T} by

$$\mathbf{T}_R = \mathbf{J} \mathbf{T} \bullet (\mathbf{F}^{-1})^T \quad (\text{A2.2})$$

where J is the determinant of \mathbf{F} . In the small strain approximation, the displacement tensor is given, to first order in the derivatives u_{ji} , by [11]

$$\mathbf{F} = \mathbf{I} + \mathbf{E} + \mathbf{\Omega} + 0(u_{i,j}^2) \quad (\text{A2.3})$$

where \mathbf{E} is the small strain tensor, (eq (3.1)), $\mathbf{\Omega}$ the small rotation tensor, and \mathbf{I} the unit tensor. To the same approximation, its inverse is given by

$$\mathbf{F}^{-1} = \mathbf{I} - \mathbf{E} - \mathbf{\Omega} + 0(u_{i,j}^2) \quad (\text{A2.4})$$

Using these equations we get

$$\begin{aligned} \mathbf{n}' \bullet (\mathbf{F}^T \bullet \mathbf{T}_R) \bullet \mathbf{n}' &= \mathbf{n}' \bullet (\mathbf{I} + \mathbf{E} - \mathbf{\Omega}) \bullet \mathbf{T} \bullet (\mathbf{I} - \mathbf{E} + \mathbf{\Omega}) \bullet \mathbf{n}' \\ &+ 0(u_{i,j}^2) \end{aligned} \quad (\text{A2.5})$$

Dropping terms of order $u_{i,j}^2$, and since, for an arbitrary 3×3 tensor

$$\mathbf{n}' \bullet \mathbf{A} \bullet \mathbf{n}' = \mathbf{n}' \bullet \mathbf{A}^T \bullet \mathbf{n}'$$

we finally obtain

$$\mathbf{n}' \bullet (\mathbf{F}^T \bullet \mathbf{T}_R) \bullet \mathbf{n}' = \mathbf{n}' \bullet \mathbf{T} \bullet \mathbf{n}' - 2\mathbf{n}' \bullet \mathbf{\Omega} \bullet \mathbf{T} \bullet \mathbf{n}' \quad (\text{A2.6})$$

Since the same reference state has been chosen for α and β , the following equalities hold

$$\begin{aligned} \rho'_i &= \rho'_0 c_i^\alpha ; \rho'_i^\beta = \rho'_0 c_i^\beta \\ J^\alpha &= \rho'_0 / \rho_0^\alpha \quad J^\beta = \rho'_0 / \rho_0^\beta \end{aligned} \quad (\text{A2.7})$$

Using (A2.7), (A2.6), and (A2.1) we finally obtain

$$\begin{aligned} V'_0 f'^\alpha - \sum M_{ik} c_i^\alpha + V'_0 [-T_{ij}^\alpha n_i'^\alpha n_j'^\alpha + 2\Omega_{ij}^\alpha T_{jk}^\alpha n_i'^\alpha n_k'^\alpha] \\ = V'_0 f'^\beta - \sum M_{ik} c_i^\beta + V'_0 [-T_{ij}^\beta n_i'^\beta n_j'^\beta + 2\Omega_{ij}^\beta T_{jk}^\beta n_i'^\beta n_k'^\beta] \end{aligned} \quad (\text{A2.8})$$

The various terms are seen to be energies per mole of lattice sites. It is then easy to make a change of reference volume (like the stress free state for each phase). To the level of approximation used in linear elasticity this won't affect the $V'_0 f'$ terms. But it does affect the terms linear in T_{ij} .

Appendix 3. Derivation of the Open-System Elastic Stiffness and Compliance Tensor

All the calculations are done at constant temperature, so that all the partial derivatives are understood to be at constant temperature. We first treat the case of a binary solid, then generalize to a multicomponent solid.

