University of Natal

Department of Mechanical Engineering

Development of the Smart Aircraft Bolt

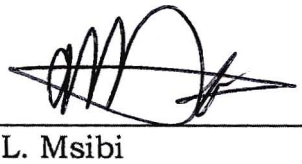
L. L. Msibi

Submitted in fulfillment of the academic requirements for the Masters degree in Mechanical Engineering.

2002

DECLARATION

I, Leonard Msibi, hereby declare that the work contained herein was solely carried out by myself. This includes inter alia, literature, material organisation and compilation thereof.



L.L. Msibi

ABSTRACT

The work contained herein is in pursuance of the Development of the SMART aircraft bolt. Failure of the bolt in the aircraft wing is taken for granted in the project, and the consequent repairs are presently very costly. The SMART material investigated in this work is the TRIP steel, and any reference to SMART material, in this work, shall at all time mean TRIP steel.

Investigation of the stresses pertaining to bolts in general is carried out, based on the bolt theories and using the finite element analysis. An optimal bolt based on impact resistance only has also been suggested. Metallurgical behaviour of materials similar to TRIP materials is also investigated, including a section dedicated only to TRIP steels.

Therefore, the work contained herein acts as a good base for further research.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Professor V. Verijenko and Professor S. Adali for their support throughout this project. Also to thank Mr. R. Bodger and Miss B. Burton for their assistance. And finally, all the PhD students for their assistance.

DEDICATION

I would like to dedicate this work to my wife, Hlakaniphile and my two children, Siphiwe and Anale.

Table of Contents

INTRODUCTION

The project herein undertaken is an integral part of a multifaceted effort to the development of a more cost saving, more predictable in failure, and a (hopefully) revolutionary smart bolt. The bolt itself is to be constructed from a special material having special clearly identifiable metallurgical transformations. These transformation also change the physical properties of the material. That is, the material changes from a non-magnetic to a magnetic material following the metallurgical transformation. The material at hand is called TRIP (transformation induced plasticity) steel. Since the induced transformation is induced through plastic deformation, the project seeks to investigate all possibilities of achieving a material with the required properties of a TRIP steel. A variety of material samples has been collected and investigated by other role players in the project, and the process is still underway. At the same time, samples were tensile tested to examine the magnetic changes. The contribution of this theses in the project entailed the investigation of the wide variety of the bolt aspects. This started with the investigation of the various bolt thread geometries, their effect on bolt overall stress concentration, including the manufacturing aspects for producing bolts with certain geometries. Some brief history of bolt thread development and standardization has been included. Also the general failure mechanisms of bolts have also been investgated. This has been considered under various types of loading conditions, including important variables under each type. The main loading types included have been static, fatigue and impact loading.

The strength of Finite Element Analysis has been utilised to further investigate geometries of interest. Also, this analysis has given more insight into areas of high stress concentration in the bolt in general, and special bolt geometries in particular. A theoretical discourse of finite

element has been provided, and the description of the particular software utilised also given.

Since the transformation behaviour of a TRIP steel is equivalent to austenitic (AISI 300 series) stainless steel when subject to deformation, a detail of all aspects of this material has been provided. The discussion touches on all aspects of transformation to martensite, from metastable austenite. Metallurgical deformation transformation are similar. Here the advantage was the preponderance availability of researched material in the topic of how metastable austenite transforms to magnetic martensite. Dislocations and stacking fault effects in FCC materials have been discussed.

The reader will hopefully knit all aspects discussed into one directed unit in the development of a proper understanding of a smart bolt in all its aspects.

CHAPTER 1

Basic thread geometry

1.1 Sharp thread (V thread) :

This thread type, shown in figure1.1, seems to be the most primitive and simple type of bolt thread without any notable mechanical advantages. However it is a geometry upon which a host of bolt geometries can be derived. The following list represents notable disadvantages:

1) Difficult to produce: Sharp crest and thread root without any amount of flatness creates some difficulty in making tools for forming the bolt, even for ductile materials. The task is even more difficult for brittle materials. More so the crest is more vulnerable to damage. The stress concentration is generally approximated at a value of k_t=3 at the root thread.

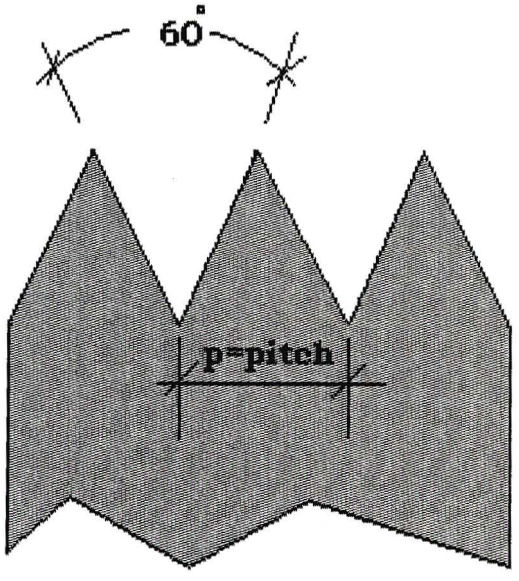


Fig1.1- Sharp V thread

1.2 American National thread :

This type of thread, shown in figure1.2, constitutes an improvement on the previous thread. It constitutes an improvement in manufacturing ability and mechanical properties. Some flatness at both the crest and the root has been introduced.

However, sharp corners at the root of the thread still constitute a weakness for fatigue. The stress concentration at these locations is approximated at k_t=2.5.

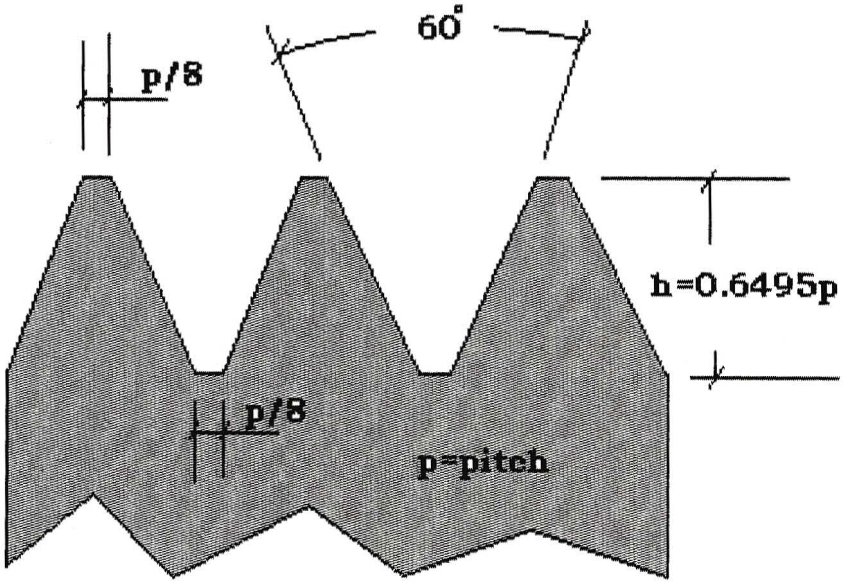


Fig1.2 – American National Thread

1.3 Whitworth thread :

The Whitworth thread, shown in figure 1.3, is rounded both at the crest and the root of the thread, thus more friendly with respect to fatigue loading. It has a 55^0 thread angle.

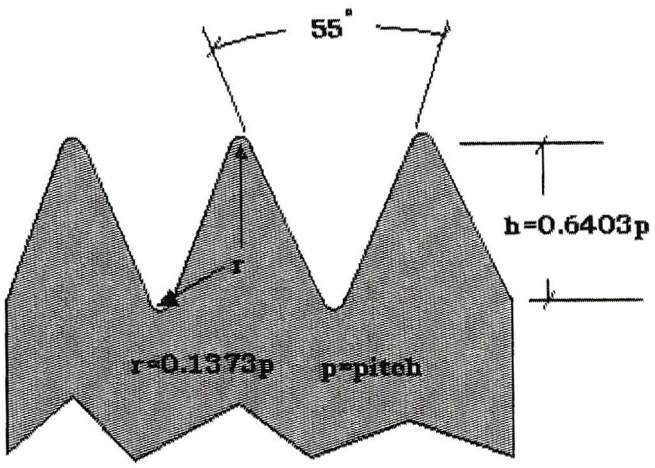


Fig 1.3 – Whitworth thread

1.4 Unified thread :

Due to the non-interchangeability of the Whitworth and the American national thread, one standard thread was adopted by Great Britain, United States, and Canada. The agreed thread type was termed the **Unified thread**, shown in figure 1.4, and it had a lot of resemblance to the American National thread.

The complete information about diameters, bolt pitch, thread dimensions etc. is all published by ASME (American Society of Mechanical Engineers) in conjunction with ANSI (American National Standards Institute).

The Unified thread is available as coarse (UNC-Unified coarse) for general use and conditions of general easy assembly.Also there is the fine series (UNF-Unified fine) for more critical situations like in the aircraft industry.

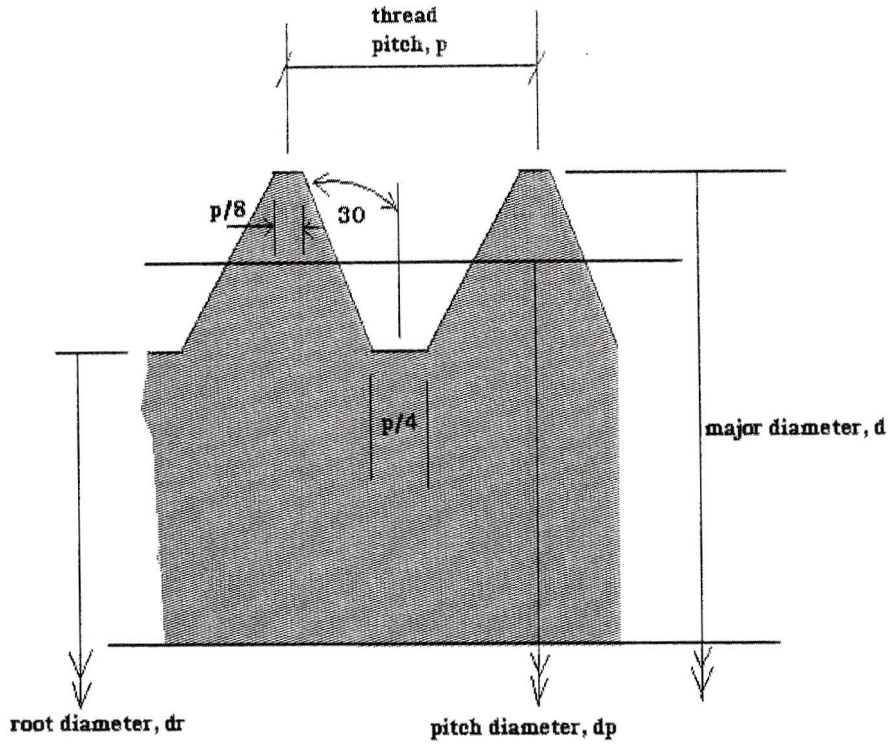


Fig 1.4 – Unified thread

Depending on the tolerances and tightness of fit required, classes 1A, 2A and 3A for external thread and 1B, 2B and 3B for internal thread are specified with tightness of fit increasing with numbers. However, classes 2A and 2B are usually specified.

1.5 ISO metric thread :

These threads have exactly the same basic form as the Unified threads, however the designations of sizes and thread classes are quite different. The designation of pitch size and other dimensions are based on millimetres, there is no reference to inches.

CHAPTER 2

Bolt loading

2.1 Tensile stress area :

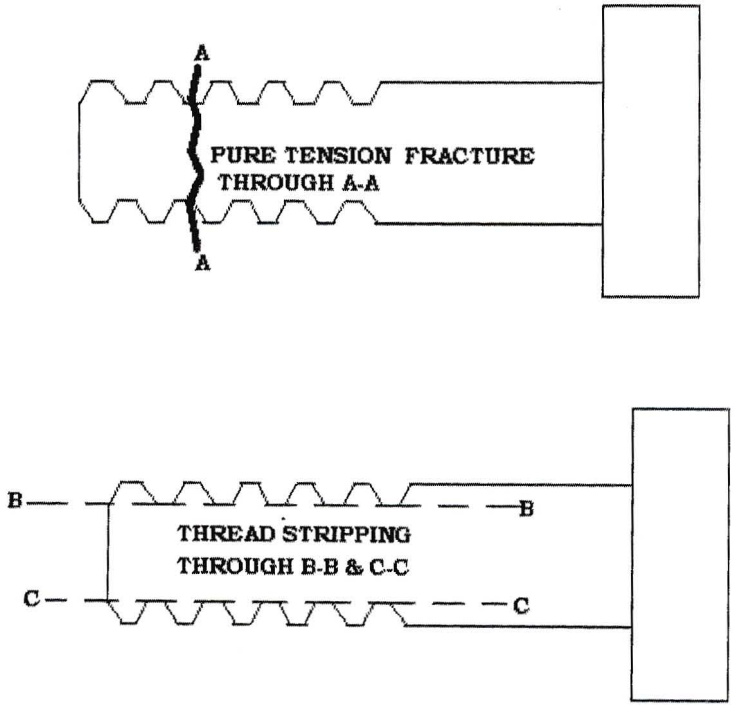
This is the area used for Force/Area stress calculations. It is based on the average of the root and the pitch diameter. It can thus be expressed as :

$$A_t = \frac{\pi}{4}\left(\frac{d_r + d_p}{2}\right)^2, \qquad (1.1)$$

The tensile stress area is tabulated in tables for Unified or Metric threads mentioned above, for various diameters.

2.2 Static Loading :

Fig.2.1 – Bolt static failure modes



There are two different ways in which a bolt can fail under static conditions, as shown in figure 2.1 above. The two ways are discussed below.

2.2.1 Pure tension :

Under this mode of failure, the bolt fails under a pure tension force.
Failure location will always take place in the tensile area A_t as defined above, upper diagram in figure 2.1.

2.2.2 Thread stripping :

Thread stripping occurs by shear in the thread root diameter, d_r , lower diagram in figure 2.1.However, a balance is usually made between thread stripping and bolt tensile yielding so that threads are not stripped before the bolt had undergone some yielding. Therefore the design criteria in this case is to increase the length of engagement of threads to compensate for for any material strength weakness of the nut.

If the bolt and nut are made of the same material, then :
Tensile force needed to yield the tensile area: $F1 = A_t S_y$ (2.1)

Shear force yield entire thread stripping area: $F2 = \pi d_r (\eta t) S_{sy}$ (2. 2)

At = Tensile stress area of the bolt
Sy = Yield strength of the bolt material
dr = Root diameter of the bolt
t = Total length of thread engagement or nut thickness
η = Total thread root area/Total threaded area
S_{sy} = Shear yield strength of the bolt.

Equating F1=F2 yields the required minimum thickness t.

2.2.3 Shear loading

This mode of loading is usually considered to be acting in the shank area of the bolt. The shank area is considered to be resisting shear if there is no high initial tension to resist the shear load by friction. If the threads are part of the the shear area the shear strength of the bolt is considered to be lowered by 25%. These type of bolts are usually found in bearing bolts for structural steel flexible joint connections .

2.3 Initial tension

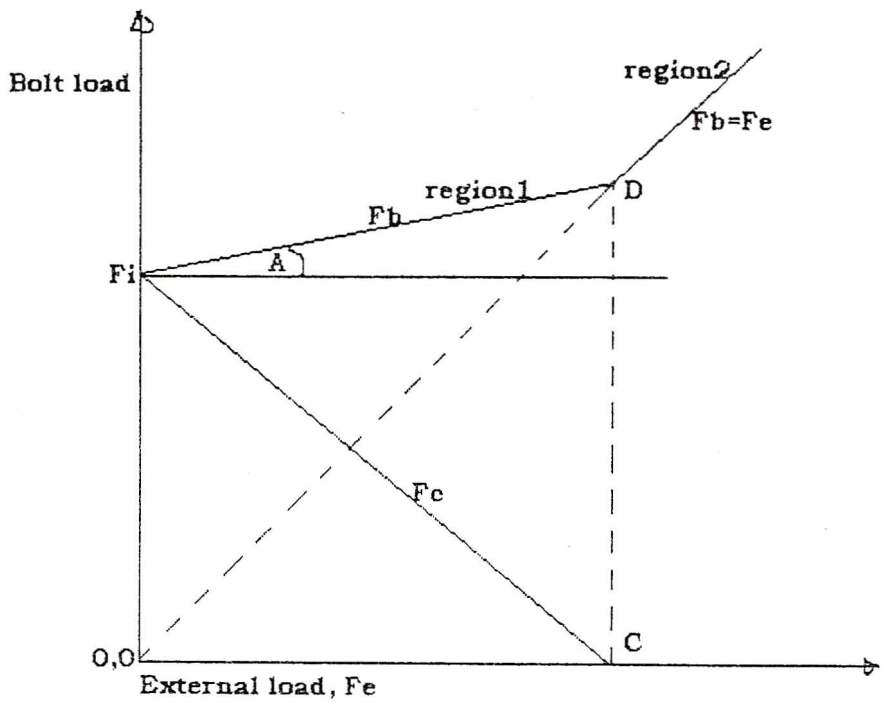


Fig. 2.2 – Bolt static load diagram

For bolts in tension, it is always necessary to tighten the bolts to some initial tension load F_i close to the yield strength of the material. This is usually the case where leak-tight joints are needed. Initial tension force is

usually a necessity for pressure vessels to ensure that leak tightness is always be maintained, and also to ensure that gaskets are properly seated . Once the external load is applied to the joint, there usually is enough clamping force F_c to ensure that there is no joint separation, point C in fig.2. Also as the external load Fe is applied, the bolt load increases slowly (region1) due to Fi, and the joint will not fail until the bolt load is equal to the external load, Fb=Fe, region2 in figure 2.1. If there were to be no initial tension force, the joint would fail immediately upon the application of load Fe, line (0,0 – D).

Initial tension also helps to increase the shear capacity of the joint because shear loads are also resisted by shear between the clamped members.

Initial tension is usually expressed as $F_i = K_i A_t S_p$,

Where :

S_p = Proof strength generally ,85% of the yield strength of the material.

K_i = 0.75 to 1, depending on the criticality of F_i. It is further suggested[2] that factors of 0.9 and 0.75 for permanent and reusable connections respectively be used.

A_t = Tensile stress area.

The bolt force in region1 is expressed as :

$$F_b = F_i + \frac{k_b}{k_b + k_c} F_e \qquad (2.3)$$

and the corresponding clamping force :

$$F_c = F_i + \frac{k_c}{k_b + k_c} F_e \qquad (2.4)$$

where k_b and k_c are the spring constants of the bolt and clamped members respectively, and they can be expressed as:

$$k_b = \frac{A_b E_b}{g} \tag{2.5}$$

$$k_b = \frac{A_c E_c}{g} \tag{2.6}$$

where :

A_b = Cross sectional area of bolt

A_c = Effective area of clamped members

E_b = Elastic modulus of the bolt

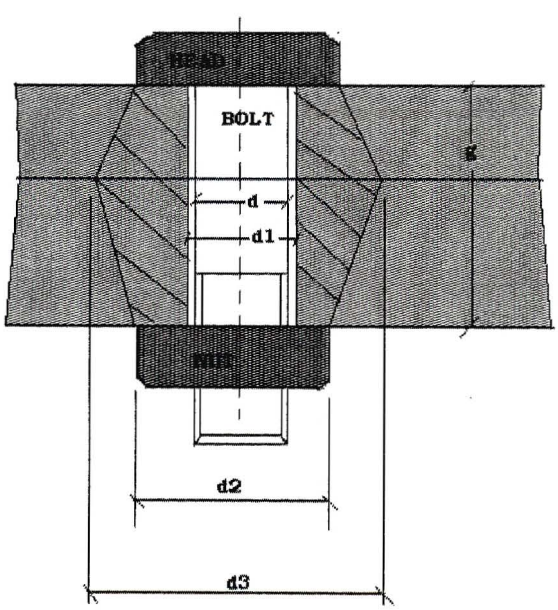
E_c = Elastic modulus of clamped members

g = grip length

For joint with a series of different clamped members, the effective spring constant , k_{eff}, is found as :

$$1/k_{eff} = 1/k1+1/k2+1/k3... \tag{2.7}$$

An empirical formula is usually used to determine the effective area of



clamped members, see figure 2.3.

$$A_c = \frac{\pi}{4}\left[\left(\frac{d_3 + d_2}{2}\right)^2 - d_1^{\,2}\right] \qquad (2.8)$$

where:

$d_1 \approx d$ (for small clearances)

$d_2 = 1.5d$ (for standard hexagonal head bolts)

$d_3 = d_2 + g \tan 30^0 = 1.5d + g \tan 30^0$,

and substituting in equation (2.8) yields

$$A_c = \frac{\pi}{16}(5d^2+6dg \tan 30^0+g^2\tan^2 30^0) \approx d^2+0.68dg+0.065g^2 \qquad (2.9)$$

2.4 Torque and initial tension

During initial bolt tightening, some torsional load is applied to the bolt. The amount of torsional load will be influenced by the coefficient of friction in the threaded region. More friction between the threads will induce more torsional load. Some shear stress induced by the torsional load will remain even after tightening. This will intend to increase the equivalent tension stress on the bolt.The increase of initial tension is more likely to increase under elastic loading conditions. The equivalent tensile stress is found from Von Mises:

$$\sigma_{eq} = \sqrt{(\sigma^2 + 3\tau^2)} \qquad (2.10)$$

However, initial tension is not always maintained under certain circumstances, especially for joints which undergo some yielding under the application of the initial load. Also some local yielding, due to some excess bearing stress under nuts and bolt heads (caused by high local spots, rough surface finish, and lack of perfect squareness of bolt and nut bearing surfaces), may lead to preload relaxation. Bolt tension also may be unevenly distributed over the threads in a joint, so thread deformation may occur, causing the load to be redistributed more evenly over the thread length. Other factors contributing to the preload relaxation are vibration, temperature cycling and creep.

2.5 Fatigue loading

2.5.1 Initial tension considerations

For bolted joints subject to fatigue loading , initial tension is of absolute necessity if the life of the bolt is to be improved. For an initial tension of Fi, figure 2.4, it can be seen that as the external load fluctuates between point (0,0) and E, the bolt load fluctuates between line1 and line 2, while the clamping force fluctuates between line 2 and line 3.Hence the bolt is subject to a lower amplitude of the fluctuating load. This, of course, assumes that the spring constants of both materials are greater than zero.

The bolt fatigue resistance is improved by ensuring that angle A in figure 2.4 is decreased. This can be achieved through increasing the spring constant k_c relative to k_b. As this happens, the bulk of the fluctuating load goes to decreasing the clamping force F_c.

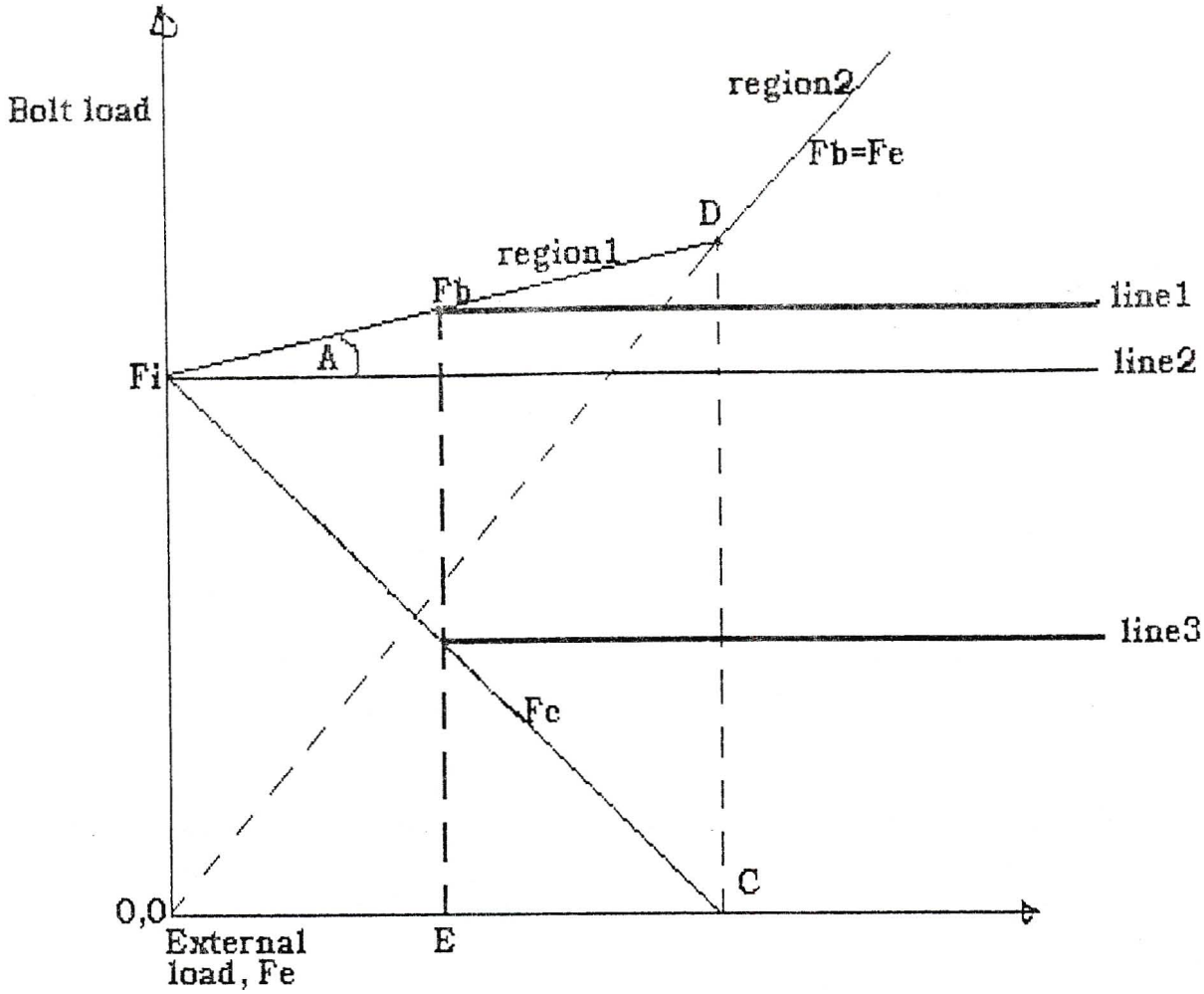


Fig. 2.4 – Bolt load diagram, showing dynamic load limits

2.5.2 Stress concentrations

For static loading, with the material that follows the ideal stress-strain graph for ductile materials, the effect of stress concentrations can be neglected. This is due to the fact that the load at the onset of yielding will never exceed the yield stress, and stresses are subsequently redistributed. Hence the bolt under static loading will still fail as described in figure 2.1 above.

However, for fatigue loading stress concentrations are the primary reasons for fatigue cracks to initiate. Those are the areas of localised plasticity, and the relative movements of slip planes usually lead to slip bands developing from which fatigue cracks have the tendency to develop. For the bolt there

are regions of stress concentrations from which bolt failure is expected to initiate. However the severity of these stress concentrations from these different regions are not the same. Well known regions of stress concentrations are :

(1) Fillet under the head
(2) End of the thread to shank transition
(3) Face of the nut.

The susceptibility to failure at these three regions is distributed as follows: 15% under the head fillet, 20% at thread to shank transition, and 65% at the face nut. However, it is also suggested that the bolt will always fail in the face of the nut, see [4]. The reasons for failure at the face of the nut is that the threaded region has the highest stress concentration for the entire bolt, and that the first thread engaged by the nut take most of the load. However, it is also suggested by [5] that the first thread takes all the load, thus suggesting a very poor distribution, if existent, of the load applied. The predominant view on this matter, however, is that all the threads engaged share in the load applied albeit unequal. Figure 2.5 shows the distribution of the load in the threads.
Part of improving the fatigue life of the bolt is to leave some few threads from the shank when engaging the threaded region because tightening to the end of the threads superimposes stress concentrations (2) and (3) above.

2.6 Manufacturing methods and stress concentration

There are various manufacturing methods for the production of bolt threads. These methods involve cold forming, cutting and grinding. Cold forming involves deformation through dies properly shaped for the formation of threads. Since the deformation is mainly compressive, therefore, residual stresses of compressive nature will result. The root of the threads will usually obtain a larger share of residual compressive stresses, and hence

improved fatigue resistance of thread roots is obtained, thus giving a favourable fatigue stress concentration than that for cut and ground threads.

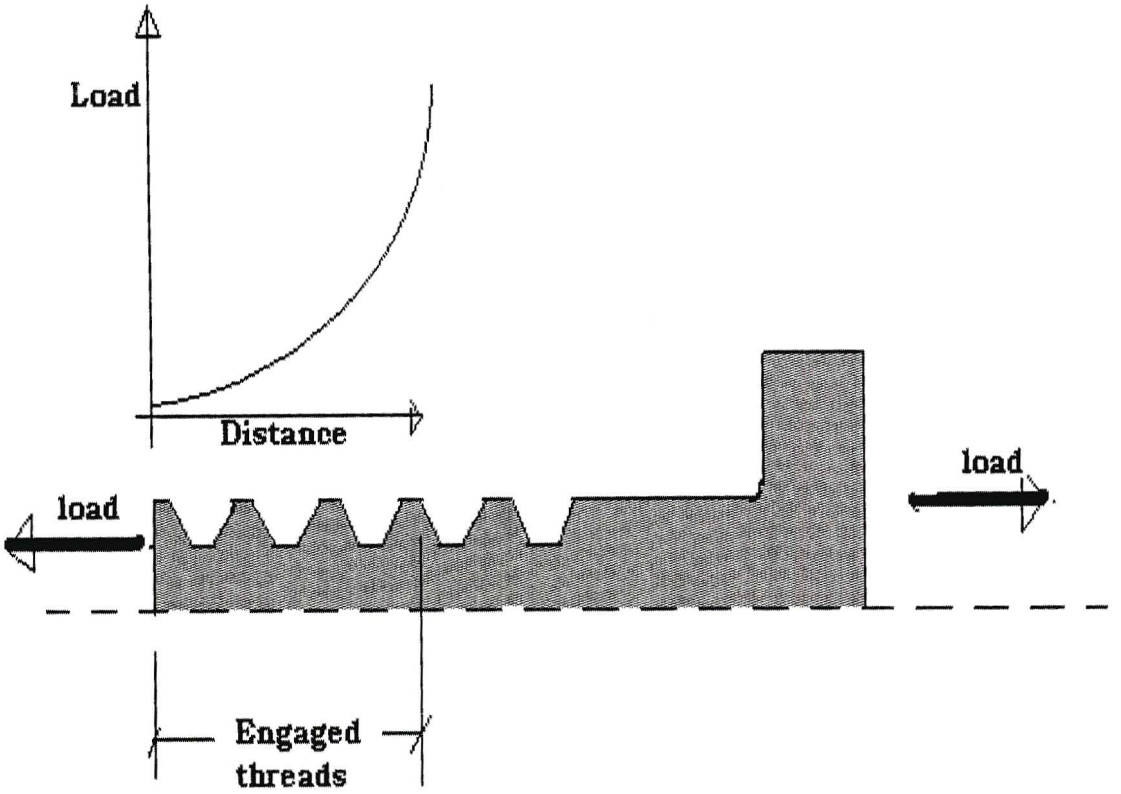


Fig. 2.5 – Load distribution in threads

2.7 Impact loading

To demonstrate the type of loadings existing under impact loading analytically, we take a case of a falling mass of weight W (Newtons) falling a height of h (metres) to a structure idealised as a spring. This assumption is premised on the fact that the structure will only undergo elastic strains. Normally the strains and the stresses experienced by a body under impact loadings will be larger than those experienced under static, and slow applications of the same weight. Further, we will assume that the

instantaneous deflection resulting from impact, is the same as the static application of the same weight multiplied by the impact factor. The equivalent static load in terms of the impact factor will be derived. We assume that the structure has a spring constant k (N/m). Also, assume the maximum value of deflection due to impact as δ, and corresponding real static deflection due to weight W (N) after energy is damped out and the body (of weight W (N)) comes to rest is δ_{st} . Therefore,

$F_e = k\delta$ (2.11)

F_e = Equivalent static force that would give same deflection.

