Brittle fracture in structural steels: perspectives at different size-scales

John Knott

School of Metallurgy and Materials, The University of Birmingham, Birmingham, UK

This paper describes characteristics of transgranular cleavage fracture in structural steel, viewed at different size-scales. Initially, consideration is given to structures and the service duty to which they are exposed at the macroscale, highlighting failure by plastic collapse and failure by brittle fracture. This is followed by sections describing the use of fracture mechanics and materials testing in carrying-out assessments of structural integrity. Attention then focuses on the microscale, explaining how values of the local fracture stress in notched bars or of fracture toughness in pre-cracked test-pieces are related to features of the microstructure: carbide thicknesses in wrought material; the sizes of oxide/silicate inclusions in weld metals. Effects of a microstructure that is ‘heterogeneous’ at the mesoscale are treated briefly, with respect to the extraction of test-pieces from thick sections and to extrapolations of data to low failure probabilities. The values of local fracture stress may be used to infer a local ‘work-of-fracture’ that is found experimentally to be a few times greater than that of two free surfaces. Reasons for this are discussed in the conclusion section on nano-scale events. It is suggested that, ahead of a sharp crack, it is necessary to increase the compliance by a cooperative movement of atoms (involving extra work) to allow the crack-tip bond to displace sufficiently for the energy of attraction between the atoms to reduce to zero.

1. Introduction

A full appreciation of the mechanisms of fracture encountered in steels used to fabricate structures such as ships, bridges, pipelines or pressure vessels covers
behaviour over a range of size-scales. The structures themselves are often extremely large: super-
tankers and ‘bulkers’ over 200 m in length; bridge spans over 1500 m; the Kazakhstan–China 
pipeline, over 2000 km. Large modern structures are almost invariably fabricated by welding, 
so that the properties of the weld metal and the ‘heat-affected zone’ (HAZ) in the ‘parent’ metal 
adjacent to the weld interface need to be included in any assessment of the structure’s integrity 
when subjected to service duty. The duty depends on application, but falls into one of two 
categories: that within the ‘design intent’ and that associated with (postulated) ‘accident’ events. 
Both must be addressed in safety justifications.

The ‘design intent’ deals with ‘normal’ service duty. A bridge has its own deadweight, must 
bear traffic loads and is subjected to wind and wave loadings. Similar loadings are experienced 
by a tanker or a bulker carrying liquids (such as oil) or solids (such as minerals or grain) across 
the oceans; by an offshore drilling rig and by an offshore wind turbine (with added effects of 
whirling blades, to which the structure must provide reaction forces). A pipeline or pressure 
vessel has to contain the pressure of the medium that it contains. In addition, thermal stresses 
may be generated: either from simple day/night/day temperature changes or from excursions 
to and from high temperature in the operation of pressure vessels in chemical plant or in 
power generation plant. All these factors are included in the initial design. In broad terms, the 
design intent is for the structure to exhibit an overall elastic response to the stresses imposed 
by service duty. This is achieved by limiting the operating stresses to a fraction of the steel’s 
yield strength, through a ‘safety factor’ specified in a design code. For pressure vessels, the safety 
factor is 2, i.e. the vessel is designed with dimensions such that the nominal stress associated with 
‘normal’ service duty is limited to 50% of the yield strength. It should be recognized that a large 
amount of detailed calculation, involving both static and dynamic loading, is involved at this 
‘macro’-level.

The second set of macroscale threats to structural integrity arises from accidental overloads, 
beyond those associated with the design intent. One cause of this is unforeseen natural forces, 
as demonstrated by the combination of tsunami and earthquakes at the Fukushima nuclear power 
plant in 2011, or by the Titanic’s collision with an iceberg in 1912. Excessive overloads (not allowed 
for in the design calculations for the completed structure) may also be encountered during the 
construction phase. Examples are given by plastic collapse failures of a number of ‘box-girder’ 
bridges in the 1970s. Here, the buckling mechanisms for the hollow trapezoidal cross-sections 
of the ‘boxes’ in bending were not fully understood and excessive bending loads were applied 
during erection. Similar overload buckling failures can occur as a result of the incautious loading 
of cargo into bulkers. These carry solid material, and it is necessary for machinery (such as cranes 
and diggers) to be able to access the cargo in an unimpeded manner. A consequence is that there 
is minimal ‘cross-bracing’ in the holds and the bulker again assumes the form of a hollow box. 
If the loading of cargo is not sequenced in a designed manner, large bending moments can be 
generated and buckling collapse occurs. A spectacular example is that of ‘Euro-bulker X’ in 2000 
at Levkandi, Greece. Guidance is now available to prevent such collapse [1], but ships may also 
fail by buckling collapse as a result of running aground, e.g. the ‘Amoco Cadiz’. Such ‘operational’ 
failures do not fall within the scope of this paper, which is devoted to fractures that may occur 
under ‘normal’ service loadings, when the overall nominal stress is designed to be a fraction of 
the yield strength.

Any such fracture necessitates the presence of a stress-concentrator in the structure, to enable 
localized yielding to occur and operate fracture mechanisms, when the nominal applied, stress is 
still a fraction of the yield strength. Such stress-concentrating features include changes in cross 
section (e.g. associated with fillets or stiffening plates), through-section penetrations or crack-like 
defects created during processing and fabrication. From a macroscopic engineering perspective, 
any fracture that occurs below the design stress is a (macroscopically) ‘brittle’ or ‘fast’ fracture and 
has to be addressed in an appropriate manner, but, at low temperatures, structural steels exhibit 
microscopically ‘brittle’ behaviour, as described below.

There are many examples of brittle fractures in service. The most cited examples are those of 
welded Liberty ships and tankers, constructed during World War II. Between 1942 and 1952, 233
welded ships suffered one or more brittle fractures of severity such that the vessels were left in a dangerous condition; 19 broke in two or were completely abandoned; another 1200 experienced brittle cracks up to 3 m in length. Seventeen bridge failures occurred in Belgium between 1938 and 1950. The King’s Bridge in Melbourne, Australia suffered a brittle fracture in 1962. Brittle failures in ships continue: three such are the World Concord in 1954, MV Kurdistan in 1979 [2] and the Selendang Ayu in 2004. Brittle fractures have also occurred in boilers and pressure vessels, even in the ‘proof test’ when 20% over-pressure is applied before the vessel enters service. One example was a boiler shell at Sizewell in 1963: another was that in 1965 of a vessel destined for use in ammonia conversion (figure 1).

