Folding

J W Cosgrove, Imperial College London, London, UK

Introduction

A fold is defined as "...a curved arrangement of originally parallel surfaces ..." and a large range of terms have been used to describe geological folds. These including folds, flexures, inflexions, bendings, plications, undulations, and crenulations. Although today most geologists use the term fold for buckle folds and many reserve flexure for layer deflections caused by bending, the terms are also used as synonyms.

Folds form on all scales, from those with a wavelength of a few mm that can only be seen in rock thin sections under the microscope, to folds with wavelengths in excess of 10 km. The mechanisms of formation are independent of size; however, for large-scale folds it is important to take into account the effect of gravity when analysing their folding behaviour.

Folds in rocks form under a wide variety of conditions. For example, folds can occur at all depths in the crust from early, near-surface folding linked to the slumping of water saturated, uncemented sediments down continental slopes, to folding of rocks under the high pressures and temperatures encountered in the lower crust. Folding of rocks occurs by various processes, the three most important being buckling, bending, and flowing. Buckling, which is the most common type of fold, requires the rock to possess a mechanical anisotropy, usually layering, and for the maximum compressive stress to act parallel or subparallel to the layering. In contrast, bending is defined as a transverse deflection of a layer or beam by a transverse couple. Flow folding occurs during the flow of a material such as lava, salt, or ice. No mechanical anisotropy is necessary for this type of folding, only passive marker bands that reveal the flow patterns within the material.

Although folds are structures that are characteristic of ductile deformation, they are often found in association with fractures (*see* Tectonics: Fractures (Including Joints)). The fractures result from the same stress field responsible for folding and the resulting fracture patterns, which are controlled by the mechanisms by which the folds form, can play an important role in hosting mineralization and in the storage of fluids such as water and hydrocarbons.

Fold Geometry

The main geometric parameters relating to the three-dimensional geometry of a fold are shown in Figure 1A. However, although folds are three-dimensional structures they are most commonly exposed as two-dimensional sections on joint surfaces. Consequently, their detailed classification is based on the geometry of a section. Because of the dependence of fold geometry on the orientation of the section (Figure 1B), a particular section is used to determine the geometry, namely the profile section, i.e., the section at right angles to the fold hinge. Historically, the problem of describing layer shape developed around two geometrical models, the parallel fold and the similar fold (Figure 2A and B). As the name implies, in parallel folds the orthogonal thickness of the layer remains constant. In similar folds, the layer thickness, parallel to the hinge surface remains constant. Although similar folds show considerable variations in layer thickness, this type of folding can produce folds which can extend indefinitely in the profile section, whereas with parallel folds this is not possible (Figure 2A and B).

As can be seen from Figure 2C, these two fold types are examples from a complete spectrum of possible geometries. The spectrum has been divided into five classes based on the pattern of dip isogons, i.e., lines

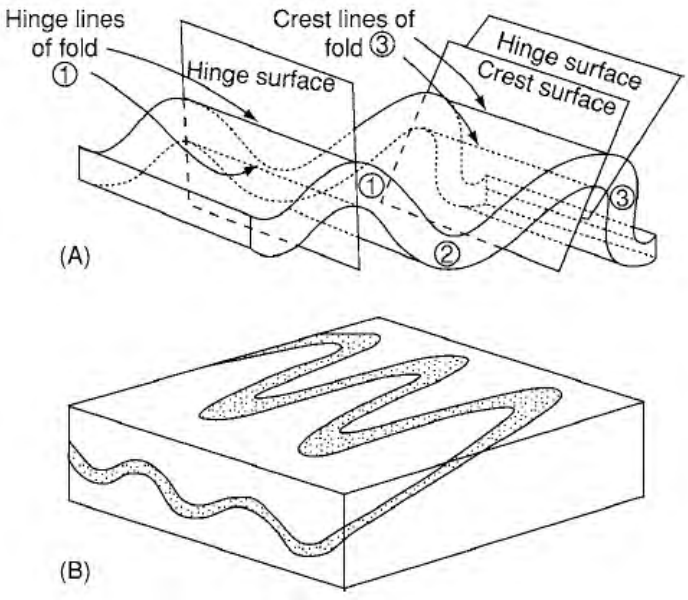

Figure 1 (A) Geometric features of a fold. Fold 1 is symmetric and folds 2 and 3 asymmetric. (B) Diagram showing the dependence of fold outcrop pattern on the orientation of the plane of exposure.

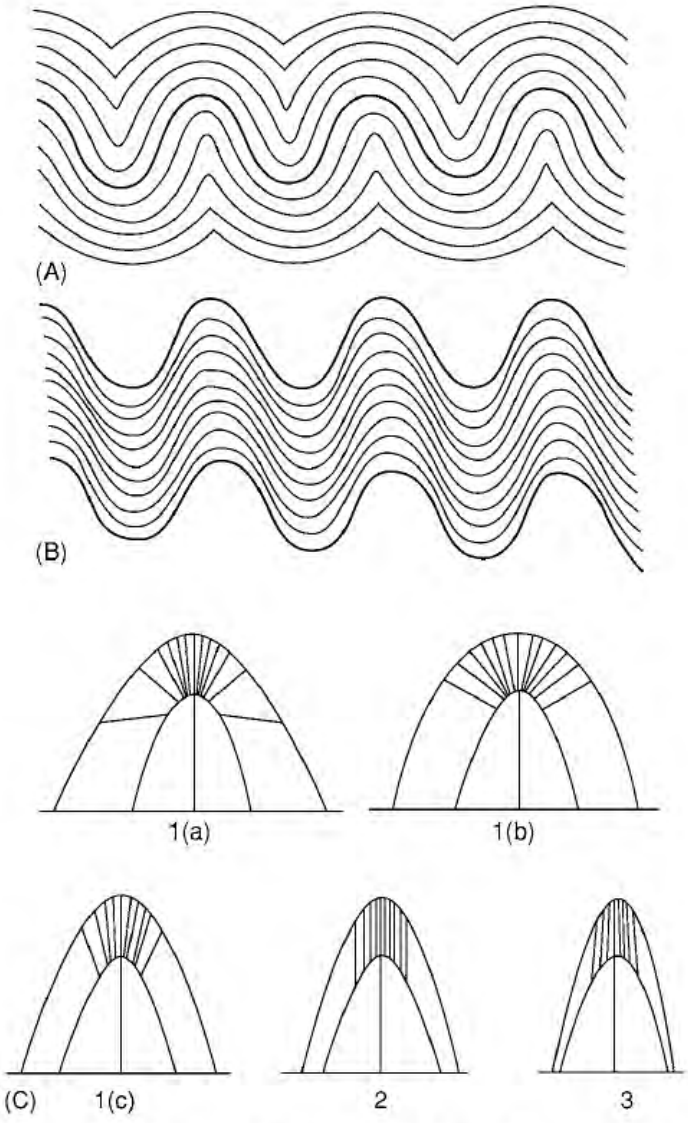

Figure 2 (A) parallel and (B) similar folds. (After Van Hise (1894).) (C) Classification of fold profiles using dip isogon patterns. 1(a) strongly convergent, 1(b) parallel, 1(c) weakly convergent, (2) similar, and (3) divergent (Ramsay (1967)).

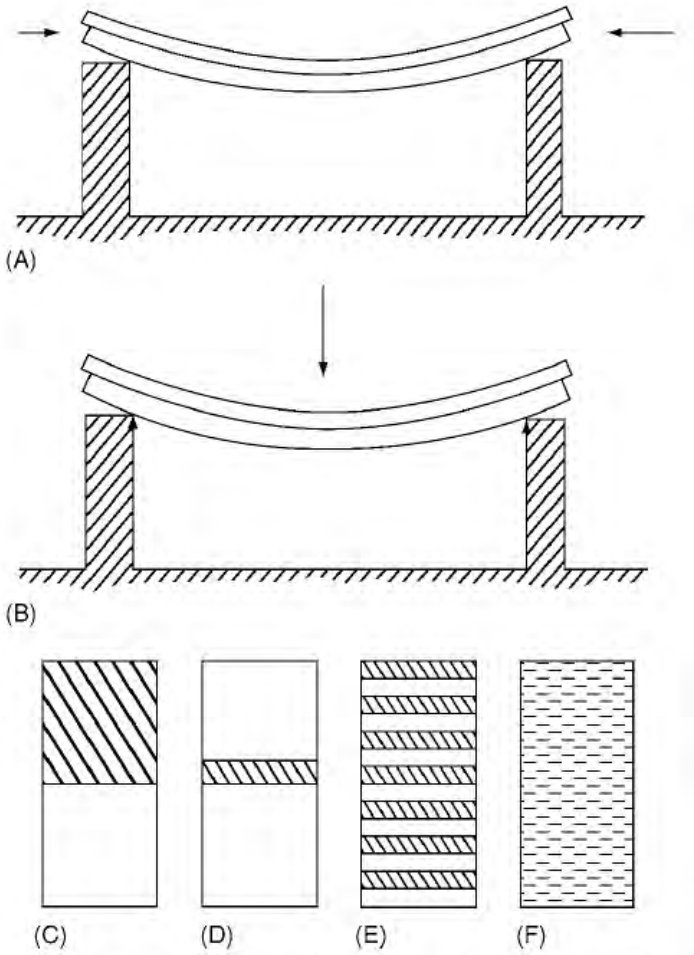

Figure 3 The orientation of the principal compression for (A) buckling and (B) bending of the layers. (C) an interface between two unlike materials; (D) a single layer; (E) a multilayer; and (F) a mechanically anisotropic material. The buckling behaviour of these systems is discussed in the text.

joining points of equal dip on the two surfaces defining the folded layer. The parallel folds and similar folds are class 1b and 2, respectively.

Mechanisms of Folding

The two most important mechanisms by which folds form in rocks are buckling, the result of compression parallel or sub-parallel to the rock layering and bending, the result of compression at a high angle to the layering, (Figure 3A and B, respectively).

Buckle Folds

The most commonly formed folds in the Earth's crust are buckle folds. In order for these folds to form, the rock must possess a mechanical anisotropy. This is generally a planar mechanical anisotropy because a variety of geological processes give rise to such materials. Rocks can be intrinsically anisotropic because of the process of their formation such as, for example, a bedded sedimentary succession, or the anisotropy can be induced as a result of subsequent deformation and metamorphism during which time they can develop planar and linear mineral fabrics as they are converted to slates, phyllites, schists, and gneisses. Buckling systems can be sub-divided into four groups, namely folds formed by: (i) the buckling of a single interface, (Figure 3C); (ii) the buckling of two interfaces, which define a single layer in a matrix (Figure 3D); (iii) the buckling of several layers (Figure 3E); and (iv) the buckling of a mineral fabric such as a slate or schist (Figure 3F).

Interface Buckling

Interface buckling occurs on many scales, from large buckles that form at the boundary between the Mesozoic cover rocks and older basement in the French Alps as a result of the collision of the African and

European plates (Figure 4A), to small-scale examples directly observable in the field (Figure 4B) and formed in analogue models in the laboratory (Figure 4C). The folds start as symmetric sinusoidal deflections but as they amplify change their geometry into the marked cusp geometry seen in Figure 4. The cusps always point into the stronger of the two materials.

Single Layer Buckling

Single layers occur commonly in nature, for example, an isolated sandstone or limestone bed in a thick shale or marl sequence or a sheet of igneous rock intruded into an unlayered matrix (Figure 5A), and they are commonly observed to buckle. Theoretical analyses