Static deflection = δ_{st}

$= W/k$ (2.12)

(After energy is damped out, and weight comes to rest).

Potential energy given up by falling mass = Elastic energy absorbed by spring

$$W(h+\delta) = 1/2F_e \quad (2.13)$$

Factor ½ means spring takes the load gradually.

$F_e = (\delta/\delta_{st})\ W$ (by defn) or $\delta/\delta_{st} = F_e/W$ (2.14)

Substituting into $(h+\delta) = \frac{1}{2}(\delta^2/\delta_{st})W$ (2.15)

$$\delta = \delta_{st}\left(1+\sqrt{1+\frac{2h}{(\delta_{st})}}\right) \quad (2.16)$$

$$Fe = W\left(1+\sqrt{1+\frac{2h}{(\delta_{st})}}\right)$$

Factor in brackets is the impact factor.

Now, expressing above equations in terms of velocity at impact :

$V^2 = 2gh$ or $h = \frac{v^2}{2g}$ (2.17)

Therefore $\delta = \delta_{st}\left(1+\sqrt{1+\frac{v^2}{g\delta_{st}}}\right)$ (2.18)

$$Fe = W\left(1+\sqrt{1+\frac{v^2}{g\delta_{st}}}\right) \quad (2.19)$$

$h \rightarrow 0$ means suddenly applied load.

For $h >>$ deflection (i.e δ_{st}), therefore implies $h/\delta_{st} >> 1$ in equations (2.18) and (2.19), therefore the equations reduce to

$$\delta = \delta_{st}\left(\sqrt{\frac{2h}{\delta_{st}}}\right) = \sqrt{2h\delta_{st}} \quad (2.20)$$

$$F_e = W\left(\sqrt{\frac{2h}{\delta_{st}}}\right) = \sqrt{2Whk} \quad (2.21)$$

And, with velocity we get:

Therefore $\delta = \delta_{st}\left(\sqrt{\frac{v^2}{g\delta_{st}}}\right) = \sqrt{\frac{\delta_{st}v^2}{g}}$, and (2.22)

$$F_e = W\left(\sqrt{\frac{v^2}{g\delta_{st}}}\right) = \sqrt{\frac{v^2kW}{g}}, \text{ where} \quad (2.23)$$

$$\delta_{st} = W/k \tag{2.24}$$

Now, using impact kinetic energy :

$$U = \frac{1}{2}mv^2 = \frac{Wv^2}{2g} \tag{2.25}$$

Substituting equations (2.25) into equations (2.22) and (2.23), we get

$$\delta = \sqrt{\frac{2U}{k}} \tag{2.26}$$

$$F_e = \sqrt{2Uk} \tag{2.27}$$

2.7.1 <u>Special case</u> of linear impact is a straight bar

- Either tensile or compression stress
- Impact applied concentrically
- Stress concentration neglected

$$\sigma = \frac{F_e}{A} \; ; \tag{2.28}$$

where $k = \frac{AE}{L}$, therefore (2.29)

$$\sigma = \sqrt{\frac{2UE}{AL}} = \sqrt{\frac{2UE}{V}}, \tag{2.30}$$

where V is the volume of the material.

Now, expressing the above equation in terms of impact energy U, we obtain

$$U = \frac{\sigma^2 V}{2E} \tag{2.31}$$

2.7.2 Elastic energy capacity :

If we take a straight rod with no variation in cross section. We must also note that a straight rod is uniformly stressed throughout its cross section.

Taking $\sigma = S_y$, impact energy is expressed as $U_a = \frac{S_y^{\ 2}}{2E}$ (2.32)

2.7.3 Evaluating variation in cross section

V= volume of full length rod.

$$U_{bl} = \frac{S_y^{\ 2}}{2E}(V/2) = \frac{1}{2}U_a \tag{2.33}$$

$$U_{bu} = \frac{(S_y/4)^2(2V)}{2E} = 1/8U_a \tag{2.34}$$

$$U_{bl} + U_{bu} = 1/2U_a + 1/8U_a = 5/8U_a \tag{2.35}$$

Thus from equation (2.35), it is clear that a varying cross section has less impact energy than a straight rod. It therefore implies that a rod or bar which is not uniformly stressed throughout its cross section has less impact capability than a straight, uniformly stressed bar. So for impact loading, the extra cross section which is less stressed under loading conditions does not add additional impact capabilities but lowers the overall impact energy.

2.7.4 Evaluating the effect of stress concentration

Assumptions :

1) Rod sufficiently long, and volume of the material in the region of stress concentrators (fillets) is a very small fraction of the total volume.
2) The material at the critical fillet location cannot be stressed in excess of material strength S.

Therefore, nearly all the material can be considered as stressed to a uniform level that cannot exceed S/k_t, where k_t is the stress concentration at the fillet regions.
Thus, a good approximation is that after considering a stress raiser, the same volume of material is involved (i.e. assumption (1)), but at a stress level reduced by a factor of k_t.
Because the stress is squared in the energy equation, the energy capacity is reduced by a factor of k_t^2.
Therefore, the effect of stress concentration is to reduce the overall impact energy of the part under impact loading.

2.7.5 Suggested optimum bolt

Following from sections 2.7.3 and 2.7.4 above, it was observed that the impact energy of a member can be increased by ensuring that a uniform stress profile is achieved throughout the part, and that all stress concentrations should be eliminated or minimised. However for a bolt and many structural components it is difficult to eliminate stress concentrations altogether. What we can try to achieve however is to maintain the same stress level throughout the bolt in spite of stress concentrations.
The suggested bolt is the one with the hole in the centre, as seen in figure 2.6.

However, suggestion in this text seeks to construct a bolt geometry whose dimensions are based on impact energy requirements as discussed above. In this construction, however, the sensor diameter would have to be known upfront (Or maybe the sensor diameter may be deduced once the other variables are known).

$$A_{se} = \pi \frac{d_{se}^{\ 2}}{4} \; ; \tag{2.36}$$

A_{se} = Area of the sensor hole.

A_r = Tensile stress area

P = applied bolt load

K_{ts} = Shank/fillet stress concentrator

K_{tr} = Thread root stress concentrator

To achieve the uniform stress throughout the shank and the threaded area, we must have:

$$\frac{P}{A_{shank}} k_{ts} = \frac{P}{A_r} k_{tr} \tag{2.37}$$

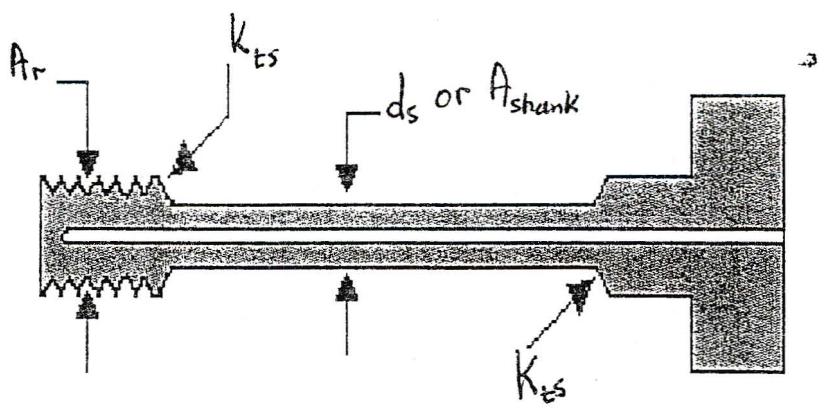


Fig. 2.6 - Suggested optimum bolt

Therefore, the required shank area to develop the uniform stress level in the bolt is represented as:

$$A_{shank} = \frac{k_{ts}}{k_{tr}} A_r , \tag{2.38}$$

Also it is interesting to note that the ratio of the stress concentrations (i.e. the machined profiles in the bolt) can actually determine the required area of the shank, with the same tensile stress area.

But the area of the shank material is represented as:

$$A_{shank} = \frac{\pi d_s^{\,2}}{4} - A_{sensor} , \tag{2.39}$$

where d_s is the outside diameter of the shank annulus.

Therefore, the outside diameter d_s is represented as:

$$d_s = 2\sqrt{\frac{A_{shank} + A_{sensor}}{\pi}} \tag{2.40}$$

CHAPTER 3

Actual bolt modelling

3.1 Nastran Software

MSC Nastran is a Windows based finite element software. It entails both the geometric capabilities and an analysis tool. For an entirely Nastran based model, the model is generated using a number of available geometric tools. There are line based geometries like an I-beam. For this, a line is available from the geometric menu. It is up to the user to assign the geometric properties of the beam which are also available. However, for this type of modelling, all the meshing and the loads are applied on a line element. In order to achieve the required geometry, Nastran offers an element property type menu from which you can choose whether you dealing with a rod, tube, bar etc. For the beam elements a shape menu is available in order to specify whether an angle beam is required or not.

Also, there is a surface type element which one can create with Nastran modelling. With this there are various methods available to achieve the objective. The most utilised methods are (i) sweeping or rotating a line object or (ii) defining a surface through a certain number of boundary lines that form an enclosure. For plane surfaces it may be preferable to sweep a line or define boundaries. However, for a curved surface like pressure vessels and their components it may be preferable to rotate a line element.

The solid element is also available. This is achieved through rotating or extruding a surface. Also there are standard solid geometries available.

Once the above steps have been successfully completed, Nastran then requires the user to specify the construction **material**. The are a lot of already available materials on the Nastran database. They are available in different units. For example, the way the numerical value for the modulus of elasticity of the material is specified will depend on the units. Once a specified material is chosen, all the available properties will be shown. In

most of the cases an isotropic (i.e. the direction in material properties is inconsequential) material, like steel, is selected.

Following the above steps, one must then supply the material **property**. This details the type of elements to be found in the analysis. It also describes in advance what kind of meshing should be created by the software. For example, if a solid property is selected then the software is told in advance that three dimensional solid elements should be used in the subsequent meshing exercise. For a plate property, the normal two dimensional elements are used.

There are other properties like axisymmetric elements. Here, in doing the meshing a global stiffness matrix with properties different than the normal two dimensional element is obtained, and the conventional rectangular co-ordinates are not used in constructing the model. This is clearly demonstrable in the corresponding global stiffness, which does not contain constants but entries which are functions of the radial variable. The structures modelled with these type of elements normally do not have varying properties in the circumferential direction. Also, these elements are used when there is a need to use the symmetry of the structure by modelling only, say, a quarter of the total structure rather than modelling the whole 360^0 structure. There is then an opportunity to increase the number of elements for critical locations and thus more accurately predict the behaviour of the structure.

Then software used by students was limited to a maximum of 5000 nodes. So this dictated the economical use of nodes. This was however limiting under certain circumstances. There were various meshing methods available under Nastran software. For a solid structure there were three possibilities: (i) Just meshing the structure. In this case the software decides to apply the best meshing for the structure, there is no control over certain critical regions. (ii) Meshing a two dimensional structure, and then extruding the elements of the two dimensional structure with a result of achieving a meshed three dimensional structure. Most of the structures were such that that method was not really suitable. However, for structures like lifting lugs

in pressure vessels it might be suitable to apply that method because of the flatness of the structure, and all dimensions would remain similar in the extruding direction. (iii) Ensuring that there are curves in certain important and critical regions. When doing the mesh, one then specifies the number of nodes per curve. For curves in critical regions, one normally specifies more nodes in those curves. This method normally ensures that more elements are created in critical regions.

The type of elements for solid **meshing** are tetragonal (4-nodes per element) mesh, and this is a normal linear three dimensional element, also one can use a hexagonal (8-nodes per element) mesh. For structures with irregular shapes the tetragonal mesh is most suitable. For most of the analysis, linear elements are used. For increasing the accuracy in certain regions, one achieves this by increasing the number of elements, and not by increasing the number of nodes per element, thus leading to elements with non-linear interpolation functions.

Having meshed the structure, one may then proceed to apply the **constraints**. There are various constraints available in Nastran. There are normal constraints just for restricting the structure, they are available in different directions. Also there are constraints just to indicate the symmetry of a structure in a particular direction. All the constraints are applied on the nodes created.

Following the above steps, **loads** are then applied on the body. There is a variety of loads that can be applied to a body, depending on the application. There are time varying loads, pressure loads applied on specific elements or surfaces. Also one can apply nodal loads directly. A selection direction of application of the loads is also provided.

Once one has applied the loads, then the structure is ready for analysis. The calculations for producing an analysed body are carried out in the background. One can also choose the views preferred. One can choose to see the bending moment distribution, force distribution, displacement distribution or solid Von Mises stress distribution. All these are visual with different colours showing different values along the analysed body. If groups

of elements have been created when constructing the analysed body, one can choose to see only certain regions during the analysis. Also doing the section cuts of the structure is possible. In order to see certain internal stress behaviour, this method has also been applied in the bolt analysis, where the shank, threaded region and the bolt head have been shown separately to clearly show the nature of the stresses in these individual components.

3.2 Actual models

This section discusses the actual models analysed. All the sketches are contained in the appendix. Some sketches show different views for better visuals of the resulting stresses.

3.2.1 Model 1 :

This is the general model showing the stresses in the entire bolt, shown figure 3.1. The objective of showing the model is for the investigation of the grip to thread transition. The model was constructed through SOLID EDGE modelling and imported to Nastran for analysis. The main purpose was to see the profile of stresses in the bolt, and identify regions of high stress concentrations. The numerical value of stresses was not necessarily of primary importance. However the loading was properly assigned to give reasonable stresses.

The material for the bolt was the high strength steel, with the yield stress approximately 1480 MPa as found in the Nastran library. The load applied was about 80% of the yield stress of the material (approximately 80% of 1480 =1184), thus obtaining one of the ranges within which proof stress of the material lies. The property assigned to the material was of course a solid property. The meshing of the material was really unpleasing, especially with a limited version of 5000 nodes. The geometry contains complicated geometry, especially at the threads. Therefore certain regions like threads

were assigned more nodes through assigning a controlled number of nodes to certain curves in certain regions, like curves in the threads.

Loads were applied at the end of the bolt, a total load applied on the surface. There was no direct loading on the threads like the other models. The solid part at the end of the bolt was for the assurance that the loads are applied axially to the bolt, and that all the nodal forces remained parallel. If the part is removed it cannot be guaranteed as the surface at the end becomes uneven because of the helix angle.

There was only one set of constraints applied. The bolt was constrained just below the head with fixed constraints to simulate a bolt resting on the mating surface through the bottom part of the head. This was the constraint applied to all the models under analysis.

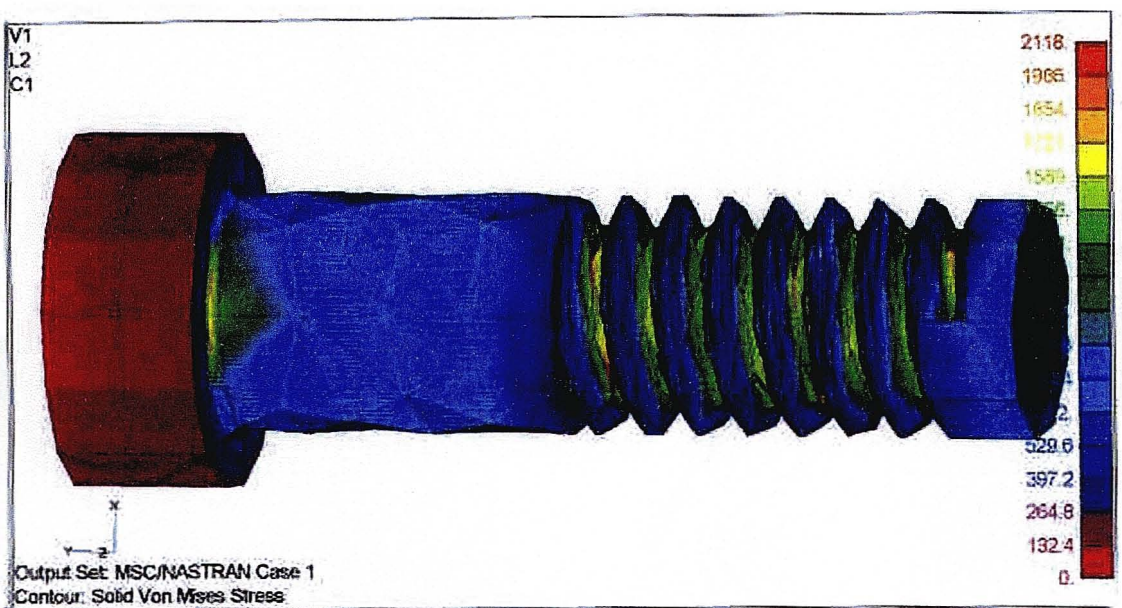


Fig. 3.1 – SOLID EDGE model, entire bolt

Following the analysis, the general pattern of stresses indicated stress amplification at the root of the threads in relation to the rest of the bolt. The stress concentration at the root of the thread, figure 3.2, indicated a general

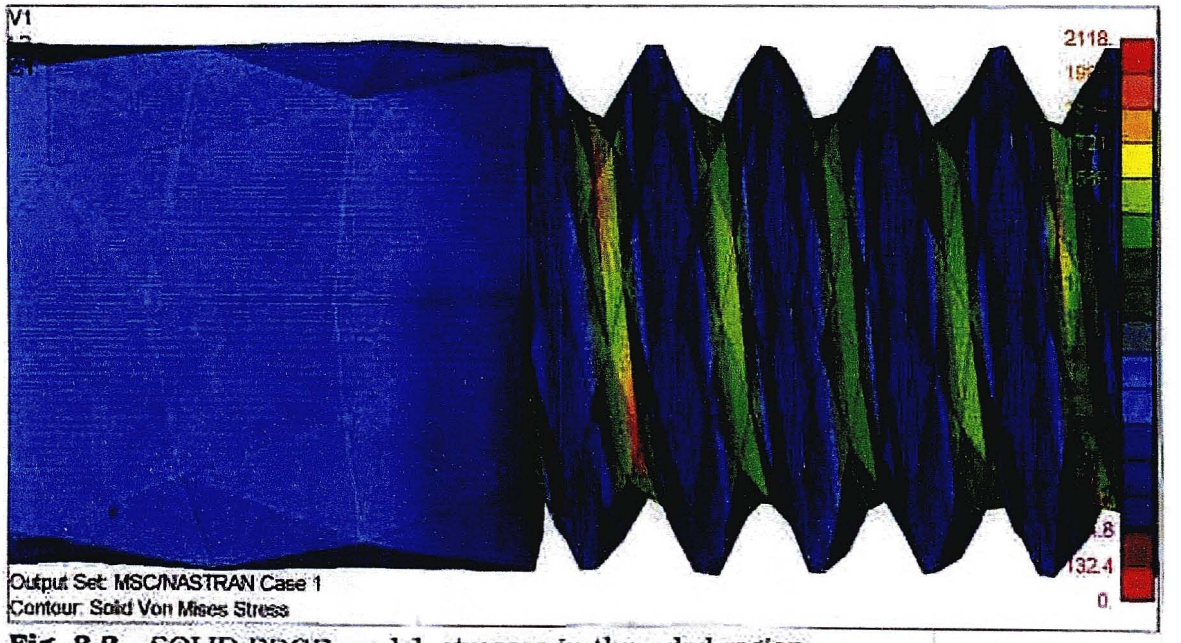


Fig. 3.2 – SOLID EDGE model, stresses in threaded region

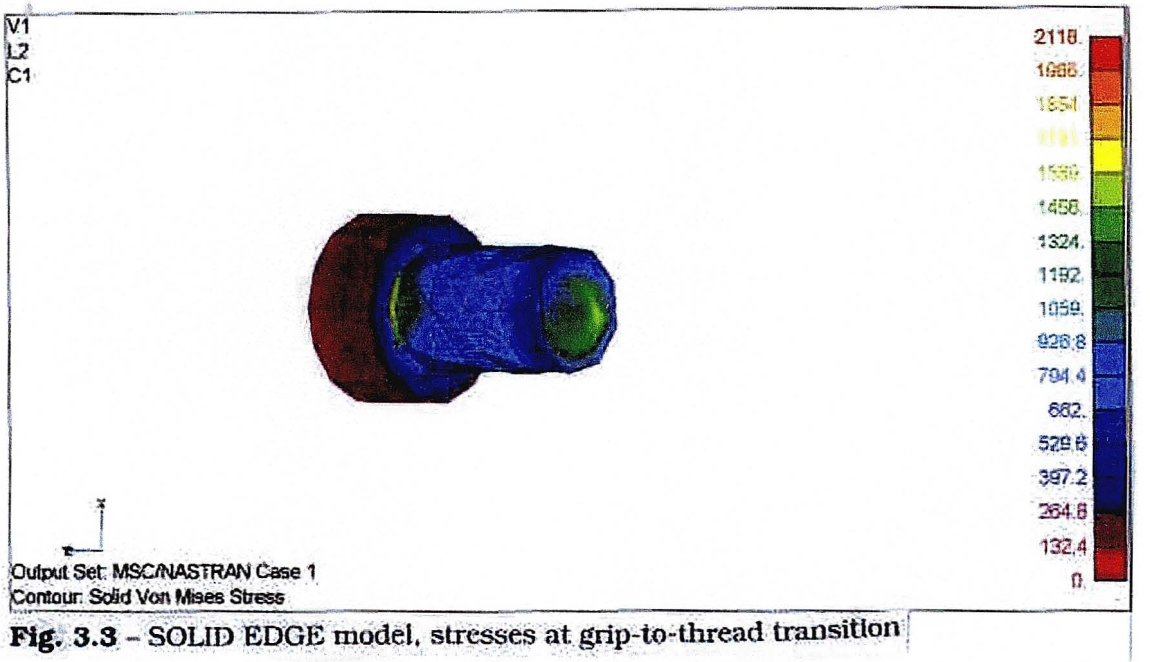


Fig. 3.3 – SOLID EDGE model, stresses at grip-to-thread transition

stress concentration of $k_t \approx 1.7$. Interestingly, and further disputing the unfounded popular belief, there is nothing special about the grip-to-thread transition region. It is not disputed that there is some measure of stress concentration there. However most of the literature referenced will point out that the root of threads at the face of the nut possess the highest stress concentration factor in the main stream threaded area, and most of the attention in improving the bolt fatigue life is geared at improving the main stream thread profile, not the grip-to-thread transition.

Also a lot of research has established that bolts threads are not equally loaded in the nut region. In fact an exponential loading situation is ascribed. The threaded region in the face of the nut is the most susceptible, and as such efforts have been made to improve nut design for the equitable distribution of loads. Refer to section on stress concentration. See figure 3.3, for the section-cut through the region of grip-to-thread transition.

As expected, the bolt head must have low stresses, especially taking into account the location of fixed constraints.

3.2.2 Model 2

The model geometry is constructed using Nastran solid modelling. The model is as approximate as the axisymmetric model (i.e two dimensional). On this score, a geometric modelling has been preferred so as to see a lot of features (i.e section cuts etc.).The bolt solid was a rotated two dimensional model. Although there would be a limitation on the number of elements due to a limited version (i.e. 5000 nodes), there would however be certain regions (i.e shank) where there would be no major variation of stresses except at the transitions, and hence a fine mesh could not warranted. A fine mesh would be biased towards the threads and transition zones, if any. This modelling sought to further investigate the valleys representing bolt threads, the shank stresses, and the shank to head transition. It would also enable us to apply an exponentially varying load through the threads engaged by the nut. Just like the axisymmetric model, the grip-to-thread transition

stress concentrator is not included. This can be safely excluded because, as pointed out in the previous model and elsewhere in this document, it is of no critical consequence to the bolt's ultimate failure mechanism.

The material of the bolt is taken from SAE class 12.9, with a yield stress of 1100 MPa, and a proof stress of 970 MPa. It is a high strength bolt, approximating a high strength material like a TRIP steel. Of course, the property is for solid elements.

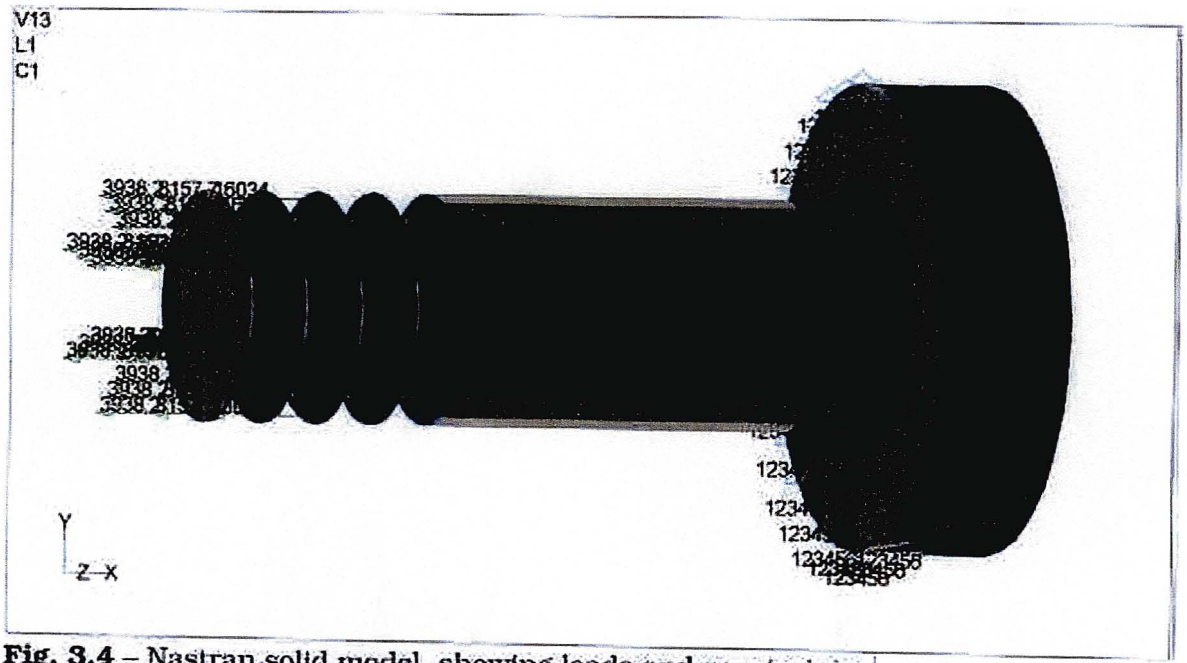


Fig. 3.4 – Nastran solid model, showing loads and constraints

Meshing the solid was by increasing the number of nodes on the thread surfaces facing the load, and having half the number of nodes on the thread surfaces opposite the application of the loads. Also the number of elements were increased on the root threads. Fewer nodes were created on the shank

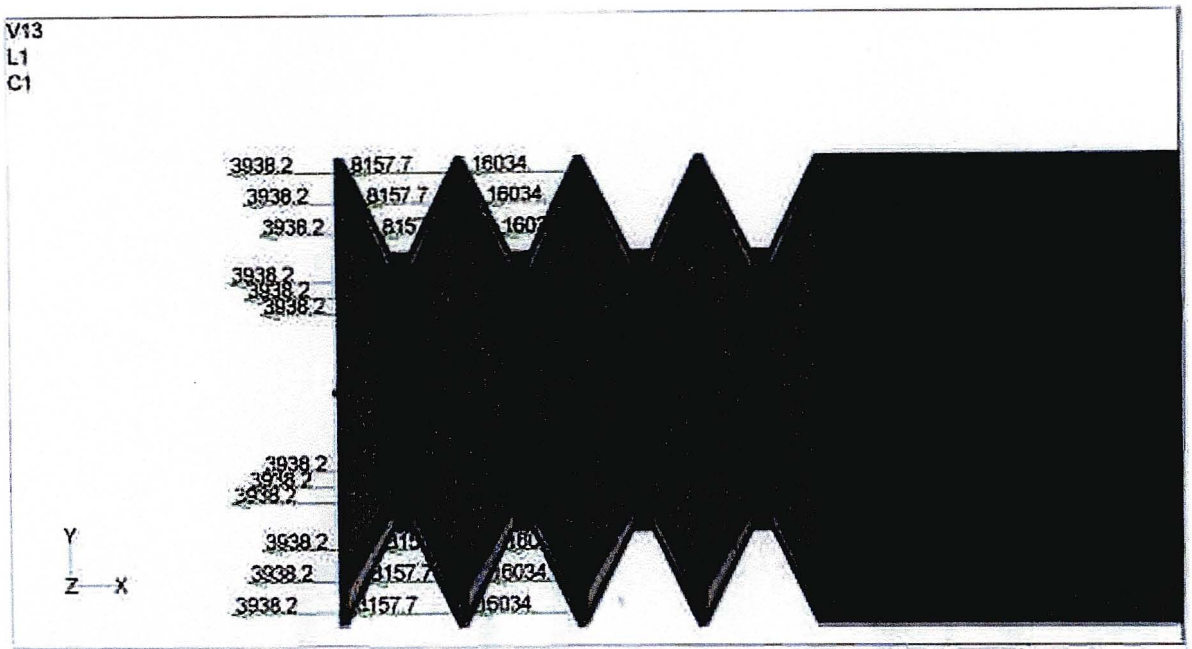


Fig. 3.5 – Nastran solid model, showing load distribution in the first three threads