The endpoint of a structural integrity assessment is the assurance that a structure will not fail under anticipated service duty, with a safety case that includes a set of postulated overload scenarios. As noted above, the structures of concern are very large, often ‘one-off’ for a specific application, and fabricated from a large number of individual components. It is not therefore possible to test full-size structures to failure, as is done, for example, for automotive engines. In some cases, specific tests may be made on ‘features’ or ‘type tests’ (e.g. fatigue tests on samples of the fabricated ‘nodes’ employed in offshore structures), but this is not always done, and many structures are constructed on the basis of design codes, measures of the material’s suitability for service as inferred from specimen testing, and non-destructive inspection (NDI), to ensure that no defects of concern are present in the structure before it enters service. Periodic NDI is employed to guard against defect growth by any ‘subcritical’ crack growth mechanism (such as fatigue or environmentally assisted crack growth) during operation. In marine applications, loss of section thickness by salt-water corrosion needs also to be monitored. This paper concentrates on three size-scales, although one of these embeds a fourth:

— Fracture at the ‘test-piece’ scale, in both notched-bars and fracture toughness test-pieces and the use of data from such tests to make large-scale structural integrity assessments.
— Fracture at the ‘microstructural’ scale, relating values of local fracture stress or fracture toughness in test-pieces to the sizes and distributions of microstructural features. This leads to a treatment of how the microstructural understanding helps to address events at the ‘meso’-scale in ‘heterogeneous’ or ‘mixed’ microstructures.
— Fracture at the ‘nano’ or ‘atomistic’ scale, considering how fracture might occur at the tip of an atomically sharp crack.

**Figure 1.** Brittle fracture of a steel pressure vessel during proof test. (The vessel walls were 149 mm thick, and a 2-tonne fragment was thrown 46 m).Courtesy TWI.
2. Assessment of structural integrity; linear elastic fracture mechanics

In recognition of the fact that a stress-concentrator is necessary for (engineering-scale) brittle fracture to occur, measurement of the resistance to fracture of structural steels employs notched or pre-cracked test-pieces. Up to some 40 years ago, material quality was assessed, using notched-bar impact testing. A common example is the Charpy test, which subjects a bar of $10 \times 10$ mm cross section, containing a $45^\circ$ V-notch of depth 2 mm and root radius 0.25 mm, to impact loading in a simple pendulum machine. The energy absorbed in fracture is measured as a function of test temperature, and is observed to undergo a transition from high-energy ‘ductile’ fracture at high temperature to low-energy ‘brittle’ fracture at low temperature. Ductile fracture involves the formation and coalescence of microvoids, centred on inclusions and other second-phase particles. A full treatment of these mechanisms would lengthen this paper unduly: details can be found in reference [3]. Two forms of brittle fracture may be observed: one is transgranular cleavage across $\{001\}$ crystallographic planes (typically observed in normalized or annealed mild steels); the other is intergranular fracture along (ferrite grain-boundaries coincident with) prior-austenite grain-boundaries (observed in quenched-and-tempered alloy steels, resulting from the segregation of trace impurity elements such as P, Sn or Sb to such boundaries): see reference [4] in the theme issue.

The main problem with notched impact tests is that the output information that they offer: energy absorption, transition temperature or fracture appearance cannot be used directly to calculate the failure stresses associated with defects in structures. These issues are discussed in more depth in reference [5]. Note, however, that notched impact values are still included in many steel quality specifications and, in the nuclear industry, due both to space limitations and to the dates at which the programmes were begun, many of the surveillance specimens used to monitor the effects of neutron irradiation on fracture properties are notched-bar impact test-pieces.

Quantification of the applied stresses required to produce an engineering-scale ‘brittle’ or ‘fast’ fracture relies on the principles of fracture mechanics (see references [6,7] for more detailed treatment). The test-pieces contain sharp cracks (grown in by fatigue) and are loaded in tension (CTS) or bending at modest strain rate. The stress distribution over a short distance, $r$, ahead of an edge crack of length $a$, or a central crack of length $2a$, lying normal to a uniform applied stress $\sigma$ is given by

$$\sigma(r) = \frac{K}{(2\pi r)^{0.5}}$$

where $K$ is the strength of the crack tip field: the crack tip stress intensity factor. For a central crack of length $2a$ normal to a uniform tensile stress $\sigma_{\text{app}}$, $K$ is given by

$$K = \sigma_{\text{app}}(\pi a)^{0.5}.$$  \hspace{1cm} (2.2)

It assumes other forms for finite bodies or for non-uniform stresses, but the square-root dependency on crack length dominates. For standard test-piece geometries, $K$ is directly proportional to stress and is related to test-piece dimensions, $a$ (crack length) and $W$ (test-piece width) through tabulated ‘compliance functions’, $Y$: $K = \sigma_{\text{app}} Y(a/W)$. The strain energy release ‘rate’ ($dU/da = G$) can be derived by a virtual work argument involving the crack tip stress and displacement fields—see [6–8]—to give, for the central crack under uniform stress, the identity

$$\frac{dU}{da} = G = \frac{K^2}{E'} = \sigma_{\text{app}}^2 \frac{\pi a}{E'},$$

where $E'$ is Young’s modulus, $E$, in plane stress, or $E/(1 - \nu^2)$ in plane strain; $\nu$ is Poisson’s ratio. This result is seen to be identical to the usually recognized Griffith expression if the critical value of $dU/da = G$ is equated to $2\gamma$ (where $\gamma$ is the surface energy) and $\sigma_{\text{app}}$ is written as $\sigma_{\text{app}(F)}$ (the macroscopically applied fracture stress).
Steels used in engineering structures are not ideal brittle solids, and the application of stress to a cracked body initially produces some local plasticity at the crack tip (before the crack propagates). Nevertheless, if the extent of such plasticity is small compared with the extent of the $K$-dominated field, it is possible to decouple the energy release rate, $G$, term from the work involved in propagating the crack, which is traditionally written as $2\gamma + \gamma_p$, where $\gamma_p$ is held to represent the ‘plastic work of fracture’. Note, however, that most of the plastic work in a fracture toughness test precedes crack propagation (setting up a local stress state ahead of the pre-crack that facilitates the propagation of microcracks formed in brittle, second-phase particles: see below).