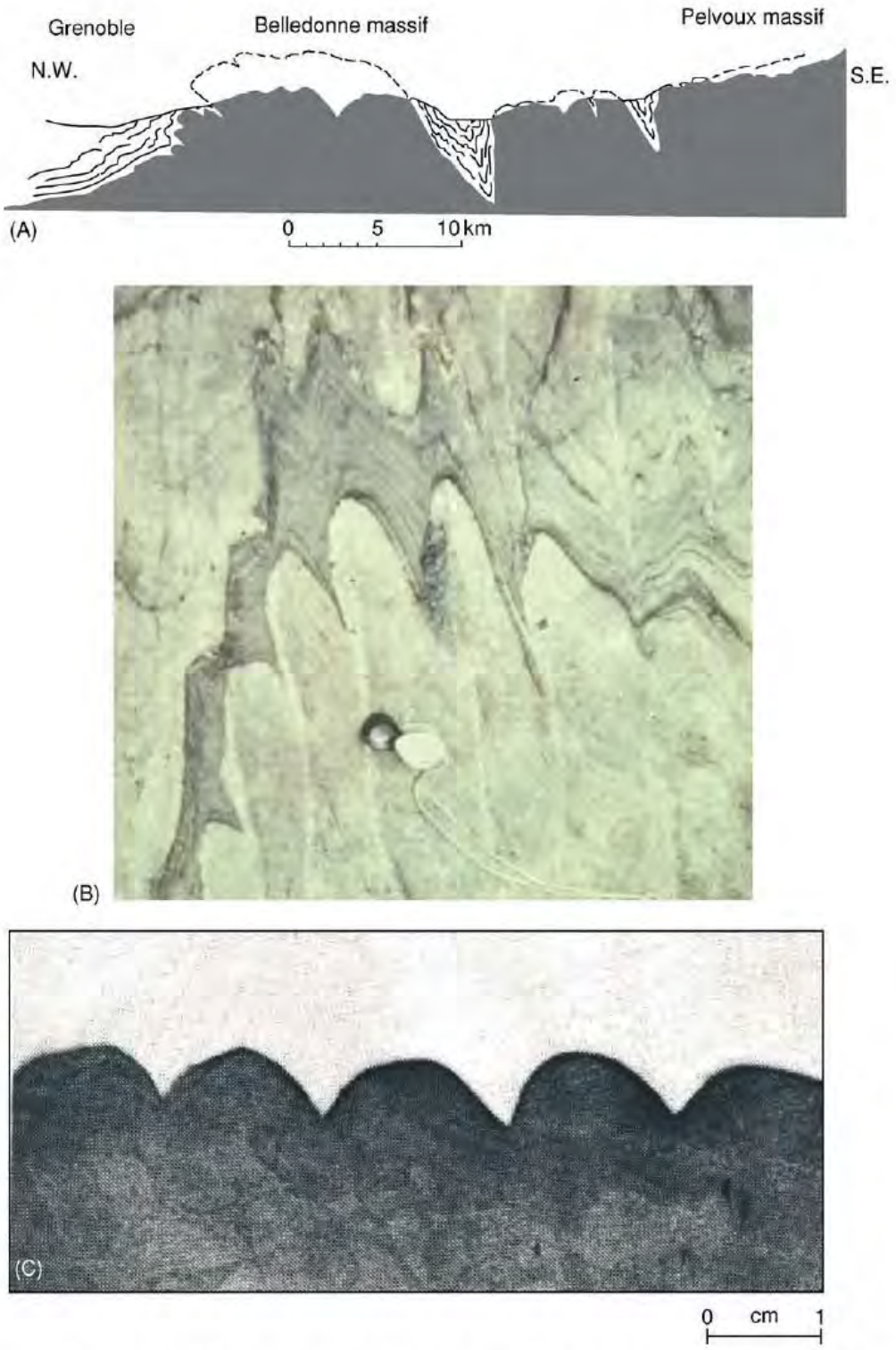

Figure 4 (A) The cuspate interface between the strong Hercynian basement rocks (black) and the weaker Mesozoic cover rocks, caused by the horizontal compression linked to the collision of the African and European plates. (B) Cusp structures formed at the interface between quartz rich (light) and mica rich (dark) bands, Loch Monar, Scotland. (C) Cusps generated by compression parallel to the interface between strong (black) and weak (light) plasticine.

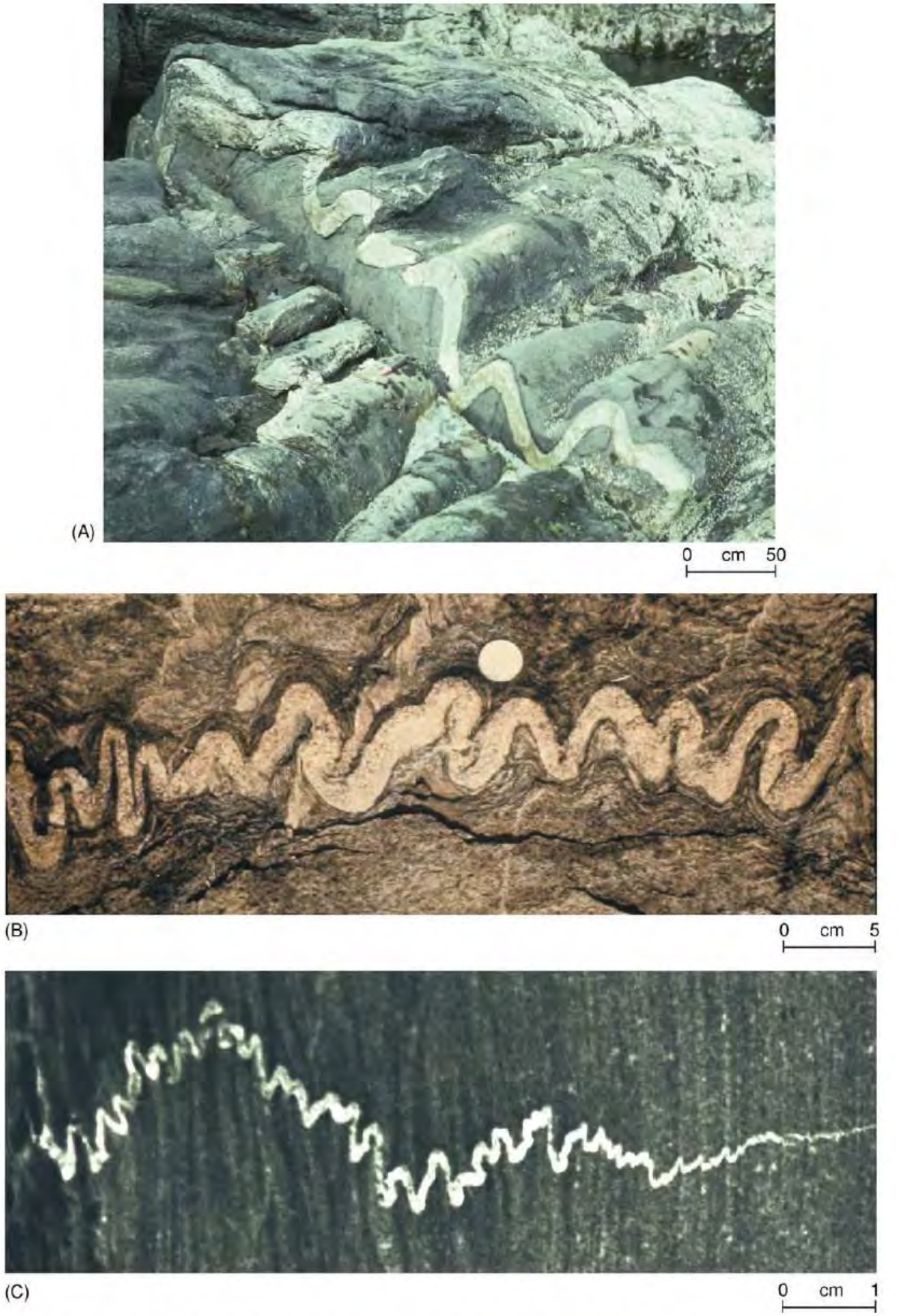

Figure 5 (A) A folded aplite dyke cutting a dolerite body, Outer Hebrides, Scotland. (B) A folded single layer, in which the folds are isoclinal. No further shortening of the layer can occur by limb rotation. Further compression may result in homogeneous flattening of the folds or in the folding of the layer into larger wavelength folds (see C). (C) A tapered quartz vein showing buckling on two scales. On both scales, the wavelength is seen to decrease with decrease in layer thickness.

predict and field observations confirm that there is a direct relationship between the wavelength (W) of folds that develop and the thickness (t) of the buckling layer (Figure 5C).

$$W = 2\pi t(\mu_1/6\mu_2)^{1/3} \quad [1]$$

The single layer buckling equation shows that in addition, the wavelength is also controlled by the ratio of the strength of the layer and matrix (μ_1/μ_2). The impact of the strength ratio (competence contrast) on fold style can be seen by rearranging eqn 1.

$$W/t = 2\pi(\mu_1/6\mu_2)^{1/3} \quad [2]$$

The wavelength/thickness ratio is determined by the relative strength of the layer and matrix. Folds with a range of W/t ratios are shown in Figure 6, which shows the control of the competence contrast between the layer and the matrix on the geometry of the buckles that form.

Once buckling has been initiated, layer shortening can continue by rotation of the fold limbs until the limbs become parallel to each other (i.e., the folds are isoclinal, Figure 5C), and no further shortening by limb rotation can occur. Shortening may continue by homogeneous flattening of the buckled layer and matrix, or alternatively continued compression may cause the buckled layer to buckle again. It has an effective thickness that is considerably greater than the original thickness of the layer and so, in accordance with the buckling equation eqn 1, will buckle with a larger wavelength (Figure 5C).

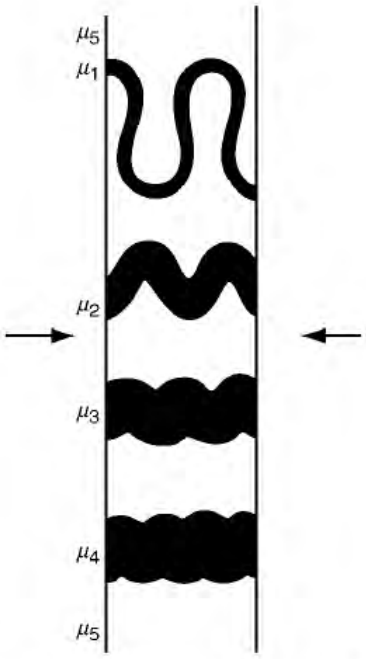

Figure 6 The effect of the ratio of the strength of the layer and the matrix on fold style. As the contrast decreases the wave length/thickness ratio also decreases. (After Ramsay (1982).)

Multilayer Buckling

The most common type of geological layering is that of a multilayer, i.e., a succession of different layers. Often the multilayer is made up of the regular alternation between two or three rock types. The regularity of these multilayers reflects the processes by which they were formed. For example, turbidites are formed as a result of the periodic, fluid-induced collapse of a sedimentary accumulation on the edge of the continental shelf. The fluidised sediments flow as dense turbidity currents down the continental slope, depositing first the heavy sands followed by the slower settling out of the fine shale particles. In this way a sequence of alternating sandstones and shales can be built up to form a turbidite.

The buckling behaviour of multilayers is controlled by the spacing of the strong units it contains. Figure 7A and B are both physical multilayers. However, their buckling behaviours are very different. The multilayer shown in Figure 7A behaves mechanically as a series of single layers, each layers folding

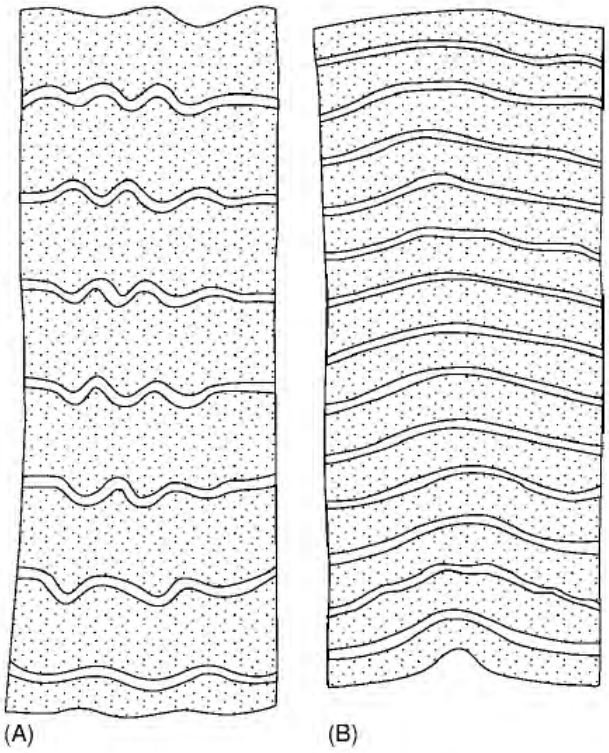

Figure 7 (A) and (B) show two compressed multilayers of alternating competent (white) and incompetent (stippled) rubber layers. The widely spaced layers (A) behave mechanically as single layers and form disharmonic folds; the more closely spaced layers behave mechanically as a multilayer and all layers form the same wavelength. (After Ramberg (1963).)

according to the single layer buckling equation, (eqn 1). This results in 'disharmonic folding'. In contrast, the multilayer shown in Figure 7B behaves mechanically as a multilayer, i.e., all the layers develop the same wavelength and amplitude. Such folding is called 'harmonic folding'.

These two types of buckling behaviour can be readily explained by considering the strain that develops in the matrix around a single, competent (i.e., strong with respect to the surrounding matrix) layer as it buckles. The zone of disturbance on each side of the buckling layer is known as the zone of contact strain. If the competent layers of a multilayer are sufficiently far apart for there to be no significant overlap of their zones of contact strain, then each layer will buckle as a single layer. If, however, the zones of contact strain of adjacent competent layers do significantly overlap, the layers can no longer buckle independently of each other. The zones of contact strain and associated zones of contact stress of adjacent layers must be compatible and, as a result, all the layers are subjected to the same stress field and develop the same wavelength.

In order to determine how close the competent layers of a multilayer must be for multilayer as opposed to single layer buckling to occur, it is necessary to know how far the zone of contact strain extends away from the layer into the matrix. For a viscous matrix it is found that the disturbance has died down to approximately 1% of its maximum value at a distance one wavelength from the layer (Figure 8).

The Buckling of Anisotropic Materials

The most complex buckling occurs in materials possessing a pervasive mechanical anisotropy. The anisotropy may be an intrinsic property of the rock resulting, for example, from the bedding parallel alignment of clay particles in a shale or induced during metamorphism when, for example, a slate or schist is formed from a mudstone. Theoretical studies and experimental work on such materials shows that there is a range of structures that can form when they are compressed parallel to the mineral fabric or layering. The two end members of this range are upright folds and box folds (Figure 9). The type of structure that forms is determined by the mechanical anisotropy of the material. As the anisotropy increases so the upright folds (Figure 9A), give way to folds with gently diverging axial planes (Figure 9B) and finally to box folds (Figure 9C).