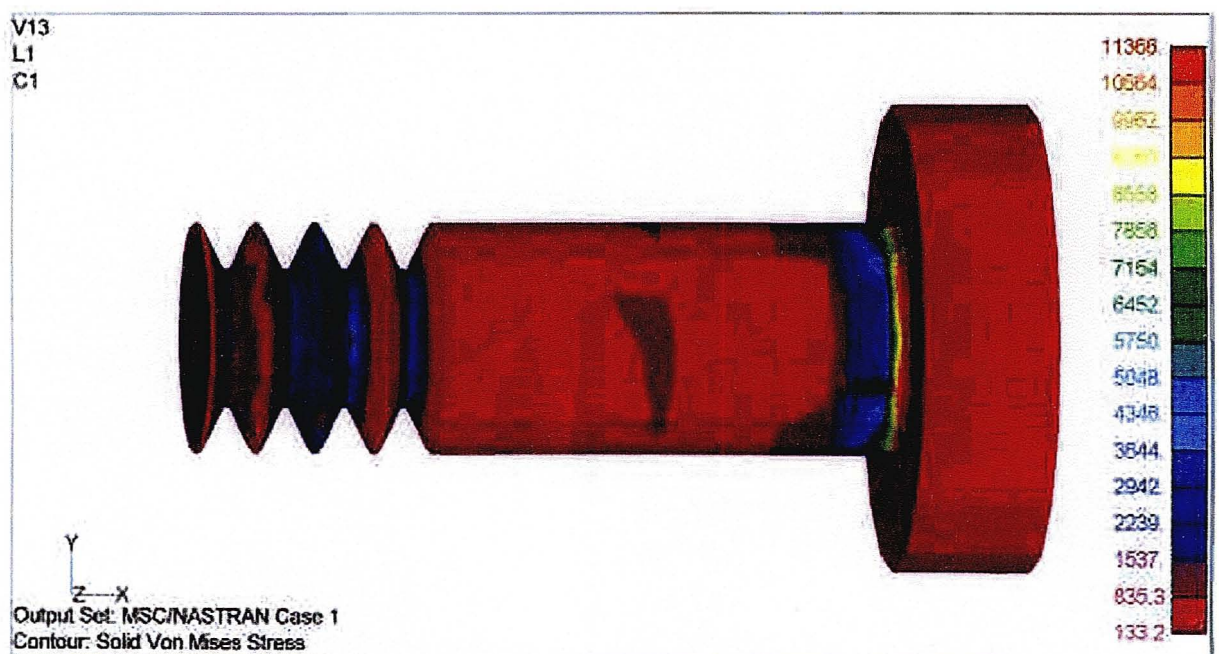


Fig. 3.6 – Nastran solid model, showing a stress profile in entire bolt

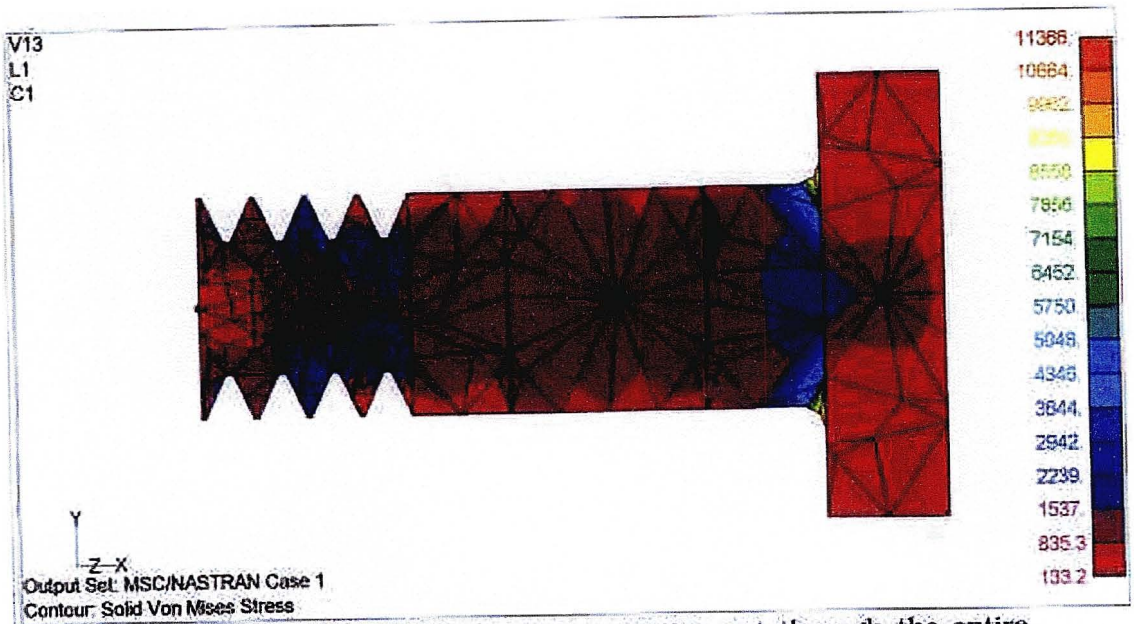


Fig. 3.7 – Nastran solid model, showing a section cut through the entire model

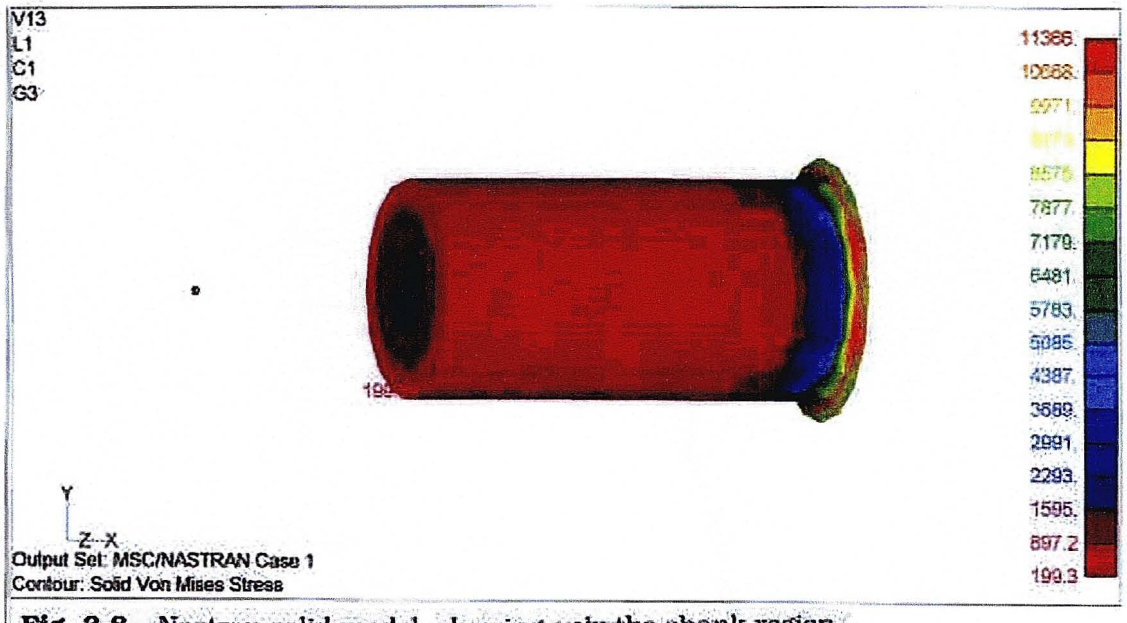


Fig. 3.8 – Nastran solid model, showing only the shank region

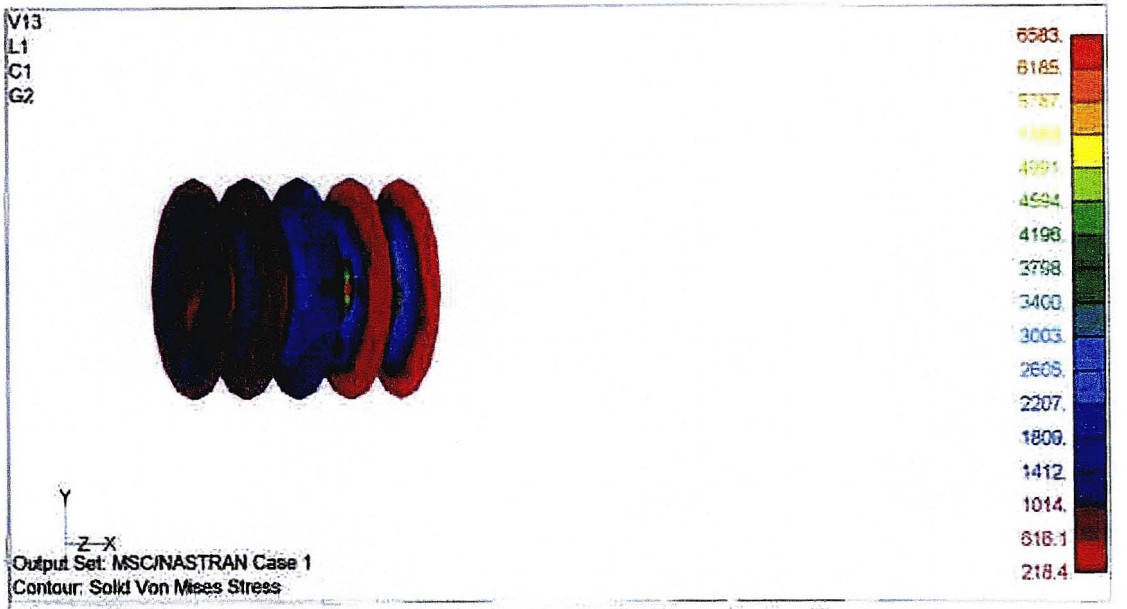


Fig. 3.9 – Nastran solid model, showing only the threaded region

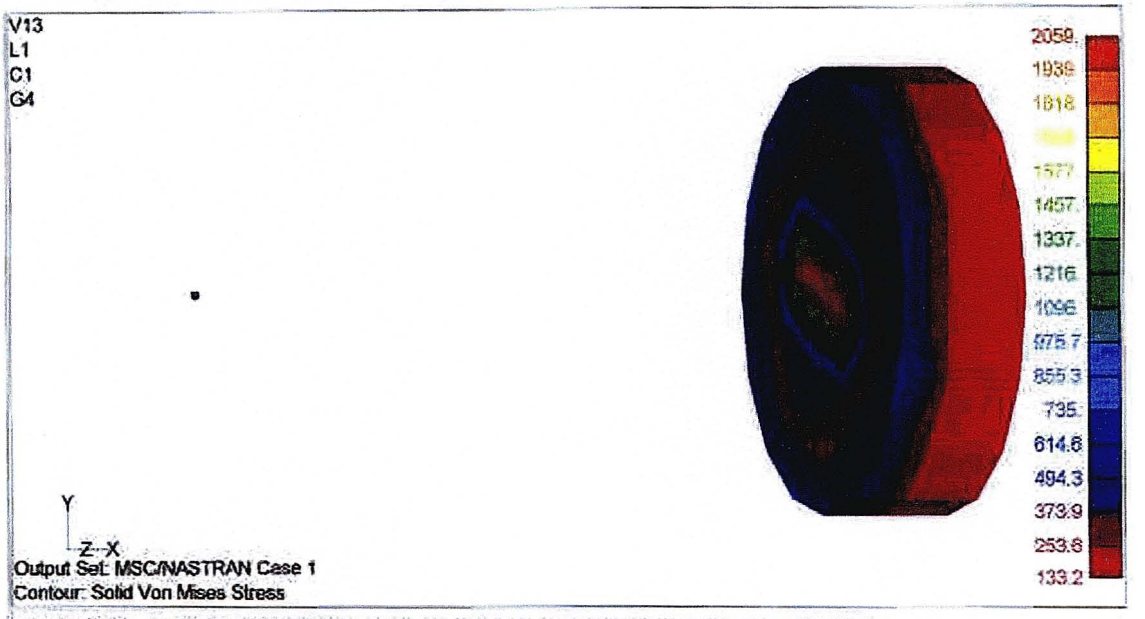


Fig. 3.10 – Nastran solid model, showing only the head region

and the bolt heads. But an effort was also made to increase the number of nodes on the fillet of the head to shank transition.

There was only one set of constraints just like the previous model on the bolt head side normally facing the support, see figure 3.4. Fixed constraints were put.

An exponentially varying load, approximately of the form $y = a^x$, was applied on the last three threads, see figure 3.5.

Similar to the previous analysis, the pattern of the stresses was of more interest. The stresses were a little bit higher at the thread roots. For this model, the observed general stress concentration at the thread root was around the value of $k_t \approx 4$, see figure 3.6. However, there was also an indication, from figure 3.9, that the stress is higher at the first thread where the highest load was applied. The abnormally high stress concentration at the head-shank transition was possibly due to non-symmetry of the nodes, and hence non-symmetry of constraints on the surface, thus increasing the stresses there. The solid was also sliced into different solid and element groups, namely, the thread, shank and head regions. Internal stresses could therefore be seen. Also see figures 3.7 to 3.10 for more features.

3.2.3 Model 3

The present geometry models an actual aircraft bolt as found in the MS-21250 specification. The bolt is of type UNJF class 3A. The solid geometry used again here was a rotated two-dimensional bolt model. But because the diameter was ¾ inch (about 19mm), a quarter of the solid model was utilized instead. The concept of symmetry was also applied in this case. That would increase the chance of a successful mesh within 5000 nodes.

The first geometry was just a plain model without any special features, see figure 3.11. The objective here was to investigate the general features. The actual UTS quoted for the material was that of the 4340 material used, and it was about 180 000 psi which translated to about 1242 MPa. The yield stress of the

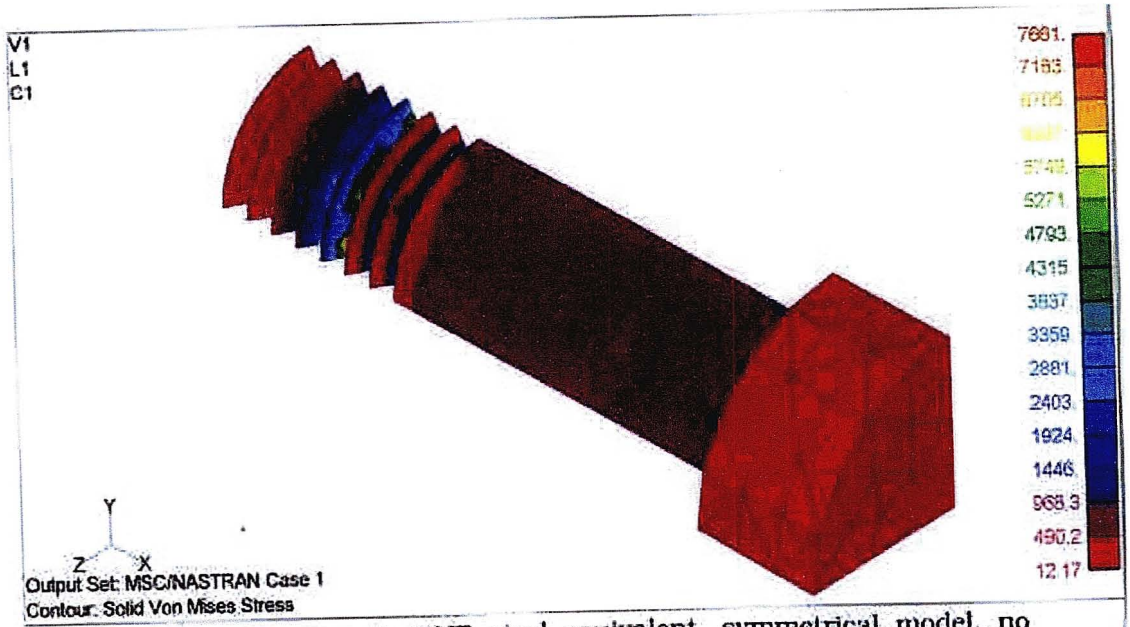


Fig. 3.11 – High strength TRIP steel equivalent, symmetrical model, no special stress concentrators

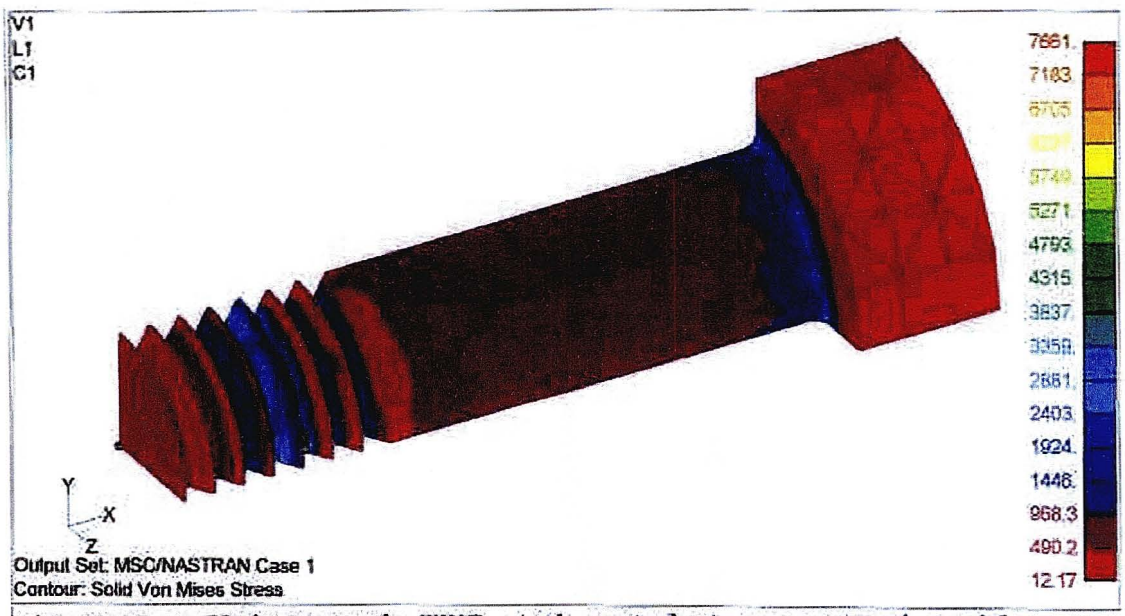


Fig. 3.12 – High strength TRIP steel equivalent, symmetrical model. Showing stresses at 1mm fillet radius for head to shank transition

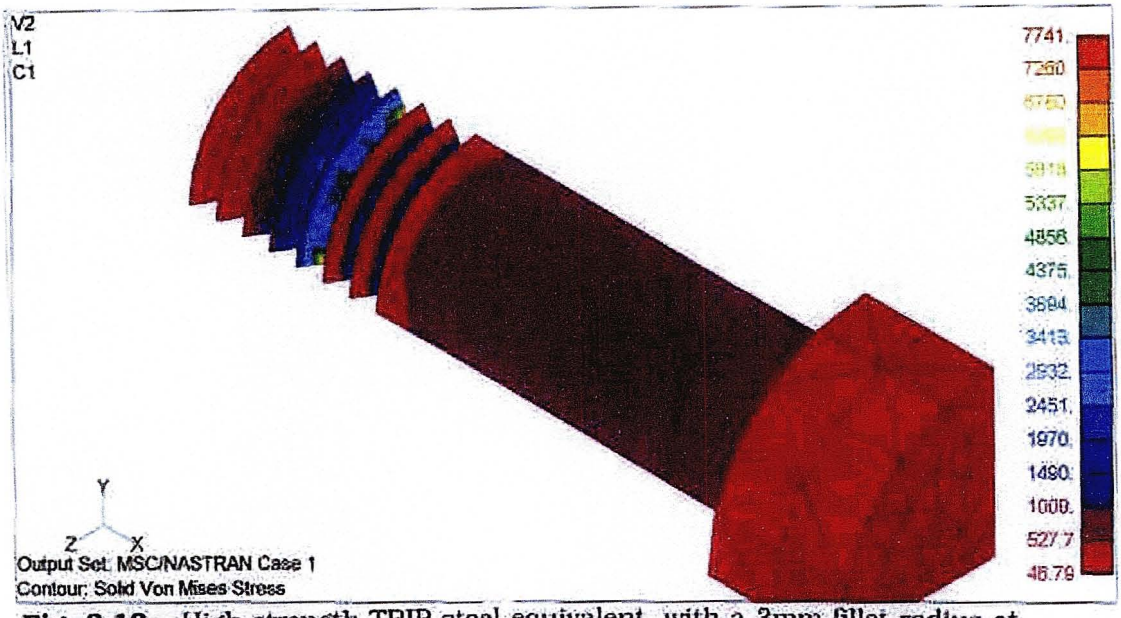


Fig. 3.13 – High strength TRIP steel equivalent, with a 3mm fillet radius at head to shank transition

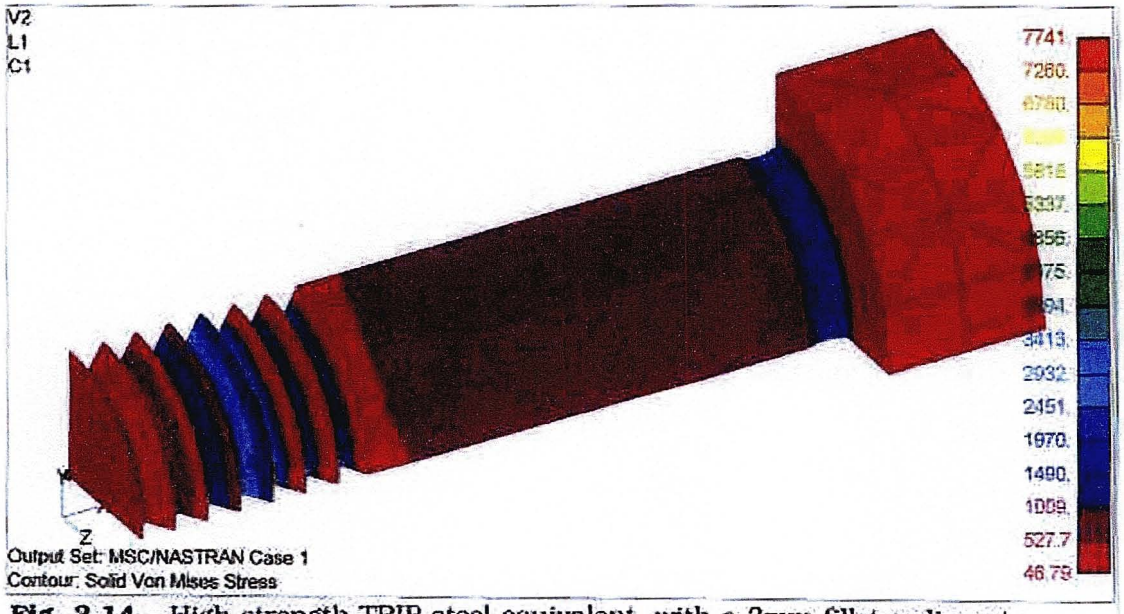


Fig. 3.14 – High strength TRIP steel equivalent, with a 3mm fillet radius at head to shank region, showing the stresses there.

the material was estimated as 90% UTS, and in turn the proof stress to be applied for loading was estimated as 90% YIELD. The property for modelling was for solid elements. A finer mesh was applied at the threads and at grip-to-head transition. An exponentially varying load, just like the previous model, was applied at the threads. As for the constraints, fixed constraints were applied at the bottom part of the bolt head in exactly the same manner as the previous models. However additional constraints were also added to take into account the fact that we were only modelling only part of the model. The model was sliced in the planes *normal* to the Y and the Z axes, and hence y-symmetry and z-symmetry constraints were applied accordingly. The sharp edge between the two planes was allowed to have both symmetrical constraints because it coincided with both faces.

As for the analysis, higher stresses were found in the head-to-grip transition and in the loaded threads, the highest stress being found in the thread carrying the highest load. This is consistent with expectations. The stress concentration under the bolt head is about $k_t \approx 2$, and at point of highest load in relation to the general threaded area is about $k_t \approx 4$. The fillet radius under the bolt head was r =1mm, see figure 3.11 to 3.12.

Another version of this bolt is the one with a groove of radius r = 3mm under the bolt head, see figures 3.13 to 3.14. For this, the stress concentration has only increased to about $k_t \approx 2.5$ there. This is the smallest groove experimented with, and there were no dramatic results. The other radii were r = 5, 7, and 10. The analyses of these model were not included because there was no major variation. In fact no useable information was to be inferred. The stress concentrations were in fact almost similar to a situation where we only had a fillet instead of a groove. In order to reach stresses closer to the thread stresses at this point, it was only possible by going deep into the material, decreasing the cross sectional area at this point. And the tip of such groove would have to possess a radius smaller than r = 0.3mm, which is the radius at the thread root. The stress pattern of the bolt was nevertheless similar to the previous bolt.

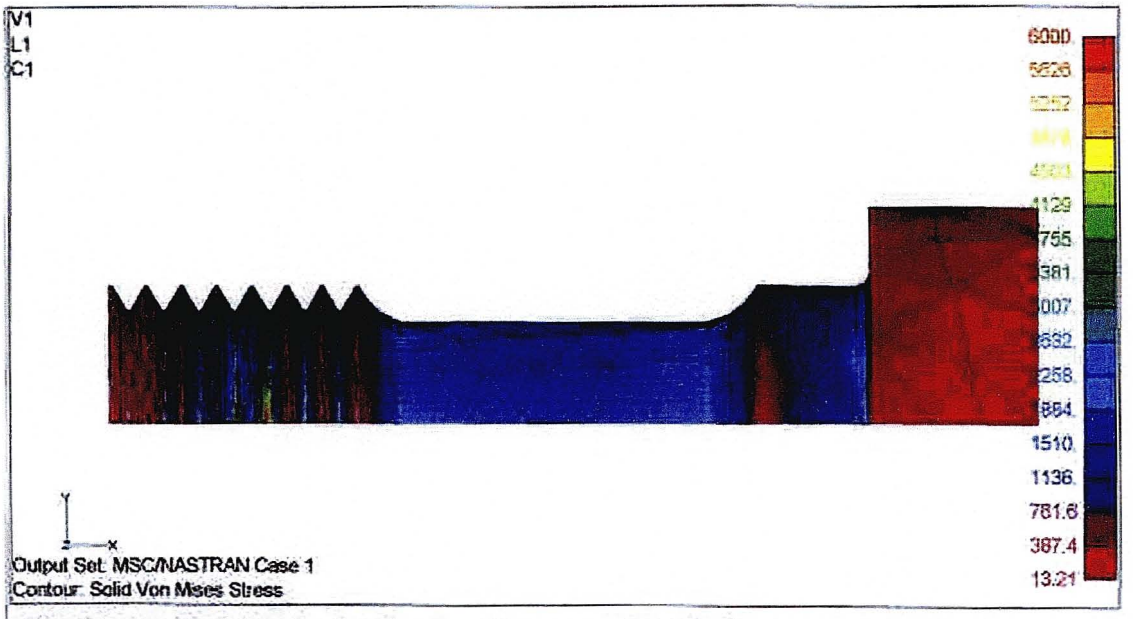


Fig. 3.15 – Model with shank transition, no sensor hole

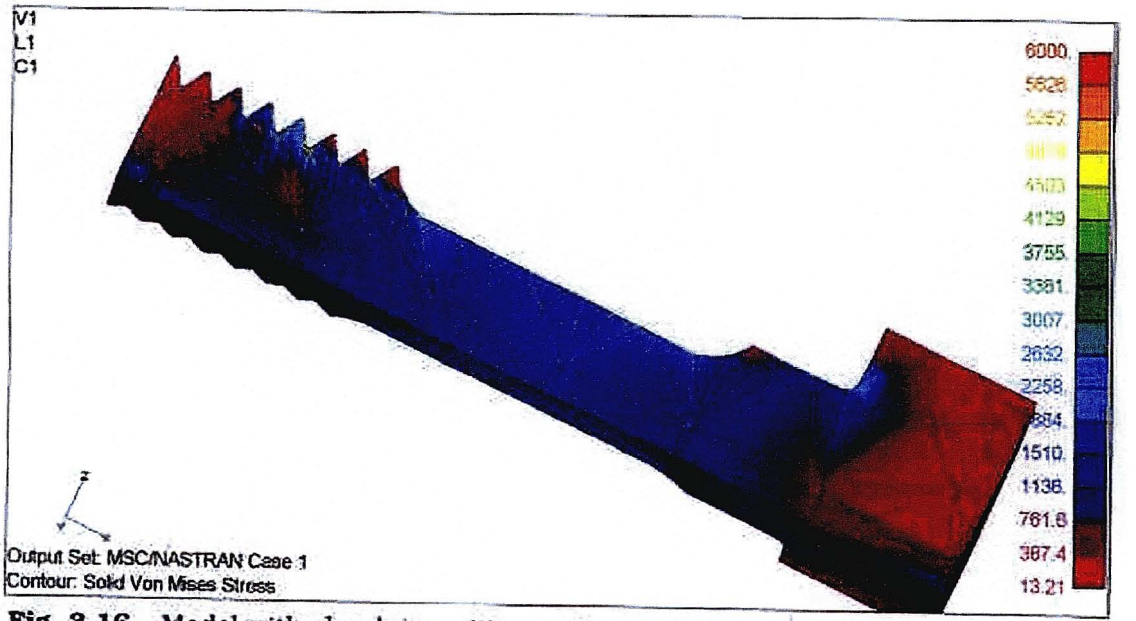


Fig. 3.16 – Model with shank transition, with a sensor hole

3.2.4 Model 4

The last bolt dealt with was the suggested bolt model (see topic on impact loading). The material, property, meshing method, constraints and loading conditions were similar to the previous two models discussed. The difference is that a central hole for the sensor was included. However figure 3.15 is just a model without a sensor hole, and this was provided to obtain an indication of the general stresses. Figure 3.16 does however have a hole for the sensor, and the visibility of the hole is slightly blurred by uniformity of stresses in the entire shank region. The shank region has got a varying radius. The grip-to-head transition has a fillet of radius r = 1mm. Also there is a radius transition from the central hole to the top of the bolt head. To be noted as well is that the central hole does not stop at the thread of interest. This was for fear of interaction of stress concentrations, and that could affect magnetic readings (if both the loaded thread and the end of the hole undergo magnetic transformation). Instead the hole has been extended to unloaded threads towards the end of the bolt. With this model there was no significant difference in stress patterns, and the area of interest in the thread still possessed higher stresses. The stress values in the central hole were those of the general bolt.

CHAPTER 4

Theoretical considerations of Finite Element Method

The finite element method involves the discretization of the structure that we intend analysing. This discretization translates into the structure constructed of nodes and elements, and each element containing a certain number of nodes. The forces and resulting displacements on a continuous structure are then translated into nodal forces and displacements. For these nodal displacements and nodal forces, we create a system of equations containing displacements and forces. This system of equations can then be arranged in the form:

$$[K][U]=[F] \tag{4.1}$$

[K] = System stiffness matrix

[U] = Vector of displacements

[F] = Vector of nodal forces

4.1 Introductory example - One dimensional problem

The derivation below gives an understanding of the variables in the finite element method. It gives a simplified form of the equation (4.1) above, and gives an understanding of what is involved in the same equation.

$$P/A_{avg} = \sigma_e \tag{4.2}$$

$$= E\varepsilon$$

$$= EU_e/l_e$$

$$= \left(\frac{EA_{avg}}{l_e}\right)U_e$$

$= K_e U_e$ (4.3)

4.2 Shape functions

The usual method for finite element analysis is the one based on shape functions. Three main assumptions precede the formulation of subsequent problems.

1) For linear interpolation, the nodal displacements vary linearly with position along the structure. Accordingly, the nodal displacement can be expressed as:

$$\text{i.e } U_i = a + bx \quad (4.4)$$

Where,

Ui =nodal displacement at node i;

x = Position along structure;

a, b = constants

2) If the problem is discretised into a certain number of nodal displacement, the solution defines the behaviour of the problem, and the solution should converge if the number of nodes is increased.
3) Each element in the structure is defined by a certain number of nodes. Normally the number of nodes per element depends on the type of interpolation.

Also

1) If the original interpolation function is linear, then the shape function is linear for the same variable

2) If the original interpolation is quadratic, the shape function is also quadratic for the same variable.

3) {u} = [N]{U} is the approximation of displacement for each element or for a single element structure.

4) For a two-node element

$$u = N_i u_i + N_j u_j \quad \text{(one element)} \tag{4.4}$$

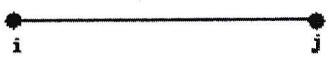


Fig. 4.1 Two-node, one-dimensional element

5) For a three-node element

$$u = N_i u_i + N_j u_j + N_k u_k \tag{4.5}$$

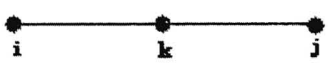


Fig. 4.2 Three-node, one-dimensional element

Corresponding interpolation functions:

1) $u = a + bx$ (4.6)

For two-node, one-dimensional element.