More generally, propagation occurs at a critical energy release rate, $G_{\text{crit}}$, or, via equation (2.3) at a critical value of the stress intensity factor, $K_{\text{crit}}$. Under tensile (‘mode I’) loading, in plane strain, this is written as $K_{1c}$ and is referred to as the material’s plane strain fracture toughness.

For engineering assessment at the macroscale, it is arguably not necessary to delve into the interpretations of $K_{1c}$, provided that consistent values represent the onset of fracture in different geometrical configurations. For structural steels, it is, however, important to pay attention to microscopically ‘brittle’ fracture modes, and to how these are affected by microstructural features and by geometrical factors: crack depth, plate thickness, etc.

For fracture under ‘small-scale yielding’ conditions—what might be termed ‘quasi-brittle’ fracture—the determination of resistance to fast fracture in a structural feature is straightforward. The fracture toughness is determined by measuring the fracture stress and fatigue crack length at failure in a test piece of standard geometry for which the compliance function $Y(a/W)$ is tabulated. For the structural feature, the applied (design) stress is known, and account can be taken of stress-concentrating features and residual stresses to derive the local stresses in areas of interest. It is then possible to assess cracks of different lengths located in these high stress regions and use the combination of local stress intensity factor (calculated from the local stress distribution and the postulated crack length) and the fracture toughness value to determine whether or not the length of postulated crack would lead to fast fracture. A number of iterations allow the critical crack length to be calculated. In most cases, further action is taken to ensure that the fabrication techniques and NDI guarantee that any crack-like defect likely to be present in the structure is much smaller than anything that would raise a concern with respect to fracture.

3. Yielding fracture mechanics

This procedure has to be modified for structural steels under normal service conditions. The ‘linear elastic’ analysis can be used only for ‘quasi-brittle’ fractures, occurring under ‘small-scale yielding’ conditions. This places severe requirements on the sizes of ‘valid’ test pieces. The standard test piece has width $W$, thickness $B = W/2$, and a crack length $a$, which has to meet the criterion $0.45 < (a/W) < 0.55$. For ‘validity’, all dimensions are incorporated in the limit set on $B: B > 2.5(K_{1c}/\sigma Y)^2$. Taking as an example nuclear reactor pressure vessel (RPV) steel, of yield strength, 400 MPa and fracture toughness 200 MPa m$^{0.5}$, $B$ is calculated as 625 mm. First, this presents major problems with respect to testing: second, the required thickness is greater even than that in thick-walled RPVs (The UK Sizewell ‘B’ RPV has walls 200 mm thick, although diagonals across the nozzle penetrations may be 350 mm). Even if it were possible to carry out ‘valid’ tests on material in the start-of-life condition (and a few very large test pieces have been fractured), it is clearly not possible to carry out ‘valid’ tests on (necessarily small) surveillance specimens to determine effects of neutron irradiation on fracture toughness.

The standard ‘validity’ criterion refers to fast fracture instability under small-scale yielding conditions. It can be relaxed if what is required in the small-scale test piece is a parameter that characterizes the onset of fracture. This can subsequently be correlated with either the onset of fracture or fracture instability in a larger-scale engineering structure. One parameter is the crack (tip) opening displacement, CTOD or $\delta$, based on the argument that the critical value of CTOD, $\delta_{\text{crit}}$, at the point of crack extension, is the same in both small test pieces and large components: see references [6,7]. In a large component, where the plasticity required to accommodate the critical value of CTOD is small compared with the component’s size, the relationships between $\delta$ and the
parameters $G$ and $K$ in plane strain are given by

$$G = 2\sigma_Y \delta = \frac{K^2}{E'}. \quad (3.1)$$

A ‘critical’ value of CTOD can therefore be equated to a ‘critical’ value of $K$, allowing defect assessment for the large component to proceed.

The second characterizing parameter is the ‘$J$-integral’; see references [6,7]. Originally defined as the equivalent of energy release rate $G$ for nonlinear elastic material, $J$ is an elastic/plastic characterizing parameter, determined from the load–displacement curve of a pre-cracked test-piece as for nonlinear elastic material, but calculating the $\frac{dU}{da}$ term from incremental areas under the loading curves obtained for different crack lengths. Criteria are laid down to ensure that fracture in the small test piece has taken place under ‘$J$-controlled’ conditions. As for CTOD, the ‘elastic/plastic’ $J$ is a characterizing parameter and does not bear any connotation of critical values of energy release rate: $J_i$ characterizes the initiation of ductile crack growth; $J_{0.2}$ characterizes 0.2 mm of ductile crack growth. For a large component, under ‘small-scale’, yielding conditions use is made of the equivalence of $J$ and $G$

$$J = G = \frac{K^2}{E'} = 2\sigma_Y \delta, \quad (3.2)$$

to obtain the relationship between ‘critical’ values of $J$ and $K$, and so proceed to defect assessment. Values of $K$ obtained in this way are written as $K_J$.

4. The failure analysis diagram

Recognizing that the two main threats to integrity are failure by plastic collapse and failure by ‘fast’ fracture, structural integrity assessments, such as the widely adopted R6 procedure [9] analyse these conjointly using a failure assessment diagram (FAD), drawn schematically in figure 2; see also reference [7]. The abscissa shows the load ratio parameter, $L_r$, which is the ratio of the applied load to the plastic collapse (or general yield) load, $L_{app}/L_{GY}$. The ordinate shows the stress intensity factor ratio, $K_r$, which is the ratio of the $K$ value at the tip of a crack in the loaded structure, $K_{app}$, to the fracture toughness of the material, $K_{1c}$ or $K_{J}$, shown in figure 2 as $K_{mat}$, hence $K_r = K_{app}/K_{mat}$. For a crack located at a stress concentrator, such as a nozzle penetration, the calculation of $K_{app}$ includes the stress concentration, the stress gradient and any residual stress associated with a nearby weld. If there were no effects of extensive plasticity on crack tip stress fields, two independent failure criteria could be contemplated: at $K_r = 1$ for all values of $L_r < 1$ and, for non-hardening material, at $L_r = 1$ for all $K_r < 1$. There is, however, a significant effect of plasticity on failure stress and the failure locus, as shown in figure 2, dips steeply as $L_r$ approaches 1. Its continuance beyond $L_r = 1$ reflects further resistance to plastic collapse as a result of strain hardening, with different ‘cut-offs’ corresponding to the different strain-hardening characteristics of different types of steel.