It should be noted that if a geological multilayer has a sufficiently high mechanical anisotropy, it will buckle to form a box fold. An example of this is shown in Figure 10A, where a box fold with a wavelength of several 100 m has developed in Carboniferous turbidites from south-west England.

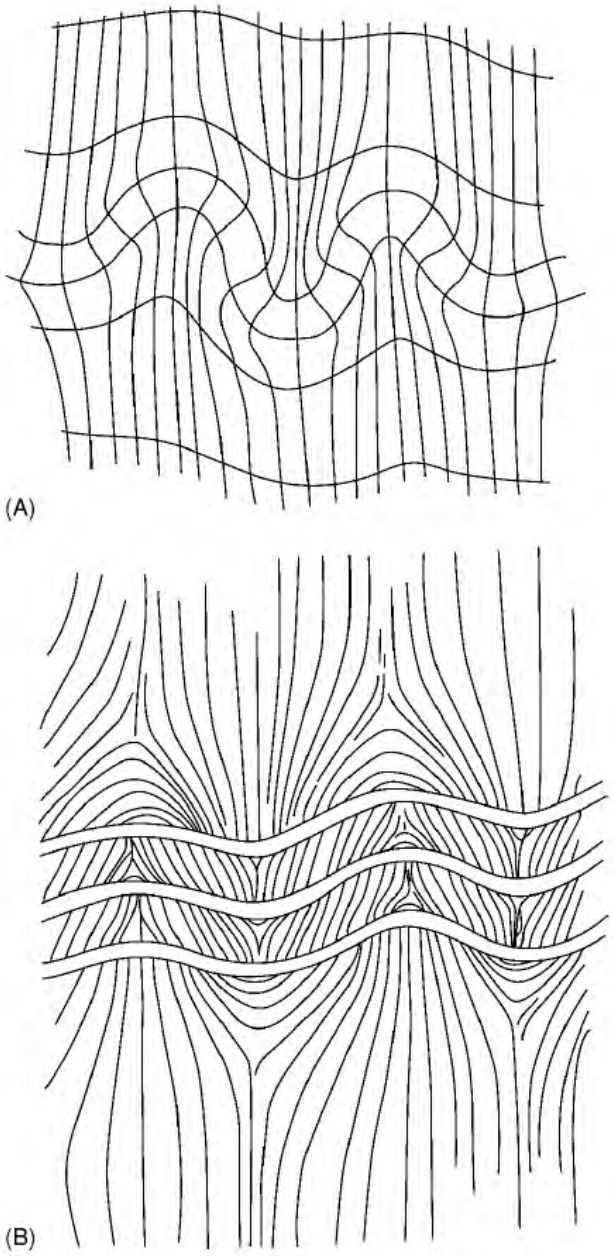

Figure 8 (A) Experimentally produced buckles in viscous materials showing the disturbance of the matrix (the zone of contact strain) by the buckling single layer (B) Folded multilayer where the proximity of the layers caused an overlap of the zones of contact stain and therefore the formation of multilayer buckles.

Experimental Work on Folds

One of the disadvantages of theoretical studies of folding is that they are often only valid for the first increment of buckling. Once buckling has been initiated, the assumptions of the analysis are violated and the theory cannot be used to predict the way in which the fold amplifies into a finite fold. In contrast, experimental work on models made from rock analogue materials such as gelatine, are ideal for the study of the amplification of folds. In addition, unlike

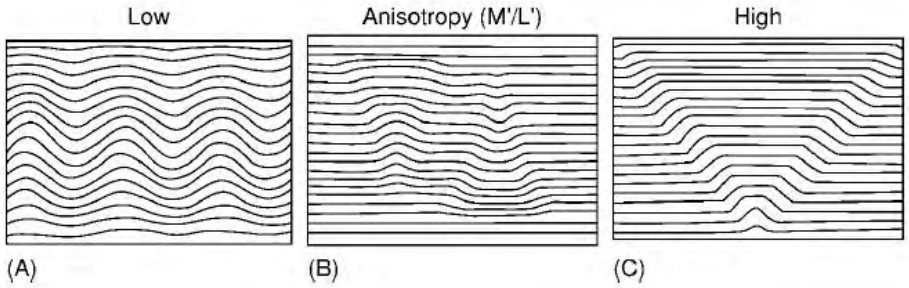

Figure 9 The structures that can form when a mechanically anisotropic material such as a mineral fabric of sedimentary layering is compressed parallel to the layering. Depending upon the anisotropy they range from upright folds with axial planes normal to the principal compression to box folds where the axial planes are inclined to the compression.

Figure 10 (A) Small scale box fold formed in a micaceous fabric and (B) large scale box fold formed by the buckling of a turbidite.

the theoretical treatments of folding which assume that the layered system is stressed instantaneously (i.e., that the process of stressing the material plays no part in the buckling), the experiments show that the process of stressing the layered system can influence the process of folding. Such experiments show that folds that form during the compression of the models do not generally develop synchronously. They form in a serial manner, either one after the other where the amplification of one fold stimulates the initiation and amplification of another next to it (Figure 11A), or one after the other at random positions within the model. Figure 11B shows such randomly positioned folds in the Jura mountains of Switzerland.

In the complex multilayers that occur in nature, it is often found that all types of buckling occur in the same multilayer. For example, in the multilayer shown in Figure 12, which is a finely laminated evaporite, examples of single layer buckling can be seen (layers 1 and 2) where the wavelength is determined by the thickness of the layers. The two layers are far enough apart for them to fold independently and they have done so producing an example of disharmonic folding. In area 3 of the multilayer, examples of multilayer folding where the folds have axial planes oriented at right angles to the layering can be seen (Figure 7B and 9A) and in area 4, box folds have formed with axial planes inclined to the layering (Figure 9C, 10A and B).

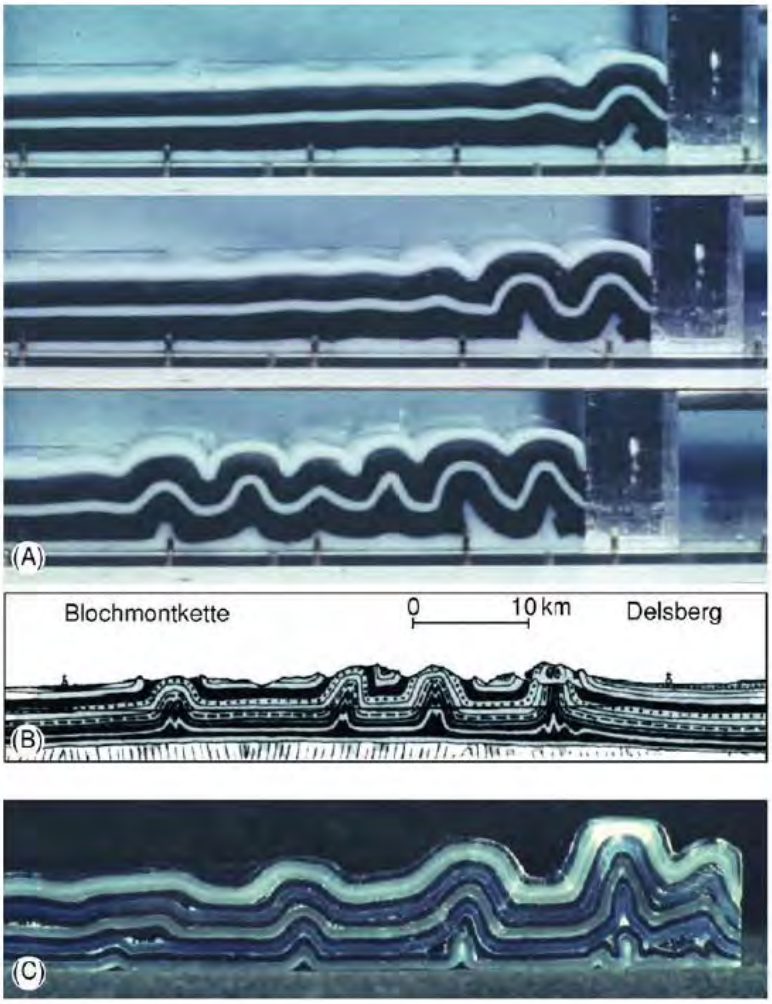

Figure 11 (A) The serial development of folds in a lubricated gelatine multilayer caused by a piston moving from right to left. (B) Isolated folds in the Jura mountains Switzerland. (C) Isolated folds in a lubricated wax multilayer.

Figure 12 Harmonic and disharmonic folding in a specimen of the Castile and Todilto evaporites, New Mexico, USA (c.f. layers 1 and 2). Some of the folding (layers 1 and 2), can be described by the theory of single layer buckling, some by the theory of multilayer buckling, (3) and some (4) by the theory of buckling of an anisotropic material.

Three-Dimensional Geometry of Buckle Folds

The three-dimensional geometry and spatial organization of buckle folds has been studied primarily by field observations and analogue modelling. These studies show that buckle folds have a periclinal geometry, i.e., have the form of an elongate dome, basin, or saddle (Figure 20A). The geometry of a pericline is often given in terms of the ratio of its half wavelength and its hinge length. This is termed the aspect ratio and although it will increase as the fold amplifies, it

is found that the majority of buckle folds in the upper crust have ratios of between 1:5 and 1:10. Typical geometries of geological folds are shown in Figure 13.

Bending

Bending is the term used to describe the flexuring of a layer induced by a compression acting at a high angle to the layering (Figure 1B). Geological flexures that are the result of bending are known as drape folds or forced folds and are frequently formed when sediments, which cover a more rigid basement, flex in response to components of vertical movement along basement faults (Figure 14). This may be normal movement, in which case the flexing of the layering involves layer parallel extension (Figure 14A), or reverse movement, in which case the folds that result will involve an element of buckling (Figure 14D).

A forced fold is defined as a fold in which the final overall shape and trend are dominated by the shape of some forcing member below. These are frequently fault blocks, movements of which produce linear fault scarps which in turn produce linear forced folds with an aspect ratio much higher that that associated with buckle folds and with different spatial organizations.

The two types of basement faults linked to the formation of forced folds are dip-slip faults (either normal or reverse Figure 14A and D respectively). These faults produce fault scarps at the basement–cover interface over which the forced folds form. When the third type of faulting, i.e., strike-slip or wrench faulting occurs in the basement, no fault scarp is produced and consequently no forced folds are produced in the overlying strata. However, movement on basement strike-slip faults can give rise to

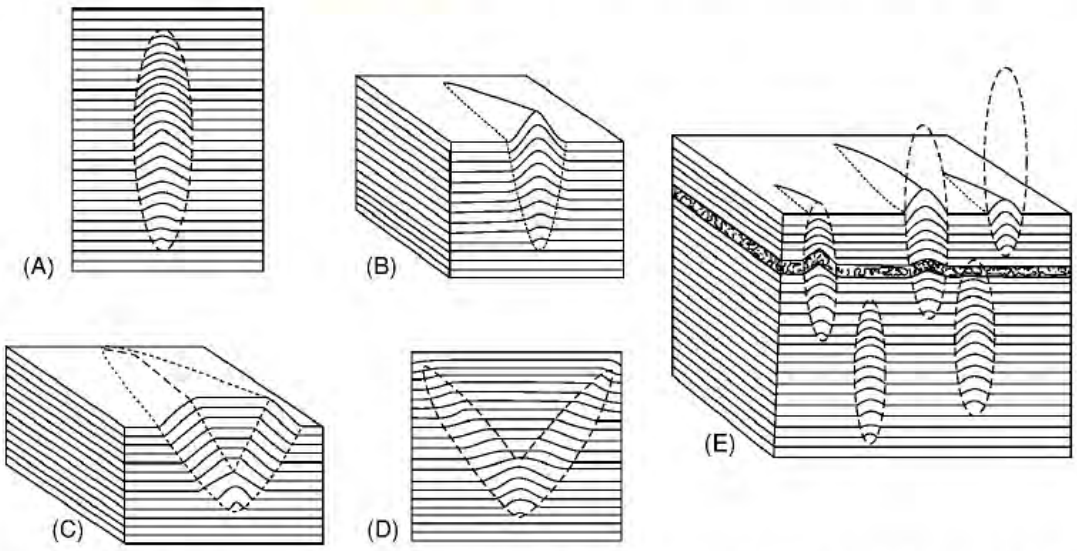

Figure 13 (A) Typical profile geometry of a fold in a multilayer. (B) Block diagram showing a fold dying out in both profile and plan view. (C) Block diagram and (D) profile geometry of a box fold. (E) The spatial organization of folds within a multilayer.

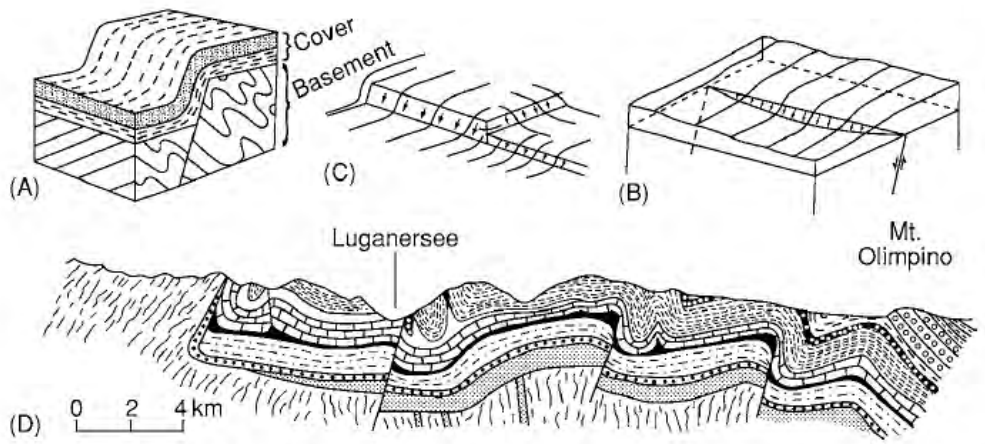

Figure 14 (A) and (B) block diagrams of drape or forced folds, the result of normal faulting in the basement. (C) shows the type of fold geometry which may be associated with block faulting in the basement. (D) Forced folds probably formed over reverse faults in the basement.