2) $u = a + bx + cx^2$ (4.7)

For three-node, one-dimensional element

3) $u = a + bx + cx^2 + dx^3$ (4.8)

For four-node, one-dimensional element

Similarly for a two-dimensional problem:

1) $u = a + bx + cy$ (4.9)

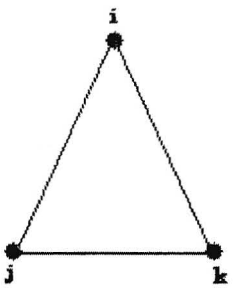


Fig. 4.3 Three-node, two-dimensional element

2) $u = a + bx + cy + dxy + ex^2 + fy^2$ (4.10)

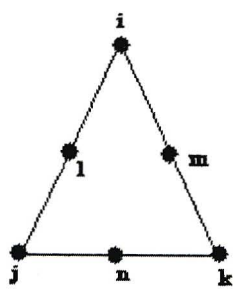


Fig. 4.4 Six-node, two-dimensional element

4.3 Derivation of Shape functions

4.3.1 One-dimensional element:

As mentioned before, the linear interpolation function for a one-dimensional element, of length L, is expressed in terms of equation (4.4)
The element has two nodes, i and j. The two nodes are located at distances x_i and x_j from the origin. By substituting the value of u at each node located at i and j we obtain

$$a = \frac{u_i x_j - u_j x_i}{x_j - x_i} = \frac{u_i x_j - u_j x_i}{L} \tag{4.11}$$

$$b = \frac{u_i - u_i}{x_j - x_i} = \frac{u_j - u_i}{L} \tag{4.12}$$

And substituting back into equation (4.4) we obtain

$$u = \frac{u_i x_j - u_j x_i}{L} + \frac{u_j - u_i}{L} x \tag{4.13}$$

Rearranging equation (4.13) in terms u_i and u_j, we obtain

$$u = \frac{x_j - x}{L} u_i + \frac{x - x_i}{L} u_j \tag{4.14}$$

$$u = N_i u_i + N_j u_j = [N]\{U\} \tag{4.15}$$

4.3.2 Two-dimensional element

Here we have triangular elements, and the linear interpolation function is of the form

$$u = a + bx + cy \tag{4.16}$$

The values of the a, b, c constants are found to be

$$a = (\alpha_i u_i + \alpha_j u_j + \alpha_k u_k)/2A \tag{4.17}$$

$$b = (\beta_i u_i + \beta_j u_j + \beta_k u_k)/2A \tag{4.18}$$

$$c = (\delta_i u_i + \delta_j u_j + \delta_k u_k)/2A \tag{4.19}$$

Where A is the area of the triangle expressed as

$$A = \frac{1}{2}\begin{vmatrix} 1 & x_i & y_i \\ 1 & x_j & y_j \\ 1 & x_k & y_k \end{vmatrix} = \frac{1}{2}(x_i y_j + x_j y_k + x_k y_i - x_i y_k - x_j y_i - x_k y_j) \tag{4.20}$$

and where

$$\alpha_i = x_j y_k - x_k y_j \quad \beta_i = y_j - y_k \quad \delta_i = x_k - x_j \tag{4.21}$$

$$\alpha_j = x_k y_i - x_i y_k \quad \beta_j = y_k - y_i \quad \delta_j = x_i - x_k \tag{4.22}$$

$$\alpha_k = x_i y_j - x_j y_i \quad \beta_k = y_i - y_j \quad \delta_k = x_j - x_I \tag{4.23}$$

Now, substituting equations (4.17), (4.18) and (4.19) into (4.16), the interpolation function can be expressed as

$$U = N_i u_i + N_j u_j + N_k u_k = [N]\{U\} \tag{4.24}$$

Where

$$N_i = (\alpha_i + \beta_i x + \delta_i y)/2A \tag{4.25}$$

$$N_j = (\alpha_j + \beta_j x + \delta_j y)/2A \quad (4.26)$$

$$N_k = (\alpha_k + \beta_k x + \delta_k y)/2A \quad (4.27)$$

So, the general form of the shape function for the two-dimensional element is expressible as

$$N_\lambda = (\alpha_\lambda + \beta_\lambda x + \delta_\lambda y)/2A \quad \lambda = i,j,k \quad (4.28)$$

4.3.3 Three-dimensional element

The principles discussed so far can be directly extended to three dimensions. The basic three dimensional element is a flat-faced tetrahedron with four nodes, such that i,j and k are labelled anticlockwise on any face as viewed from the vertex opposite the face , which is labeled m. The general interpolation polynomial for the element is of the form

$$U = a + bx + cy + dz, \quad (4.29)$$

Which can be expressed in terms of nodal values of u and the nodal shape functions as

$$U = N_i u_i + N_j u_j + N_k u_k + N_m u_m = [N]\{U\} \quad (4.30)$$

Where the shape functions are given by

$$N_\lambda = (\alpha_\lambda + \beta_\lambda x + \delta_\lambda y + \mu_\lambda z)/6V \qquad \lambda = i,j,k,m \quad (4.31)$$

The constants α, β, δ, μ are functions of the nodal coordinates, and are expressible as

$$\alpha = \begin{vmatrix} x_j & y_j & z_j \\ x_k & y_k & z_k \\ x_m & y_m & z_m \end{vmatrix} \qquad \beta = \begin{vmatrix} 1 & y_j & z_j \\ 1 & y_k & z_k \\ 1 & y_m & z_m \end{vmatrix} \tag{4.32}$$

$$\delta = \begin{vmatrix} x_j & 1 & z_j \\ x_k & 1 & z_k \\ x_m & 1 & z_m \end{vmatrix} \qquad \mu = \begin{vmatrix} x_j & y_j & 1 \\ x_k & y_k & 1 \\ x_m & y_m & 1 \end{vmatrix} \tag{4.33}$$

4.3.4 **Variational formulation**

We begin by defining the following symbols:

$\prod$ = Potential energy of the system

Λ = Strain energy of the system

W = Work done by the external forces acting on the body

The potential energy $\prod$ equals the strain energy Λ in the body less the work done W by the external loads acting on the body:

$$\prod = \Lambda - W \tag{4.34}$$

Now, considering a differential element of volume dV, the strain energy is given by

4.3.5 Strain energy

$$d\Lambda = \frac{1}{2}\{\varepsilon\}^T\{\sigma\} - \frac{1}{2}\{\varepsilon_0\}^T\{\sigma\} \quad (4.35)$$

$\{\varepsilon\}$ = Vector of total strains

$\{\varepsilon_0\}$ = Vector of initial strains

Typically $\varepsilon_0 = \alpha\Delta T$ (4.36)

$\{\sigma\}$ = Vector of stress components

Therefore total strain energy for the finite volume is given by

$$\Lambda = \int_V \frac{1}{2}(\{\varepsilon\}^T\{\sigma\} - \{\varepsilon_0\}^T\{\sigma\})\, dV \quad (4.37)$$

For a plane stress or plane strain problem, the stress can be expressed as

$$\{\sigma\} = [D]\{\varepsilon\} - [D]\{\varepsilon_0\} \quad (4.38)$$

$$\Lambda = \int_V \frac{1}{2}(\{\varepsilon\}^T[D]\{\varepsilon\} - 2\{\varepsilon\}^T[D]\{\varepsilon_0\} + \{\varepsilon_0\}^T[D]\{\varepsilon_0\})\, dV \quad (4.39)$$

If we use a strain vector, then the strains can be equated to nodal displacements by

$$\{\varepsilon\} = [B]\{U\} \quad (4.40)$$

To determine the [B] matrix, consider, for example, the displacement in the x-direction in a two dimensional element, which is given previously by the following equation :

$$u = N_i u_i + N_j u_j + N_k u_k \tag{4.41}$$

where u_i, u_j, and u_k are the nodal displacements in that direction and form part of {U}. The strain in the x direction is calculated as

$$\varepsilon_x = \frac{\partial u}{\partial x} = \frac{\partial N_i}{\partial x} u_i + \frac{\partial N_j}{\partial x} u_j + \frac{\partial N_k}{\partial x} u_k \tag{4.42}$$

Hence [B] is a matrix of terms derived by the proper differentiation of the shape functions. Consideration of the general format of shape functions of two-dimensional elements in equations (4.25), (4.26) and (4.27) shows that this differentiation is trivial, resulting in a constant. For higher order elements the calculation is similarly straight forward, but yields linear, quadratic or higher order functions.

Substituting equation (4.38) and (4.40) into equation (4.39) , we get the strain energy $\Lambda^{(e)}$ for a single element as

$$\Lambda^{(e)} = \frac{1}{2} \int_{V^{(e)}} (\{U^{(e)}\}^T [B^{(e)}]^T [D^{(e)}][B^{(e)}]\{U^{(e)}\} - 2\{U^{(e)}\}^T [B^{(e)}]^T [D^{(e)}]\{\varepsilon_0^{(e)}\} + \{\varepsilon_0^{(e)}\}^T [D^{(e)}]\{\varepsilon_0^{(e)}\}) \, dV \tag{4.43}$$

The strain energy of the whole volume can then be found by summing the strain energies of all the individual elements.

4.3.6 Work done on element by applied external forces

External loads can be divided into three parts, and the work done by each is equal to the magnitude of the force multiplied by the distance moved:

Body (inertia) loads

If the body forces are **X**, **Y**, **Z** per unit volume then, for a general element, the work done by the body forces is

$$W_B^{(e)} = \int_{V^{(e)}} (u\mathbf{X}^{(e)} + v\mathbf{Y}^{(e)} + w\mathbf{Z}^{(e)})dV \tag{4.44}$$

$$= \int_{V^{(e)}} \{U^{(e)}\}^T[N^{(e)}]^T \begin{Bmatrix} X^{(e)} \\ Y^{(e)} \\ Z^{(e)} \end{Bmatrix} dV \tag{4.45}$$

Distributed (pressure) loads

Pressure forces will act over the surface, and hence the work done will be found from

$$W_P^{(e)} = \int_{S^{(e)}} (up_x^{(e)} + vp_y^{(e)} + wp_z^{(e)})dS$$

$$= \int_{S^{(e)}} \{U^{(e)}\}^T[N^{(e)}]^T \begin{Bmatrix} p_x^{(e)} \\ p_y^{(e)} \\ p_z^{(e)} \end{Bmatrix} dS \tag{4.46}$$

where p_x, p_y and p_z are distributed pressure loads parallel to the co-ordinate axes, and are considered positive when acting in the positive co-ordinate direction.

Nodal (concentrated) loads

If $\{P^{(e)}\}$ is a vector of nodal loads, then the work done will be

$$W_C^{(e)} = \{U^{(e)}\}^T\{P^{(e)}\} \tag{4.47}$$

Therefore the potential energy of a single element is given by

$$\prod^{(e)} = \Lambda^{(e)} - W_B^{(e)} - W_P^{(e)} - W_C^{(e)} \tag{4.48}$$

The potential of the whole system is then equal to the sum of the energies of all the elements in the model.

$$\prod = \sum_{e=1}^{E} \Big[\int_{V^{(e)}} \frac{1}{2} \{U^{(e)}\}^T[B^{(e)}]^T[D^{(e)}][B^{(e)}]\{U^{(e)}\}dV$$

$$- \int_{V^{(e)}} \{U^{(e)}\}^T [B^{(e)}]^T[D^{(e)}]\{\varepsilon_0^{(e)}\}dV$$

$$+ \int_{V^{(e)}} \frac{1}{2} \{\varepsilon_0^{(e)}\}^T[D^{(e)}]\{\varepsilon_0^{(e)}\}dV$$

$$- \int_{V^{(e)}} \{U^{(e)}\}^T[N^{(e)}]^T \begin{Bmatrix} X^{(e)} \\ Y^{(e)} \\ Z^{(e)} \end{Bmatrix} dV$$

$$- \int_{S^{(e)}} \{U^{(e)}\}^T[N^{(e)}]^T \begin{Bmatrix} p_x^{(e)} \\ p_y^{(e)} \\ p_z^{(e)} \end{Bmatrix} dS - \{U^{(e)}\}^T\{P^{(e)}\}\Big] \tag{4.49}$$

For the body to be in equilibrium, its potential energy must be equal to zero. If there are n degrees of freedom, that is n possible nodal displacements, then we have

$$\frac{\partial \Pi}{\partial u_1} = \frac{\partial \Pi}{\partial u_2} = \ldots = \frac{\partial \Pi}{\partial u_n} = 0 \tag{4.50}$$

This gives n equations in n unknowns, and is expressed more conveniently as

$$\frac{\partial \Pi}{\partial \{U_S\}} = 0 \tag{4.51}$$

where is $\{U_S\}$ is the vector of all the degrees of freedom in the model, and consists of all elemental displacement vectors. And differentiation of equation (4.49) by standard methods yields

$$\frac{\partial \Pi}{\partial \{U_S\}} = \sum_{e=1}^{E} \left[\int_{V^{(e)}} [B^{(e)}]^T [D^{(e)}][B^{(e)}] dV \right] \{U_S\}$$

$$- \sum_{e=1}^{E} \left[\int_{V^{(e)}} [B^{(e)}]^T [D^{(e)}]\{\varepsilon_0^{(e)}\} dV + \int_{V^{(e)}} [N^{(e)}]^T \begin{Bmatrix} X^{(e)} \\ Y^{(e)} \\ Z^{(e)} \end{Bmatrix} dV \right.$$

$$\left. + \int_{S^{(e)}} [N^{(e)}]^T \begin{Bmatrix} p_x^{(e)} \\ p_y^{(e)} \\ p_z^{(e)} \end{Bmatrix} dS \right] - \{P_S\} = 0 \tag{4.52}$$

where $\{P^{(e)}\}$ has been taken outside the summations and replaced by $\{P_S\}$, the complete vector of nodal loads.

Equation (4.52) may be written as

$$[k]\{U_S\} = \{F\} \tag{4.53}$$

where

$$[k] = \sum_{e=1}^{E} \left[\int_{V^{(e)}} [B^{(e)}]^T [D^{(e)}][B^{(e)}] dV \right] = \sum_{e=1}^{E} [k^{(e)}] \tag{4.54}$$

and is known as the global (or system) matrix, and the global force vector is

$$\{F\} = \sum_{e=1}^{E} \left[\int_{V^{(e)}} [B^{(e)}]^T [D^{(e)}]\{\varepsilon_0^{(e)}\} dV + \int_{V^{(e)}} [N^{(e)}]^T \begin{Bmatrix} X^{(e)} \\ Y^{(e)} \\ Z^{(e)} \end{Bmatrix} dV \right.$$

$$\left. + \int_{S^{(e)}} [N^{(e)}]^T \begin{Bmatrix} p_x^{(e)} \\ p_y^{(e)} \\ p_z^{(e)} \end{Bmatrix} dS \right] + \{P_S\} = \sum_{e=1}^{E} \{F^{(e)}\} \tag{4.55}$$

If the model has n degrees of freedom, then [k] will be an [n x n] square matrix, while a single element with m degrees of freedom will result in an [m x m] matrix for $[k^{(e)}]$. The global force vector, on the other hand, will be a vector of length n, while the element vector will be of length m.

4.4 One-dimensional boundary value problems

Here we begin with a solid mechanics example to demonstrate the most common form of differential equation for our finite element models. The figure shows a variable-cross-section bar supported at one end against axial motion and loaded axially by distributed load q(x) and by typical concentrated loads P1 and F. P1 is applied at an interior location and F at the end of the bar. The variation of cross section is gradual so as to avoid stress concentrations; it is assumed that D << L, where D is the diameter at the large end, and L is the length of the bar. The bar is composed of linear

elastic material. The development of the governing equations is based on kinetics or balance of momentum (Newtons second law), kinetics or strain displacement relations, and constitutions or stress-strain relations. As mentioned above, we develop the equations for analysing the relations between the displacement and the axial force. Here, the aim is to reach the derivation of a stiffness matrix using a different approach, showing the Garlekin finite element model.

From the free body diagram in figure 4.5, we balance the forces as :

$$P + \Delta P + q\Delta x - P = 0, \tag{4.56}$$

Where P represents internal forces due to the external loads above, $q(x)$ is distributed load as seen in figure4.5.

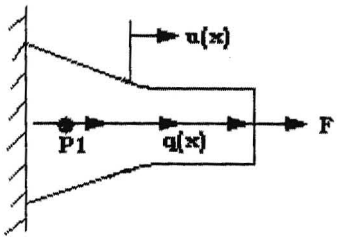


Fig. 4.5 – Load diagram

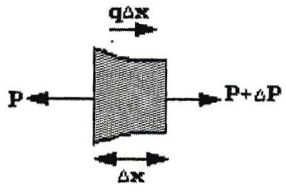


Fig. 4.6 – Free body diagram

$$\Delta P + q\Delta x = 0$$

$$\frac{\Delta P}{\Delta x} + q\frac{\Delta x}{\Delta x} = 0$$

$$\frac{\Delta P}{\Delta x} + q = 0$$

$$\lim_{\Delta x \to 0} \frac{\Delta P}{\Delta x} + \lim_{\Delta x \to 0} q = 0$$

$$\frac{dP}{dx} + q(x) = 0 \qquad (4.57)$$

$$P = \int_{Area} \sigma \, dA, \qquad (4.58)$$

Where σ is the axial stress arising from the internal force P.

$$\varepsilon = \frac{du}{dx} = u' \text{ (Strain)} \qquad (4.59)$$

$$\sigma = E\,\varepsilon \text{ (E Young's modulus)} \qquad (4.60)$$

$$\therefore P = \int_{Area} \sigma \, dA = \int_{Area} (E\,\varepsilon)\, dA$$

$$= \int_{Area} (E\,u')\, dA$$

$$= E u' \int_{Area} dA$$

$$= E\,u'\,A \text{ or}$$

$$= AEu' \qquad (4.61)$$

$\therefore\ P' + q(x) = 0$

$$(AEu')' + q = 0 \tag{4.62}$$

This is a 2nd order differential equation requiring two boundary conditions, namely

u (0) = 0 and
P (L) =F (4.63)

Or

u (0) = 0
P (L) = $AEu'(L) = F$ (4.64)

4.4.1 Thin film lubrication

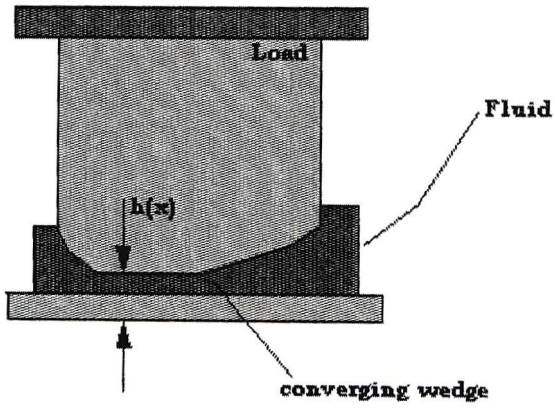


Fig. 4.7 – Film lubrication

$$\frac{d}{dx}\left(\frac{h^3}{12\eta}\frac{\partial P}{\partial x}\right) = \frac{d}{dx}\left[\frac{h}{2}(u_2 - u_1)\right] \tag{4.65}$$

P (0) = P_1 and P (L) = P_2 i.e assuming pressures are known at the ends of the film

4.4.2 Heat transfer applications

$$\frac{d}{dx}(kA\frac{dT}{dx}) - hP(T - T_0) + q_0 A = 0 \qquad (4.66)$$

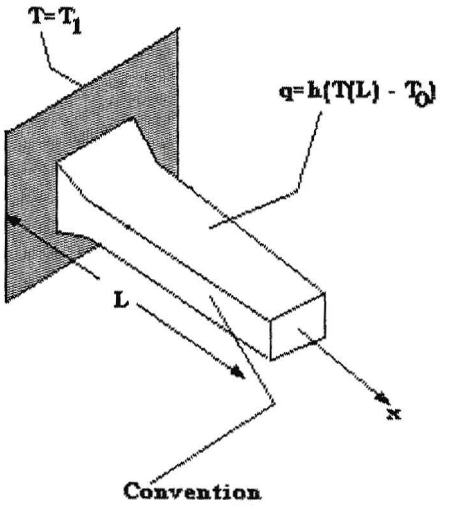


Fig. 4.8 – Heat transfer through a fin

T (0) = T1 and $-kAT'|_L = h(T(L) - T_0)$ (4.67)

4.4.3 Torsional vibration applications

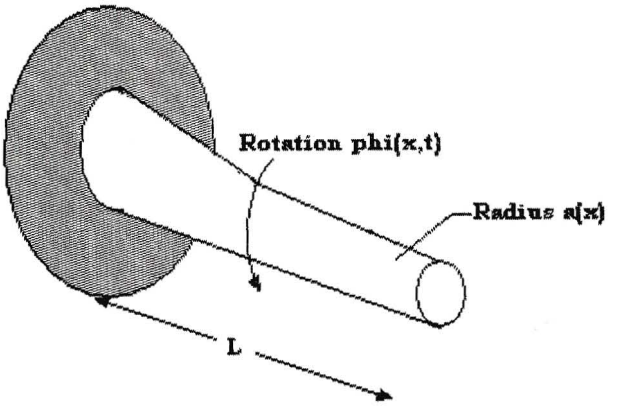


Fig. 4.9 – Torsional vibration shaft

$$(JG\psi')'+P\omega^2 J\psi = 0 \tag{4.67}$$

$$\psi(0) = 0 \text{ and } T(L) = JG\psi'(L) = 0 \tag{4.68}$$

In the above cases, the governing equations are 2nd order differential equations with two boundary conditions. The task in each case is to find a function $f(x)$ satisfying the boundary conditions. F(x) is an approximation to $f(x)$; it consists of straight lines, i.e. a piecewise linear function. By increasing the number of elements on F(x), we will more approximate the function $f(x)$.

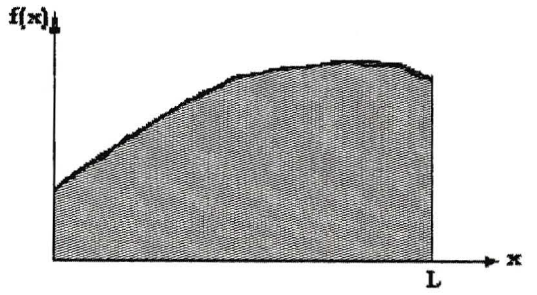
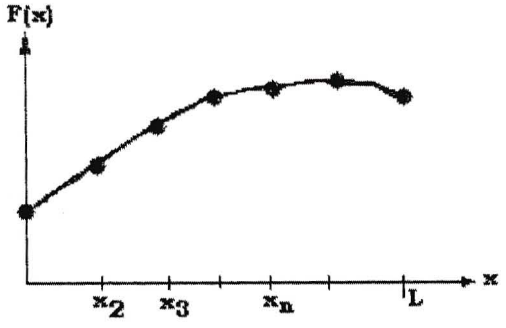


Fig. 4.10 – Actual smooth function

Fig. 4.11 – Approximate function



4.4.4 Two approaches for finite element model

4.4.4.1 Garlekin method

With this method we approach the differential equation directly.
The other approach is that we develop the corresponding "energy", represented as an integral over the interval. This energy , which for a conservative mechanical system is essentially the total potential energy, is then used as the basis for constructing a **Ritz finite element** model which is a system of linear algebraic equations.
Ordinary differential equations considered above may be considered as the specific example of the class of boundary value problems defined by the differential equation

$$(pu')'+(\lambda r - q)u + f = 0 \qquad a < x < b \qquad (4.69)$$

and the two boundary conditions

$$\begin{aligned} -p(a)u'(a) + \alpha u(a) &= A \\ p(b)u'(b) + \beta u(b) &= B \end{aligned} \qquad (4.70)$$

4.4.4.2 Assumptions

q, r, f : piecewise continuous function of x,

p : Has a continuous derivative,

p, q, r : Are positive in the basic interval $a < x < b$,

λ : Is a parameter

α, β, A : Are constants

The collection of differential equations and the boundary conditions above is called the Sturm – Liouville system.

4.5 Galerkin finite element model

1) Discritization

2) Interpolation

3) Elemental formulation

4) Assembly

5) Constraints

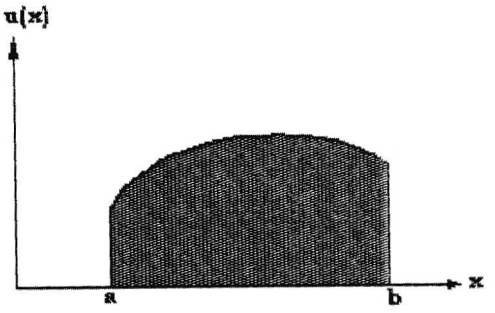
6) Solution

7) Computation of derived variables

4.6 Discretization

As mentioned before, we expect u (x) to be a smooth function as in fig. 4.12

Fig. 4.12 – Actual smooth function defined between a and b



The range of the independent variable is $a \leq x \leq b$

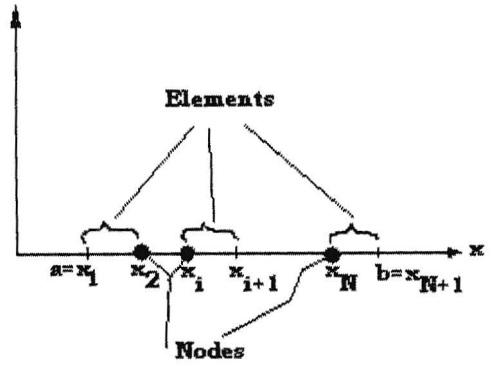


Fig. 4.13 – Nodal distribution between a and b

Discretization consists in choosing points x_i, i = 1, 2, 3..., N+1 as in fig.4.13

i) The points are referred to as nodes. The intervals between the nodes are referred to as elements.

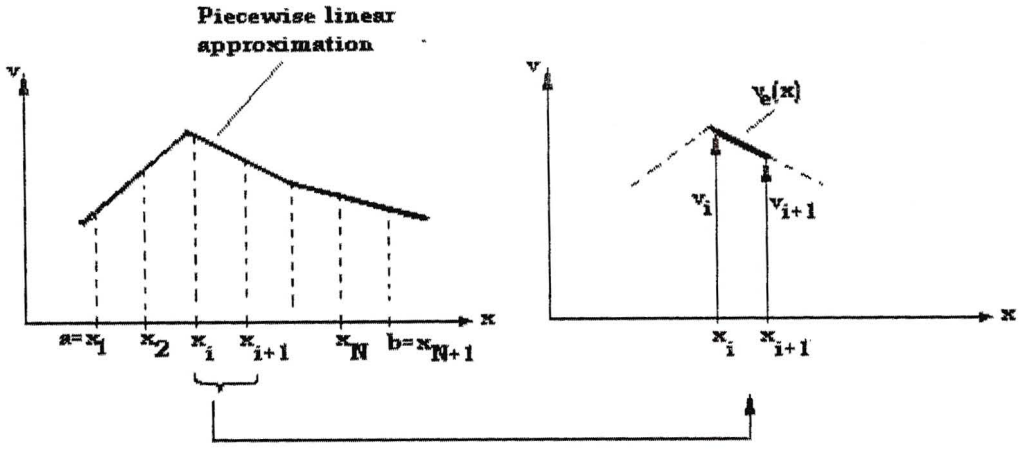


Fig. 4.14 – Approximate function $v_e(x)$

ii) Nodes and elements are collectively referred to as a mesh.

iii) Nodes are usually chosen to be uniformly spaced, so that all elements have equal length.

iv) In regions where rapid changes in the solution $u\ (x)$ are expected, the nodes can be chosen such that the element lengths are smaller, thus enabling the approximate finite solution $v\ (x)$ to more easily represent the rapid changes.

4.7 Interpolation

After the location of the nodes have been chosen, an approximate solution $v(x)$ is assumed to be represented in a piecewise linear fashion as in fig.4.14. As mentioned previously, we expect $v(x)$ to approach the exact solution $u(x)$ as the number of nodes is increased.

On a typical element $x_i \le x \le x_{i+1}$, the form for $v(x)$ is :

$v_e(x) = \alpha + \beta x$, where α, β are solved by requiring that $v_e(x_i) = v_i$ and $v_e(x_{i+1}) = v_{i+1}$, and hence

$$v_e = \frac{x_{i+1} - x}{x_{i+1} - x_i} v_i + \frac{x - x_i}{x_{i+1} - x_i} v_{i+1} \qquad (4.71)$$

$v_e(x) = N_i v_i + N_{i+1} v_{i+1}$, where N_i and N_{i+1}, shown in fig.4.14, are called elemental linear interpolation functions.
$l_e = x_{i+1} - x_i$ is the corresponding length of the element.
On the entire interval $a \le x \le b$ the approximate function $v(x)$, shown in fig.4.14 is represented according to :

$$v(x) = \sum_{i=1}^{N+1} v_i n_i(x) \qquad (4.72)$$

Where $n_i(x)$, are referred to as nodal linear interpolation functions. Each of the $n_i(x)$ satisfies

$$N_i(x_j) = 1 \text{ if } i = j$$
$$N_i(x_i) = 0 \text{ if } i \neq j \qquad (4.73)$$

In other words n_i has the value 1 at the node located at x_i and n_i has the value 0 at all the other nodes.

Thus v (x_j) = $\sum_{i=1}^{N+1} v_i n_i(x) = v_j$, (4.74)

thus showing that the coefficients v_i are the nodal values of the approximate solution v (x).

Also

$n_i = N_{i+1}$ (from element i-1) for $x_{i-1} \le x \le x_i$

$n_i = N_i$ (from element i) for $x_i \le x \le x_{i+1}$ (4.75)

Note, the *nodal interpolation functions* n_i are used in developing the finite element model using the **Galerkin** method, whereas the elemental interpolation functions N_i are used in the developing of the finite element model using the Ritz method.

4.8 Elemental formulation and assembly

$$(pu')'+(\lambda r - q)u + f = 0 \qquad a \le x \le b \tag{4.76}$$

$$-p(a)u'(a) + \alpha u(a) = A$$
$$p(b)u'(b) + \beta u(b) = B \tag{4.77}$$

where $f(x)$, A and B are usually zero when λ appears in the differential equation. The Galerkin's finite element model for the non-eigenvalue problem is as follows

$$(pu')'-qu + f = 0 \qquad a \le x \le b \tag{4.78}$$

$$-p(a)u'(a) + \alpha u(a) = A$$
$$p(b)u'(b) + \beta u(b) = B\,, \tag{4.79}$$

which can be applied to the analysis solids, fluids and the heat transfer applications. We will develop the model for the case of natural boundary conditions.

As mentioned above, the approximate solution is:

$$v(x) = \sum_{i=1}^{N+1} v_i n_i(x), \tag{4.80}$$

substituting into the differential equation, we get

$$(pv')'-qv + f = \left(p\sum_{i=1}^{N+1} v_i n'_i(x) \right)' - q\sum_{i=1}^{N+1} v_i n_i(x) + f$$

$$= E(x, v_1, v_2, ..., v_{N+1}), \tag{4.81}$$

where E is the error arising from the fact that the approximate solution v (x) does not (in general) satisfy the differential equation.

Galerkin's method requires that the error function E be integrated over the interval from a to b against each member of the set of N+1 independent nodal interpolation functions $n_k(x)$ used in representing the solution $v(x)$, or

$$\int_a^b En_k(x)dx = 0, \quad k = 1,2,...,N+1 \tag{4.82}$$

The pv' terms evaluated at the boundary are replaced by using the boundary conditions satisfied by the approximate solution v to yield

$$n_k(a)[B - \beta v(b)] - n_k(a)[\alpha v(a) - A]$$
$$-\int_a^b [n'_k\, pv' + n_k qv + n_k f]dx = 0 \tag{4.83}$$

$$k = 1,2,...,N+1$$

Second derivatives of $v(x)$ are eliminated, and integration by parts has enabled the boundary conditions to be satisfied by $v(x)$ to be included.

Finally, substituting for $\sum v_i n_i$ and accounting for the fact that $n_k(b)$ is zero except when $k = N+1$

And that $n_k(a)$ is zero except when k=1, these equations can be expressed as :

$$\sum_{i=1}^{N+1}\left[\int_a^b (n'_k\, pn'_i + n_k q n_i)dx\right]v_i + \beta v_{N+1}\delta_{kN+1} + \alpha v_1\delta_{k1}$$

$$= \int_a^b n_k f dx + B\beta_{kN+1} + A\delta_{k1} \qquad (4.84)$$

$$k = 1,2,\ldots,N+1$$

where δ_{ij} is the Kronecker delta given by

$\delta_{ij} = 0$ if i≠j

$\delta_{ij} = 1$ if i = j (4.85)

The term δ_{kN+1} indicates the occurrence of a term in only (N+1)st equation and δ_{k1} indicates the occurrence of a term in only the first equation.