An assessment proceeds by focusing on a particular size of crack-like defect in the structure and the two ratios $L_r$ and $K_r$ are calculated. The values $K_r$ and $L_r$ are then taken as the coordinates of the assessment point on the FAD, point P in figure 2. If this point lies within the failure locus, the system is classed as safe. Points lying outside the failure locus would not normally be encountered unless a failure investigation were being carried out or the defect size or load were of postulated magnitude, chosen to explore excursions beyond the design intent. At any given time throughout life, the margin of safety is the ratio $OQ/OP$ where $Q$ is the point at which the extrapolation of OP (for the position of P at a given time) intersects the failure locus. If the crack length increases, owing to subcritical crack growth, both $K_r$ and $L_r$ increase, because $K_{app}$ increases and $L_{GY}$ decreases: the assessment point moves closer to the failure locus. A worked example of how a critical defect size can be deduced by carrying out a series of assessments for different postulated defect sizes (for a defect in the stress field of a large nozzle) is given in
reference [7], and an example of the use of an FAD analysis based on CTOD data is to be found in reference [2], describing the failure analysis of MV Kurdistan.

In service, a number of possible changes in material properties may occur and affect the position of P. At high temperature, material could become softened, increasing L. Hardening will be encountered, through irradiation hardening in RPVs or through strain ageing. Hardening decreases L, but there is also an associated increase in K (see below). An insidious form of strain ageing is brought about if marine vessels suffer minor collisions in warm climates, producing local plasticity that subsequently becomes aged. This renders those local regions more susceptible to brittle fracture if a major collision should occur in arctic conditions. Grain-boundary segregation decreases K\textsubscript{mat} and hence simply increases K.

5. Slow notched-bend tests and fracture toughness values

Much information on the criteria applicable to microscopically brittle fracture has been gained by studying the fracture behaviour in slow-bend of specimens containing blunt notches: see reference [6]. A particular geometry (45°V-notch, 0.25 mm radius, a/W = 1/3), subjected to four-point bending, has been analysed using finite-element techniques [10], and it is possible to gain full information of the stresses and strains in a local plastic zone ahead of the notch simply from the fraction of general yield load (L/L\textsubscript{GY}) at which the specimen breaks. The crucial result is that brittle fractures are observed to occur at a critical value of tensile stress, σ\textsubscript{F}, existing just behind the plastic/elastic interface. The tensile stress in this region is higher than the uniaxial yield strength, as a result of the development of a high hydrostatic component of stress in the constrained plastic zone. It is found that σ\textsubscript{F} is sensibly independent of, or only weakly dependent on, test temperature.

The fracture criterion in notched bars can be expressed simply as

\[ Q\sigma_Y = \sigma_F. \] (5.1)

Here, σ\textsubscript{Y} is the uniaxial yield stress and Q is a ‘stress intensification factor’, which increases in monotonic manner with L/L\textsubscript{GY}. The interpretation of σ\textsubscript{F} is that it represents the local tensile stress required to propagate a microcrack nucleus, of half-length or radius, c, initiated in a brittle, second-phase particle (carbide, carbonitride, oxide, silicate) by the high stresses induced by the formation of arrays of dislocations around the particle. The propagation condition is given...
Figure 3. Variation with temperature of local tensile stress beneath a notch at general yield and fracture stress (the inset shows a brittle fracture that has occurred just after general yield of the net section).

by applying the Griffith equation to the microcrack. For a through-thickness crack (e.g. grain boundary carbide film)

$$\sigma_F = \left( \frac{E'\gamma_p}{\pi c} \right)^{0.5}, \quad (5.2)$$

and, for a penny-shaped crack (e.g. cracked spheroidal carbide, inclusion)

$$\sigma_F = (\pi E'/c)^{0.5}, \quad (5.3)$$

where $\gamma_p$ is now defined as the effective work of fracture required for the (dynamic) brittle microcrack to extend into and across the ferrite matrix. From this model, it can be seen that the value of $\sigma_F$ is decreased by increase in $c$, the size of the brittle nucleus, or by a decrease in $\gamma_p$.

Referring to equation (5.1), it is clear that the ratio of $\sigma_F$ to $\sigma_Y$ determines the required value of $Q$, and hence of $L/L_{GY}$, at which a specimen breaks. It is convenient to define a ‘transition’ at the point of general yield (plastic collapse), $L/L_{GY} = 1$. The value of $\sigma_Y$ increases with decrease in test temperature and a ‘transition temperature’, $T_{GY}$, can then be defined such that $Q_{GY}\sigma_Y = \sigma_F$, where $Q_{GY}$ is the value of $Q$ for $L/L_{GY} = 1$ (using the Von Mises yield criterion). ‘Embrittlement’ is characterized by an increase in $T_{GY}$. This may be achieved in two distinct ways (figure 3). First, $\sigma_Y$ may be increased, (to $\sigma_Y + \delta$) by mechanisms such as strain-ageing, precipitation-hardening, irradiation-hardening or by testing at higher strain rate, and an increase in $T_{GY}$ is obtained, because it is possible to satisfy the expression $Q_{GY}\sigma_Y = \sigma_F$ at a higher temperature, $\Delta TT_1$ in figure 3. Second, $\sigma_F$ may be decreased (through increase in $c$ and/or decrease in $\gamma_p$) giving rise to a lower value of $\sigma_Y$ (higher temperature) at which $Q_{GY}\sigma_Y = \sigma_F - (the decreased value of $\sigma_F$), $\Delta TT_2$ in figure 3. It is, of course, possible to obtain an increased shift owing to the combination of the two effects $Q_{GY}\sigma_Y = \sigma_F - (the decreased value of $\sigma_F$), $\Delta TT_3$ in figure 3. A combination of this sort appears to be operative in the observed embrittlement of some VVER RPV steels: $\sigma_Y$ is increased through irradiation hardening and copper precipitation; $\sigma_F$ is decreased through phosphorus segregation to grain boundaries [11].