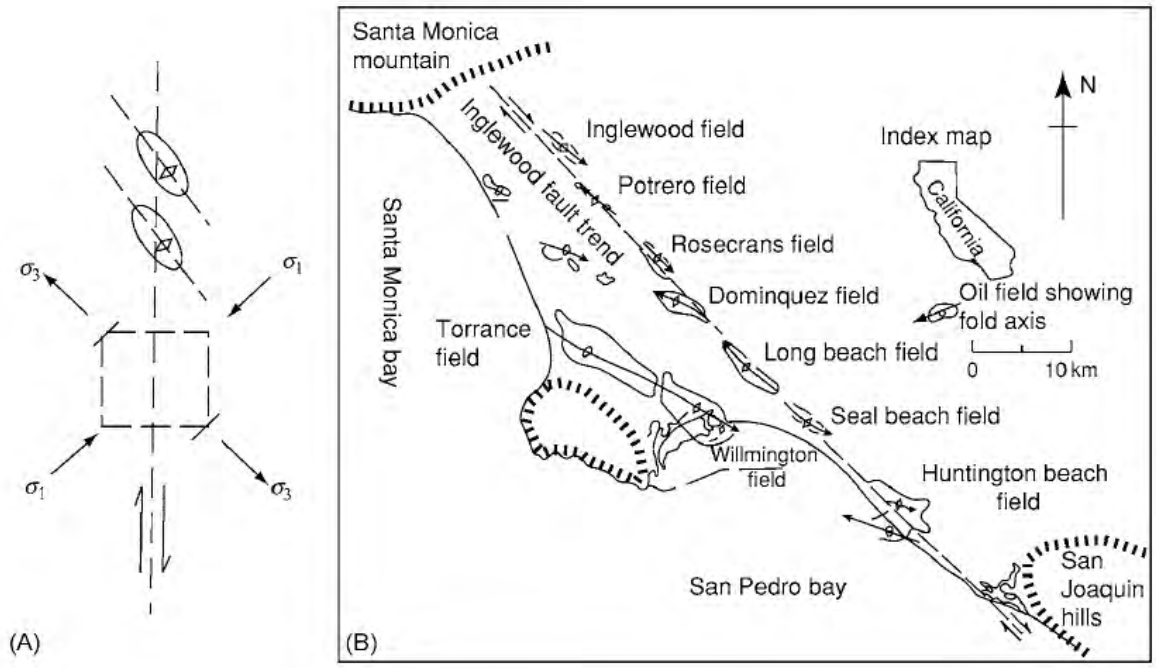

Figure 15 (A) Movement along a wrench fault in the basement produces horizontal compression in the cover rocks and the formation of a linear train of offset folds above the fault. (B) Such en echelon folds above the Inglewood fault, California.

folding in the overlying cover rocks. The stress generated in the cover rocks above a basement strikeslip fault is shown in Figure 15A. A local horizontal compression, inclined at 45° to the basement fault, is developed in the cover and this can give rise to a variety of structures, depending on the rheological condition of the cover rocks. If they behave in a ductile manner, then a series of folds may develop. They will form with their axial planes at right angles to the local maximum compression (σ_1) and will be arranged in an offset manner along the trace of the basement fault (Figure 15A). Natural examples of these linear arrays of buckle folds (see, for example, Figure 15B, which shows folding induced in the cover rocks above the San Andreas wrench fault) are excellent markers for locating major hidden faults.

Fault-Bend Folds

The forced folds shown in Figure 14 occur in cover rocks which respond in a ductile manner in response to movement on faults in a relatively rigid basement. Other well-documented relationships between faults and folds are known, one of the most familiar being fault-bend folding (Figure 16). In this type of forced folding, the folding is not the result of the movement of rigid fault blocks in the basement but rather the result of fault movement within the cover rocks. As can be seen from Figure 16, faults are not perfectly planar surfaces of slip. They generally have gentle undulations and may display substantial curvature or sharp bends. For example, the thrust fault shown in Figure 16 is made up of two horizontal portions (flats) linked by an inclined portion (ramp). As the two fault blocks slip past one another there must be deformation in at least one of them because rocks are not strong enough to support large voids. For this reason, many major folds within layered rock exist within the hanging-wall fault blocks, formed by bending the fault blocks as they slip over non-planar fault surfaces. This mechanism of folding is termed Fault-bend folding.

Flow Folding

In addition to the mechanisms of buckling and bending, folds can also be produced by flow. Impressive examples occur during the outpouring of lavas and during the slower flow of ice and salt. Salt is less dense than most rocks and therefore tends to rise diapirically though overlying strata. The resulting salt domes often emerge at the Earth's surface (Figure 17) where they can form salt glaciers (Namakiers). Figure 18 shows large flat lying flow folds formed as a result of flow within a salt glacier.

Implications of Folds Regarding the Properties of the Rock

The Earth's crust is characterized by an upper seismically active zone where brittle failure dominates, resulting in the formation of fractures and loss of continuity of the rocks, and a lower aseismic zone associated with ductile deformation, i.e., deformation that

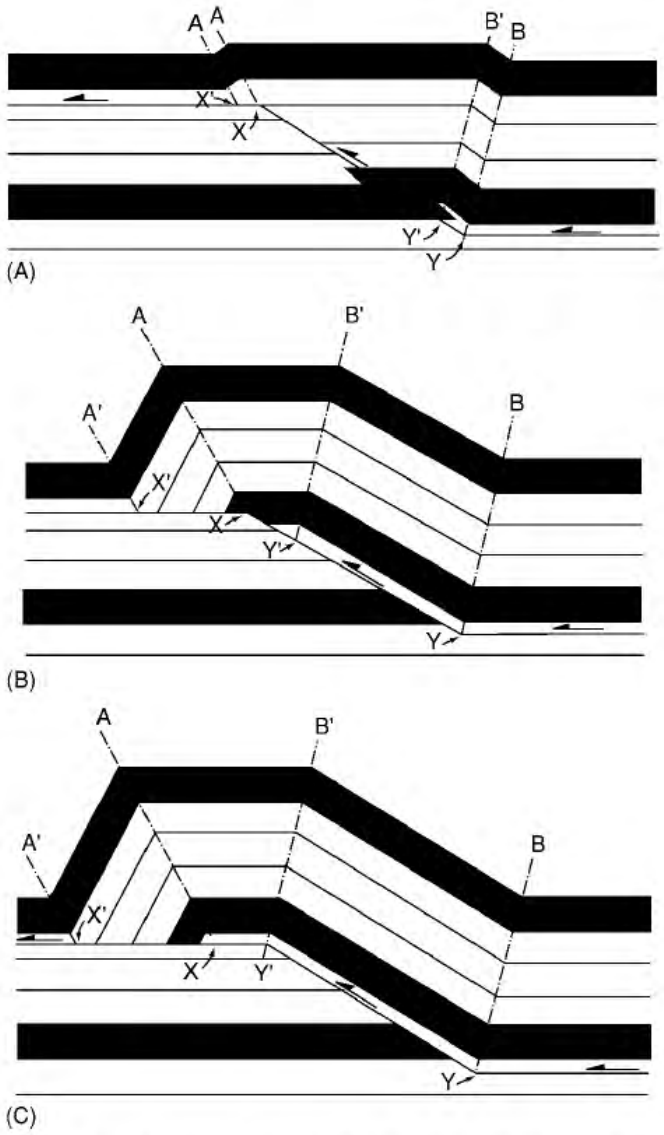

Figure 16 The development of a fault bend fold as a thrust sheet rides over a ramp in the detachment horizon. (After Suppe (1983).)

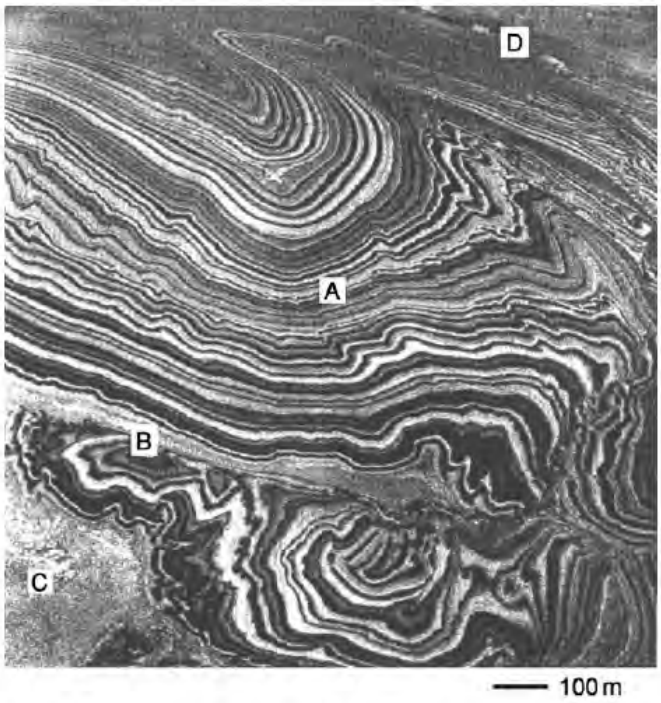

Figure 17 Oblique aerial view of a salt dome from the Great Kavir, Iran, showing the agate like regularity of bedding in the younger salt (A) and a fault contact (B) with the more massive older salt (C).

occurs without the loss of continuity of the material. Folding tends to occur within this ductile zone although, as noted earlier, folds can form at all depths in the crust. In addition, fractures can form in association with folding and the orientation of the fractures indicates clearly that they are formed by the same stress field that operated during folding (Figure 19).

Strain Within a Folded Layer and Associated Fracturing

The strain distribution (and therefore the fracture pattern) within a folded layer is dependant on the layer properties. In a homogenous isotropic layer, such as a uniform sandstone or limestone bed, the strain distribution is likely to be similar to that shown in Figure 20B, in which a layer parallel extensional field associated with the outer arc is separated from a layer parallel compression field associated with the inner arc by a neutral surface along which there is no strain. This model of strain distribution is termed Tangential Longitudinal strain folding. In contrast, a homogeneous anisotropic layer, such as a well-bedded shale, may fold by bedding parallel slip which results in the strain distribution shown in Figure 20C. This is known as flexural flow folding if the shear strain parallel to the layer boundary is uniformly distributed across the layer, and flexural slip folding if it is concentrated along distinct bedding planes. It is interesting to note that both models of folding (Figure 20B and C) produce parallel folds, i.e., folds with a constant orthogonal thickness. This illustrates the fact that the strain state within a fold cannot be deduced from the geometry of its profile section. However, the different strain patterns within the two models reflect the fact that they have very different stress fields within them which may lead to the formation of characteristic fracture patterns which enable the two fold types to be recognized in the field. For example, the extensional fractures in the outer arc of the pericline shown in Figure 20A and the shear fractures in the inner arc, indicate outer arc extension and inner arc contraction, respectively, i.e., a pattern compatible with the deformation associated with the Tangential Longitudinal Strain fold (Figure 20B).

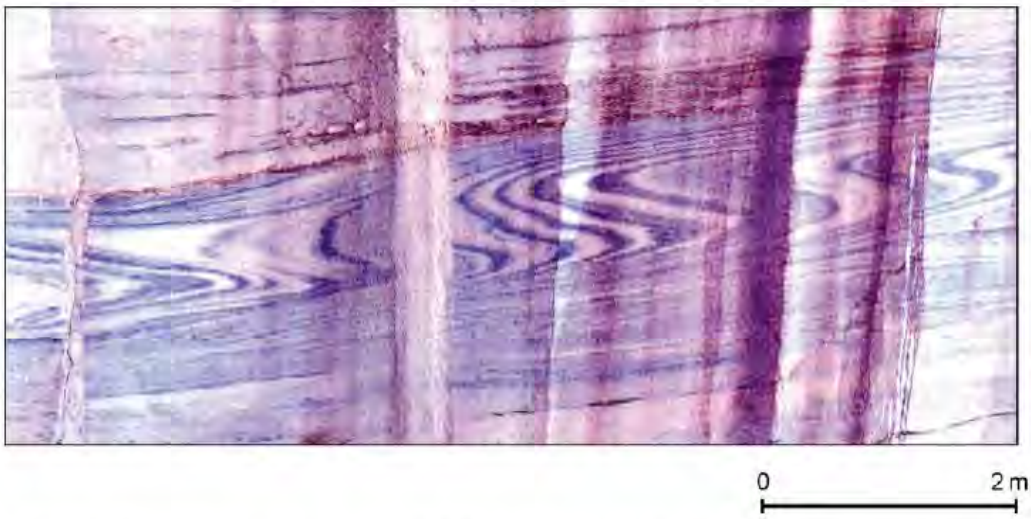

Figure 18 Large flat lying flow folds formed as a result of flow within a salt glacier, Iran.