Above equation can be written as

$$\sum_{i=1}^{N+1} K_{ki} v_i = F_k \quad , \; k = 1,2,\ldots,N+1$$

or $\underline{K}\,\underline{v} = \underline{F}$, where , (4.86)

$$K_{ki} = \int_a^b (n'_k\, pn'_i + n_k qn_i)dx + \beta\delta_{ki}\delta_{iN+1} + \alpha\delta_{ki}\delta_{i1} \tag{4.87}$$

are the co-ordinates of the global stiffness matrix

Note that when i and k are interchanged on the right hand side of the equation there is no change in the value of the coefficients K_{ki} so that the matrix $\underline{K}$ is symmetric. The F_k given by

$$F_k = \int_a^b n_k f dx + B\delta_{kN+1} + A\delta_{k1} \tag{4.88}$$

are the coefficients of the global load matrix

For $k \neq 1$ or $k \neq N+1$, $n_k > 0$ only in

$x_{k-1} \leq x \leq x_k$ and $x_k \leq x \leq x_{k+1}$, so that the Kth equation is

$$\sum_{i=1}^{N+1} \int_{x_{k-1}}^{x_{k+1}} (n'_k\, pn'_i + n_k qn_i)dx v_i = \int_{x_{k-1}}^{x_{k+1}} n_k f dx \tag{4.89}$$

Then with only the n_i for $i = k-1, k, k+1$ being non-zero in the intervals $x_{k-1} \leq x \leq x_k$ and

$x_k \leq x \leq x_{k+1}$, the Kth equation can be expressed as :

$$\int_{x_{k-1}}^{x_k} (n'_k\ pn'_{k-1} + n_k qn_{k-1})dxv_{k-1}$$

$$+ \int_{x_{k-1}}^{x_{k+1}} (n'_k\ pn'_k + n_k qn_k)dxv_k \qquad (4.90)$$

$$+ \int_{x_k}^{x_{k+1}} (n'_k\ pn'_{k+1} + n_k qn_{k+1})dxv_{k+1} = \int_{x_{k-1}}^{x_{k+1}} n_k f dx$$

That is, only the unknowns v_{k-1}, v_k, v_{k+1} appear in the k^{th} equation . In a completely similar fashion it follows that only the unknowns v_1 and v_2 appear in the first equation and that only the unknowns v_N and v_{N+1} appear in the (N+1)st equation. Thus the global stiffness matrix $\underline{K}$ has the form :

$$K = \begin{bmatrix} K_{11} & K_{12} & \cdots & \cdots & 0 \\ K_{21} & K_{22} & K_{23} & \cdots & 0 \\ 0 & K_{32} & K_{33} & K_{34} & \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ 0 & \cdots & K_{N,N-1} & K_{N,N} & K_{N,N+1} \\ 0 & \cdots & & K_{N+1,N} & K_{N+1,N+1} \end{bmatrix} \qquad (4.91)$$

CHAPTER 5

Metallurgical considerations

5.1 Ordinary carbon steel

As an introduction to the structure of stainless steel and or/and TRIP steel, the basic structure of stainless steel will be discussed below.

A large variety of different steels, including stainless steel can be explained on the basis of iron-carbon equilibrium diagram. The diagram, figure 5.1, is divided into a number of phase fields, each occupied either by a single phase or a mixture of two phases. The curve ABCD is the liquidus line above which there is only one liquid phase consisting of iron and dissolved carbon. The curve AEPGCH is the solidus line below which the various iron-carbon compositions are completely solid. The regions between these two lines represent the mixtures of solid and liquid.

Both alpha iron, having a body-centered cubic structure, and gamma iron, with face-centred cubic structure, are capable of dissolving certain amounts of carbon depending on temperature. The solid solution of carbon in alpha iron is called ferrite, whereas the solid solution of carbon in gamma iron is called austenite. The terms alpha and gamma iron are also used to designate ferrite and austenite respectively. Austenite is capable of dissolving far more than ferrite, and it may contain up to 2% carbon (point G), whereas ferrite can only dissolve up to 0.025% carbon (point L). The solubility of carbon changes with temperature both in austenite and ferrite in the manner indicated by the curves GK for austenite and LN for ferrite. The third solid phase existing in the diagram is cementite, which is compound of the composition Fe_3C, corresponding to 6.67% by weight of carbon. Cementite is hard and brittle, whereas ferrite is relatively soft and malleable.

The transformation of one solid phase into another, as from austenite to ferrite or cementite, occurs at certain critical temperatures indicated on the diagram by lines LKM, IK, and GK. These lines are usually referred to as A1,

called the lower critical temperature line, and A_3 and A_{cm}, called upper critical temperature lines. These lines represent transition temperatures both for upward and downward transformation occurring under equilibrium conditions. In practice, however, equilibrium is seldom obtained and the critical temperature.

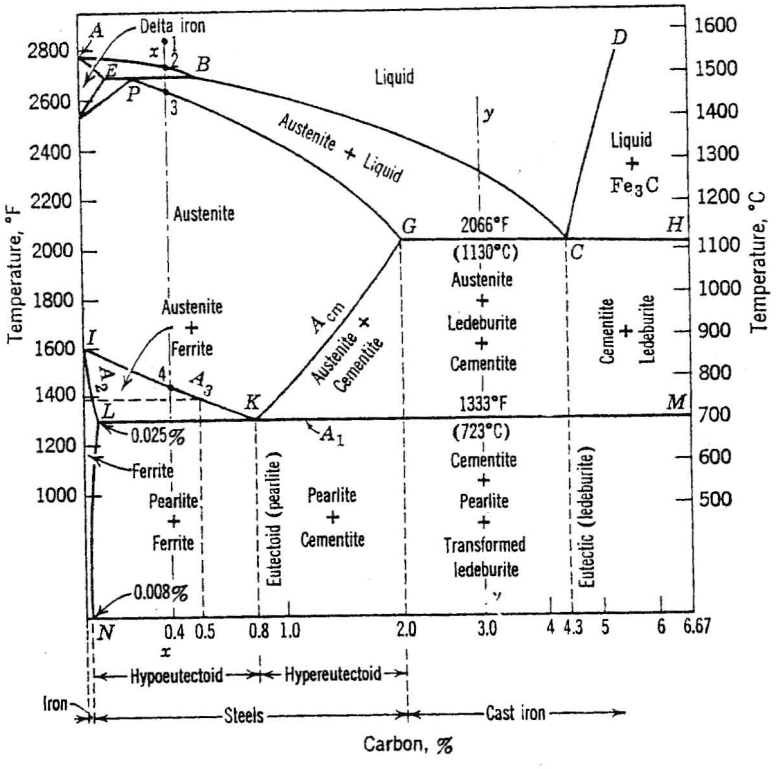


Fig. 5.1 – Iron carbon equilibrium diagram

lines will be shifted to higher positions on heating and to lower positions on cooling.

The diagram also indicates the existing ranges of composition for iron, steel, and cast iron. Compositions up to 0.008% carbon are regarded as commercially pure iron, those from 0.008% to 2% carbon represent steel, and those above 2% carbon represent cast iron. Steels are further

subdivided into hypoeutectoid steel up to 0.8% carbon and hypereutectoid steel from 0.8% up to 2% carbon.

Austenite is not stable below the lines A_3 (IK) and A_{cm} (GK). For compositions containing less than 0.8% carbon, austenite on cooling begins to transform into ferrite, and the remaining austenite increases its carbon content along the line A_3 until the point K (0.8% carbon) is reached. For compositions between 0.8% and 2% carbon, cementite will separate out and the austenite composition will vary along the Acm (GK) until again K is reached. At point K, called the eutectoid, austenite will transform into pearlite, which is an intimate mixture of ferrite and cementite. This is called the eutectoid transformation, characterised by the decomposition of a solid solution into two different solid phases according to the reaction

Solid phase A $\longrightarrow$ Solid phase B + Solid phase C

There are three phases, austenite, ferrite, and cementite, existing in equilibrium at the eutectoid point. It follows from the phase rule that the number of degrees of freedom is zero, F=C-P +1=2-3+1=0, and the eutectoid point, like that of eutectic, is nonvariant. The eutectoid reaction is one of the most important transformations occurring in the iron-carbon system; it is also found in many other alloy systems.

At a temperature of 1130^{0}C, the eutectic transformation (point C) takes place. The eutectic liquid with 4.3% carbon begins to freeze to a eutectic mixture of austenite and cementite called ledeburite. On further cooling, the eutectic austenite transforms gradually to cementite, and its composition changes along line GK until again the eutectoid point is reached. Here all the remaining austenite will finally transform to pearlite. Another type of transformation is the peritectic which occurs at the temperature of 1499^{0}C (point P). The peritectic reaction involves the transformation of a liquid phase of different composition according to the reaction

Liquid (0.5% C) + Delta iron (0.08% C) $\longrightarrow$ gamma iron (0.18% C)

The peritectic reaction affects only the solidification of steels with less than 0.55% carbon, and it is of little practical importance.

5.1.1 Time-temperature transformation diagram

The iron-carbon equilibrium diagram shows only the phases and the resulting microstructure corresponding to equilibrium conditions. In practice, however, equilibrium is seldom obtained since the cooling rates are

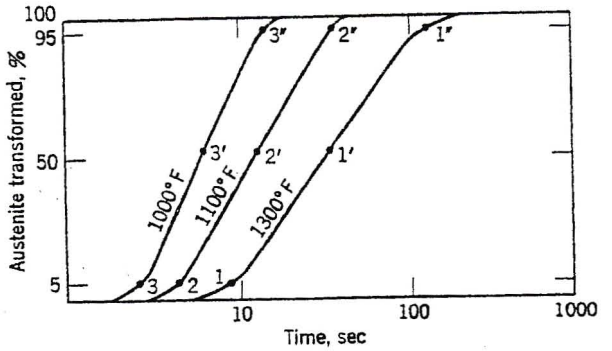


Fig. 5.2 – Isothermal reaction curve

usually much higher than those necessary to obtain equilibrium. As the cooling rate increases, the experimentally observed transformation temperatures are lowered and the metastable (non-equilibrium) phases may be found. For example, at very high rates of cooling in the steel range a

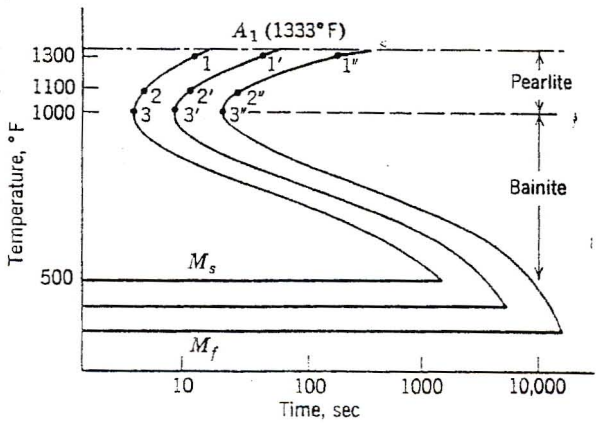


Fig 5.3 – Temperature-Time-Transformation diagram

metastable phase called martensite can develop which, of course, has no place in the equilibrium diagram. The presence of certain alloying elements also affects considerably the transformation range of various phases and may result in completely new diagram.

The Eutectoid reactions are also greatly affected by the rate of cooling. The course of eutectoid transformation may be represented for each temperature by a series of isothermal curves. These are obtained by plotting the per cent of austenite transformed to pearlite versus time for any specified temperature (fig. 5.2). A series of such isothermal reaction curves can be used to plot a time-temperature transformation diagram as shown in fig.5.3. The resulting curve is known as the Bain or S-curve. It is seen from the S-curve that the greatest speed of transformation corresponds to the nose of the knee of the S-curve. The rate of transformation decreases below the temperature of the nose, owing to the slow down of atomic diffusion, whereas above this temperature it decreases because of increased stability of austenite. Above the critical temperature A_1, austenite will be stable and

Fig. 5.4 – Typical austenite microstructure

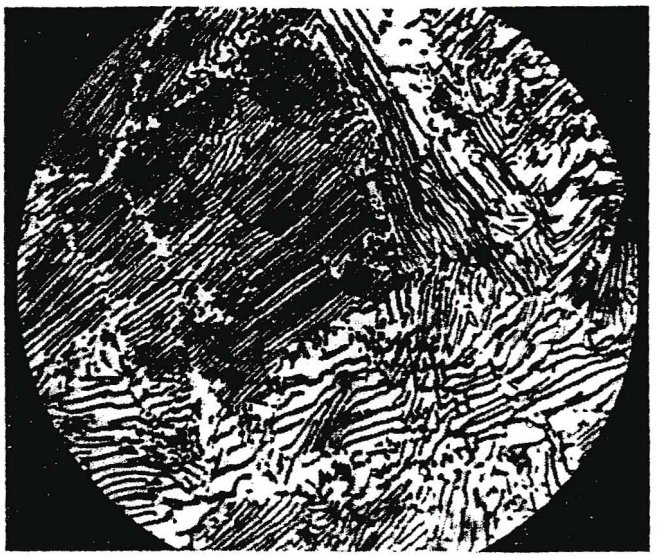


Fig. 5.5 – Typical pearlite microstructure

Fig. 5.6 – Typical martensite microstructure

there will be no transformation to pearlite. In addition to the variations in the rate of transformation with temperature, there are variations in the structure of the transformation products. Transformation at temperatures between approximately 723 and 538^{0}C result in the characteristic lamellar microstructure of pearlite, figure 5.5. At temperature just below the A_1 line nucleation of cementite from austenite will be very slow, but diffusion and growth of nuclei will proceed at maximum speed, so that there will be few large lamellae and the pearlite will be coarse. However, as the transformation temperature is lowered, the pearlite becomes increasingly fine. Transformation at temperatures between 538 and 250^{0}C results in the formation of bainite, consisting like pearlite of a ferrite matrix in which particles of cementite are embedded. The individual particles are much finer than in pearlite, and bainite cannot be resolved by ordinary microscopic examination. Bainite has an acircular microstructure, and it is harder than pearlite and tougher than martensite.

Martensite, figure 5.6, is regarded as a supersaturated solid solution of carbon in ferrite but it has a body-centred tetragonal structure with carbon atoms located interstitially. Martensite is formed by diffusionless transformation of austenite on rapid cooling to a temperature about 260^{0}C designated as M_s. Since the martensite transformation is accompanied by an increase in volume, the temperature must be lowered in order to cause the transformation to proceed to completion. Finally, at a temperature designated as M_f, austenite will be wholly transformed to martensite. The beginning temperature of the martensite transformation M_s and the ending temperature M_f are represented on the T-T-T diagram as horizontal lines, figure 5.3. The position of these lines varies somewhat with the carbon content of the steel. The martensite transformation differs from the other transformations in that it is not time-dependent and occurs almost instantaneously; the proportion of austenite transformed depends only on the temperature to which it is cooled.

Martensite as formed in steels is distinguished by very high hardness, but it also has tendency to brittleness. The hardness depends on the carbon

content; steels containing 0.15% carbon or less do not appreciably harden on quenching. With higher carbon contents above 0.2% the martensite is so hard that the steel is brittle unless properly tempered.

5.1.2 Cooling-transformation diagrams

For a complete picture of transformations, it is necessary to add to the time-temperature-transformation diagram curves showing the rate of cooling. Such a type of diagram is called the cooling –transformation diagram (fig.5.7). The minimum cooling rate required to produce martensite in a given steel is determined by the position of the nose of the S-curve. The cooling rate required to avoid the nose of the S-curve is called the critical cooling rate; any cooling greater than that will cause complete transformation of austenite to martensite. If the cooling rate is slower than the critical cooling rate, the cooling curve (B) will intersect the pearlite zone. This results in the transformation of some austenite to pearlite, and the remaining austenite will only transform to martensite when the Ms temperature is reached. The microstructure of such steel will then be composed of fine pearlite and martensite. The cooling curves C and D show that at such cooling rates
austenite will transform completely to pearlite before any martensite can be formed.
The diagram also shows that the bainite cannot be obtained in any appreciable quantity by continuous cooling to room temperature, because it is sheltered by an overhanging pearlite nose. Thus austenite will be completely transformed to pearlite before the temperature of transformation to bainite is reached. The bainite microstructure can only be obtained by cooling the steel from the austenite range to a temperature between 510 and 260^{0}C at a rate greater than the critical cooling rate. This is accomplished in practice by quenching the steel in a bath of molten salt or a low melting point alloy and holding it at this temperature for some definite time. Then the steel is removed from the quenchant and cooled to room temperature.

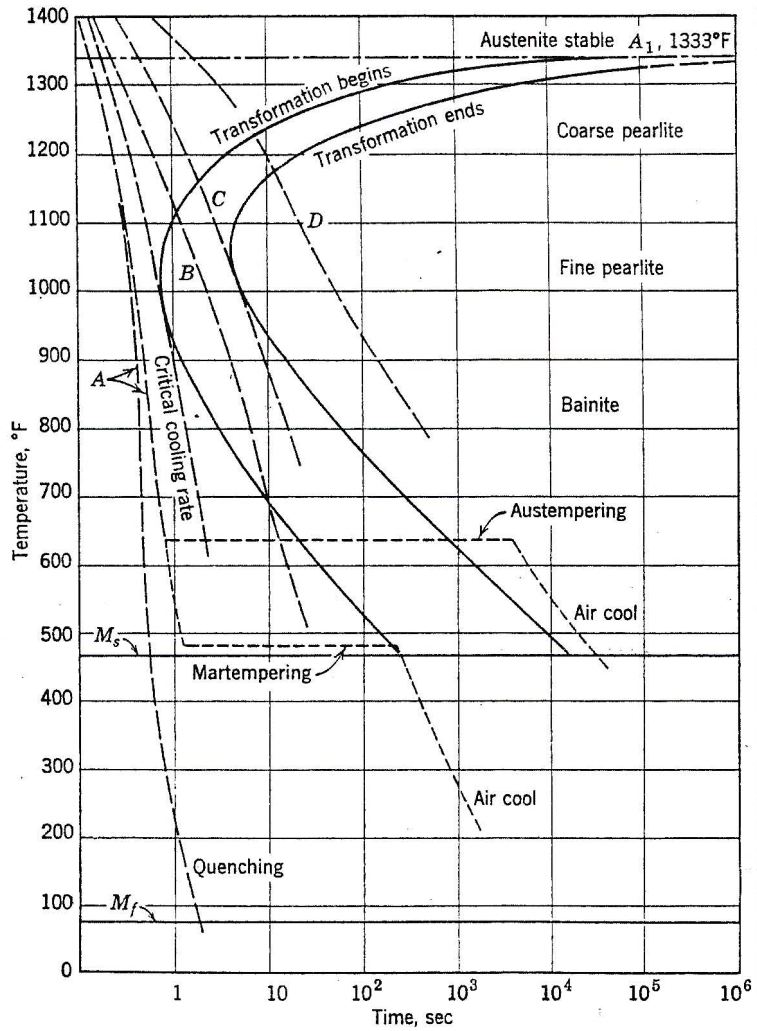


Fig. 5.7 – Cooling-transformation diagram

5.2 Stainless steels

The main alloying element in stainless steel is chromium, which must be present in sufficient quantity to make the steel corrosion resistant. Investigation has shown that for complete corrosion resistance the minimum chromium content should be over 11%. Other alloying elements are also frequently added to enhance certain specific properties of the steel.

Stainless steels are usually divided into ferritic, martensitic, and austenitic types, according to their characteristic microstructure. The ferritic and the martensitic stainless steels are straight-chromium steels designated as 400

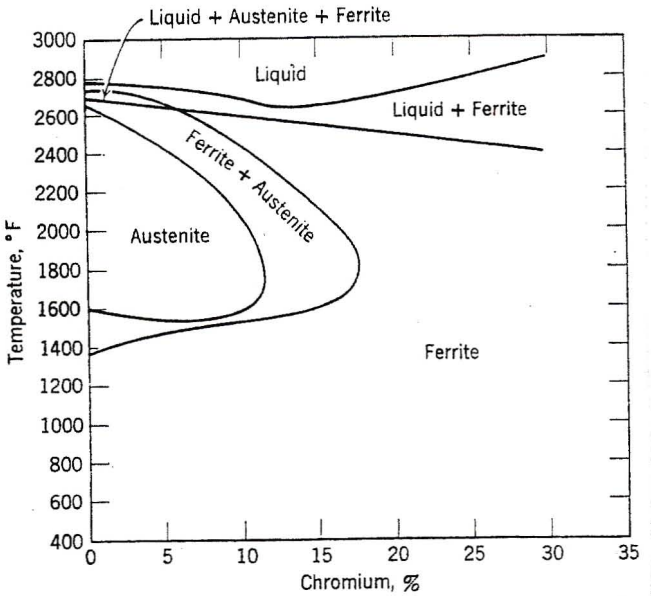


Fig. 5.8 – Iron-Chromium system

series. The austenitic stainless steels are chromium-nickel alloys designated as type 300 series.

Martensitic stainless steels contain from 11.5% to 18% chromium and from 0.15 to 1.20% carbon. The content of chromium and carbon, also sometimes of nickel (only present in types 414 and 431 in an amount from 1.25 to 2.25%) are so proportioned as to permit hardening by heat treatment. Ferritic stainless steels are characterised by the chromium-to-carbon ratio that does not permit hardening by heat treatment. Chromium content varies from 11.5% to 27% and carbon from 0.08 to 0.2% maximum. Both martensitic and ferritic stainless steels are magnetic under all conditions and offer good resistance to oxidation and corrosion. In ferritic stainless steels the main microconstituent is ferrite, present both at high and low temperatures. Consequently, there is no phase change on cooling or heating the alloy through the critical temperature range, and ferritic steels are not hardenable by heat treatment. In the martensitic stainless steels the same phase changes take place as in carbon steels and low-alloy engineering steels during their hardening. Austenite is a stable phase at higher

temperatures and transforms on rapid cooling to martensite, making the steel hardenable by suitable heat treatment.

The contents of carbon and chromium are the principal variables in determining the stability of austenite and ferrite at various temperatures. This is illustrated on the equilibrium phase diagram of the iron-chromium system with 0.1% carbon, as shown in figure 5.8. Chromium and iron are mutually soluble in the liquid state and on cooling from a continuous series of solid solutions. The stability of austenite increases appreciably with the carbon content, whereas the stability of ferrite increases with chromium content. Hence, the gamma loop representing the austenite region and the two phase boundary region containing austenite and ferrite become larger as the carbon content increases. At zero carbon content the limit for the existence of austenite is 12% chromium and it is somewhat larger (13%) chromium for the two phase boundary region. As the carbon content increases up to 0.35%, the austenite becomes stable up to about 15% chromium content, whereas the two-phase boundary region extends to even more than 20% chromium. It follows that to obtain ferritic stainless steel the ratio of chromium to carbon must be sufficiently high to suppress the formation of austenite. Consequently, ferritic steels are characterised by high chromium contents with a reasonably low carbon content not exceeding 0.07% to 0.15%. On the other hand, martensitic stainless steels are characterised by a lower chromium-to-carbon ratio, so that austenite will be formed at high temperatures, thus allowing hardening to martensite on cooling. Figure 5.8 shows that a steel with 18% chromium and 0.1% carbon remains essentially ferritic at all temperature ranges, so that hardening by quenching is not possible. A steel with 15% chromium is in a two phase region above 1010^{0}C and, when cooled from this temperature, will produce a microstructure made up of martensite and whatever ferrite exists. With still lower chromium content, the steel will consist only of austenite phase above 1010^{0}C and will produce a completely martensitic microstructure when quenched.

Both ferritic and martensitic stailess steels are magnetic under all conditions. Generally, the corrosion resistance and oxidation resistance of ferritic stainless steels are higher than those of martensitic stainless steels, because of the higher chromium and lower carbon contents. The martensitic stainless steels show their maximum corrosion resistance in the fully hardened condition, because then all chromium carbides are dispersed uniformly in the matrix, but ferritic steels exhibit their maximum corrosion resistance in the annealed state.

5.2.1 Chromium nickel austenitic stainless steels

Austenitic stainless steels are essentially chromium-nickel-iron alloy, generally of a composition varying from 16 to 26% chromium and 6 to 22%

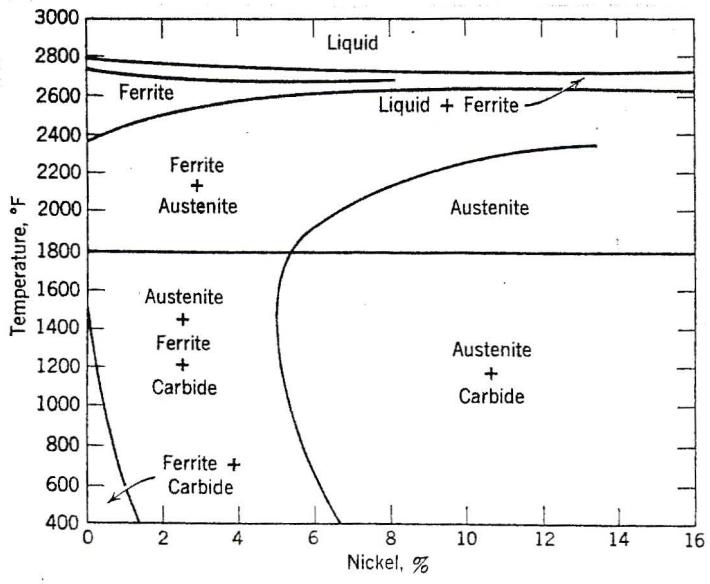


Fig. 5.9 – Iron-carbon-chromium phase diagram

nickel, with a maximum of up to 0.25% carbon. Other alloying elements such as molybdenum, vanadium, and titanium are also added to develop or intensify certain specific properties. The most widely used grade of austenitic stainless steel is the type "18/8" containing from 17 to 19%

chromium, 8 to 10% nickel and up to 0.2% carbon. When from 2 to 4% molybdenum is added to the basic composition, types 316 and317 are obtained, which show an improved corrosion resistance against pitting and a higher temperature strength.

As nickel is progressively added to a ferritic stainless steel, containing 18% chromium, the stability of the austenite phase is increasing until the steel becomes austenitic at all temperatures. For a steel with 0.1% carbon, a complete austenitic structure is obtained with 7% nickel. The structure however, becomes complicated because of the presence of chromium carbides. These have very small solubility below a temperature of 982^{0}C and precipitate preferentially along the grain boundaries. The microstructure of the steel with less than 7% nickel will consist of austenite, ferrite chromium carbides, whereas that of steel with a minimum 7% nickel and more will consist of austenite and precipitated carbides. Since chromium carbides, as $Cr_{27}C_3$, contain appreciable quantities of chromium, the areas immediately adjacent to the grain boundaries will be impoverished in chromium content and corrosion resistance of the steel will be considerably decreased. When the 18-8 steel is heated again to temperature above 982^{0}C, the carbides will pass in to solution, resulting in a single phase austenitic region. On quenching in water the precipitation of carbides can be suppressed, and the steel will have a completely austenitic structure. However, the austenite retained at room temperature is in a metastable condition and tends to decompose into ferrite and finely dispersed carbides. This may occur as a result of intensive cold work or heating of the steel to a temperature between 427 and 871^{0}C. The quenching of the austenitic steel from a temperature above 1800^{0}F to room temperature is called full annealing, and steel in such a condition is the softest and most malleable material, while also showing maximum resistance to intergranular corrosion. The stability of austenite increases as the nickel content is increased and decreases with increase in chromium content or with the addition of other strongly carbide -forming elements such as molybdenum, titanium, vanadium, and columbium. Austenitic

stainless steel is non-magnetic in a fully annealed condition, but the effect of cold work makes the steel become highly magnetic through the formation of some ferrite. As the result of the instability of austenite, various fabricated pieces of equipment may need to be annealed after fabrication, particularly if localised sections of the steel are heated, as during welding operation. The phenomenon known as welding decay is the result of intensive carbide precipitation in the regions adjacent to the weld, where the temperature of the steel is in the range from 427 and 871^{0}C. Such steel can only be restored to its normal condition by full annealing. However this is impractical in many circumstances, or even impossible when the steel has to be used in this temperature range. Therefore, it is important to prevent the precipitation of chromium carbides in such cases; otherwise the material would find only limited application. This is accomplished by adding such elements as titanium and columbium. These, being very much stronger carbide formers than chromium, combine with all the carbon in the steel, thus preventing the precipitation of chromium carbides. Carbide precipitation itself is not eliminated, but the undesirable effects of chromium carbide precipitation are minimised. In order to obtain complete stabilization titanium additions should be at least five to six times the carbon content of the steel and columbium additions eight to ten times the carbon content.

Another method of obtaining a stable material would be to eliminate completely the carbon from the steel. This is impractical and too expensive, but lowering the carbon content below 0.03% is possible and will decrease the danger of chromium carbide precipitation. Recently extra-low-carbon austenitic stainless steels (such as 316E.L) containing a maximum of 0.03% carbon have been developed, which give satisfactory performance without full annealing of the fabricated and welded pieces.

Another structural phase observed in austenitic stainless steels and in some ferritic stainless steels is the "sigma" phase (σ). The sigma phase is a hard and brittle intermediate compound of variable composition that affects adversely the ductility and the corrosion resistance of the stainless steel. It is believed that the sigma phase is formed from ferrite, preferably at the edge

of ferrite grains. Steels with chromium content above 20%, containing such ferrite elements as molybdenum, columbium, titanium, and tungsten, are susceptible to the sigma phase formation. On the other end , an increase in nickel retards the sigma-phase formation because of its stabilizing effect on austenite. Consequently, the nickel content of type 316 stainless steel, containing from 2.5 to 3% molybdenum, must be raised to 10% in order to prevent the formation of sigma phase.

Austenitic stainless steels harden appreciably during cold working, but cannot be hardened by heat treatment.

5.2.2 Austenitic manganese steel

Austenitic manganese steel contains about 1.1% carbon and 12% manganese. This steel exhibits exceptional hardening at the surface when subjected to impact or heavy cold working, for austenite, because of its instability, transforms spontaneously to wear-resistant martensite, whereas the inner core of the steel remains austenitic. The combination of the extremely hard surface with a relatively soft but tough core makes the steel particularly suitable for jaw crushers, grinders, rails, and the teeth of power shovels which require high abrasive resistance and high hardness as well as high impact resistance. Because of this extreme work hardening, austenitic manganese steel is very difficult to machine and is generally used in the form of castings.

5.3 Magnetism

The magnetic properties of materials arise from the spin of electrons and the orbital motion of electrons around the atomic nuclei. Since a moving charge sets up a magnetic field around it, the moving and the spinning electrons act as tiny magnets. In many atoms the opposite spins neutralise one another, but when there is an excess of electrons spinning in one direction,

a magnetic field is produced. Except for ferromagnetic materials, which can form permanent magnets, all substances exhibit magnetic effects only when subject to external electromagnetic field.

The external field is characterised by the magnetic field strength H and the magnetic induction B. The magnitude of the magnetic induction is defined by a number of hypothetical induction lines per unit area normal to their direction and is measured in gauss units. The magnetic induction is also referred to as flux density, because the induction at a point is the flux per unit area.

The magnetic field strength, called also the magnetic intensity or magnetising force, can be represented, like the magnetic induction, by lines of magnetisation.

Magnetisation, or magnetic moment per unit volume, is the product of the magnetic moment of each atom and their number in a unit volume. The ratio of the magnetisation to the applied magnetic field is called magnetic susceptibility,

$$\chi = M / H \tag{5.1}$$

The magnetisation M, the magnetic field H, and the induction B are related through the equation

$$B = H + 4\pi M \tag{5.2}$$

and the permeability μ and the susceptibility χ will be related by an analogous relation,

$$\mu = 1 + 4\pi\chi \tag{5.3}$$

From equations (5.1) and (5.2) the relation between B, H and χ is obtained as

$$B = H + 4\pi\chi H = H(1+4\pi\chi) \qquad (5.4)$$

and the combination of equations (5.3) and (5.4) gives

$$B = \mu H \quad \text{or} \quad \mu = B/H \qquad (5.5)$$

Permeability is a measure of the ability of the material to be magnetised or the ease with which the magnetising force H can induce a flux density B in a magnetic circuit.

5.3.1 Magnetisation curve

The changes of B with varying H are best represented by the magnetisation curve, figure 5.10. Three characteristic regions can be distinguished

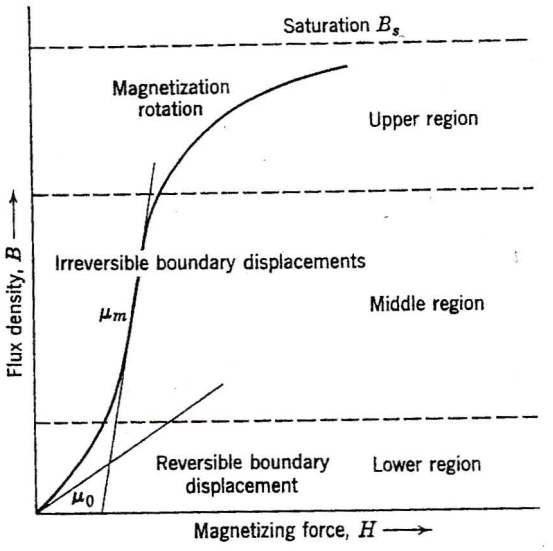


Fig. 5.10 – Magnetisation curve

in the magnetisation curve. The initial region has a definite slope called the initial permeability. The middle portion is the steepest, and in this region the permeability has its greatest value. The upper portion bends over and approaches the horizontal line marked B_s, representing saturation magnetisation of the material. Above this line any further increase in H causes no useful increase in B.