The critical tensile stress criterion has been used to explain the values of fracture toughness, $K_{1C}$, obtained in pre-cracked specimens failing by transgranular cleavage or brittle intergranular fracture. Because the dimensions of $K_{1C}$ are [stress][length]^{0.5} and those of $\sigma_F$ are [stress], it is clear that a length dimension must be incorporated into any relationship. In the original Ritchie, Knott and Rice (RKR) model [12], this ‘critical distance’, $X_o$, was found experimentally to be of the order of two grain diameters in normalized and annealed mild steel. More detailed investigations by
Curry & Knott [13–15] showed that the ‘critical distance’ was not a fixed multiple of grain size, but should be regarded as a statistical average, relating to the distribution of brittle (carbide) particles.

The ‘general yield’ transition used above to introduce the separate effects of increase in $\sigma_Y$ and decrease in $\sigma_F$ is for convenience and does not equate to the ‘cleavage/fibrous’ (brittle/ductile) transition observed in notched impact tests. This occurs at larger strains: it is possible to obtain cleavage fracture in notched bars above general yield, so long as yield is confined to the (work-hardening) net section: see reference [6]. Constraint is, however, lost when gross-section yielding occurs, and the strains increase rapidly, leading to ductile initiation. In pre-cracked test-pieces, a point is reached, as temperature is raised, at which the local maximum tensile stress (located at 1.9δ ahead of the crack tip) becomes equal to $\sigma_F$. Beyond this point, the model becomes unbounded, and the crack-tip would stretch open limitlessly were ductile initiation not to ensue. This point is then the transition.

6. Microstructural features of transgranular cleavage

The microstructure of normalized or annealed low-carbon steel is one of equiaxed ferritic grains with grain-boundary carbides. Distributions of spheroidal carbides may be obtained in a range of plain-carbon steels of different carbide contents by quenching and tempering at rather high tempering temperatures. In normalized or annealed higher carbon steels, the carbides may be present in pearlitic form, but these microstructures are of less relevance to structural steels, where the C contents are rarely more than 0.25 wt%, owing to weldability issues. Early measurements of $\sigma_F$ indicated that it increased with decrease in grain size, $d$; plots of $\sigma_F$ versus $d^{-0.5}$ showing a monotonic, but not linear, increase. This presented some difficulties when comparisons were made with the theoretical analysis of Smith [16], which predicted no effect of grain size, per se. The dilemma was resolved by a set of experiments [13] carried out to demonstrate that, in simply-cooled mild steels, carbide thickness increased monotonically with grain size, so that the variation of $\sigma_F$ with ‘grain size’ was, in fact, a reflection of the variation of $\sigma_F$ with carbide thickness, $2c$ as predicted by equation (5.2). Further studies on spheroidal carbide distributions showed that $\sigma_F$ was related in a linear fashion to $c^{-0.5}$ as predicted by equation (5.3), and that the ‘effective work of fracture, $\gamma_F$', was approximately 8–14 Jm$^{-2}$. The significance of this figure in terms of mechanisms of crack-tip separation at the atomic or ‘nano’-scale, will be discussed below, where it is argued that ‘shuffling’ of atoms in the region of the microcrack tip is required to effect the full separation of bonds.

Higher strength structural steels have been generally developed by quenching-and-tempering routes (e.g. A533B/A508, HY80, 2.25Cr1Mo), or by controlled-rolling to produce ultrafine grain size (pipeline steels, ULCB steels). The quenched-and-tempered steels may be employed to construct thick pressure vessels in which cooling-rates may be slow, so that higher-temperature transformation products, such as upper bainite, could be formed if hardenability were limited. Too high a carbon-equivalent; however, results in the HAZ next to a weld becoming susceptible to hydrogen cracking. If pre-heat is applied to reduce the possibility of cracking, this can reduce the cooling rate in the HAZ, again, possibly, promoting the formation of higher temperature transformation products. Coarse, upper-bainite products can also be obtained in the coarse-grained region of the HAZ in weldments. A concern with ultrafine grain size UCLB steels is that the particles that pin grain boundaries are ineffective at temperatures of order 1200–1350°C in the HAZ, leading to coarse grain size and a marked deterioration in properties. There is also a concern that such coarse-grained material, subsequently heated into the intercritical region in a subsequent pass, may develop coarse carbides or martensite/austenite/carbide (MAC) product.

Bowen and co-workers [17,18] made a detailed study of the effects of microstructure on cleavage fracture in A533B steel (figure 4). Broadly, the values of $\sigma_F$, which are, at most, weakly dependent on temperature, fall into two classes: some 3200–3800 MPa for the fine distributions found in auto-tempered martensites and lower bainites; 1200–2200 MPa for the coarser upper bainites and other ferrite/carbide mixtures. These results demonstrate clearly the effect of carbide thickness (‘c’ in equations (5.2) and (5.3)) on the value of local fracture stress. A similar range of
values has been obtained by Balart [19] for the RPV steel DIN 22NiMoCr37, used to generate
the European fracture toughness database. These figures compare with values of order 700–
1400 MPa for mild steels with ferrite grains and grain-boundary carbides (figure 4). Zhou Wei has
measured values in the range 1400–2200 MPa for carbon–manganese (C/Mn) steels in a granular
bainite condition, and 2600–2900 MPa for 50D steel in the quenched and tempered condition
[20]. Some effects of temper embrittlement on local fracture stress in structural steels are given
in [21]: in 2.25Cr1Mo steel, the value reduced from 2950 to 2500 MPa; in A533B, from 2400
to 2200 MPa.

Bowen showed that carbide thickness is the controlling feature, even for particles in the nm size
range. The $\sigma_F$ results can be reconciled with a value of $\gamma_F$ in the range 8–14 Jm$^{-2}$. The transition
temperatures for the microstructures depend on the factors implicit in equation (5.1): not only is $\sigma_F$
high for fine microstructures, but $\sigma_Y$ is also high. Bowen’s work encompassed also the effects
of microstructure on fracture toughness, $K_{1c}$, values in A533B steel [22]. The conclusions here were
that the fracture toughness was also controlled by carbide size. At low temperatures, where tensile
stresses are high, even in a small ‘process zone’ ahead of the crack-tip, the size corresponded to the
mean of the distribution, but at higher temperatures, as the tensile stresses decreased, the carbide
sizes corresponded to those at the top end of the size distribution. As for the findings in weld
metals described below, this could be an effect of the ‘warm pre-stressing’ introduced during the
(room temperature) fatigue pre-cracking procedure.