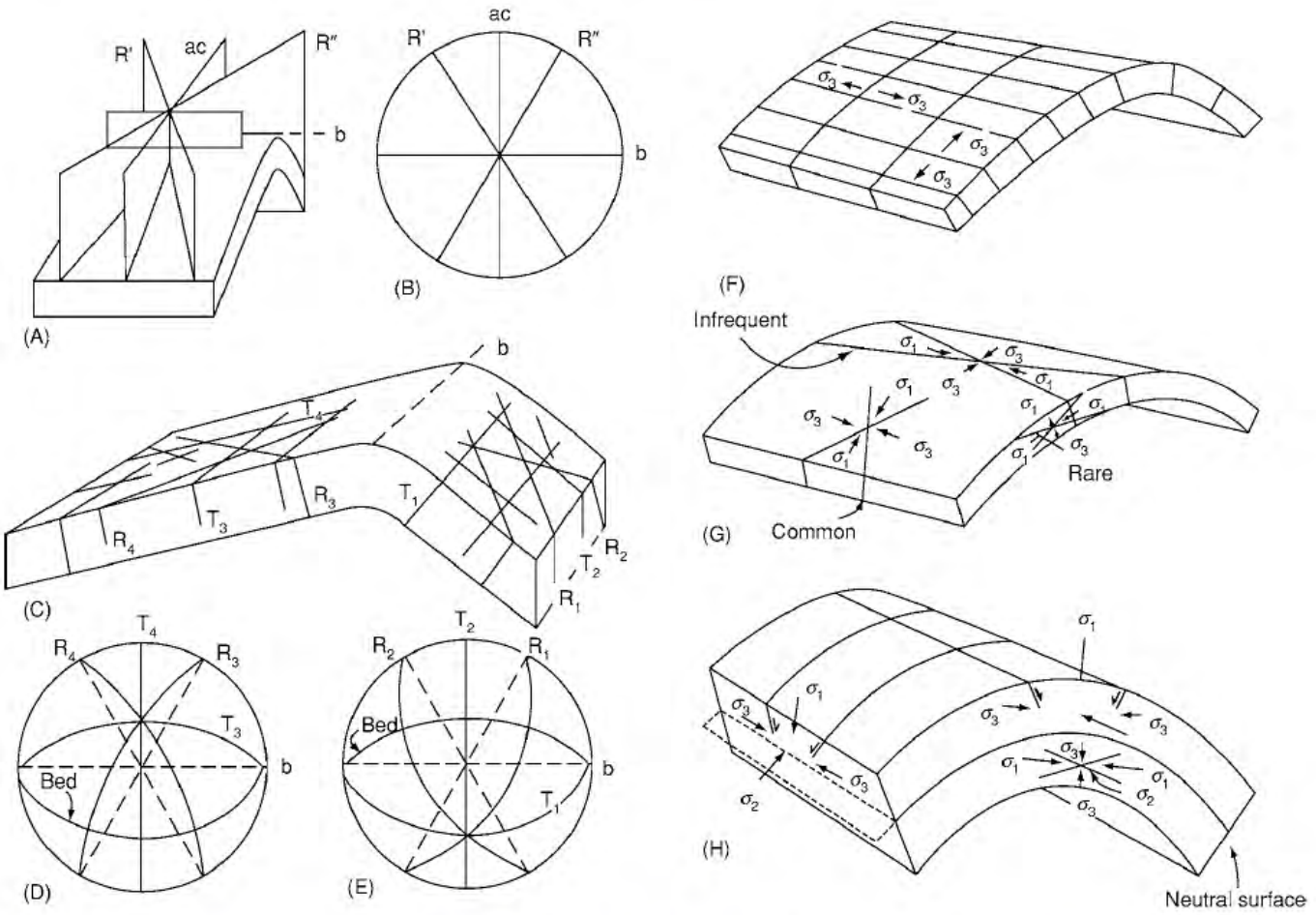

Figure 19 (A) Ideal relationship of master joints to a relatively small fold. (B) Stereographic plot of fractures shown in (A). (C) Trend of minor fractures in a folded competent unit. (D) and (E) Stereographic plots of fractures in the two limbs. R and T are shear and extension fractures, respectively. (F) Typical relationship of extensional fractures to a fold. The orientation of the least principal stress associated with each set (which are of different ages) is shown. (G) Typical orientation of shear fractures in a thin bedded layer, with associated stress systems. (H) Typical orientation of normal faults and thrusts which may develop in a thick flexed unit.

Thus, fractures formed in association with buckle folds may be the result of the local stresses generated as a result of buckling. Alternatively, they may be caused by the regional stress field. The existence of a buckle is likely to disturb the regional stress field. For example, folds are often accompanied by bedding plane slip, implying that the bedding planes cannot sustain a large shear stress. It follows, therefore, that

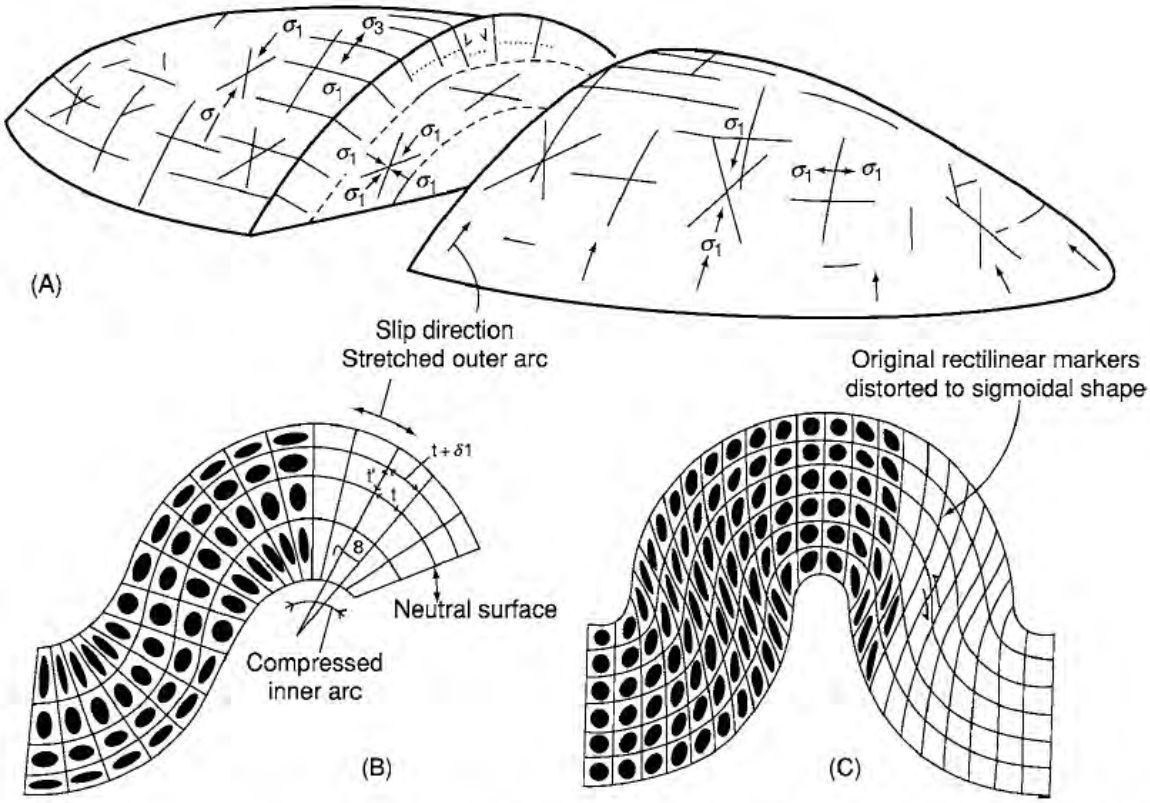

Figure 20 (A) Various fractures associated with a pericline. (After Stearns (1978).) (B) Strain distribution in a tangential longitudinal strain fold and (C) a flexural flow fold. (After Ramsay (1967).)

the principal stresses must be constrained to being either subparallel or subnormal to the folding layers. As a result of this stress deflection, the fractures also form normal to bedding. This is illustrated in Figure 19, which shows the predicted orientation of the shear and extensional fractures that would form in response to the regional compression generating the fold (Figure 19A), together with their projection on a stereographic plot (Figure 19B), and the frequently observed orientation of these fractures on the limbs of the fold which occurs as a result of the principal compressive stress following the layering (Figure 19C–E). The types and orientations of fracture found in association with buckle folds, which form as a result of both the regional and local stresses, are summarized in Figure 19F–H.

Fractures (Including Joints)

J W Cosgrove, Imperial College London, London, UK

Introduction

Fractures are the result of brittle failure which is the general term given to failure during which continuity of the material is lost. Deformation that does not involve loss of continuity is termed ductile. Two modes of brittle failure have been recognized, namely Shear failure and Tensile failure and these can be distinguished from each other on the basis of: (i) the orientation of the fractures with respect to the principal stresses that caused them; and (ii) the relative motion of the rock on each side of the fractures (Figure 1).

If the rock moves parallel to the fracture (Figure 1A) the fracture is a shear fracture and if it only moves normal to the fracture (Figure 1B) it is a tensile fracture.

Shear fractures in rocks are called faults and tensile fractures joints.

Mechanism of Formation of Fractures

The current understanding of brittle failure in rocks is the result of a combination of field observation, experimental work, and theoretical study. Field observations show the two types of fractures, and experimental work reveals that shear fracture occurs when the principal stresses are all compressive, and tensile failure when the least principal stress is tensile and equal to or greater than the tensile strength of the rock.

Shear Failure

Experiments have been performed in which cylindrical rock samples, surrounded by a hollow jacket into which a fluid can be injected to provide a confining stress (σ_3), are subjected to an axial loading (σ_1) until they fail. This results in a body of data recording the principal stress necessary for the rock to fail under a wide range of confining stress. These data can be represented graphically in two ways. The first is to plot the axial load against the confining stress (Figure 2A). For many rocks this plot produces a straight line whose intersection with the axial stress axis gives the uniaxial strength, i.e., the stress a rock can sustain with no confining stress. The graph

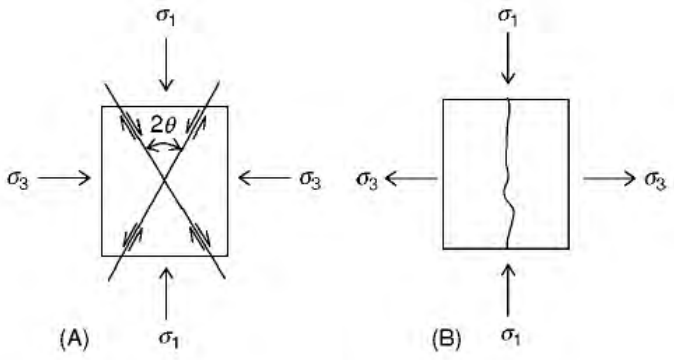

Figure 1 The two modes of brittle failure (A) Shear failure and (B) Tensile failure. They can be distinguished from each other on the basis of: (i) the orientation of the fractures with respect to the principal stresses that caused them; and (ii) the relative motion of the rock on each side of the fracture.

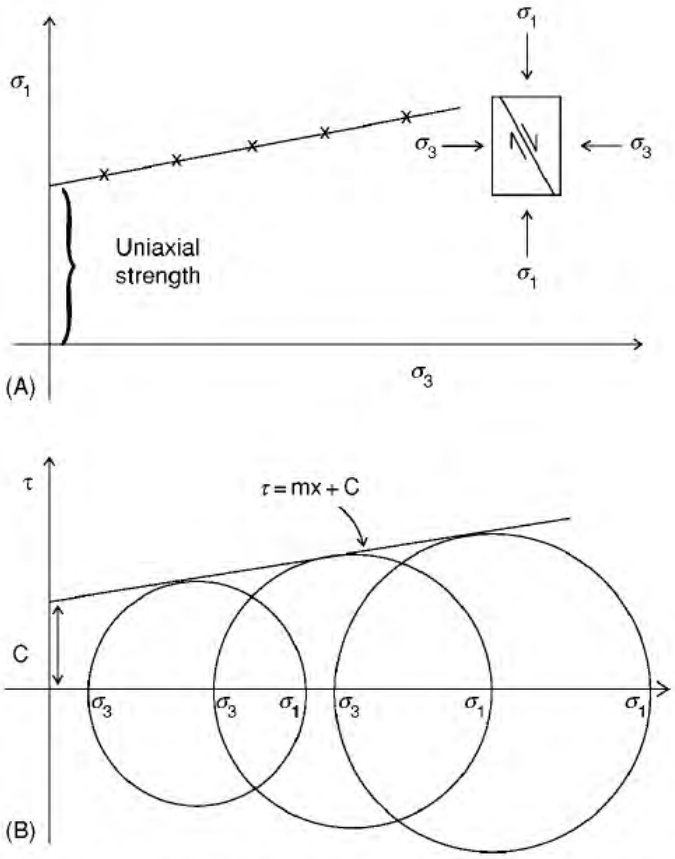

Figure 2 The graphical expression of the experimental data on shear failure. These data can be represented graphically in two ways. (A) is a plot the axial load against the confining stress at failure, (B) shows the same data plotted as a series of Mohr stress circles. The tangent to these circles represents the failure envelope for shear failure.