The permeability μ of ferromagnetic material is a function of several variables. The variation of permeability with the magnetic induction B is shown on the $B\mu$ curve, figure 5.11.

5.3.2 Hysterisis

Hysterisis can be defined as the lag in the changes of magnetisation behind variations of the magnetic field. If a ferromagnetic material is subjected to increasing or decreasing magnetic fields, the changes in the magnetic

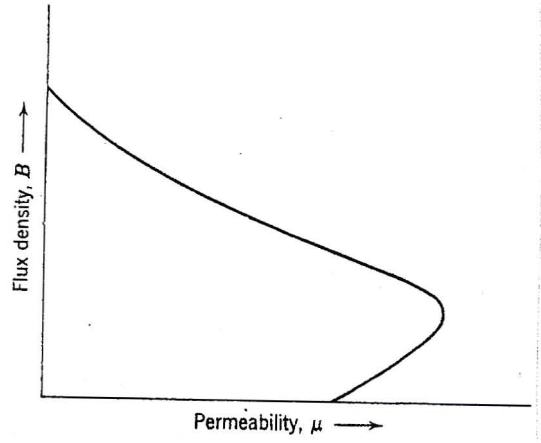


Fig. 5.11 – Permeability curve

induction plotted against the magnetic field result in the hysterisis loop. Upon increasing the magnetic field H, the magnetic induction B will also increase, and the magnetisation curve will follow the line OB (fig.5.12). On reducing the field H, the magnetisation curve does not coincide with the originally rising curve OB, but it follows the line BC. At point C, where the field H is zero, the magnetic induction does not become zero but assumes a value B_r, known as the residual induction. If the magnetic field is now applied in the negative direction, the point D is reached where the magnetic

induction is zero and the magnetising force is negative, having a value of H_c. This negative value of the magnetic field causes demagnetisation, and it is called the coercive force. A further increase in the magnetic field in the negative direction results in a maximum induction, but in the opposite direction (point E). By continuing in this way the curve EFGB is obtained and the cycle is completed. The enclosed area is called the hysterisis loop and represents the energy loss during that cycle.

When the applied magnetic field varies cyclically at some finite speed, eddy current effects may become important. The magnitude of these currents depends on the frequency and the flux density imposed by the application of the magnetic field, and on the specific resistance and the thickness of the

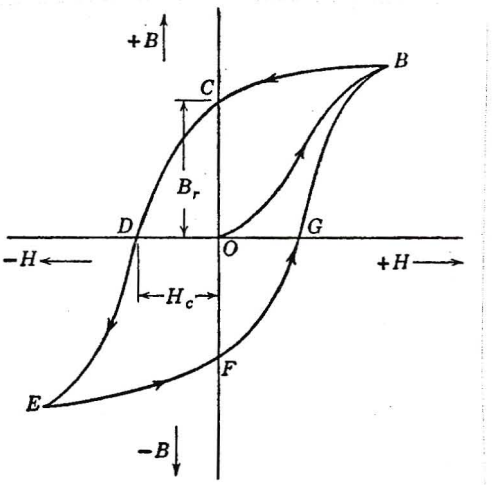


Fig. 5.12 – Hysterisis curve

core material. The effect of eddy currents is to increase the magnetic field H for the same induction, thus causing a much wider hysterisis loop. The sum of the hysterisis loss and eddy current loss is known as the total core loss.

5.3.3 Curie temperature

At very low temperatures, all ferromagnetic materials are nearly completely saturated, but when the temperature increases the saturation decreases, at first slowly and then rapidly, until a temperature is reached at which the material becomes practically nonmagnetic, . This temperature is known as the Curie temperature and has a characteristic value for each particular magnetic material. For example, the Curie temperature is 353°C for Nickel, 768°C for iron, and 1120°C for cobalt. Through suitable alloying of ferromagnetic elements, or by adding a nonmagnetic element to a ferromagnetic element, it is possible to produce a magnetic material of any desired value of Curie temperature, figure 5.14. This principle has found wide application in many engineering devices. For example, in motors, generators, transformers, relays, and permanent magnets, the material used must have a Curie temperature well above the operating temperature; otherwise the magnetic properties will be destroyed. On the other hand, the use of nickel and iron alloys in cathode ray tubes, mass spectrometers, and compass housings requires nonmagnetic behaviour. Hence the Curie temperature must be below the lowest operating temperature. The changes

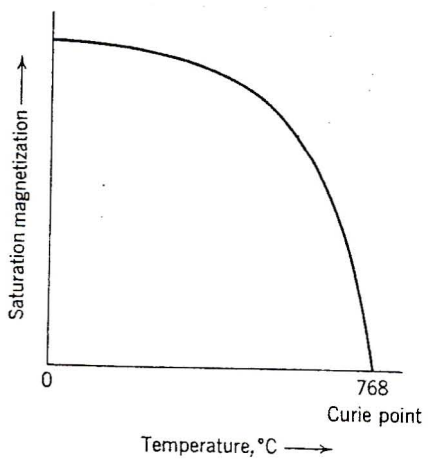


Fig. 5.13 – Saturation magnetisation versus temperature

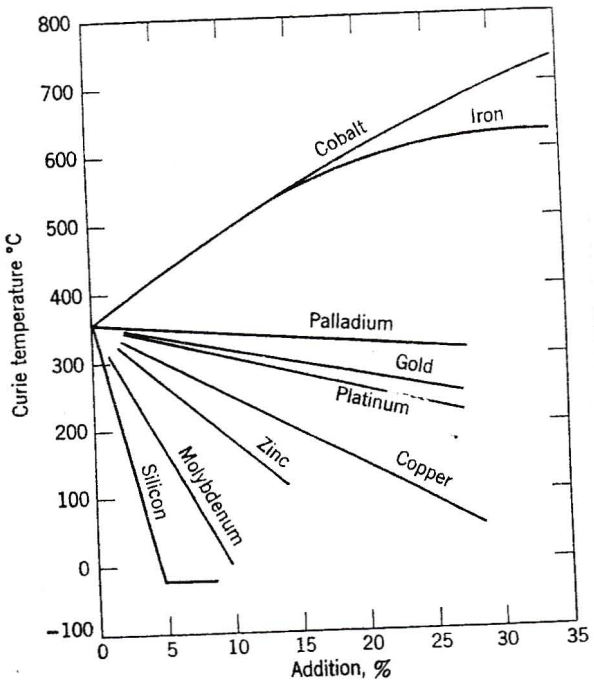


Fig. 5.14 – Curie temperature

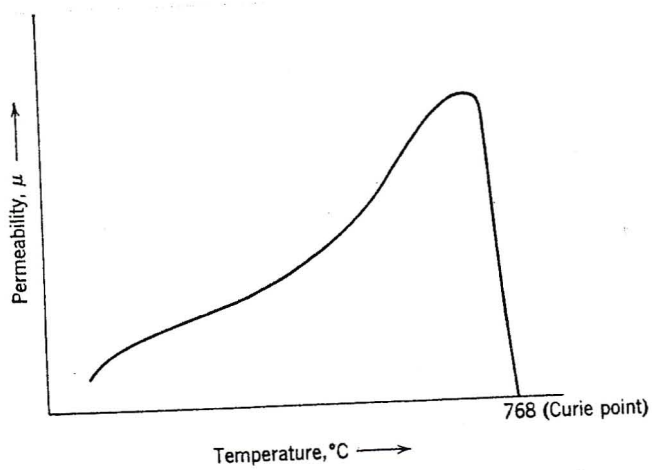


Fig. 5. 15 – Initial and maximum permeabilities for iron

of magnetic properties with temperature are also utilised in various temperature compensating devices, or in producing special compensating alloys. Permeability will rise with increasing temperature, reaching its maximum value close to the Curie temperature; then it abruptly falls to zero at Curie temperature, figure 5.13.

5.3.4 Domain theory

The domain theory, states that all ferromagnetic materials are composed of a number of submicroscopic regions called domains. Each domain contains parallel atomic moments and is always magnetised to saturation, but the directions of magnetisation of different domains are not necessarily parallel. In the absence of the applied magnetic field, adjacent domains may be

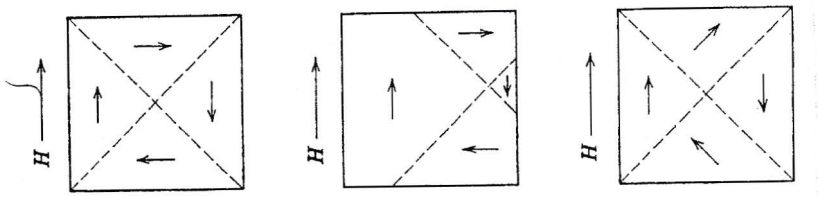


Fig. 5.16 – Fundamental magnetisation process

orientated randomly in any of the several directions, called the directions of easy magnetisation, which depends on the geometry of the crystal. The resultant effect of all these various directions of magnetisation may be zero, as is the case with an unmagnetised specimen, figure 5.16. When a magnetic field is applied, the domains most nearly parallel to the direction of the applied field grow in size at the expense of the others. This is called boundary displacement of the domains or the domain growth, figure 5.16. When this is completed a further increase in the magnetic field causes the domains to rotate and align parallel to the applied field. The material reaches a point of saturation magnetisation, and no further change will take place on increasing the strength of the magnetic field.

The three regions of the magnetisation curve may be explained in terms of the domain theory. The initial portion of the curve corresponds to the

reversible boundary displacement, the middle region to the irreversible boundary displacement, and the upper region to the domain rotation.

5.4 Dislocations and stacking faults

5.4.1 Stacking faults and dislocations in FCC structures

Closed packed planes in FCC atoms are the (111) planes. It can be shown that the (111) planes are stacked in the sequence ABC...ABC... .The burgers vector $b_1 = (a_0/2)[101]$ defines one of the observed slip directions. If atoms are considered as hard spheres, it is easier for an atom on type B planes to move along a zigzag plane b_1+b_2 in the valleys instead of moving over the hump that lies in the path of vector b_1. The dislocation reaction is given by $b_1 \rightarrow b_2+b_3$. Slip by a two stage process creates a stacking fault ABC AC | ABC in the stacking sequence. The dislocation with burgers vector b1 has been dissociated into two partial dislocations b_2 and b_3. These two dislocations are imperfect in that they do not produce any lattice translations. The two dislocations combined together are collectively known

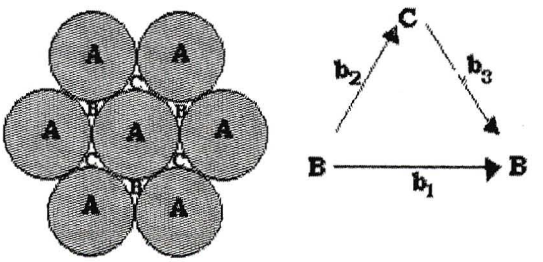


Fig. 5.17 – Atom arrangement in FCC structure

as extended dislocations. The partial dislocations will settle at an equilibrium separation determined primarily by the stacking fault energy. The stacking fault energy can vary considerably for different FCC metals and

alloys and this in turn can have an important influence on their deformation behaviour.

Dissociation of unit dislocations is independent of the character (edge, screw or mixed) of dislocations. However, unlike the unextended screw dislocation, the extended screw dislocation defines a specific slip plane, the (111) plane of the fault, and it will be constrained to move to move in this plane. The partial dislocations move as a unit maintaining the equilibrium width of the faulted region. Because of the restriction to a specific slip plane, an extended screw dislocation cannot cross slip or climb unless the partial dislocations combine into a perfect dislocation.

The deformation process involves slip or, particularly at high temperatures, climb. As mentioned previously, extended dislocations cannot readily slip past each other or climb from one plane to another. In this way work hardening behaviour or the strength of the material is increased. These areas of crystallographic discontinuity or their interaction can facilitate the formation of transformation products, and precipitation of second phases can also occur on stacking faults.

5.4.2 Dislocation energy and stacking faults

The regularity of the crystal lattice results from each atom taking up a position which minimises its potential energy. Therefore a dislocation represents the rising up of potential energy of all the atoms affected by the dislocation. Thus an energy may be ascribed to a dislocation which, physically represents the strain energy built into the crystal by displacement of atoms from their regular positions.

We may estimate the energy by considering the screw dislocation in a cylinder of length l. At radius , r, the deformation in a thin annulus of thickness dr is given by magnitude ,b, of the Burgers vector so that the shear strain, γ, is $b/2\pi r$, and using Hooke's law for shear, the average stress, τ, will be given by $Gb/2\pi r$, where G is the elastic shear modulus.

The elastic strain energy dw of a small volume element dv is given by $\frac{1}{2}\tau\gamma dv$

So we have $dw = \frac{1}{2}\frac{Gb}{2\pi r}\frac{b}{2\pi r}dv$ (5.6)

$$= \frac{1}{2}G\left(\frac{b}{2\pi r}\right)^2 dv$$

The volume of the annular element is $2\pi r/dr$, hence

$$dw = \frac{Gb^2 l}{4\pi}\frac{dr}{r} \quad (5.7)$$

The elastic strain energy obtained within a cylinder of radius R around the dislocation is obtained by integrating this eqution up to radius R. However, the lower limit of integration is not taken as zero for two reasons: (1) the calculation assumes the material is an isotropic continuum, but in the region near to the dislocation this assumption is unrealistic and it is necessary to consider the displacements of, and the forces between, individual atoms; and (2) the strains near the dislocation are large and Hooke's law, on which the calculation is based, fails at large strains. Thus the lower limit of integration is taken as some small value, r_0 (which is often taken as equal to b). The region inside r_0 is referred as dislocation core. We therefore have

$$E = \frac{Gb^2 l}{4\pi}\int_{r_0}^{R}\frac{dr}{r} + E_c l \quad (5.8)$$

Where $E_c l$ is the strain energy within the radius 0 to r_0, i.e, the core energy. The calculation for the core energy is complex but a reasonable estimate is

Gb2l/10. Thus on integrating to obtain the elastic strain energy we get for the total strain energy.

$$E=\frac{Gb^2l}{4\pi}\ln\frac{R}{r_0}+\frac{Gb^2l}{10} \tag{5.9}$$

Since generally R>>r_0, the logarithmic term in this expression only varies slowly with R/r_0. As an approximation, ln(R/r_0) may be taken as 4π, then

$$E= Gb^2l+\frac{Gb^2l}{10} \tag{5.10}$$

Two important features of this formula are : (1) the total energy is proportional to b^2; and (2) the core energy is only one-tenth of the elastic strain energy.

A dislocation of Burgers vector $\mathbf{b_1}$ can split into two, as explained above. This behaviour can be explained in terms of energies of dislocations involved. We know that the energy of a dislocation is proportional to the Burgers vector squared, therefore a dislocation will split, if it is *geometrically* possible, when

$$b_1^2 > b_2^2 + b_3^2 \tag{5.11}$$

Alternatively, two dislocations will combine if

$$b_1^2 < b_2^2 + b_3^2 \tag{5.12}$$

From the FCC stacking sequence, in figure 5.17, it can be realised that the included angle between the two dislocations is 60^0, therefore from the vector diagram

$$b_2 = \frac{b_1}{2\cos 30^0} = \frac{b_1}{\sqrt{3}} \qquad (5.13)$$

Similarly, $b3 = \frac{b_1}{\sqrt{3}}$. Therefore the partial dislocations are energetically favourable, since

$$b_1^2 > b_2^2 + b_3^2 = \left(\frac{b_1}{\sqrt{3}}\right)^2 + \left(\frac{b_1}{\sqrt{3}}\right)^2 = \frac{2b_1^2}{3} \qquad (5.14)$$

The partial dislocations are responsible for slip in FCC structures. Usually, the leading dislocation creates a stacking fault, and the following dislocation recreates a normal FCC stacking sequence.

5.4.3 Stacking fault energy

As mentioned previously, a stacking fault is created between two partial dislocations, and the equilibrium separation of these dislocations is determined by a balance between the elastic repulsive force between these dislocations (i.e. angle between the partial dislocations is 60^0) and the surplus free energy per unit area of the fault, the stacking fault energy.
The width of the stacking fault η is expressed in terms of the stacking fault energy as :

$$\eta = G(b_1.b_2)/2\pi r\gamma, \qquad (5.15)$$

where G is the shear modulus, b_1 and b_2 are the magnitudes of the Burgers vectors of the two partial dislocations, and γ is the stacking fault energy.

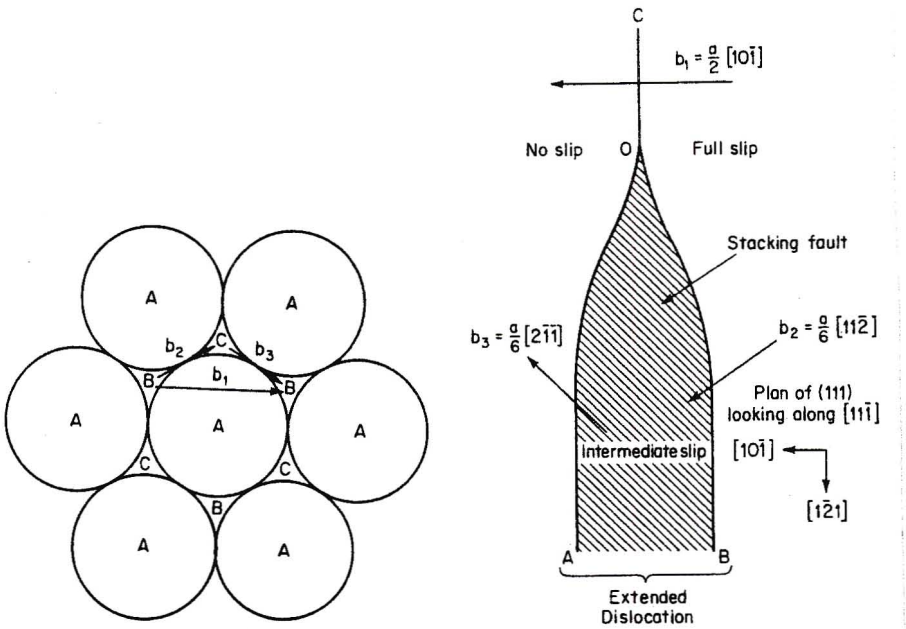


Fig. 5.18 – Extended dislocations

5.4.4 Stacking faults in FCC stainless steels

Many of the austenitic stainless could properly be termed metastable austenitic steels. From Figure 5.19 , one can see that several transformation products can be obtained. Some of these products arise from different types of discontinuities possible under the FCC austenite. A brief discussion follows. The closed-packed planes in FCC lattice are the cell diagonal (111) planes. The atoms in one such plane are shown in figure 5.17. In the FCC lattice, atoms in successive planes are in positions marked ABCABC.... A hexagonal-closed packed (hcp) structure is very similar, but the stacking arrangement is ACACA... instead.

The favoured slip direction in the FCC lattice is [110]. Producing such slip is not easy, so slip is usually visualised as occuring in a zigzag fashion shown by b2 and b3 in figure 5.17.

If the net slip between the (111) FCC planes is from B to C instead of B to B, the ABCABC sequence will be disrupted. It then becomes ABC/CAB, thus generating an HCP sequence. Depending on the frequency of faulting, either an HCP structure (fault on every other plane) or a twin structure (fault on every plane) can result. A metastable austenite, when cooled or deformed,

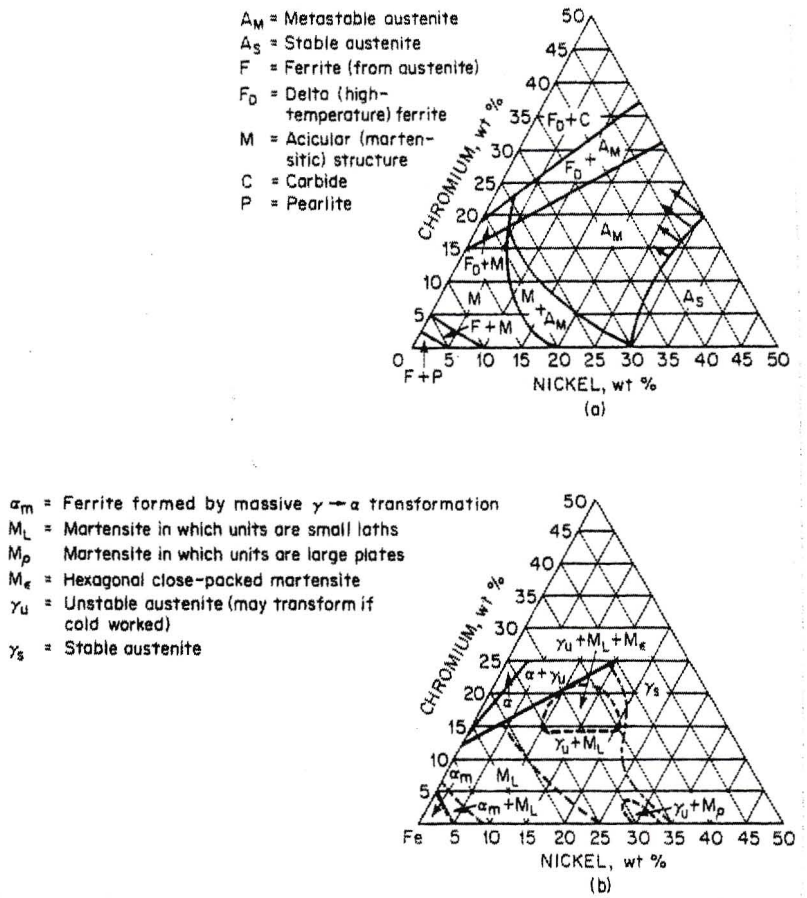


Fig. 5.19 – Fe-Cr-Ni system at room temperature

can fault or undergo transformation to bcc martensite. As will be seen, a combination of these processes may also occur.

The energy associated with faulting is often considered in terms of dislocation theory. A common representation, figure 5.18, shows a dislocation split into two partial dislocations with faulted material between

them . A film analogy is often used: faulted material is considered to have "surface tension" which tends to pull the partial dislocations together. This

Table 5.1- Stacking fault energy of Fe-Cr-Ni alloys

Reference	Stacking-fault energy, mJ/m²	Composition, %				
		Cr	Ni	C	N	
Whelan et al.[63]	15–20	18–20	8–11	<0.08		(AISI 304)
Whelan[64]	30	18–20	8–11	<0.08		(AISI 304)
Swann[65]	14	17.9	7.1	0.06	0.11	
	16	17.6	7.9	0.06	0.04	
	30	18.1	12.8	0.02	0.12	
	36	17.6	12.7	0.02	0.004	
	44	17.8	17.8	0.03	0.004	
Breedis[68]	29.2	19.3	11.2			
	40.7	17.3	11.0			
	51.1	16.0	12.1			
	76.1	13.0	14.2			
	98.0	10.4	16.2			
Douglass et al.[66]	28–41	18.74	9.43	0.07		(AISI 304)
	58	20	20	0.013		
	131	20	40	0.012	0.006	
Dulieu and Nutting[67]	~50	8.5	11.84	0.05		
	23–28	18.3	10.28	0.079	0.02	
	28–32	22	9.87	0.079		
	23–28	18	9.87			
	20–25	16	9.87			
	30	18	8			(Probably AISI 304)
	>46	22.0	34.0	0.014		
Silcock et al.[66]	53	15.3	15.9	0.02		
	64	15.8	23.0	0.007	0.006	
	48	15.9	15.8	0.017	0.004	
	70.2	15.4	24.7	0.011	0.007	
Clement et al.[69]	19	16.6	9.5	0.028		
Thomas and Henry[70] and Thomas[71]	53	17.8	14.1	0.01		
Vingsbro[72]	~8	18	13	0.02		
Fawley et al.[73]	23	20	10	0.006	0.005	
	32	20	15	0.018	0.003	
	40	20	20	0.015	0.006	
	38	20	25	0.009	0.006	
	34	20	30	0.011	0.006	
	53	10	20	0.012	0.004	
	40	15	20	0.019	0.005	
	45	25	20	0.022	0.005	
	57	30	20	0.036	0.005	
	23	20	10	0.012	0.044	
	34	20	15	0.010	0.035	
	38	20	20	0.018	0.051	
	38	20	25	0.009	0.033	
	34	20	30	0.010	0.029	
	48	10	20	0.012	0.013	
	44	15	20	0.012	0.011	
	47	25	20	0.017	0.047	
	43	20	20	0.027	0.003	
	47	20	30	0.050	0.010	
Latanision and Ruff[74]	23.1 ±1.7	18.7	15.9			
Murr[75]	21	18.43	9.52	0.058		(AISI 304)

Table 5.1- Stacking fault energy of Fe-Cr-Ni alloys (contd)

Reference	Stacking-fault energy, mJ/m²	Composition, %			
		Cr	Ni	C	N
LeCroisey and Thomas[76]	45	17.8	14.1	0.01	
	24	15.9	12.5	<0.01	
Latanision and Ruff[77]	16.4 ±1.1	18.3	10.7	0.005	
	23.6 ±0.9	18.7	15.9	0.005	
Butakova et al.[78]	60	5	20	0.04	
	30	10	15	0.04	
	20	15	10	0.04	
	16	18	8	0.04	

configuration is one of relatively high energy. Its formation and stability will depend on the magnitude of the "surface tension" or stacking fault energy. Deformation processes involve dislocation slip or, particularly at high temperatures, climb. Extended dislocations, such as those represented in figure 5.18, cannot readily slip past each other or climb from one plane to another. In this way work hardening behaviour or strength of the material can be increased. These areas of crystallographic discontinuity or their interaction can facilitate the formation of other transformation products. Precipitation of second phases (such as carbides) can also occur on stacking faults and precipitation on such sites is also often more advantageous.

Table 5.2 – Effect of elements on strength of austenite in alloys approximating AISI Type 302 stainless steel

Solute	Type	Strength coefficients	
		For 0.2% proof stress	For tensile strength
N	Interstitial	32	55
C		23	35
Cb	Substitutional, Ferrite Stabilizer	2.6	5.0
Ti		1.7	3.0
Al		.8	2.4
Si		1.3	1.2
V		1.2	0
Mo		0.9	0
W		0.3	0
Cr		0.2	0
Ni	Substitutional, Austenite Stabilizer	0	−0.1
Mn		0	0
Cu		0	0
Co		0	

Table 5.3 – Stacking fault energy of several commercial austenitic alloys

Alloy	Composition, wt %									Stacking-fault energy,* mJ/m²
	Ni	Cr	Mn	Mo	C	Si	P	S	N	
AISI 304L	8.28	18.31	0.82	0.02	0.025	0.30	0.010	0.007	0.0183	18
AISI 305	11.85	18.02	1.64	0.10	0.074	0.32	0.13	0.002		34
AISI 310S	18.8	24.7	1.73	0.44	0.047	0.56	0.022	0.005		94
AISI 316	13.01	17.15	1.40	2.09	0.055	0.53	0.027	0.008		78
21-6-9 (melt 1)†	7.11	21.00	8.75	0.03	0.027	0.43	0.011	0.003	0.31	65
21-6-9 (melt 2) †	6.48	20.30	9.55		0.034	0.13	0.022	0.012	0.26	41
22-13-5†	12.34	21.57	5.17	2.20	0.041	0.40	0.020	0.004		64
Hadfield	4.1		15.7		0.91	1.29	0.009	0.017		21

5.4.5 Stacking fault energy in FCC stainless steels

Schramm and Reed cited in [17] recently surveyed published stacking fault energy data obtained from Fe-Ni-Cr alloys. Their tabulation is shown in Table 5.1. The four dimensional linear-regression analysis equation is given in figure 5.21, and was obtained by using the data in table 5.2. (Silicon and manganese contents, though not listed in table 5.1, were included in the analysis.)

Schramm and Reed cited in [17] also independently determined the stacking fault energy of eight commercial austenitic steels. The results are shown in 5.3. These results conflict with some of those given in figure 5.21. Chromium is shown to either decrease or increase stacking fault energy. In either case, the coefficients are small relative to some other elements. The quantitative effects of other elements also differ. Among the elements, the observed effect of Nickel in increasing the stacking fault energy has been most consistent.

Cobalt as an element has a low stacking fault energy and is recognised as reducing the stacking fault energy in Fe-Ni-Cr austenitic alloys. Results of Dulieu and Nuttig cited in [17] are presented in Table 5.4.Qualitative considerations suggest that the relative effect of cobalt should perhaps be greater than that indicated by Dulieu and Nuttig cited in [17]. Cobalt has been utilised in alloys like the trip steels and multiphase alloys. Rationalization of its role in such alloys strongly implies significant reduction of stacking –fault energy.

The conflicts in the various studies originate from the difference and uncertainties associated with the experimental techniques. Measurement of stacking -fault dimensions using transmission electron microscopy has often been used. X-ray diffraction techniques utilising line broadening and shifts also have been applied.

Fe-Ni-Cr austenites are, as a class, relatively low stacking fault energy materials. Energetically, they are borderline at room temperature. They do not ordinarily spontaneously fault extensively, but can be induced to do so with the observed effects being sensitive to specific alloy chemistry and /or temperature. Stacking fault energy varies approximately linearly with temperature in the vicinity of room temperature.

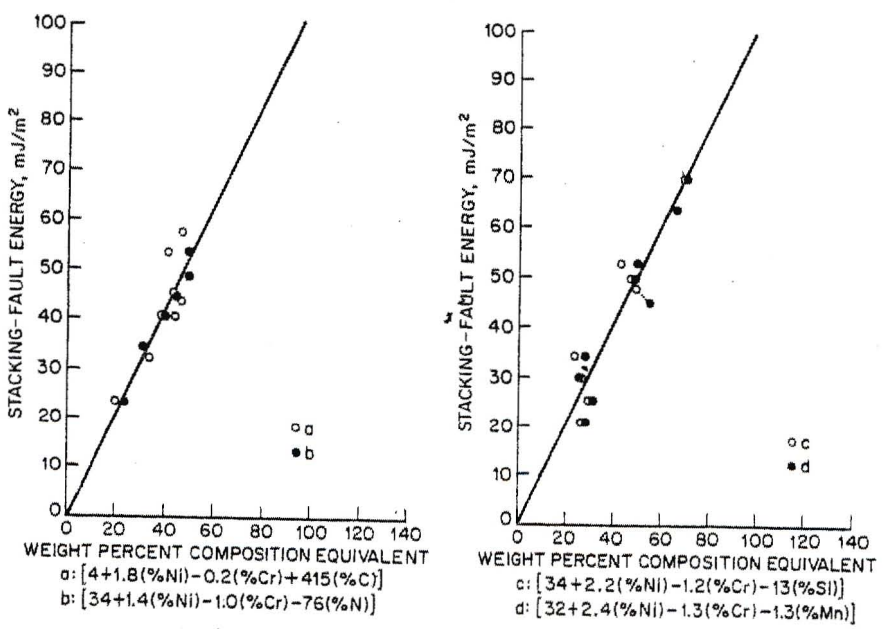


Fig 5.21 – Linear regression analysis results from table 5.1

5.4.6 Martensitic transformation in austenite stainless steels

The transformation of Fe-Ni-Cr stainless steel is a complicated subject and has attracted a great deal of attention. It has been known for a very long time that the learner FCC austenites (γ) transform to BCC martensite (α') at or below room temperature. As mentioned previously and noted in figure 5.19, there are distinctions to be drawn between the types of martensite formed. Considerable effort and discussion has been directed towards defining these types with regard to morphology and modes of formation.
The types of martensite are chemistry related, as shown in figure 5.19. The large plate morphology falls close to the Fe-Ni binary system in alloys not represented by the AISI 300 series of alloys.