A3.1 Binary Solid

To simplify the notation we take ρ' to be ρ'_i . The differential of the stress can be written

$$dT_{ij} = \left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{\rho'} dE_{kl} + \left(\frac{\partial T_{ij}}{\partial \rho'} \right)_{E_{kl}} d\rho' \quad (\text{A3.1})$$

or

$$dT_{ij} = \left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{M_{12}} dE_{kl} + \left(\frac{\partial T_{ij}}{\partial M_{12}} \right)_{E_{kl}} dM_{12} \quad (\text{A3.2})$$

The differential of the diffusion potential is

$$dM_{12} = \left(\frac{\partial M_{12}}{\partial \rho'} \right)_{E_{ij}} d\rho' + \left(\frac{\partial M_{12}}{\partial E_{ij}} \right)_{\rho'} dE_{ij} \quad (\text{A3.3})$$

Replacing $d\rho'$ from (A3.3) into (A3.1) yields

$$dT_{ij} = \left[\left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{\rho'} - \left(\frac{\partial T_{ij}}{\partial \rho'} \right)_{E_{kl}} \left(\frac{\partial M_{12}}{\partial E_{ij}} \right)_{\rho'} / \left(\frac{\partial M_{12}}{\partial \rho'} \right)_{E_{ij}} \right] dE_{kl} + \left[\left(\frac{\partial T_{ij}}{\partial \rho'} \right)_{E_{kl}} / \left(\frac{\partial M_{12}}{\partial \rho'} \right)_{E_{kl}} \right] dM_{12} \quad (\text{A3.4})$$

and the coefficient of the term dE_{kl} is the $(ijkl)$ component of the open-system stiffness tensor.

Using the stress-strain relationship (4.14) and the Maxwell relation

$$\left(\frac{\partial T_{ij}}{\partial \rho'} \right)_{E_{kl}} = \left(\frac{\partial M_{12}}{\partial E_{ij}} \right)_{\rho'} \quad (\text{A3.5})$$

one gets

$$\left(\frac{\partial M_{12}}{\partial E_{ij}} \right)_{\rho'} = -C_{ijkl} \eta_{kl} + \frac{dC_{ijkl}(E_{kl} - E_{kl}^c)}{d\rho'} \quad (\text{A3.6})$$

The value of M_{12} as a function of E_{ij} rather than T_{ij} is obtained from (4.14) by using

$$S_{ijkl} C_{klmn} = \delta_{im} \delta_{jn} \quad (\text{A3.7})$$

Neglecting strain dependent terms, we finally get

$$C_{ijkl}^* = C_{ijkl} - \frac{1}{\rho'_0 \chi} C_{ijmn} C_{klpq} \eta_{mn} \eta_{pq} \quad (\text{A3.8})$$

Because of the linearity, we have

$$S_{ijkl}^* C_{klmn}^* = \delta_{im} \delta_{jn} \quad (\text{A3.9})$$

where S_{ijkl}^* are the open-system compliances. Combining (A3.8) and (A3.9) gives

$$S_{ijkl}^* = S_{ijkl} + \rho'_0 \chi \eta_{ij} \eta_{kl} \quad (\text{A3.10})$$

where η_{ij} are defined by (4.4).

A3.2. Multicomponent Solids

We follow the same derivation as above. The differential of the stress tensor is

$$dT_{ij} = \left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{\rho_i} dE_{kl} + \sum_{1 \neq K} \left(\frac{\partial T_{ij}}{\partial \rho'_{1K}} \right) d\rho'_{1K} \quad (\text{A3.11})$$

The differentials of the potentials are

$$dM_{1K} = \sum_{J \neq K} \left(\frac{\partial M_{1K}}{\partial \rho'_{JK}} \right) d\rho'_{JK} + \left(\frac{\partial M_{1K}}{\partial E_{ij}} \right)_{\rho_i} dE_{ij} \quad (\text{A3.12})$$

$d\rho'_j$ can be obtained from this system of linear equation by Kramer's rule

$$d\rho'_j = \left[\sum_{l \neq k} (dM_{lK} - \left(\frac{\partial M_{lK}}{\partial E_{ij}} \right)_{\rho_l} dE_{ij}) A_{lj} \right] / D \quad (\text{A3.13})$$

where D is the determinant

$$D = \left| \frac{\partial M_{lK}}{\partial \rho'_{lK}} \right|$$

and A_{lj} is the minor of the $(\partial M_{lK} / \partial \rho'_{lK})$ term of D . Replacing $d\rho'_l$ by its value in (A3.11) and using the Maxwell relation

$$\left(\frac{\partial T_{ij}}{\partial \rho'_i} \right)_{E_{kl}} = \left(\frac{\partial M_{lK}}{\partial E_{ij}} \right)_{\rho_l} \quad (\text{A3.14})$$

we get

$$\left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{M_{lK}} = \left(\frac{\partial T_{ij}}{\partial E_{kl}} \right)_{\rho_l} - \frac{1}{D} \sum_{l \neq K} \sum_{j \neq K} \left(\frac{\partial T_{ij}}{\partial \rho'_i} \right)_{E_{pq}} \left(\frac{\partial T_{ij}}{\partial \rho'_i} \right)_{E_{pq}} A_{lj} \quad (\text{A3.15})$$