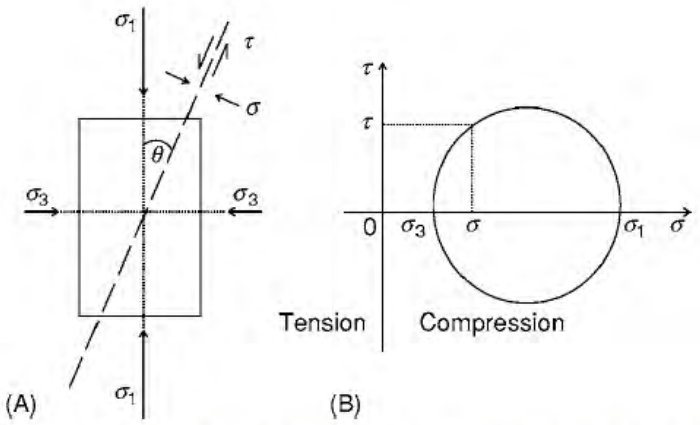

Figure 3 (A) The state of stress (σ and τ) on a plane inclined at $\theta°$ to the maximum principal compression, is given by the biaxial stress equations. (eqns [1] and [2]). The graph of σ against τ for values of θ between 0° and 180°; (B) defines a circle, the Mohr stress circle, which defines the state of stress.

indicates clearly that the strength of a rock (i.e., its ability to sustain a load without permanent deformation) is not a fixed value but depends (amongst other things) on the confining stress. The higher the confining stress, σ_3, the greater the axial load needed to cause failure. A rock's strength will, therefore, increase with increasing depth in the crust.

The second method of plotting the experimental data is to plot the stress state for each experiment (i.e., a stress state that caused the rock to fail) as a Mohr circle. The state of stress on any plane inclined at $\theta°$ to the maximum principal compression (Figure 3) is given by the biaxial stress equations:

$$\sigma = \sigma_1 \cos^2\theta + \sigma_3 \sin^2\theta \qquad [1]$$

$$\tau = (\sigma_1 - \sigma_3) \cos^2\theta \sin^2\theta \qquad [2]$$

These equations can be represented graphically by calculating σ and τ for values of θ between 0° and 180° and plotting the results on a graph of σ against τ. The resulting points define a circle, the Mohr stress circle (Figure 3B), which defines the state of stress. The diameter of the circle, which is a measure of the differential stress, $(\sigma_1 - \sigma_3)$, is determined by the values of the principal stresses, σ_1 and σ_3. As the experimental data represented in Figure 2A consists of values of the principal stresses that caused the rock to fail, these data can be plotted as a series of Mohr circles (Figure 2B). The tangent to these circles represents the failure criterion for shear failure.

For many rocks, this tangent is a straight line whose equation is:

$$\tau = m\sigma + C \qquad [3]$$

where m is the slope of the line and C its intersection with the shear stress axis.

Having established the shear failure criterion experimentally, it is important to compare it with the theoretically derived criterion. This was developed independently by Navier and Colomb, who argued that in order for a shear fracture to develop, the shear stress τ acting along the potential fracture plane (Figure 3A) must be sufficiently large to overcome the cohesion along that plane, C_0, plus the resistance to shear along the plane once it had formed. The resistance to slip is given by Amonton's law of frictional sliding which states:

$$\tau = \mu\sigma \qquad [4]$$

where τ and σ are the shear and normal stresses, respectively acting on the fracture plane and μ the coefficient of sliding friction. μ is defined as the tangent of the angle of sliding friction φ. Hence the complete criterion can be expressed in the form:

$$\tau = \sigma \tan\varphi + C_0 \qquad [5]$$

This is known as the Navier–Colomb criterion of shear failure and is identical to the criterion established experimentally (eqn [3]). The orientation of the planes where this condition is first met can be determined by substituting the biaxial stress equations (eqns [1] and [2]) into the shear failure criterion (eqn [5]) and solving for the minimum. The optimum orientations for the shear fractures are:

$$\theta = + \text{ or } - [45° - \varphi/2] \qquad [6]$$

where θ is the angle between the maximum compressive stress (σ_1) and the shear fracture (Figure 1A). Note that two fracture orientations are predicted, inclined at $45° - \varphi/2$ each side of σ_1. They are termed conjugate shear planes and although the magnitude of the shear stress along them is the same, the sense of shear is different. Faults are the geological expression of shear failure, and conjugate small-scale faults in a sequence of alternating sandstones and shales are shown in Figure 4.

Classification of Faults

As can be seen from Figure 1A, the orientation of a fault is controlled by the orientation of the principal stresses that generate them. Field observations reveal that faults fall into three classes, normal faults, wrench (or strike slip) faults, and thrust (or reverse) faults which, as can be seen from Figure 5, correspond to stress states where σ_1, σ_2, and σ_3 are vertical.

Figure 4 Conjugate normal faults in Carboniferous turbidites from Bude, Cornwall, England.

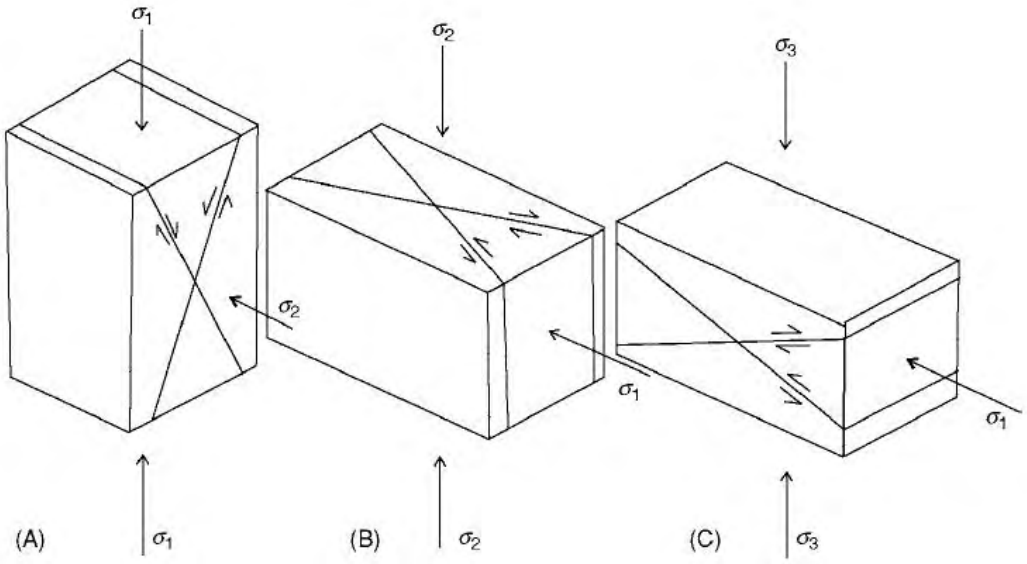

Figure 5 The orientation of shear fractures that form when (A) σ_1, (B) σ_2 and (C) σ_3 are vertical. The resulting faults are termed, normal faults, wrench (or strike slip) faults, and thrust (or reverse) faults, respectively.

This observation implies that the principal stresses tend to be oriented in one of these three orientations. In 1951, Anderson argued that this was because the Earth's surface is a free surface which cannot sustain a shear stress. Thus, in order not to generate a shear stress parallel to this surface, the principal stresses are constrained to remain either parallel or normal to it. As is discussed later, the three classes of faults characterize three different tectonic regimes.

Tensile Failure

Joints and veins are the most common geological expression of tensile failure. Experiments show that this type of failure generates fractures normal to σ_3 (Figure 1B). The theory of tensile failure was developed by Griffith (1925) who argued that in an ideal material the tensile strength of a material would be determined by the strength of the

inter-atomic bonds. However, experiments revealed that the measured tensile strength of a material is usually several orders of magnitude lower than that calculated on the basis of their inter-atomic bond strength. Griffith argued that this was because the materials contain small flaws or microfractures and that these resulted in a local stress magnification at the crack tips. This is illustrated diagrammatically in Figure 6, which shows the stress configuration of a plate subjected to a horizontal extension. In the first plate (Figure 6A), the stress is evenly distributed through the plate. In the second plate, which contains a flaw represented by an elliptical crack (Figure 6B), the tensile stress is locally magnified at the fracture tips. The amount of magnification depends primarily on the orientation and eccentricity of the crack. The greater the eccentricity, the greater the magnification. Griffith argued that by this means, a relatively low applied stress could be locally amplified at fracture tips within the material to the point where it reached the stress required to break the atomic bonds within the material causing the material to fail. He developed the following failure criterion (the Griffith criterion of tensile failure) based on this model:

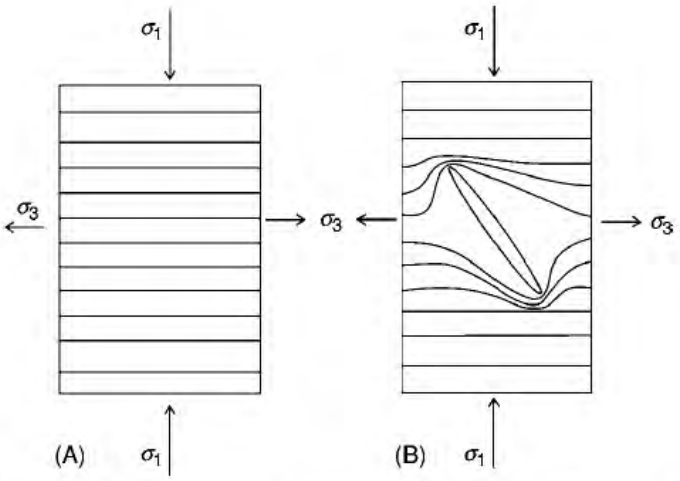

Figure 6 (A) A uniform stress field represented by uniformly spaced stress trajectories in a stretched layer. (B) The concentration of tensile stress at crack tips in a uniformly extended layer.

$$\tau^2 + 4T\sigma - 4T^2 = 0 \qquad [7]$$

where τ is the shear stress, σ the normal stress, and T the tensile strength of the material. The graphical expression of this failure criterion is shown in Figure 7A. It has the form of a parabola whose intersection with the normal stress axis gives the tensile strength and with the shear stress axis the cohesion.

The complete criteria for brittle failure, (the Griffith, Navier–Colomb criteria), is obtained by linking the two criterion (eqns [5] and [7]) at the point where their slopes are identical (Figure 7B). Any stress state can be represented on this graph as

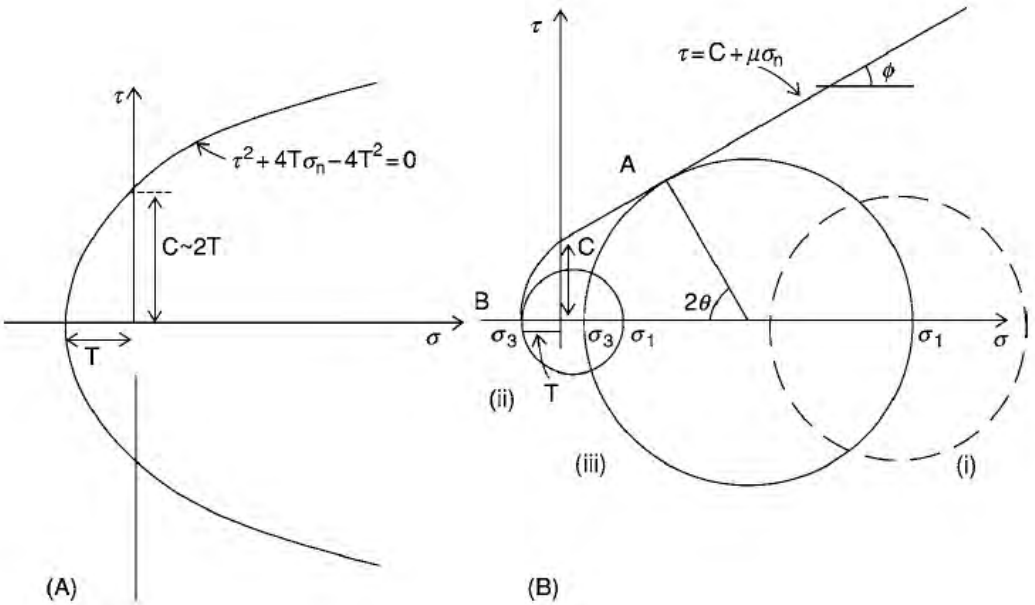

Figure 7 (A) The graphical expression of the Griffith criterion of tensile failure (eqn [7]). It has the form of a parabola whose intersection with the normal stress axis gives the tensile strength of the material and with the shear stress axis the cohesion. (B) The complete criteria for brittle failure, (the Griffith, Navier Colomb criteria), is obtained by linking the two criteria (eqns [5] and [7]) at the point where their slopes are identical.

a Mohr circle whose position is determined by the values of the principal stresses (σ_1 and σ_3). If a stress state, when plotted on the graph, does not touch or intersect the failure envelope, the stress state is stable, i.e., will not cause the rock to fail, e.g., stress field (i) Figure 7B. If, however, it does touch the envelope, failure will occur, either by tensile fracturing if the contact is with the tensile part of the envelope (Figure 7B (ii)), or shear fracturing if it is with the shear part (Figure 7B (iii)).