In thick section, it is possible, owing to limited hardenability, to obtain mixtures of coarse and
fine microstructures. The rolling of thick plate can also give rise to ‘banded’ microstructures. The
consequences of this on distributions of fracture toughness and extrapolations to low probabilities
follow from the work of Zhang & Knott [23,24] and are further discussed in references [25–27];
supporting evidence is given in references [28,29]. Zhang produced probabilistic distributions of
fracture toughness in A533B at −80°C for auto-tempered martensitic and bainitic microstructures.
Both conformed to Gaussian distributions (the cumulative distribution functions, CDFs, plotting
as straight lines on normal probability paper, with standard deviations of order 5 MPam$^{-0.5}$).
Results were also obtained for a mixed microstructure: nominally 30% bainite, 70% martensite,
but exhibiting some banding. All data points for the mixed microstructure lay between the two
linear CDFs for the individual microstructures, but the spread was much wider, because the
pre-crack tips in some samples were located in bainite: in others, in martensite. (The ‘nominal’
standard deviation was 22 MPa m$^{-0.5}$). The importance of these observations relates to attempts
to obtain lower bound values of fracture toughness for safety-case assessments. These are often
obtained at the $10^{-4}$ (0.01%) level. When this limit was applied to Zhang’s results, the lower
bound for martensite was 65 MPa m$^{0.5}$ and that for bainite was 22 MPa m$^{0.5}$. If the bounding CDF
‘tramlines’ for the individual microstructures results were not available or known to the analyst,
the data points for the mixed microstructure might be treated as belonging to a single population.
Extrapolation to the 0.01% level then gives a negative value! An effect similar to this was obtained
when determining the lowerbound ‘cut-off’ level when ‘free-fitting’ Weibull distributions to the
fracture toughness results contained in the (‘round Robin’) Euro dataset: a number of these were
also found to be negative [30]. A pragmatic decision was taken to fix a lower bound at 20 MPa m$^{0.5}$
(quite similar to Zhang’s 0.01% bound for bainite, but only one-third of that for martensite). This
is now incorporated in the master curve approach, but note that if a Weibull modulus of 4 is held
to apply to the raw dataset, it cannot have the same value of 4 for the dataset with values reduced
by 20 MPa m$^{0.5}$; and vice versa.

A broader point arising from issues such as hardenability in thick sections (or spatial
segregation of both alloying and impurity elements during casting and forging, which affects
both hardenability and susceptibility to embrittlement) is the need to extract fracture toughness
specimens from appropriate regions of forgings (usually from a prolongation), with thought as
to how the fatigue pre-crack tips should be best located. Logically, this should not be such as to
give the ‘best’ result in terms of a high fracture toughness value, but to give the most appropriate
measure of fracture toughness relevant to the perceived failure scenario. The Sizewell RPV walls
are, for example, approximately 200 mm thick and the ‘rule-book’ dictates that the specimen
should be extracted with its crack tip corresponding to the 1/4T position, i.e. 50 mm below the
surface. Arguably, the main threat to the system is adjudged to be the ‘loss-of-coolant-accident’,
in which, as a result of flooding the vessel with cold water to prevent over-heating of the core,
high thermal stresses are generated at the inside wall of the vessel. Early (albeit erroneous: see
‘brief case study’ in reference [5]) calculations suggested that defects of some 6–7 mm in ‘height’
might propagate into the vessel wall. The relevance to this scenario of fracture toughness data at
50 mm depth in a steel of limited hardenability is not obvious.

A detailed study of spatial variability in the toughness of multipass C/Mn welds was carried
out by Todinov et al. [31]. Here, the scale of ‘meso’ structure was typically a few mm (between
coarse-grain, as-deposited material, and fine-grain, grain-refined material), and the effects were
demonstrated in Charpy specimens, having a 0.25 mm radius notch. As for Zhang’s ‘mixed
microstructure’ dataset, it was found that the scatter associated with data for ‘all specimens’ could
be greatly reduced if the spatial effects were taken into account. The advantage of doing this is
that such specimens are used as surveillance specimens in nuclear reactors, to estimate the shift
of the transition temperature as a function of neutron fluence. In many cases, both the original
(non-irradiated) dataset and the irradiated dataset exhibit such a wide scatter that it is difficult
to attach high confidence to the trends. Knowledge of the microstructural state at every notch or
crack tip would reduce scatter and improve precision in estimation.

Studies of the microstructural and fractographic features associated with cleavage fracture in
(well characterized) C/Mn weld metals have made major contributions to the understanding of
basic mechanisms. The welds generally contain a high volume fraction of deoxidation products,
such as oxides or silicates. One effect of these is to provide a number of intragranular sites
for ferrite nucleation. The microstructure of a deposited C/Mn weld deposit consists of a
combination of grain-boundary ferrite, ferrite with aligned carbide and acicular ferrite, nucleated
on intragranular sites, which may often be inclusions. Overall carbon contents are low (typically,
less than 0.13 wt% C) but very small regions of MAC (even pearlite) can be detected in some
deposits. In alloy steel deposits, the structures are more generally (low carbon) martensite
or bainite.

For a range of weld metal deposits (C/Mn, A533B, 2.25Cr1Mo), it is found that cleavage
fracture nucleation is associated with oxide/silicate inclusions, which are brittle particles, firmly
Figure 5. Cumulative distributions of fracture stress in weld metals (Wu).

‘clamped’ into the ferrite matrix by the stresses induced by thermal expansion mismatch [32,33]. The great advantage of this is that the inclusions can be identified on the fracture surface without etching; their sizes can be measured, and their locations fixed, with respect to the position of maximum stress ahead of a notch or pre-crack. The relationship between \( \sigma_F \) and \( c^{-0.5} \) is again linear, with a slope conforming to a \( \gamma_p \) value in the range 8–14 Jm\(^{-2}\). In blunt-notched bars, the location of the crack nucleus is always within the plastic zone, at, or just behind, the position of maximum tensile stress. In these tests, the observations tend to relate to inclusions of a size corresponding to the top end of the distribution and there is therefore a ‘trade-off’ between the size of inclusion, \( c \), equation (5.3) and the value of \( Q \), equation (5.1), required to give the critical condition. A comprehensive set of data for inclusion-initiated fracture is given in reference [34] and is shown schematically in figure 5. Two conditions were explored: ‘as-received’ and ‘degraded’ (with coarse grain-size, pre-strained to simulate irradiation hardening and then subjected to a temper-embrittling treatment). Observations were made of the sizes of crack-initiating inclusions, found to lie generally in the range 2.2–2.5 \( \mu m \) diameter. For the median fracture stress in as-received material of 1555 MPa, an average size of 2.35 \( \mu m \) implies a local \( \gamma_p \) work-of-fracture of approximately 8 Jm\(^{-2}\). The ‘degraded’ mean fracture stress of 1400 MPa implies a \( \gamma_p \) value of approximately 6.4 Jm\(^{-2}\) (and the fractures were intergranular). These values will be discussed in §7.