Table 5.4 – Effect of elements on stacking fault energy

Element	Effect on stacking-fault energy ergs/(cm²)(atomic %)
Cr	+0.5
Ni	+1.4
Si	−3.4
Co	−0.55
Cu	+3.6
Cb	+3.2
Mo	+0.1

A characteristic of Fe-Ni-Cr alloy transformations attracting considerable attention is the apperance of the HCP phase, ε. Interest in this HCP phase has increased in recent years partly because techniques for studying such constituents have progressed and partly because of recent research and development work on cobalt-bearing low stacking fault energy materials like TRIP steels. The ε martensite is closely related to stacking faults discussed in the previous section. Since the ε martensite is commonly found in conjunction with α' martensite, researchers have tried to determine if ε is an intermediate to the formation of α' martensite or a separate constituent which could form as a result of strain generated by the $\varepsilon \rightarrow \alpha'$ transformation . Since ε martensite is non magnetic and is difficult to define clearly by light microscopy, its identification requires diffraction techniques.

Venables and Olson and Cohen cited in [17] have shown the formation of α' needles at the intersection of ε martensite plates. Cina, Reed, and Breedies and Robertson also reported the HCP structure to be an intermediate in the formation of α' martensite. Dash and Otte reported instances of independent formation of the two phases and summarized various possibilities as shown in figure 5.22. From subsequent work it appears ε is usually involved in the transformation process to α'.

As discussed by Christian and others, there also has been a speculation that the HCP structure could be a transition phase; this possibility is also noted in figure 5.23. The peaking nature of the data in figure 5.23 would appear to support this notion at least at large strains.

The formation of ε martensite should depend on temperature and composition. One example of this is shown in figure 5.23, which compares the volume fraction of ε in type 304 austenite formed at 25^0 and -196^0. Olson and Cohen cited in [17] have described the formation of transformation products by correlating austenite plastic strain and martensite formation by an expression

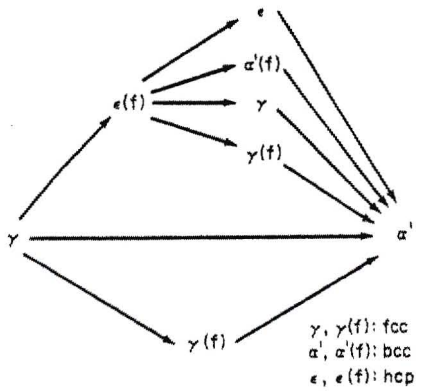


Fig. 5.22 – Various Hypothetical paths for formation of α' martensite from γ

which includes separate terms related principally to either shear band energy (stacking fault energy dependent) or chemical driving force. Figure 5.23 is illustrative of the temperature dependance of the shear band energy

Alloys which failed to transform at -196°C were further cooled to 40K, and no transformation took place.

Schumann studied the formation of ε and α' in a series of alloys with a varying manganese, nickel, and chromium contents with the results shown

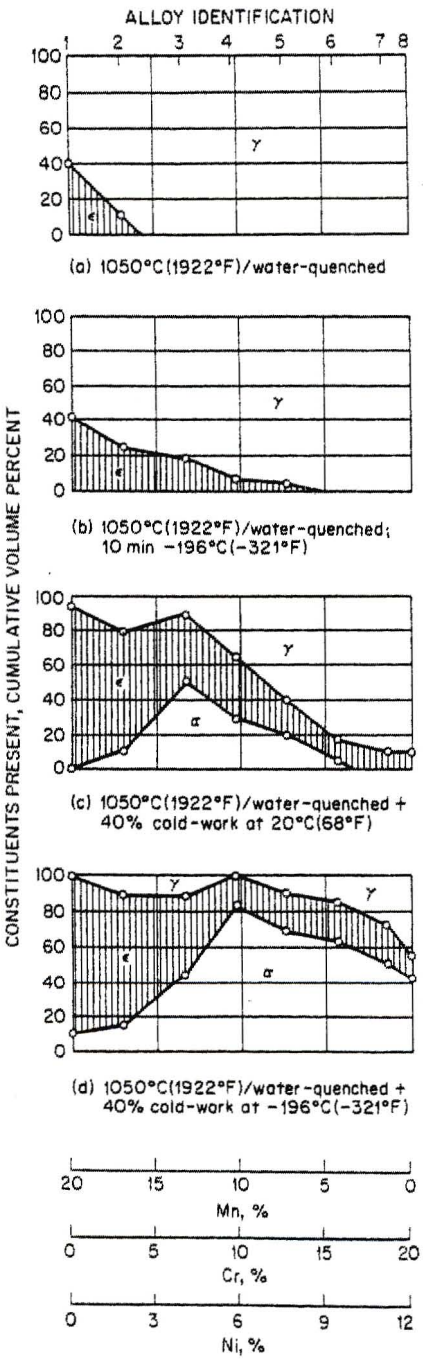


Fig. 5.25 - Transformation products in Fe-Cr-Ni-Mn alloys

in figure 5.25 . Alloy1, an Fe-Mn binary, formed only ε except when cold-worked at -196°C . As Nickel and chromium were added, the amount of ε formed with diminished rapidly. In alloys 7 and 8, which approximated conventional AISI composition ranges, no transformation at all occurred upon cooling. Working the alloy at ambient temperature produced some ε. The transformation behaviour of Fe-Cr-Co-Ni-Mo austenites has been examined by deBarbadillo. Although not specifically determined, these alloys should have lower stacking fault energies than normal AISI austenitic grades. Transformation arising from cold rolling at various temperatures is shown in figure 5.26. Alloys with higher nickel equivalent contents exhibited

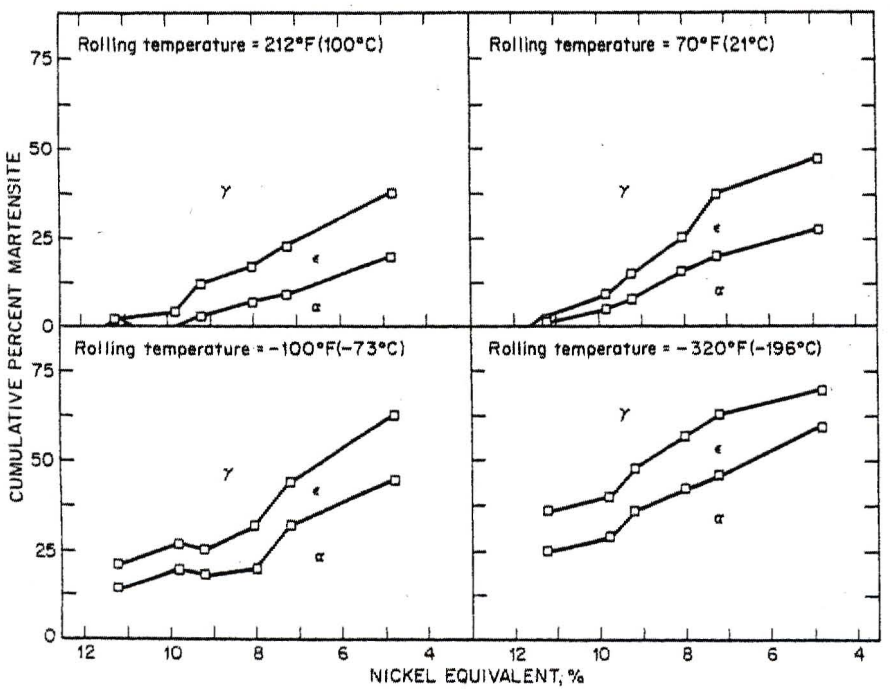


Fig. 5.26 - Effect of composition and rolling temperature on transformation of Fe-Cr-Co-Ni-Mo austenites

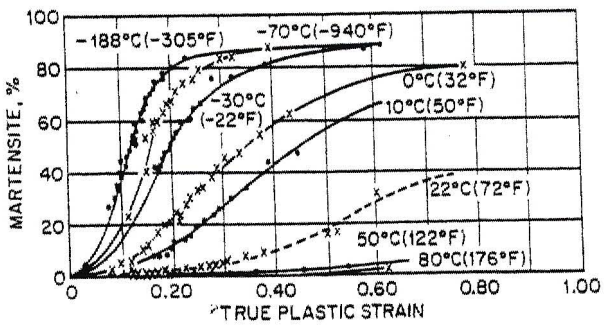


Fig. 5.27 – Formation of α' (BCC) martensite in 18-8 stainless steels

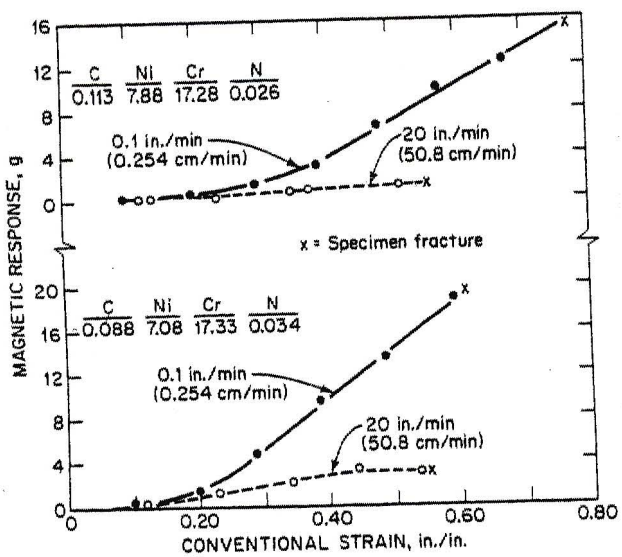


Fig. 5.28 – Effect of strain rate on α' martensite formation in Type 304 stainless steel deformed in air

less transformation and the ε formation was reduced, especially at higher rolling temperatures.

The trends observed with respect to ε martensite are also apparent when only α' martensite is considered. Angel's results from an 18Cr-8Ni stainless steel is shown in figure 5.27. Increasing strain and decreasing temperature both cause increased transformation. (The balance of this discussion will

concern effects on formation of BCC α' martensite. The ε martensite has not been considered in most of the past work. This does not impair the value of such work, as the contribution of this constituent to the properties of cold-worked AISI austenite grades is still uncertain. It is sufficient to remember that strengthening effects not directly attributable to α' martensite may be due to in part to the formation of the HCP constituent).

Angel's curves fit (figure 5.27) the following autocatalytic-type equation:

$$\ln(1/f - f) = n\ln(truestrain) + k$$

where f is the fraction transformed and n and k are constants. By studying a series of alloys, Angel cited in [17] found the value of n to be about 3 and independent of both alloy composition and temperature. The value of k on the other hand, depended on both composition and temperature.

In addition to austenite composition, strain, and temperature, strain rate effects have been examined. The results of Bressanelli and Moscowitz[17] are shown in figure 5.28. The absolute martensite contents were not determined, but the reduced transformation at higher strain rates is obvious. These authors point out that this effect is largely influenced by specimen heating at high strain rates, where temperatures of 93^{0}C can be reached. The effect of preventing specimen heating by use of water bath is shown in figure 5.29. Thus it can be seen that the primary influence is still one of temperature dependence. This effect is largely to be operative any time high-strain-rate deformation is applied to metastable austenite steels and attempts are made to correlate observed effects with probable transformation.

5.4.7 Austenite stabilization

The preceding discussions have concerned material initially cooled from normal austenizing temperatures. Under some circumstances transformation can be suppressed, a phenomenon which has been called

austenite stabilization. This term refers to changes in behavior different than bulk composition effects.

One obvious instance of stabilization is the effect of existing α' martensite on further $\gamma \rightarrow \alpha'$ transformation. In figure 5.27, it can be seen that transformation proceeds to a "saturation " value of less than 100%α'. This is typical of Fe-Ni-Cr alloys. Angel cited in [17] analysed curves of the type shown in figure 5.27 and found that the rate of reaction per unit of austenite as a function of martensite contents passes through a maximum at intermediate martensite contents. This behaviour (figure 5.30) clearly shows stabilisation effects during the last stages of transformation. Angel cited in [7] attributes this stabilization to an increase in the surface energy of the martensite such that finally exceeds the effect of externally applied stress. He also states that the deformation energy per increment of martensite volume increases with temperature, thus rationalizing why the

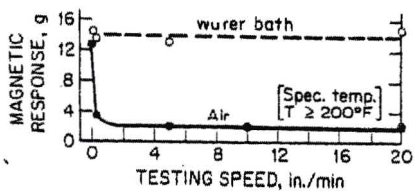


Fig. 5.29 – Effect of test medium and resultant temperature on martensite formation in Type 301 deformed in ambient temperature bath

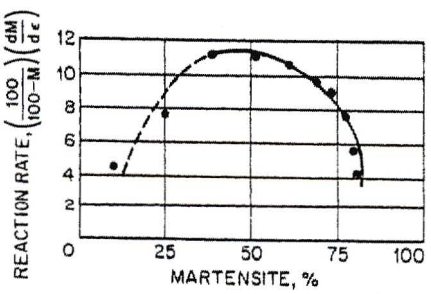


Fig. 5.30 – Rate of reaction per unit of austenite as a function of martensite content at -70°C

"saturation" volume fraction of martensite decreases with increasing temperature.

Increasing austenizing temperature also can result in an apparent stabilization of austenite. Otte and Cina cited in [17] reported that alloys which normally transform do not do so when quenched from extraordinarily high temperatures. In Otte's case, austenizing a 15Cr-15Ni alloy at 1320°C completely suppressed transformation all the way to to 4°K. When quenched from 1050 to 1150°C, the same alloy exhibited both α' and ε at −196°C. While it is known that increasing grain size reduces M_s somewhat, this effect is more pronounced. Breedis cited in [17] has examined the effect of reheating a high purity 16Cr-12Ni alloy after it was quenched to 196°C, which then initially formed 14%α' martensite. His results are shown in figure 5.31. Specimens reheated at 500°C and then requenched formed more martensite, whereas alloys reheated at 600, 700, and 800°C generated less martensite. The increased martensite after 500°C exposure was related to the relief of internal strains. Reheating to higher temperatures partially

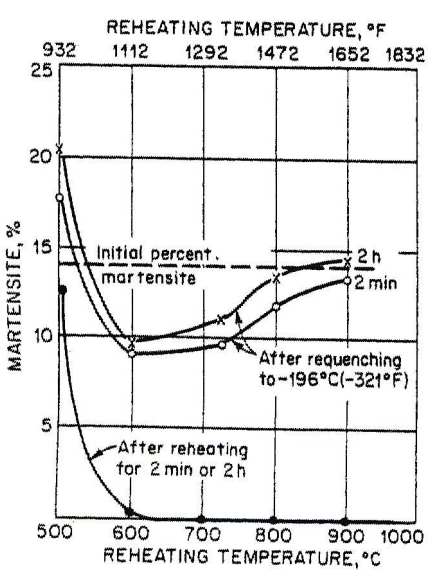


Fig. 5.31- Effect of reheating a previously transformed 16Ni-12Cr alloy on austenite reversion and subsequent transformation

stabilizes austenite, because it is postulated, the remaining more-altered defect structure is such that it hinders subsequent development of

martensite laths. Using transmision microscopy, Breedis cited in [17] reported the prior martensite areas were still identifiable even after they had become austenitic. Volumes have been written concerning austenite stabilization in other related iron-based systems, i.e., Fe-C, Fe-Ni, Fe-Ni-C. An important aspect of stabilization mechanisms in some of these other systems is solute redistribution (especially carbon). Review of these very numerous references has been entirely eliminated because, as discussed previously, the martensite in the Fe-Ni-Cr alloys of interest here is different. In, addition the carbon contents of Fe-Ni-Cr austenite stainless steels , while not so low that the role in transformation process is unequivocally nonfunctional, are generally low enough to make meaningful comparisons with results on related systems uncertain.

5.4.8 Composition and austenite stability

Austenite in metastable alloys can transform spontaneously to martensite either on cooling or as a result of deformation. The temperature at which transformation to α' martensite occurs on cooling is called the M_s (martensite start) temperature. The temperature below which α'

Table 5.5 – Approximate relationship

Approximate Relationship	
(g)	Percent of α'
1	20
4	50
9	70
16	90

Table 5.6 – Transformation temperatures for an Fe-18Cr-7Ni-0.18C stainless steel

	Transformation temperature, °K
E_d	420
M_d	360
E_s	≥ 300
M_s	190

martensite will form under deformation conditions is called the M_d temperature. More uncommon are E_d and E_s temperatures. E_d is defined as

the maximum temperature at which transformation to ε martensite can be deformation-induced. Es is defined as the temperature where γ→ε occurs spontaneously. The relative relationship between these various

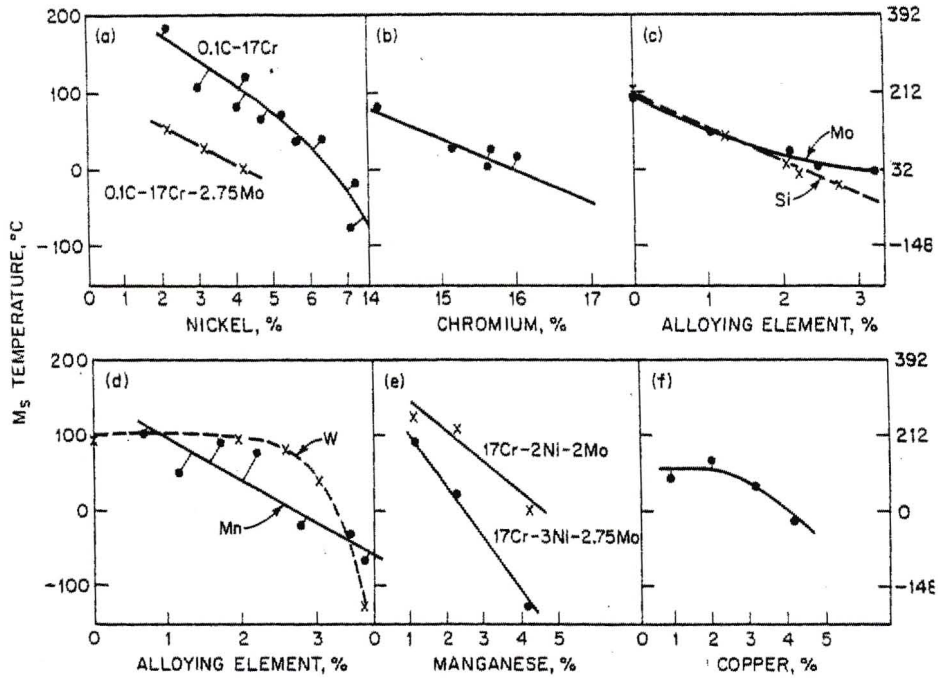


Fig. 5.32 - Effect of alloying elements on M_s temperature

transformation temperatures is shown by the Remy and Pineau data given in table 5.6 [17].

Several expressions relating austenite stability and chemistry have been developed. These are given in table 5.7. Irvine, et al, cited in [17], presented their results graphically, as shown in figure 5.32. These relationships pertain to stability in terms of the $\gamma \rightarrow \alpha'$ martensite transformation. All elements cause increased stability, thus lowering transformation temperatures. An exception to this trend would probably be observed for cobalt if there were a similar relationship for E_d and E_s.

Post and Eberly developed their Δ factor to indicate when an alloy would resist transformation at cold reducing up to 80%. Modifying this expression to include copper and nitrogen, Griffiths and Wright cited in [17] showed

that the stability factor adequately described trends in composition versus work-hardening behaviour. Also, as shown in figure 5.33, the onset of α' martensite formation seemed to correlate with the modified factor.

Table 5.7 – Expression relating austenite stability and alloy chemistry

Quantity computed	Relationship (elements in weight %)
Stability factor Δ	$Ni - \left[\frac{(Cr + 1.5Mo - 20)^2}{12} - 0.5Mn - 35C + 15\right]$
Stability factor Δ modified to include copper and nitrogen	$Ni - \left[\frac{(Cr + 1.5Mo - 20)^2}{12} - 0.5Mn - 35C - Cu - 27\right]$
M_s (°F)	$75(14.6-Cr)+110(8.9-Ni)+60(1.33-Mn)+50(0.47-Si)+[0.068-(C+N)]$
M_s (°F)	$2160-66(Cr)-102(Ni)-2620(C+N)$
M_{d30} (°C)	$413-462(C+N)-9.2Si-8.1Mn-13.7Cr-9.5Ni-18.5Mo$
Stability factor S	$Ni + 0.68Cr + 0.55Mn + 0.45Si + 27(C+N)$
Stability factor S	$Ni + 0.68Cr + 0.55Mn + 0.45Si + 27(C+N)+Mo+0.2Co$

Using the data shown in figure 5.34, Angel cited in [17] pointed out that M_d temperature is somewhat ambiguous. For very large strains the onset of transformation is fairly abrupt. The M_d temperature is not at all well defined for small strains. For this reason, his expression was defined for transformation at 30% strain, hence M_{d30}.

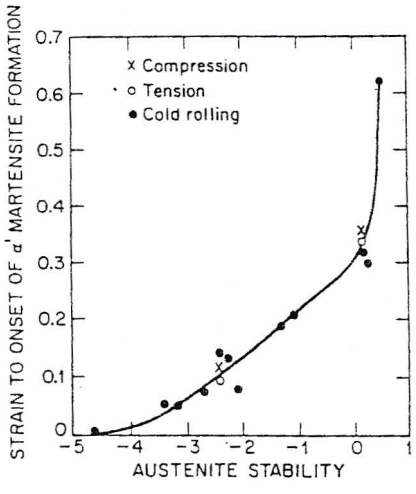


Fig. 5.33 – Effect of austenite stability on threshold strains for α' martensite formation

Relationships of the type shown in Table 5.7 have often served well in characterising the behaviour of stainless steels. However, it should be remembered that they are approximations since possible interactions are not taken into account. Figure 5.35, for example , shows that the effect of copper or carbon on work hardening behavior (or stability) depends on the nickel content.

5.4.9 Work hardening

The AISI austenite stainless steels are only strengthened by cold working. The effect of cold working on room temperature properties is shown in figure 5.36.

The primary cause of work hardening is the transformation to martensite. Figure5.37 shows the increase in α' content concurrent with strengthening in Type 301. The effects of deformation and temperature are consistent with the earlier discussions of transformations. The effect of rolling temperature has been further explored by Floreen and Mihalisin cited in [17]. As shown in figure 5.38, much higher strengths were obtained by rolling at subzero temperatures. The yield strengths obtained generally correlated with the amount of α' martensite obtained.

The strength of subzero-rolled steel is further enhanced by heat treatment at 427^{0}C. This is a general effect, defined by previous investigators.

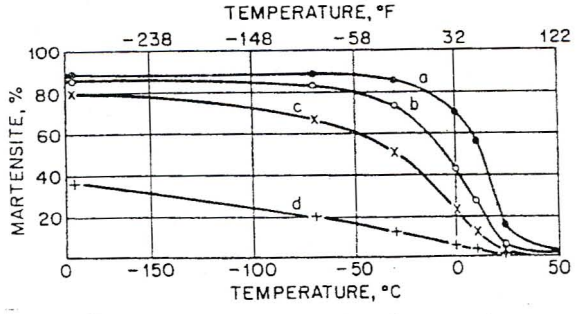


Fig. 5.34- Temperature dependance of martensite formation at true plastic strains

Figure 5.39 compares the effect of cold working on the properties of types 301 and 305 stainless steels. The much more stable Type 305 exhibits a small strengthening increment compared with metastable Type 301. That very limited transformation occurs in Type 305 is demonstrated by the permeability data of table 5.8. It can also be seen that some work hardening is obtained from lattice deformation processes which precede the $\gamma \rightarrow \alpha'$ transformation.

Work hardening can also be considered from the standpoint of process occuring during tensile testing. The behaviour of the tensile test specimen is determined to a considerable degree by the work hardening behaviour of the material. A simple representation of the true-stress-true-strain curve is given by

$$s = k\varepsilon^n$$

where s = true stress

ε = true strain

n = work hardening coefficient

This expression defines the straight line segment at the beginning of the true-stress-true-strain curve. Figure 5.40 (from Barclay cited in [17]) shows that the point of departure from the initial straight line and subsequent increased strengthening coincide fairly closely with the onset of measurable martensite formation. Figure 5.40 also shows the stabilising effect of Nickel.

In order to define the full stress-strain curve, Griffiths and Wright cited in [17] developed a more complicated quadratic equation:

$$\log s = C_1 + C_2 \log \varepsilon + C_3 (\log \varepsilon)^2$$

They found all the constants were influenced by austenite stability as expressed by the modified Post and Eberly cited in [17] stability factor (Table 5.7). This is shown in figure 5.47.

Bressanelli and Moscowitz cited in [17] examined a number of type 301 stainless steel heats and examined the relationship between apparent

martensite formed during test and tensile characteristics. Those alloys developing the most martensite also exhibited the greatest work hardening (figure 5.48). The variations in chemistry were well within those defined by the AISI

Table 5.8 – Permeability of stainless steels as affected by cold working

AISI type	Cold reduction, %	Magnetic permeability	
		$H = 50$	$H = 200$
301	0	1.0027	1.0028
	19.5	1.148	1.257
	55	14.8	19.0
305	0	1.0032	1.0044
	18.5	1.0040	1.0054
	34.5	1.017	1.020
	52.5	1.049	1.063
	84.0	1.093	1.142

range for Type 301, indicating why metastable alloys for critical forming applications are often very tightly controlled in chemistry.

As has been shown, the $\gamma \rightarrow \alpha'$ transformation contributes significantly to work hardening. Since lower temperatures favour transformation , difficult forming operations are sometimes performed at subzero temperatures. The role of the transformation also explains why cold worked stainless steel does not subsequently work harden as much as annealed material. True-stress-true-strain curves for a 17.5Cr-7Ni alloy are shown in fig. 5.49.

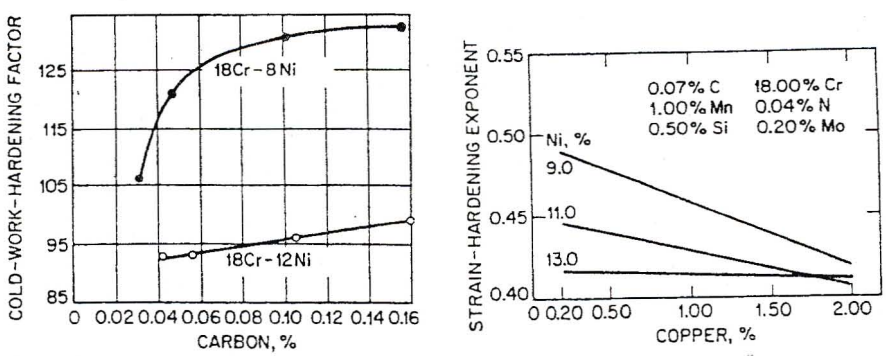


Fig. 5.35 – Examples of interactions affecting stability of stainless steels

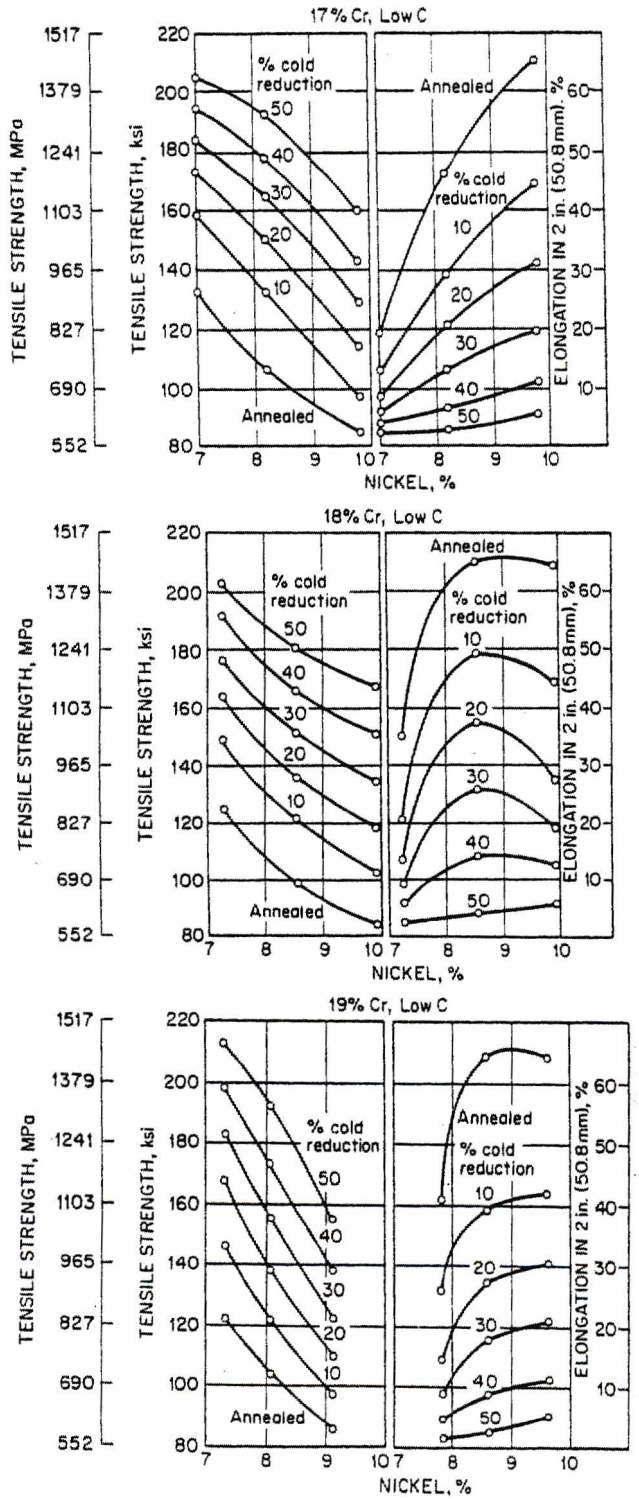


Fig. 5.36 – Effect of chromium and nickel on tensile strength and elongation of cold rolled steels containing 0.05% carbon

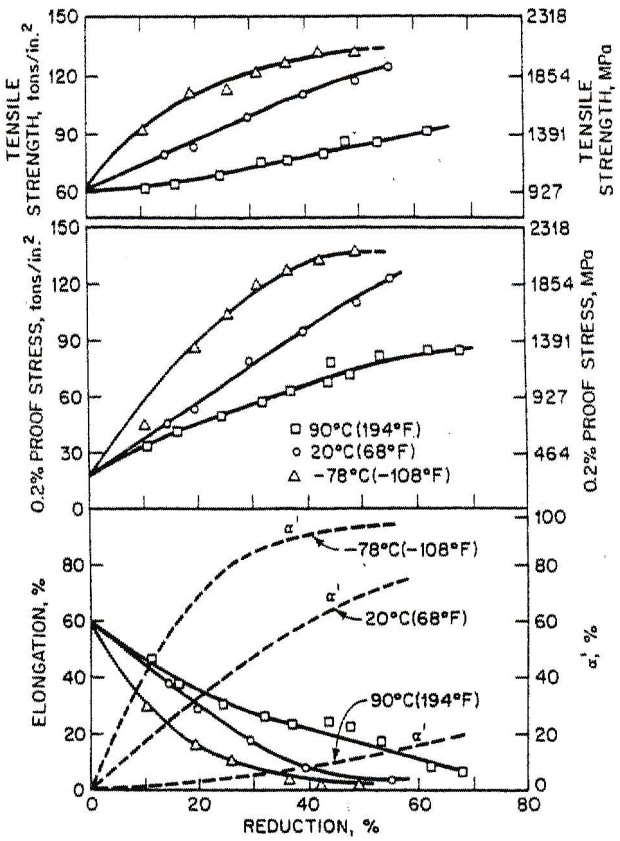


Fig. 5.37 – Effect of cold rolling on Type 301 stainless steels

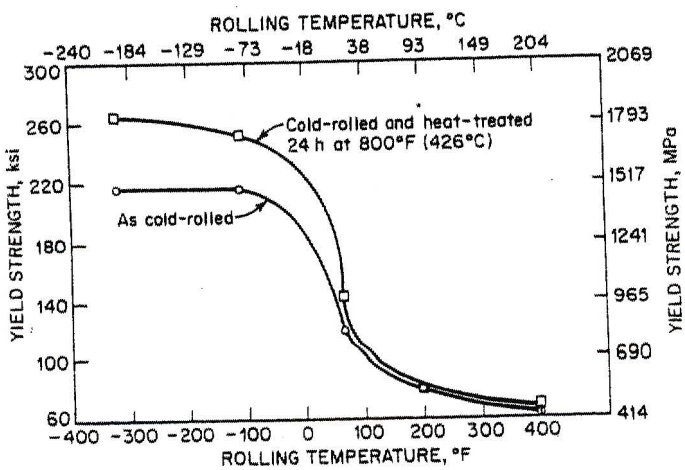


Fig. 5.38 – Effect of rolling temperature on yield strength of 18Cr-7.6Ni-0.9Mo-0.09Si-0.05C alloy