What Determines Whether Tensile or Shear Fractures Form?

It can be seen from Figure 7B that shear failure is associated with a large differential stress ($\sigma_1 - \sigma_3$) i.e., the Mohr's stress circle must be large in order to intersect the shear failure envelope, and that tensile failure is associated with a low differential stress, i.e., the Mohr's stress circle must be small in order to intersect the tensile failure envelope. The precise conditions necessary for the formation of the two types of fractures are:

$$\text{For tensile failure to occur} \quad (\sigma_1 - \sigma_3) < 4T \qquad [8a]$$

$$\text{For shear failure to occur} \quad (\sigma_1 - \sigma_3) > 4T \qquad [8b]$$

where T is the tensile strength of the material.

The geometrical relationships between the principal stresses and the fractures they produce (i.e., a conjugate set of shear fractures symmetrically about σ_1 and a single set of tensile fractures at right angles to σ_3) is shown in Figure 1 and, as noted below, the understanding of these relationships provides a powerful tool in fracture analysis.

It follows therefore that the orientation of the fractures that form in response to a stress field is determined by the orientation of the principal stresses (Figure 1), and the type of fracture (shear or tensile) by the magnitude of the differential stress.

The Effect of a Fluid Pressure on Fracturing

Fluid-Induced Failure

The state of stress in the crust is dominantly compressional. For example, in a nontectonic environment, the stress at any depth is generated by the overburden which produces a compressive vertical stress which induces a compressive horizontal stress. Thus, at any depth the Mohr stress circle will plot in the compressive regime in Figure 7B and there will be no possibility of tensile failure. Geologists were, therefore, perplexed to find that large numbers of tensional fractures occur in the crust. This paradox was resolved when the importance of fluid pressures within a rock was understood. Pore fluid pressure within a rock increases as the rock is buried. (*see* Tectonics: Hydrothermal Activity). The stress state within the pores is hydrostatic and the pressure acts so as to appose the lithostatic stress caused by the overburden. This effect can be shown diagrammatically by representing the lithostatic stress as an ellipse with the stress acting compressively and the fluid pressure as a circle with the pressure acting outwards (Figure 8A). The fluid pressure reduces all the lithostatic stresses by an amount P_{fluid} to give an effective stress. Thus, the principal stresses σ_1 and σ_3 become ($\sigma_1 - P_{fluid}$) and ($\sigma_3 - P_{fluid}$). This new stress field can be plotted as a Mohr stress circle (Figure 8B). It can be seen that the original lithostatic stress circle is moved towards the tensile regime but that the diameter of the circle, i.e., the differential stress, remains unchanged.

The amount of migration of the stress circle is determined by the magnitude of the fluid pressure. Thus, as the fluid pressure gradually increases during burial, the stress circle is pushed inexorably towards the failure envelope. When it hits the envelope, failure occurs. Such failure is termed fluid induced or hydraulic fracturing. In this way an originally compressional stress regime can be changed so that one or more of the principal stresses becomes effectively tensile and the conditions for tensile failure can be satisfied.

The Expression of Fluid-Induced Failure

In the example shown in Figure 8B the lithostatic stress had a small differential stress (i.e., less that 4T (see eqn [8]) and as a result the induced hydraulic fractures were tensile fractures. If it had been greater than 4T, shear fractures would have formed.

The Organization of Tensile Fractures

The Mohr circles shown on Figure 9 all intersect the failure envelope in the tensile regime, i.e., the differential stresses are all less than 4T and will all therefore result in tensile failure. Their differential stresses vary from just less than 4T (circle (i) Figure 9), to zero (circle (iv), Figure 9). Note that when the stress state is hydrostatic, the Mohr circle is reduced to a point.

As noted above, tensile fractures form normal to the minimum principal compressive stress σ_3 (Figure 1B), i.e., they open against the minimum compressive stress. The stress state represented by Mohr circle (i) in Figure 9, has a relatively large differential stress and there is, therefore, a definite direction of easy opening for the fractures. The fractures would

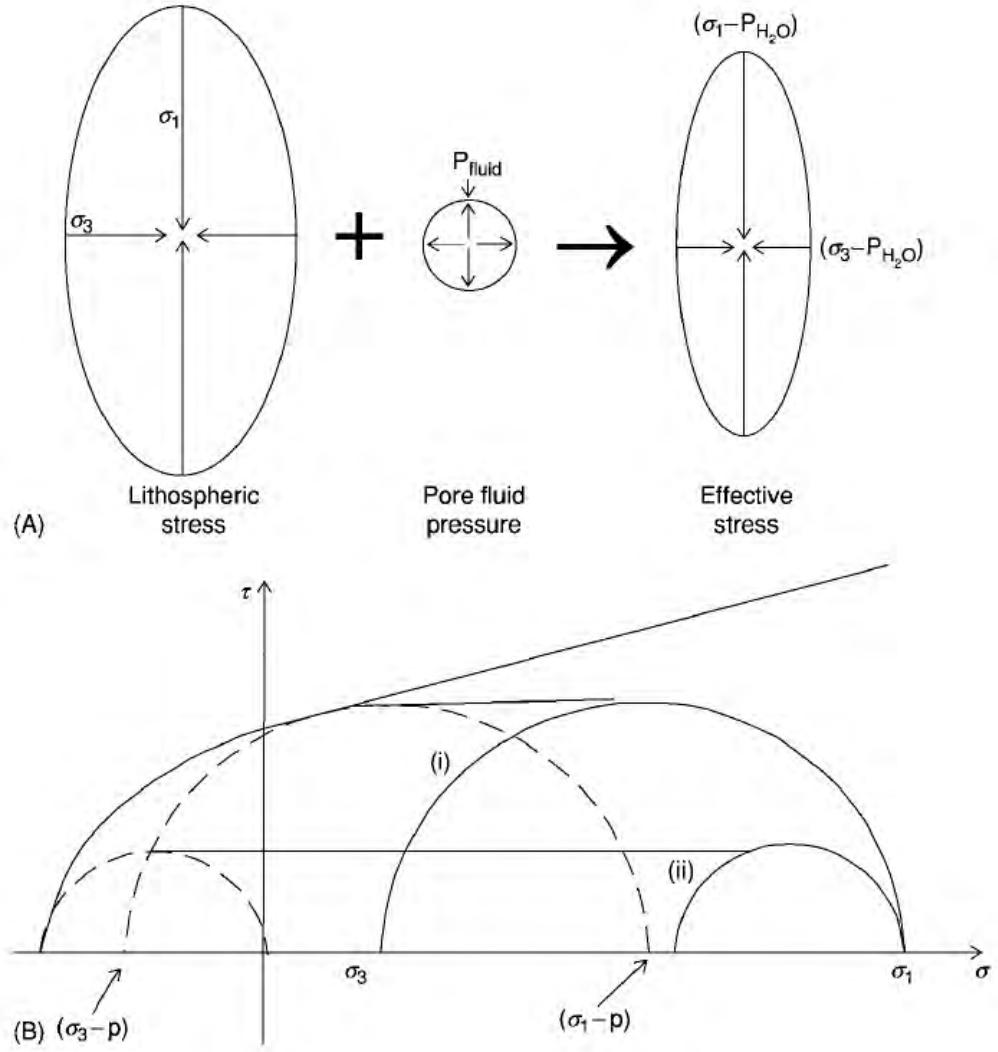

Figure 8 (A) Diagramatic representation of the effect of a fluid pressure (the circle with the outwardly acting stress) on the stress state in a rock (the ellipse with the inwardly acting stresses). All normal stresses are reduced to an effective stress (σ p_{fluid}) but the differential stress (σ_1 σ_3) remains unchanged. The effect is to cause the Mohr stress circle to move to the left by an amount equal to the fluid pressure; (B). Thus depending on the magnitude of the differential stress the induced fractures will be either shear (stress state (i)) or tensile (stress state (B)) (*see* **Tectonics:** Folding).

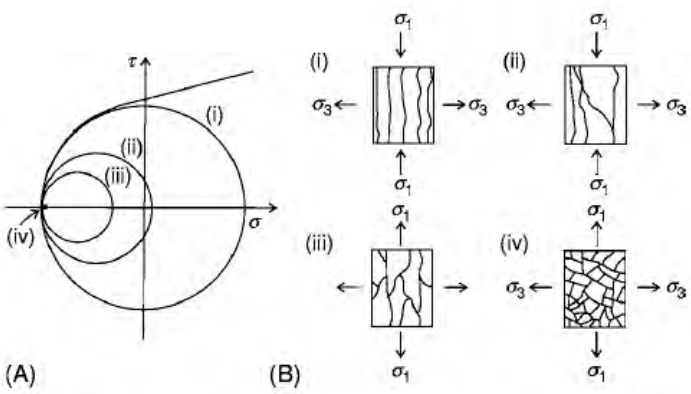

Figure 9 (A) Mohr stress circles (i) (iv) representing a range of stress states, all of which will lead to tensile failure. NB the Mohr circle (iv), that represents hydrostatic stress is a point. (B) Pat terns of tensile failure generated by the corresponding stress states shown in (A).

therefore exhibit a marked alignment normal to this direction (Figure 9B (i)). However, for the stress states represented by the Mohr circles (ii–iv), the differential stress becomes progressively smaller until, for the hydrostatic stress represented by circle (iv), the differential stress is zero. In a hydrostatic stress field the normal stress across all planes is the same and there is, therefore, no direction of relatively easy opening for the fractures. Thus, they will show no preferred orientation and, if they are sufficiently closely spaced and well developed, will produce a brecciation of the rock (Figure 9B (iv)). It can be seen that as the differential stress becomes progressively lower so the tendency for the resulting tensile fractures to form a regular array normal to σ_3 decreases. Tensile fracture systems ranging from well-aligned fractures to randomly

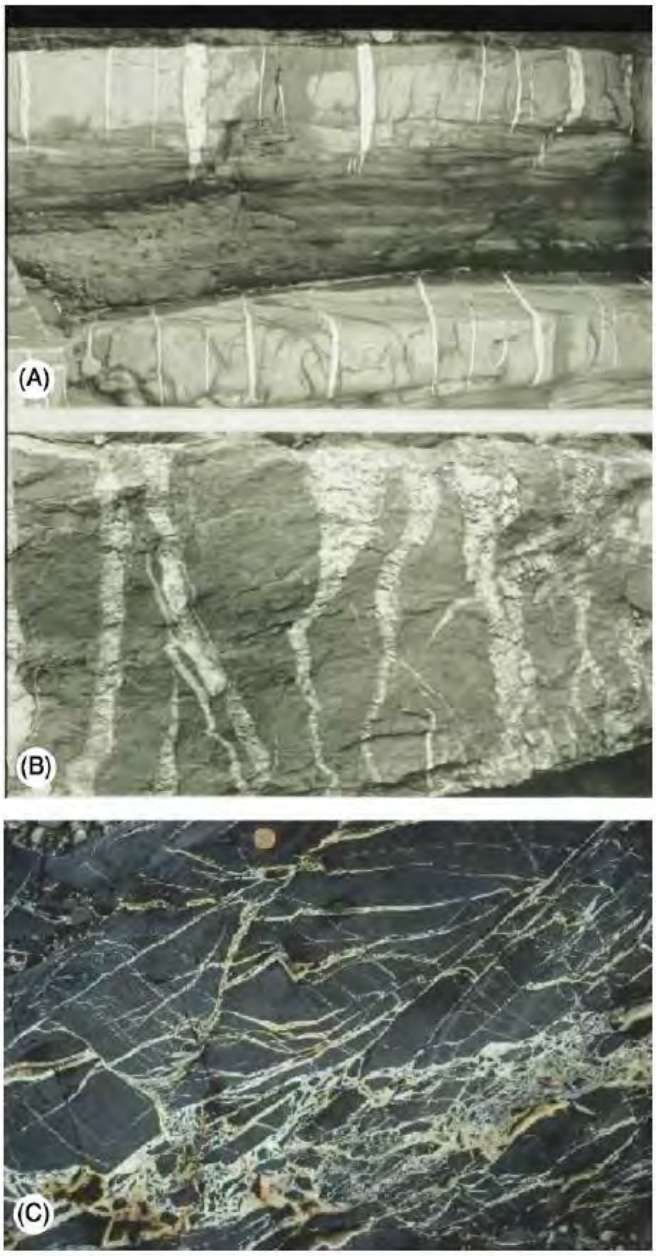

Figure 10 (A) A regular array of tensile fractures exposed on a bedding plane in Carboniferous sandstone, Millook, Cornwall, England. (B) Less well organized tensile fractures cutting Devon ian sandstones, St. Anne's Head, South Wales. (C) Carboniferous sandstone cut by randomly oriented tension fractures.

oriented fractures are to be expected in rocks, and field observations support this idea, Figure 10.

As noted above, the problem of forming tensile fractures in the compressive stress field that generally characterizes the Earth's crust can be solved by appealing to high fluid pressures. However, tensile failure can occur in rocks without the aid of a high internal fluid pressure, for example, during the contraction of a layer as a result of desiccation of a sediment (Figure 11) or the cooling of an igneous body (Figure 12).

Figure 11 Polygonal arrays of tensile fractures cause by the desiccation of a silt layer.

Figure 12 Polygonal arrays of tensile fractures cause by the cooling of a lava flow. (Giant's Causeway Northern Ireland).