It is observed that inclusions also serve as the sites for microcrack nuclei in fracture toughness (\( K_{IC} \)) tests and it appears therefore that the combination of fracture criterion, stress analysis and fractographic observation should be sufficient to calculate values of ‘critical distance’. When such calculations are made, agreement with RKR/Curry, Knott theory [12,15] is not good at the lower temperatures and it appears that the ‘pre-strain’ or ‘warm-pre-stressing’ (WPS) effect on \( \sigma_F \) induced by the (room-temperature) fatigue pre-cracking process employed for \( K_{IC} \) tests is affecting low-temperature behaviour. Better agreement is obtained with theory at higher temperatures, when the process zone associated with fracture is significantly larger than the fatigue ‘pre-strained’ zone [33]. Although the effect of WPS is related to the residual stress distribution created on unloading, a further interesting observation is that the sizes of crack-initiating inclusions in specimens subjected to WPS are not those at the top end of the size distribution but are more typical of the ‘average size’ [35]. The more virulent microcrack nuclei appear to have been blunted out in the high-temperature ‘pre-stress’. 
7. Fracture at the nanoscale: the ‘effective work of fracture’

The picture that emerges from the detailed microstructural studies is one in which a microcrack is initiated in a brittle particle such as a carbide or oxide/silicate by the local stress generated at the head of an array of slip dislocations. The condition that needs to be met for this microcrack to propagate in a brittle manner across the surrounding ferrite grains, however, is that a critical value of tensile stress must be attained in the plastic zone ahead of a blunt notch or (an initially sharp, but, at the point of fracture, blunted) fatigue pre-crack tip. The hydrostatic component in this constrained plastic zone is such that, for an initially sharp crack, the tensile stress can be up to 2.5 (Tresca) or approximately 3 (Von Mises) times greater than the uniaxial yield stress [6]. The fact that the microcrack travels across the brittle particle at high velocity is of importance, because this means that there is insufficient time for dislocation sources in the ferrite matrix close to the pre-crack tip to operate. Reliable values for the critical stress $\sigma_F$ are obtained from blunt-notched test-pieces (as in figure 5), in which no warm pre-stressing effects are involved.

Treating the inclusion-initiated microcracks as penny-shaped cracks, the median fracture stress for the ‘as-received’ material in figure 5 (1555 MPa), may be combined with an average inclusion diameter of 2.35 $\mu$m to give a local $\gamma_p$ work-of-fracture of approximately 8 Jm$^{-2}$. Other results, derived from the fracture stresses in wrought steels containing carbides, suggest a range of order 8–14 Jm$^{-2}$ [14,18]. It should be emphasized that the very high amounts of work associated with failure in standard fracture toughness tests relate to the precursor work required to develop a sufficiently large plastic zone ahead of the fatigue pre-crack to generate the required level of tensile stress to propagate a microcrack. The value, 8–14 Jm$^{-2}$, is a few times that of the generally accepted value for the energy of two surfaces of iron: $2\gamma = 2$ Jm$^{-2}$; although Kelly, Tyson and Cottrell (KTC) [36] suggest 4 Jm$^{-2}$ at 0 K.

It is of value to examine two rather different approaches to these criteria. The first relates to the theoretical strength of an uncracked body. Here, the usual approach is to start from the energy of interaction between two atoms, figure 6a, and then to derive from this, by differentiation, the atomic force–displacement curve (figure 6b) [6]. This exhibits a maximum at the point of inflection of the energy/distance curve; but note that the energy of interaction between atoms at this point is still negative, i.e. there is still attraction between the atoms. What is then done, in simple treatments, is to approximate the atomic force–displacement curve to a sine law (not critical, but it makes integration easier) and to equate the area under the curve to 2$\gamma$ $\sigma_{\text{max}} = \left(\frac{E\gamma}{b_0}\right)^{0.5}$, (7.1)

where $b_0$ is the original lattice spacing. Substitution of appropriate values gives a result of order $\sigma_{\text{max}} \sim E/10$. More sophisticated forms can be used for the energy of interaction, but a rather crucial assumption is that the work of fracture is equated to just that of two free surfaces and no more. For a defect-free lattice subjected to force control, the system will become unstable at the force maximum (rather like the fracture of a violin string, if over-tautened during tuning, or the failure of a tensile specimen at its UTS), but instability is by no means obvious if the bond that is to fracture is located at the tip of an atomically sharp crack, where all the surrounding atoms have positive stiffness. This is analogous to the situation in a displacement-controlled testing system, in which a neck can form in the tensile specimen and develop to a degree in a stable manner: even more so if the system is unloaded after the neck has started to form and a stiff ‘collar’ is attached, so that further displacement involves increasing load, but no instability. Under these circumstances, the neck geometry can be taken to high reductions-of-area without failure. At a crack tip, the bond that is first extended to the maximum stress is, at this point, surrounded by a ‘collar’ or an ‘iron cage’ of atoms, with positive stiffness (figure 7).

Nanoscale treatments of the balance between brittle fracture or plastic blunting ahead of an initially sharp crack have generally followed the approach developed by KTC [36] and subsequently developed by Rice & Thomson (RT) [37]. KTC considered a range of materials, and compared the value of theoretical tensile strength, $\sigma_{\text{max}} \sim E/10$ (which assumes a work of
Figure 6. (a) Energy of interaction between two atoms as a function of distance of separation. (b) Atomic force–displacement curve derived from (a).