Using (A3.9), Hooke's law, and neglecting strain dependent terms we finally get

$$S^*_{ijkl} = S_{ijkl} + \frac{1}{\chi} \sum_{l, j \neq K} \left(\frac{\partial E_{ij}}{\partial c_l} \right) \left(\frac{\partial E_{ij}}{\partial c_l} \right) A_{lj} \quad (\text{A3.16})$$

where χ is the determinant

$$\chi = \rho'_0 \left| \frac{\partial M_{lK}}{\partial c_l} \right|$$

and A_{lj} the minor of the (lK) term of χ / ρ'_0 .

Appendix 4. A Common Tangent Construction

Let ξ^k be three unit vectors normal to each other, such that ξ^3 is the normal to the interface, with components ξ^k_j . The vectors λ^k are defined by

$$\lambda^k_i = E_{ij} \xi^k_j \quad (\text{A4.1})$$

Since the determinant $|\xi^k_j|$ has the value 1 the system of equations (A4.1) constitute a valid linear change of variable. Using the chain rule, we obtain, considering the ξ^k as fixed

$$\left(\frac{\partial f'}{\partial E_{ij}} \right)_{E_{kl}, \rho_l} = T_{ij} = \left(\frac{\partial f'}{\partial \lambda^k_i} \right)_{\lambda^m} \xi^k_j \quad (\text{A4.2})$$

After multiplication by ξ^k_j and summation on j one gets

$$T_{ij} \xi^k_j = \left(\frac{\partial f'}{\partial \lambda^k_i} \right)_{\lambda^m} \quad (\text{A4.3})$$

Let us define the free energy ζ' by

$$\zeta' = f' - \lambda_j^3 \left(\frac{\partial f'}{\partial \lambda_j^3} \right) \quad (\text{A4.4})$$

$$= f' - T_{ij} n_j E_{ik} n_k \quad (\text{A4.5})$$

and it is easy to show that

$$M_{\text{IK}} = \left(\frac{\partial f'}{\partial \rho'_{\text{IK}}} \right)_{\varepsilon_{ij}} = \left(\frac{\partial \zeta'}{\partial \rho'_{\text{IK}}} \right)_{\lambda_j^3, T_{ij}, n_j} \quad (\text{A4.6})$$

The conditions for equilibrium at an incoherent interface (eq (3.24)) can be written

$$\begin{aligned} \hat{f}^\alpha - \Sigma c_1^\alpha \left(\frac{\partial \hat{f}^\alpha}{\partial c_1} \right) - T_{ij}^\alpha n_i^\alpha n_j^\alpha V_0^\alpha \\ = \hat{f}^\beta - \Sigma c_1^\beta \left(\frac{\partial \hat{f}^\beta}{\partial c_1} \right) - T_{ij}^\beta n_i^\beta n_j^\beta V_0^\beta \end{aligned} \quad (\text{A4.7})$$

where quantities such as \hat{f} are just $f' V_0'$, i.e., quantities per mole of lattice sites. If the normal pressure is zero, so that $T_{ij} n_j = 0$ it becomes equivalent to

$$\hat{\zeta}^\alpha - \Sigma c_1^\alpha \left(\frac{\partial \hat{\zeta}^\alpha}{\partial c_1} \right)_{\lambda_j^3, \lambda_j^2} = \hat{\zeta}^\beta - \Sigma c_1^\beta \left(\frac{\partial \hat{\zeta}^\beta}{\partial c_1} \right)_{\lambda_j^3, \lambda_j^2} \quad (\text{A4.8})$$

which together with

$$M_{\text{IK}}^\alpha = M_{\text{IK}}^\beta$$

which can then be written

$$\left(\frac{\partial \hat{\zeta}^\alpha}{\partial c_1} \right)_{\lambda_j^3, \lambda_j^2} = \left(\frac{\partial \hat{\zeta}^\beta}{\partial c_1} \right)_{\lambda_j^3, \lambda_j^2} \quad (\text{A4.9})$$

show that c_{IK} can be obtained by a tangent construction to $\hat{\zeta}$, which, in this case is just equal to \hat{f} .