In both these examples, the fractures are organized into polygonal arrays showing that the tensile stresses generated were the same in all directions.

Fracture Sets

Generally, the state of stress in the Earth's crust is not hydrostatic. Consequently a single episode of deformation is likely to generate a set of fractures with the same orientation. However, most rocks experience several different stress regimes during their history with the result that several fracture sets are frequently found superimposed on each other to produce a fracture network (Figure 13). The interaction of late fractures with early fractures is illustrated in Figure 14.

The effect of early fractures on later ones is to arrest their propagation and to modify their orientation. It can be seen from Figures 14A and B that one

fracture often ends abruptly against another. This abutting relationship gives the relative age of the fractures, i.e., the later fracture abuts against the earlier fracture. If an early fracture is an open fracture then it will represent a free surface within the rock and, as discussed in the above section on classification of faults, will be unable to support a shear stress. Consequently, the principal stresses will reorient as they approach it into a position either normal or parallel to the fracture. This effect can be clearly seen in Figure 14, where later fractures curve into an orientation at right angles to the earlier fracture as they approach it.

Figure 13 A fracture network in a Liassic limestone bed from Lilstock, North Somerset, England. It was produced by the super position of individual fracture sets.

Fracture Networks

Structural geologists study the cross-cutting relationships of different fracture sets in order to determine their relative age. A variety of rules have been established to help in this task. It is found that early fractures tend to be long and relatively continuous and, as noted above, later fractures abut against these and are consequently shorter. Some of these features can be seen in Figure 15, which shows a fractured limestone pavement containing several fracture sets. The

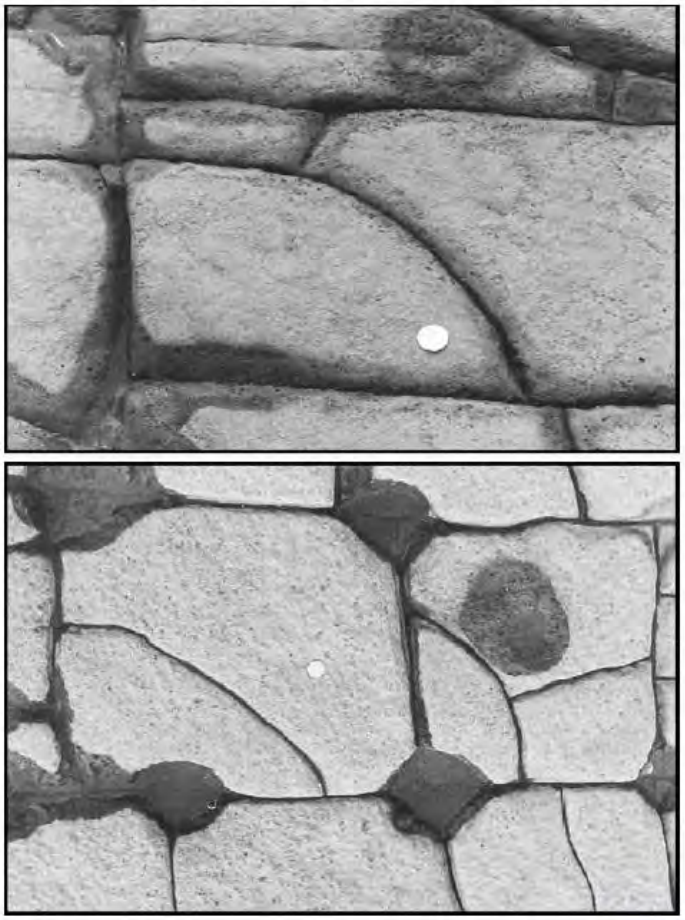

Figure 14 Details of the limestone pavement shown in Figure 13 illustrating the interaction of late fractures with early fractures. The effect of early fractures on later ones is to arrest their propagation and to modify their orientation. It can be seen that the later fractures are deflected by and abut against the earlier fractures.

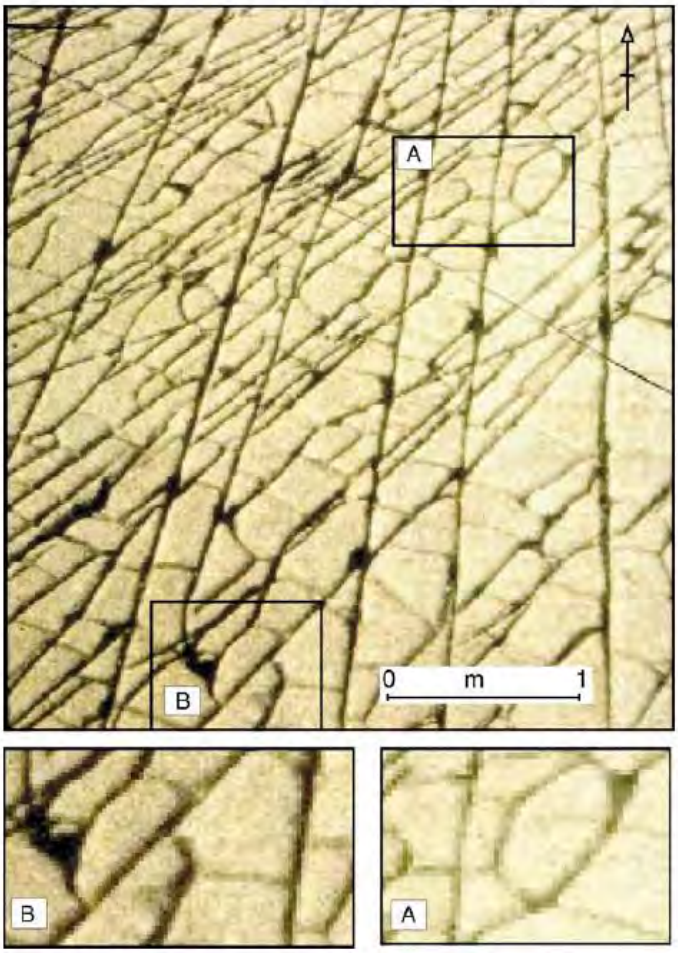

Figure 15 Fracture patterns in a limestone pavement at Lilstock, North Somerset, SW England. The older fracture sets are the most continuous and, as the sets become progressively younger, they become less continuous and less well oriented.

longest and most continuous runs approximately N–S. These are the oldest fractures and are crosscut by several younger sets which become progressively less continuous and less aligned as the regional stress fields responsible for their formation becomes progressively modified by the pre-existing fractures. The fracture set trending approximately NW–SE, the second set to form, shows a remarkable degree of continuity, being only affected by the N–S fractures; its orientation is related directly to the regional stress field.

However, as more fracture sets develop in the rock mass, modification of the stress orientation by the pre-existing fractures may result in there being a poor correlation between the fracture orientation and the regional stress field responsible for its formation. This is well illustrated in subarea A in Figure 15 which has been enlarged in the bottom left-hand corner of the figure. The influence of the pre-existing fractures on the orientation of the late fracturing is so marked that the later fractures display a polygonal organization and cannot be linked directly to the regional stress field responsible for their formation.

Fracture Analysis

A fracture analysis is the study of a fractured rock mass in order to: (i) establish the detailed geometry of the fracture network; (ii) determine the sequence of superposition of the different fracture sets that make up the fracture network; and (iii) deduce the stress regime associated with the formation of each fracture set. The reason why a detailed knowledge of the geometry of the fracture network is so important is that the bulk properties (e.g., strength, permeability) of a fractured rock mass (and most natural rocks are fractured) are generally determined by the fractures they contain rather than by the intrinsic rock properties.

Stages (ii) and (iii) of a fracture analysis are carried out using the principals outlined above relating to the interaction of fractures and the relationship between the stress field and fracture orientation (Figure 1).

Types of Faults a Plate Margins

The 'type' of plate margin is controlled by the relative motion of the two adjacent plates. They can be subdivided into three classes, convergent, divergent, and strike-slip. Convergent margins lead to compressional regimes at the plate margins which results in the formation of mountain belts. The stress regime is that appropriate for thrusts to form, namely a horizontal maximum principal compressive stress and a vertical minimum stress (Figure 5C). Divergent plate margins result in the formation of oceans and the separation of plates. The initial stage of this process is the fracturing of the lithosphere and the formation in the upper crust of major rift systems such as the East African Rift (*see* Tectonics: Rift Valleys). The stress regime of a horizontal minimum principal compressive stress and a vertical maximum stress is appropriate for the formation of normal faults (Figure 5A). When plates move parallel to each other at different velocities, conditions are appropriate for the formation of major wrench (strike-slip) faults (Figure 5B) such as the San Andreas Fault zone of California which separates the Pacific and North American plates.

Thus it can be seen that each of the three types of plate margins is characterized by a different types of fault.

Scale of Fracturing

Fractures occur on all scales within the Earth's crust, ranging from major faults that define plate margins, through faults that can be seen on seismic sections (*see* Tectonics: Seismic Structure At Mid-Ocean Ridges), down to faults that can be observed directly in the field, e.g, Figure 4, to microscopic fractures only visible under the microscope. Detailed studies of the microfractures in rocks at different stages of the evolution of tensile fractures show, as predicted by Griffith's theory of stress magnification (1925) outlined above, that the microfractures grow by tensile failure at the crack tips and that suitably located microfractures link to form larger fractures oriented normal to σ_3, the minimum compressive stress (*see* Tectonics: Faults).

More remarkably, when the growth of shear fractures are studied in the same way, it is found that

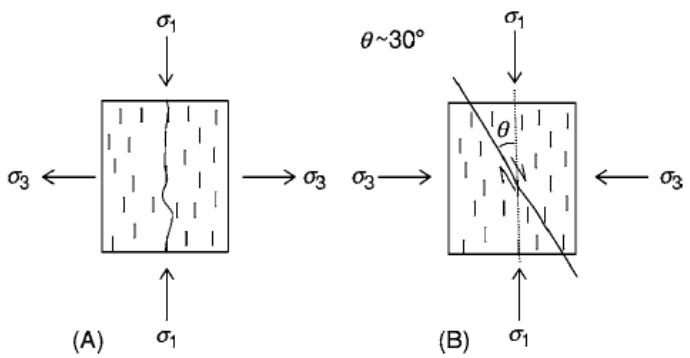

Figure 16 Randomly oriented micro fractures within a material and their growth by tensile failure and subsequent linkage to form (A) macroscopic tensile fractures and (B) macroscopic shear fractures.

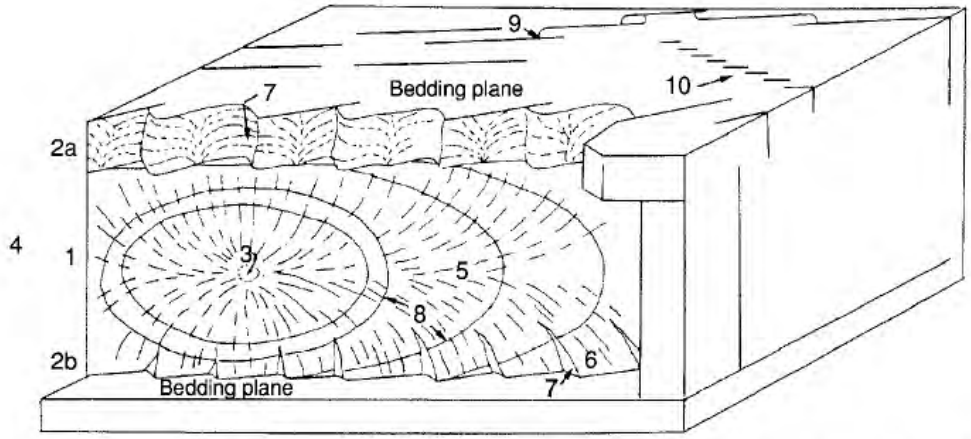

Figure 17 A block diagram illustrating the different types of surface structures (patterns) and fracture trace architecture. (Based on Kullander *et al.*, 1990.) (1) Main joint face, (2a) abrupt twist hackle fringe, (2b) Gradual twist hackle fringe, (3) Origin of fracture, (4) Hackle plume, (5) Plume axis, (6) Twist hackle face, (7) Twist hackle step, (8) Rig marks (front line of the fracture), (9) hooking, (10) En echelon fractures.

initially fracturing occurs by the growth of microfractures at their tips and in an orientation normal to σ_3. However, the macroscopic shear fractures are formed by the linking of offset microfractures, as shown in Figure 16. Thus, it can be seen that despite the two types of fractures having independent failure criteria and different orientations with respect to the principal stresses, they are nevertheless fundamentally linked on a microscopic scale. They are both the result of the growth of microfractures by tensile failure and differ only in the way in which these fractures are linked to a macroscopic fracture.

Surface Features of Fractures

Fractures display a variety of surface features and Figure 17 is a summary diagram showing some of these. Fractography is the science which deals with the description, analysis, and interpretation of fracture surface morphologies and links them to the causative stresses, mechanisms, and subsequent evolution of the fractures. It has been demonstrated that the diverging rays of the plumose structures (Figures 17 and 5) always remain parallel to the direction of propagation of the fracture. Thus, by constructing lines at right angles to these rays, the position and shape of the fracture front at different times of its evolution can be determined (Figures 17 and 8). When the exposures are sufficiently good, it is found that the fracture fronts form a series of concentric 'ellipses', the centre of which marks the site of fracture initiation.