Figure 7. The ‘iron cage’ of atoms surrounding the crack-tip bond.

fracture of $2\gamma$) with that required to blunt the crack by the generation at, and movement away from, the crack tip of dislocations, at a theoretical shear strength of $\mu/10$, modified by appropriate geometrical factors to allow for the operative slip systems. The KTC approach provides a rationale for why some materials are intrinsically brittle, and others are intrinsically ductile. The conclusion for iron is that it is a ‘borderline’ material, consistent with the fact that it exhibits a ductile/brittle transition temperature. RT approached the problem in a somewhat different manner, by analysing the emission of dislocations from a crack tip at the Griffith stress. (This also assumes that the work of fracture is $2\gamma$). Again iron was found to be a borderline material. Further work of this sort has been reviewed by Xu [38].

Both KTC and RT assume implicitly that the work-of-fracture is that of two free surfaces, $2\text{Jm}^{-2}$ (KTC $4\text{Jm}^{-2}$ at 0 K). Experimental values for the work-of-fracture, derived from $\sigma_F$ values, lie in the range 8–14 $\text{Jm}^{-2}$ and it is of interest to explore this discrepancy. Even setting aside the restructuring of surfaces after fracture, it should be noted that the Griffith energy balance,
in equating energy release rate to the energies of two surfaces, treats two states: a before state, comprising the strained bonds; and an after state, comprising the two surfaces. It does not address the question of how the system proceeds from one state to the other, i.e. whether it might be necessary to surmount any activation energy barrier. If such a barrier were to exist, it would be measured as the work-of-fracture and would be greater than two surface energies. What the discussion above suggests is that the simple concept of the bonds in a defect-free solid, under force control, flying apart at the maximum stress has to be re-examined for the highly strained bond at the crack tip, which is surrounded by the ‘iron cage’ of atomic bonds still possessing positive stiffness (figure 7). Note that at the maximum stress (figure 6) there is still a negative energy of interaction (i.e. attraction) between atoms. What is suggested is that it is necessary for some further cooperative movement (‘shuffling’) of the atoms surrounding the crack-tip bond to allow that bond to be displaced sufficiently for the attraction between the atoms to fall to zero. The extra work associated with this ‘shuffling’ would then represent an activation energy and be measured as a ‘work-of-fracture’.

Pictorially, this could be envisaged as dislocations generated at the crack tip moving one or two Burgers vectors on either side to facilitate the bond extension (figure 8). The work done by a stress of order 0.1 \( \mu \) moving through such distances is sufficient to cover the discrepancy between the experimental ‘work-of-fracture’ and two surface energies. Interestingly, if the dislocations did not have to exceed the RT [37] ‘saddle-point’ configuration, as soon as the crack-tip bond fractured and a free surface was created, the dislocations would be drawn back to the free surface by image forces and disappear. A similar concept, involving ‘shuffles’ around crack tips in glass, was proposed by Marsh [39] to explain why dry glass did not behave as a perfectly brittle solid in the Griffith sense. It might also be noted that when Obreimoff [40] fractured mica in vacuum, rather than laboratory air, he did not achieve ‘reversible’ cleavage, but instead observed a plasma glow as a result of exposing charged bonds.

This paper has made only slight mention of intergranular fracture, because that is treated elsewhere in this theme issue [4], but it is clear that segregation of elements such as P, Sn to (ferrite boundaries coincident with) prior austenite grain boundaries reduce the values of \( \sigma_F [21] \) and hence the work of fracture. The suggestion is that the impurity element reduces the strength of the iron–iron bond (by forming a metalloid bond, or by affecting the distribution of the three-dimensional bonding electrons between the iron atoms). Any electronic perturbation around an impurity atom arguably occurs where-ever one is present, and it is the accumulation of many of them close together at grain boundaries that produces the low ‘work-of-fracture’ path. From the tip of a sharp crack, the crack path will follow the embrittled boundary, but any dislocations moving away (at right angles) from the tip move into non-segregated material. If the separation processes occur at lower stress levels in embrittled material, the shear stresses around the crack tip will also be lower, and there will be less movement of, and hence less work done by, the ‘accommodating’ dislocations.
8. Conclusion

This paper has attempted to summarize the main features associated with brittle cleavage fracture in structural steel over a wide range of size-scales. In the space available, it has been necessary to summarize much detailed research, which can be found in extenso in the references. Structural integrity assessment at the macroscale rests on both structural analysis and the application of fracture mechanics. Critical inputs to these assessments are materials’ test data, but there are issues relating to how test-pieces should be extracted from thick sections and how estimates of data at low failure probabilities should be made. These are of particular concern when there are spatial variations of microstructure at the mesoscale. At the microstructural scale, brittle cleavage fracture can be well characterized in terms of the attainment of a critical value of the local tensile stress in a plastic zone ahead of a blunt notch or an initially sharp fatigue pre-crack. The pre-crack blunts as its plastic zone develops and the plastic work associated with a conventional fracture toughness value is precursor work, required to develop the critical value of local tensile stress in the plastic zone.

The values of the critical tensile stress can be related to the thicknesses or diameters of brittle second-phase particles in which microcracks are initiated by dislocation arrays, following a Griffith-type analysis. It is found that the local work of fracture is in the range 8–14 J m\(^{-2}\), a few times the energy of two free surfaces. It is argued that the conventional Griffith analysis considers two states: a ‘before’ state and an ‘after’ state, but does not consider whether there might be any activation barrier to proceeding from one state to the other. A suggestion as to how this might occur is proposed.

Acknowledgements. I have discussed the topics covered in this paper with scores, if not hundreds, of people over the last 50 years, starting with the late Sir Alan Cottrell, Dr Alan Wells and Professor Ted Smith; also with Professor Jim Rice and Professor Charles McMahon. The main inputs here, however, have come from ex-research students and others in my research teams at Cambridge and Birmingham. Special mention should be made of Dr Richard Dolby, Professor Rob Ritchie, Dr David Curry, Dr James Tweed, Dr Yukito Hagiwara, Professor Paul Bowen, Dr David McRobie, Dr David Neville, Professor Philippa Reed, Dr XiaoZhong Zhang, Professor SuJun Wu, Dr Maria Balart, Dr Rengen Ding and Dr Milorad Novovic. I have drawn on their results freely, but they can in no sense be held responsible for my more speculative proposals.

The references cited are those related to technical issues raised in books, refereed journals, conference proceedings or documents issued by authorized bodies. Information on failures mentioned in the Introduction section may be obtained by typing in the appropriate name on a web search engine.

References