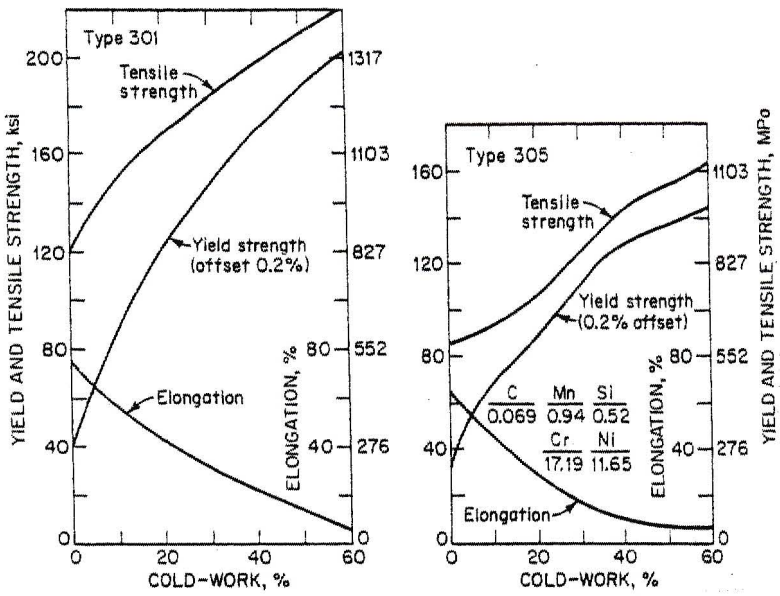


Fig. 5.39 – Effect of cold working on mechanical properties of Types 301 and 305 stainless steels

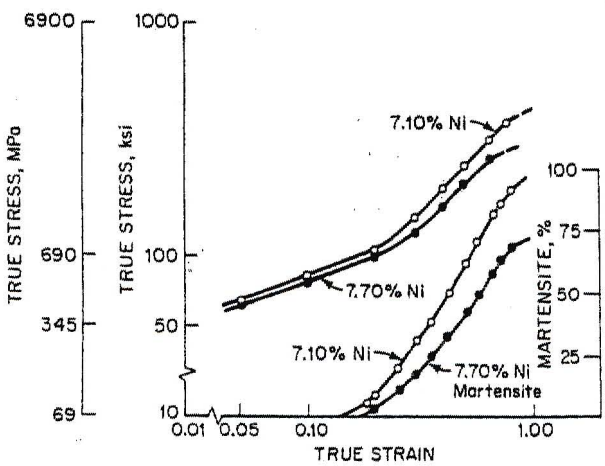


Fig. 5.40 – Relationship between alloy stability, strain hardening, and martensitic formation in Type 301 stainless steels

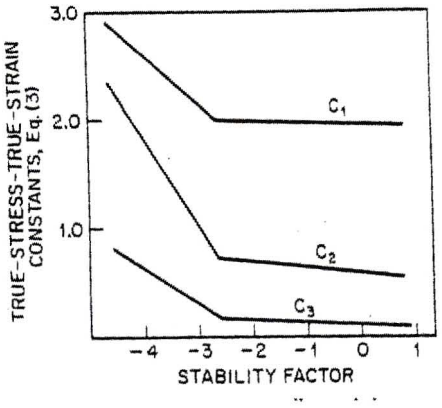


Fig. 5.41 – Relationship between alloy stability and true-stress-strain equation constants

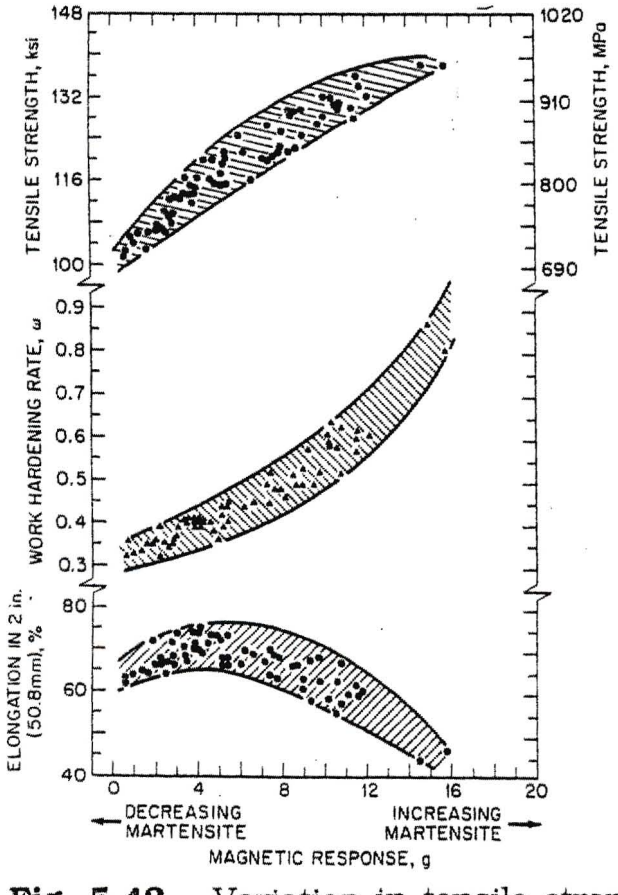


Fig. 5.42 – Variation in tensile strength, work-hardening rate, and tensile elongation with martensite content of Type 301. Heats tested at room temperature.

5.5 Trip steels

Transformation induced plasticity (TRIP) steels were originally developed as ultra high strength structural steels with improved toughness as compared with the high strength, low alloy steels. TRIP steels display a solid state, strain dependent phase transformation from a metastable, austenitic (face centred cubic crystal structure) parent phase to the thermodynamically stable, martensitic product phase.

The strain induced phase transformation is responsible for large values of ductility and strain hardening that are observed in these steels. The phase transformation is triggered as the material begins to yield and enters the plastic deformation regime. The extent of transformation increases with applied strain resulting in increased martensitic fraction. Observation during tensile testing indicates that transformation occurs inhomogeneously along the gauge length of the tensile specimen, regions where martensite has formed become stronger than the rest of the gauge length thereby delaying necking of the specimen. Once transformation progresses to a significant degree along the entire gauge length the specimen deforms homogeneously throughout the entire gauge length until failure. Failure typically occurs with very little or no necking. Specimen elongations of 40 to 60% were observed for materials with yield strengths of 1379MPa. These levels of ductility are high compared with those of high strength steels where elongations are typically in the range of 10 to 15%.

Besides the advantages offered by the above mentioned structural properties, the transformation from an austenitic face-centered cubic crystal structure parent phase to a martensitic body-centered cubic product phase results in a change in the magnetic properties of the steel. The magnetic signature of the steel depends on the fraction of martensite present and thus changes as transformation progresses with increasing levels of strain. This behaviour provides the principal basis for structural health monitoring system.

5.5 Trip steels

Transformation induced plasticity (TRIP) steels were originally developed as ultra high strength structural steels with improved toughness as compared with the high strength, low alloy steels. TRIP steels display a solid state, strain dependent phase transformation from a metastable, austenitic (face centred cubic crystal structure) parent phase to the thermodynamically stable, martensitic product phase.

The strain induced phase transformation is responsible for large values of ductility and strain hardening that are observed in these steels. The phase transformation is triggered as the material begins to yield and enters the plastic deformation regime. The extent of transformation increases with applied strain resulting in increased martensitic fraction. Observation during tensile testing indicates that transformation occurs inhomogeneously along the gauge length of the tensile specimen, regions where martensite has formed become stronger than the rest of the gauge length thereby delaying necking of the specimen. Once transformation progresses to a significant degree along the entire gauge length the specimen deforms homogeneously throughout the entire gauge length until failure. Failure typically occurs with very little or no necking. Specimen elongations of 40 to 60% were observed for materials with yield strengths of 1379MPa. These levels of ductility are high compared with those of high strength steels where elongations are typically in the range of 10 to 15%.

Besides the advantages offered by the above mentioned structural properties, the transformation from an austenitic face-centered cubic crystal structure parent phase to a martensitic body-centered cubic product phase results in a change in the magnetic properties of the steel. The magnetic signature of the steel depends on the fraction of martensite present and thus changes as transformation progresses with increasing levels of strain. This behaviour provides the principal basis for structural health monitoring system.

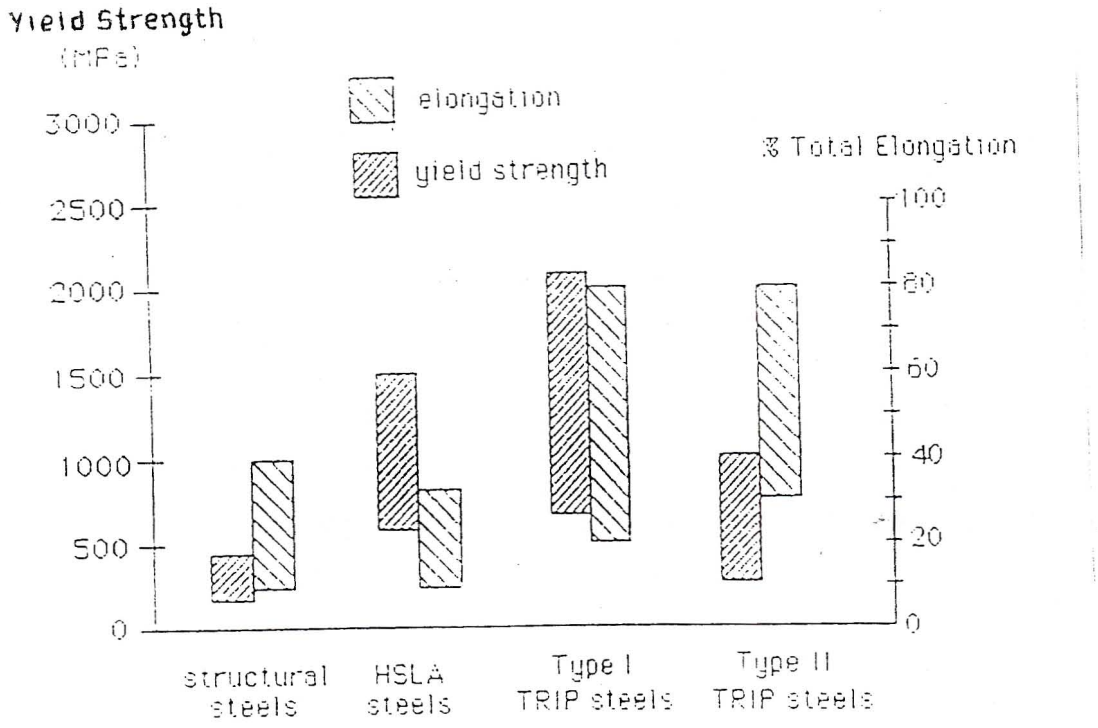
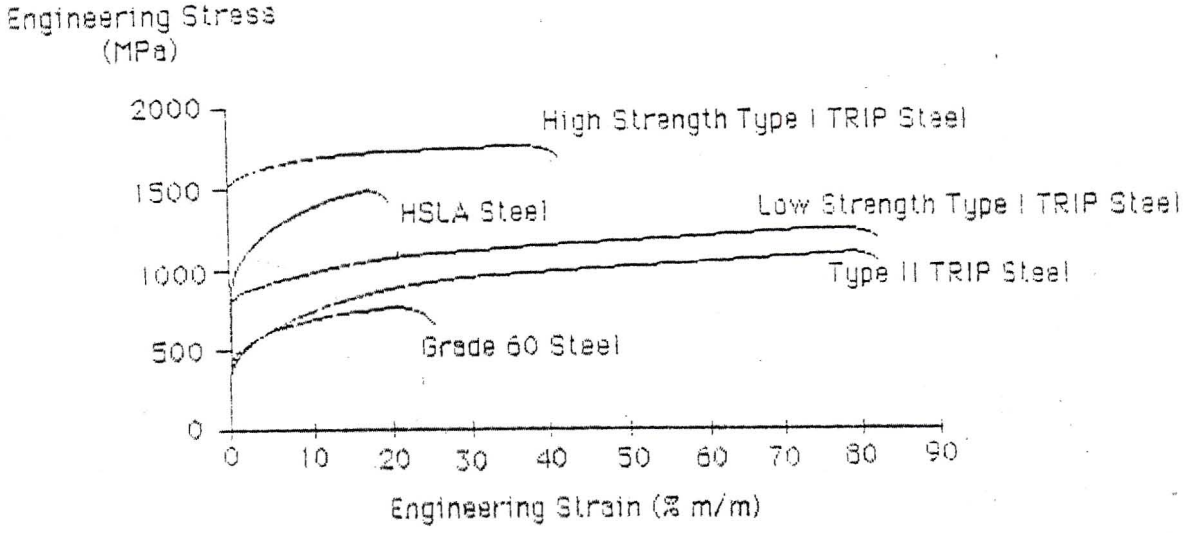


Fig. 5.43 – Comparison of the tensile properties of structural steels

Studies evaluating the behaviour of Fe-Mn and Fe-Mn-Cr alloys have been conducted in parallel with those involving the conventional austenitic stainless steels (AISI 300 series). It is possible to develop an austenitic microstructure from in the Fe-Mn-Cr system and attempts were made to develop austenitic austenitic stainless steels from this base alloy system. Technically, Fe-Mn-Cr steels are not TRIP steels, but they do transform sequentially from FCC to HCP to BCC. The phase transformation can be triggered by lowering the temperature below the martensite start temperatures, referred to here as ε (epsilon) and α' (alpha prime) temperatures. The same transformation sequence is exhibited when

the material is mechanically strained provided that the phase stabilities are adjusted properly through control of the alloy chemistry.

5.5.1 Structural characteristics of TRIP steels



al

Fig. 5.44 – Representative Engineering Stress-Strain curves for structural materials

Fig. 5.45 - Stress-strain curves for structural grade steel and low-strength

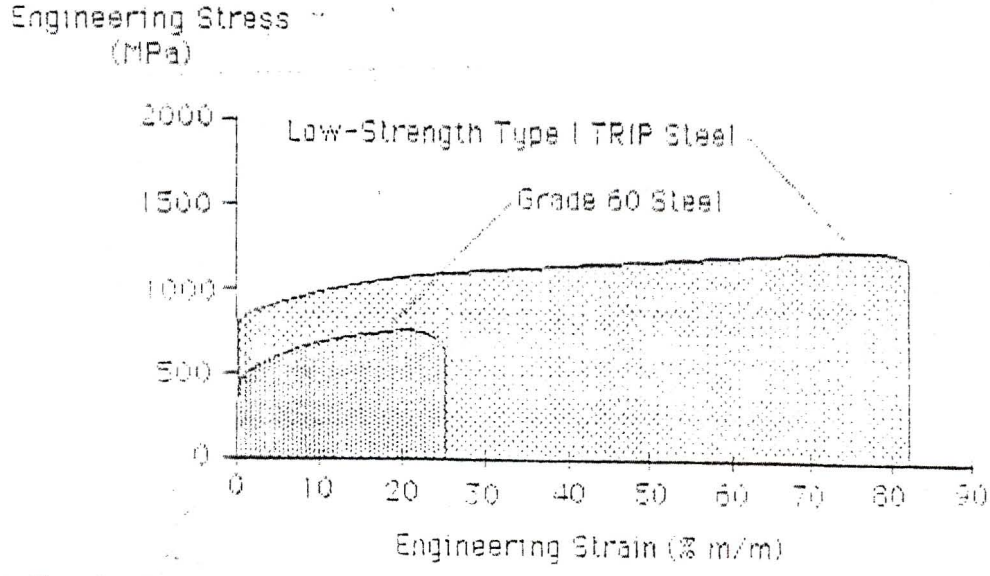


Fig. 5.45 - Stress-strain curves for structural grade steel and low-strength TRIP steel

Figure 5.43 compares the tensile properties of high-strength and low-strength variants of TRIP steels (referred to here as Type I and Type II respectively) with those of structural steels (Bridges, civil engineering applications) and high strength low alloy (HSLA) steels (aircraft landing gear, high stress applications). While the strongest TRIP steels compare favourably with HSLA steels with regards to strength, the ductility values and energy absorption capacities of the TRIP steels are higher than those offered by HSLA steels. Figure 5.44 compares representative stress-strain curves for TRIP steels as compared to HSLA and structural steels and figure 5.45 compares stress-strain curves for for low-strength TRIP steel and

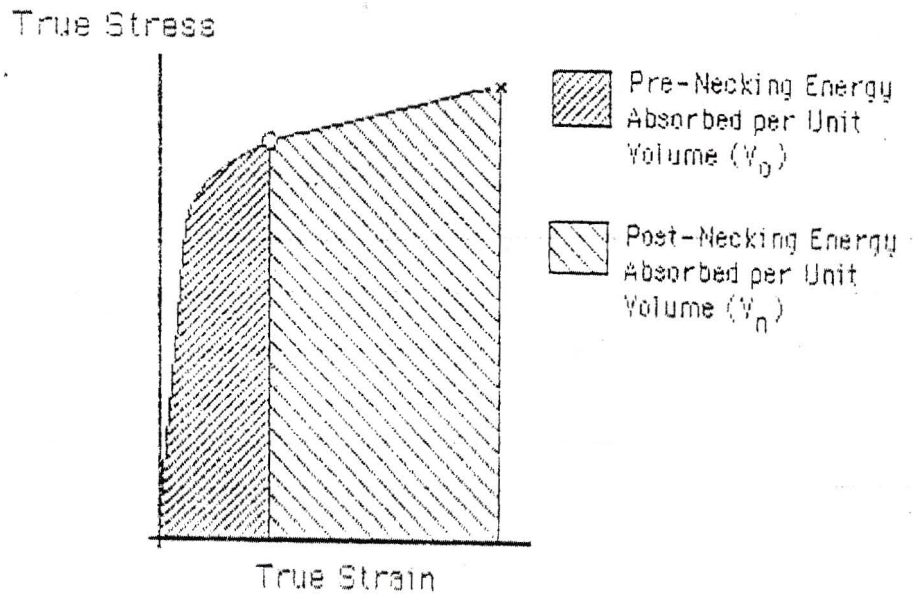
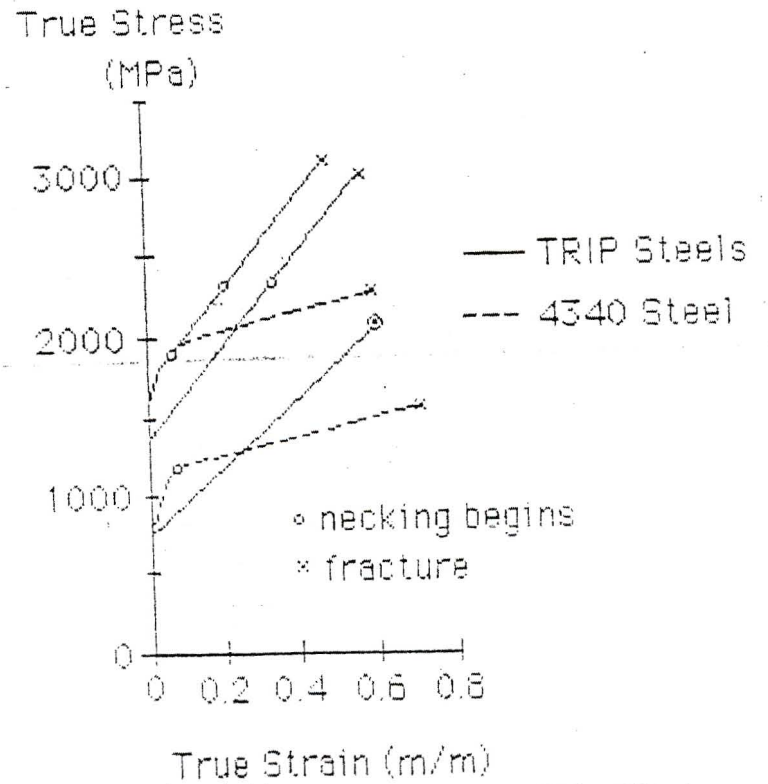


Fig. 5.46 – True-stress-strain curve divided into pre and post necking regimes

grade 60 structural steel. The high levels of toughness exhibited by TRIP steels are as a direct consequence of the phase transformation of the austenitic parent phase into the stronger martensitic phase. Although the Type II steels, which include the metastable Fe-Mn-Cr alloys are not apparently as strong as the Type I steels, it should be noted that these steels have not been evaluated in a condition other than the as-austenitised condition using the simplest heat treatment schemes. To obtain the highest yield strengths depicted in figure 5.43 for the Type I TRIP steels it is necessary to thermo-mechanically process the material using complicated and expensive schemes. The Type II materials were austenitised and either air-cooled or quenched to room temperature. The Type I TRIP steels have much lower yield strengths if not thermo-mechanically processed but still retain high levels of uniform elongation. Type I TRIP steels in this weaker



pe I

for

Fig. 5.47 – True-stress-strain curves for HSLA steel (AISI 4340) and Type I TRIP steel (different curves represent different processing conditions for materials)

condition will be referred to as Low Strength Type I TRIP steels. The lower yield strengths of the Type II TRIP steels, more typically of those in standard structural grade steels reflect the lack of thermo-mechanical processing. High strain hardening rates and elongations have been observed in both Type I and Type II steels with ultimate tensile strength ratios in the range of three to four being common.

The corrosion resistance of the Type II TRIP steels is superior to that of Type I materials due to higher levels of chromium added to their chemical formulations. In addition, both the Type I and Type II TRIP steels experience the strain induced phase transformations responsible for delaying the mechanical instability and necking in uniaxial tensile testing. The increased plastic deformation observed in these alloys corresponds to increased levels of energy absorption capacity.

A direct measurement of the energy absorption capacity of TRIP steels may be made by considering the stress-strain curves in order to calculate the energy absorbed during fracture, i.e.,

$$E_v = \int_0^{\varepsilon_f} \sigma(\varepsilon) d\varepsilon$$

where, Ev = energy absorbed per unit volume

ε = true strain

ε_f = true failure strain

$\sigma(\varepsilon)$ = true stress is expressed as a function of true strain

Where, $\sigma(\varepsilon)$ represents the true stress expressed as a function of true strain.

Energy absorption should be considered in terms of pre-neck and post-neck regimes as shown in figure 5.46. While the energy absorbed per unit volume

is generally greater after the neck forms, the actual volume of the material participating in the deformation process should be taken into account. TRIP steels generally display uniform elongation, i.e., non-localised deformation throughout the entire gauge length, and in so doing, the entire gauge length is contributing to energy absorption. If a mechanical instability initiates during plastic deformation any further straining is concentrated within the unstable region. This would lead to the formation of a true neck in conventional steel materials. The unstable region in a TRIP steel component however transforms from a weaker austenitic phase to a stronger martensitic phase resulting in the region becoming locally stronger than the surrounding material so that the deformation continues without formation of the true neck. This process is repeated throughout the deformation gauge section of the tensile specimen or deformation region of the component. Many TRIP steels therefore display little localised deformation prior to failure.

Figure 5.47 shows the true stress-strain curves for TRIP steel processed using three different treatments, the stress-strain curves for two different tempers of 4340 steel are also included. A circle on each of the respective curves shows the necking point for test. Any further deformation beyond the necking strain is considered to be locally concentrated within the necked region of the tensile specimen gauge length. At strains below the necking strain the entire volume (either gauge or component) absorbs deformation energy whilst beyond the necking strain limit, only the necking volume absorbs energy. The true strain at necking sometimes referred to as the strainhardening exponent n is therefore useful parameter in modelling the deformation response of a material. The fracture energy absorption capacity of a material is therefore directly related to the strain-hardening exponent. TRIP steels have a combination of high strength and ductility thus accounting for their increased fracture resistance.

The total fracture energy per unit volume (E_v) is the sum of the energy absorbed per unit volume prior to and after necking.

$$E_v = \int_0^{\varepsilon_{UTS}} \sigma(\varepsilon)d\varepsilon + \int_{\varepsilon_{UTS}}^{\varepsilon_f} \sigma(\varepsilon)d\varepsilon$$

where, ε_{UTS} =true strain at the ultimate tensile strength

- Other symbols are same as defined before

Where ε_{uts} represents the true strain at ultimate tensile strength (necking) and ε_f represents the true failure strain. The total energy absorption during fracture is obtained by multiplying the respective integrals by the corresponding volumes participating in the deformation before (V_0) and after (V_n) necking.

$$E = V_0 \int_0^{\varepsilon_{UTS}} \sigma(\varepsilon)d\varepsilon + V_n \int_{\varepsilon_{UTS}}^{\varepsilon_f} \sigma(\varepsilon)d\varepsilon$$

Table 5.9 compares energy absorption characteristics of the materials. The volume of the material associated with necking was assumed to be one-tenth of the total gauge length volume, i.e. $0.1V_0$. As mentioned above it is assumed that the entire gauge length volume absorbs energy during the uniform plastic deformation at true strains less than the strain hardening exponent n whereas only the necking volume is absorbing energy at true strain levels greater than n. The values contained in the table indicate that the three TRIP steel variants absorbed considerably more energy during fracture than the two variants of HSLA steels. Although both types of steels can be processed to achieve a range of stress strain behaviour. The effectiveness in delaying necking of the deforming gauge section is readily apparent in the levels of energy absorbed. One aspect evident in the data is that the lower strength TRIP steel variant, which has a true strain at necking equal to 0.60, absorbed considerably more energy than the high-strength steels including the high strength TRIP steels. These estimates of energy absorption indicate the enhanced structural safety that could be

realized through selective utilization of TRIP steels as structural components. The additional energy absorption would be particularly beneficial in structures subjected to cyclic deformation where the energy absorption would delay structural failure thereby enhancing structural safety.

Table 5.9 – Energy absorption characteristics of TRIP and HSLA steels

Steel	σ_{yield} [MPa]	n	E_V ($\varepsilon < n$) [MJ/m^3]	E_V ($\varepsilon > n$) [MJ/m^3]	E_{total} [MJ/m^3]
TRIP	760	0.60	858	-	858
TRIP	1400	0.35	433	56	489
TRIP	1600	0.25	395	104	499
4340	800	0.10	100	85	185
4340	1600	0.08	138	104	242

CONCLUSION

The foregoing discussions have touched on a variety of important topics which are an important necessity in the understanding of the smart bolt technology. At the time of compiling this thesis the project was in its infancy. A lot of aspects of the project are still under development, and a lot of good ideas and samples are developed from time to time. This thesis covers a lot of basic aspects which are under content review. A lot of metallurgical contribution from this thesis will certainly be useful. Also the optimum bolt suggested in the theses will also certainly contribute to the construction of the ultimate bolt diameter based solely on optimum mechanical properties, taking into account all failure modes. Also, this thesis, and all the reading undertaken cements a good foundation for the successful continuation if the student continues further with the research at hand. The finite element analysis undertaken was of great benefit both to the project development and the writer. The magnetic aspects investigated were useful in that some understanding was also generated at micro level. Also the Curie temperature examined could explain why at certain temperatures austenite may not transform into magnetic martensite, however this idea was not developed to an unequivocal conclusion in the thesis, some further research is needed on this aspect. Also the idea of stacking faults contribution to martensite development is there only as inferences, but not conclusively developed. A limited amount of research on this aspect is also evident. More literature must be reviewed and more experiments need to be carried out. However, there also seems to be an inference that the elements added to develop a TRIP steel successfully lower the stacking fault energy, thus increasing the propensity of faulting in FCC structures. And in turn this leads to much needed preponderance of HCP structure which is a transitional condition for the development of BCC magnetic martensite. As mentioned before, these ideas are more inferred than conclusively confirmed through experiment.

REFERRENCES

1) Avallone, A.E ; Baumeister ,T. : *MARK'S standard handbook for mechanical Engineers* (Theodore Baumeister III editor, 9th edition – 1994)
2) Juvinall, R.C.; Marshek, K,M : *Fundamentals of machine component Design* (John Wiley & Sons, New York – 1991)
3) Juvinall, R.C.: *Stress, Strain and Strength* (McGraw- Hill Book company, New York – 1967)
4) Phelan, R.M. :*Funamentals of mechanical design* (McGraw – Hill, New York, 1970)
5) Shigley, J.E.: *Mechanical Engineering design – Second edition* (Mcgraw-Hill Kogakusha, Ltd, London - 1972)
6) Bickford,W.B : *Finite element method* (Richard D. Irwin, Inc., Illinois, 1994)
7) Grandin, H Jnr : *Fundamentals of the finite element model* (Macmillan Publishing Company – 1986)
8) Owen, D.R.J ; Hinton,E : *A simple guide to finite elements* (Swansea: Pineridge press, 1990)
9) Rocky, K.C; Evans,H.R;Griffiths,D.W;Nethercot,D,A : *The finite element method* (Granada, London, 1983)
10) Davis,A.J : *The finite element method (Oxford, Clarendon press, 1980)*
11) Ross, C.T.F : *Finite element methods in structural mechanics* (Chichester, Ellis Horwood, 1985)
12) Owen,D.R.J ; Hinton,E : *Finite element programming* (Academic press, London, 1977)
13) Norrie, D.H.; Devries G: *An introduction to finite element analysis* (Academic press, New York, 1978)
14) Lovel, M.C.; Avery, A.J.; Vernon,M.W.: *Physical properties of materials* (Van Nostrand Reinhold, New York, 1975)
15) Burke, H. E: *Handbook of magnetic phenomena* (Van Nostrand Reinhold, New York, 1986)

16) Tebble, R.S; Craik, D, J: *Magnetic materials* (Wiley – Interscience, london, 1969)

17) Peckner, D; Bernstein, I, M: *Handbook of stainless steels* (McGraw – Hill, New york, 1977)

18) Jastrzebski, Z.D: *Nature & Properties of engineering materials* (Wiley international edition, John Wiley & Sons, Inc., New York/London –1959)

19) Anderson, J.C; Leaver, K.D; Rawlings, R.D; Alexander, J.M.: *Material science, Third edition*(Chapman & Hall, London – 1990)

20) Chikazumi, S; Charap, S.H.: *Physics of magnetism* (John Wiley & Sons Inc., New York, 1964)

21) Olsen, E: *Applied magnetism, a study in quantities* (Granada Publishing Limited, London – 1975)

22) Bhandarkar D., Zackay V.F., Parker E.R.: *Stability and mechanical properties of some Austenitic steels* (Metallurgical Transactions, vol. 3, 1972, pp2619 – 2631)

23) Lesroisey F., Pineau A.: Martensitic *Transformation induced by Plastic Deformation in the Fe-Ni-Cr-C system* (Metallurgical Transactions, 1972, pp 387-396)

24) Angel T.: *Formation of martensite in stainless steels. Effects of deformation, temperature, and composition* (Journal of the Iron and Steel Institute, May 1954, pp 165-174)

25) Fiedler H.C., Averbach B.L., Cohen M.: *The effect of deformation on the martensitic transformation in austenitic stainless steels* (Transactions of the ASM, vol. 47, 1955, pp 267 285)