SMART STRUCTURAL HEALTH MONITORING OF MINING SUPPORT UNITS

JASON APSEY

2003

DURBAN, SOUTH AFRICA

Submitted in fulfilment of the academic requirements for the degree of Doctor of Philosophy in the Department of Mechanical Engineering, University of Natal.

ABSTRACT

In the South African mining industry, the design of tunnel support systems is generally based on empirical methodologies that consider rockmass characteristics as well as the type of loading (e.g. seismic) that the excavation experiences. The design methodologies are by no means infallible, and work is continually being conducted to improve the classification of excavation conditions and thereby improve the selection of a suitable support system.

This study is concerned with finding a means to monitor the installed support units rather than with improving the classification methodologies. It is postulated that with the extraction of accurate information describing the state of any support unit at any given time, areas of instability in the tunnel can be readily identified and strengthened. Also, the information gathered as to the behaviour of the support units in a particular region can be used to assist in understanding the environmental characteristics of that region (rockmass, loading, etc.).

A material survey was conducted to identify suitable candidates that could feasibly be used in either a passive (feedback when interrogated) or active (constant feedback) structural health monitoring system. The preferred candidates identified in this study are the group of passive smart materials referred to as TRIP steels, which are a subset of strain memory alloys. TRIP steels exhibit microstructural changes from paramagnetic austenite to ferromagnetic martensite as a function of increasing deformation at a given temperature. The strength of the magnetic field at critical locations provides an indication as to the health state of the component. Because of their high strengths and ductility, TRIP steels can be used as what amounts to a self-monitoring support unit (interrogation apparatus required).

Finite element methods are a practical means of predicting the mechanical and magnetostatic behaviour of TRIP steel structural members once material equations have been established by experiment.

PREFACE

I declare that this dissertation is my own unaided work, except where due acknowledgement is made to others, and is being submitted for the Degree of Doctor of Philosophy to the University of Natal, Durban, having not been submitted previously for any other degree or examination.

Jason Apsey

September, 2003

ACKNOWLEDGEMENTS

A big thank you to Professor Victor Verijenko, for the assistance and the motivation that went beyond all duties or expectations. Thank you also to my parents, who only ever wanted the best for me, and continue to make sure I never stop working for it.

TABLE OF CONTENTS

ABSTRACT II

PREFACE III

ACKNOWLEDGEMENTS IV

TABLE OF CONTENTS V

LIST OF TABLES X

LIST OF FIGURES XI

LIST OF SYMBOLS AND ABBREVIATIONS XV

1. INTRODUCTION 1

1.1. OVERVIEW 1

1.2. PARAMETERS OF THE INVESTIGATION 3

2. CURRENT MINE TUNNEL SUPPORT SYSTEM DESIGN METHODOLOGIES 4

2.1. OVERVIEW 4

2.2. TYPICAL ENVIRONMENTAL CONDITIONS 7

2.3. CONTROLLING HIGHLY STRESSED TUNNEL BEHAVIOUR 7

2.4. DESIGN OF TUNNEL SUPPORT SYSTEMS 8

2.5. ANALYSIS OF STRUCTURAL STABILITY 9

2.6. EMPIRICAL DESIGN METHODOLOGIES 16

2.6.1. Local Estimations for Highly Stressed Environments 17

2.6.2. The NGI Q System 17

2.6.3. The Rockwall Condition Factor Classification System 22

2.6.4. The Geomechanics Classification System 24

2.6.5. The Modified Rock Mass Rating System 28

2.6.6. U.S. Corps of Engineers 32

2.6.7. Farmer and Shelton 32

2.6.8. Guidelines for Deep Mining 33

2.7. SUMMARY 36

3. STRUCTURAL HEALTH MONITORING 37

3.1. OVERVIEW 37

3.2. TRADITIONAL SHM SYSTEMS 37

3.3. NON-DESTRUCTIVE TESTING 38

3.4. SMART STRUCTURAL HEALTH MONITORING VS. NON-DESTRUCTIVE TESTING 39

3.5. SMART MATERIALS AND STRUCTURES FOR SHM 40

3.6. SMART SHM TECHNOLOGY 42

3.6.1. Passive Systems 43

3.6.2. Active Systems 44

3.6.3. Passive vs. Active SHM 45

3.6.4. Selecting a Suitable Smart Sensor/Material *45*
3.6.5. Sensor Arrays *46*
3.6.6. Signal Processing *46*
3.6.7. Damage Identification and Interpretation *46*
3.6.8. System Integration *47*
3.7. INTRODUCING SMART SHM SYSTEMS INTO NEW ENVIRONMENTS 47
3.8. MINIMISING STRUCTURAL INSPECTION AND REPAIR COSTS 48
3.9. SUMMARY 49
4. STEEL MICROSTRUCTURES 50
4.1. BASIC STEEL PROPERTIES 50
4.2. CRYSTALLINE PHASES AND PHASE EQUILIBRIA 51
4.2.1. Overview of Unit Cell Structures and Terminology *51*
4.2.2. Grains (Crystals) and Phases *54*
4.2.3. Solutions and Mixtures *54*
4.2.4. Imperfections in Crystalline Solids *58*
4.2.5. Nucleation *58*
4.2.6. Homogenisation *59*
4.3. STRENGTHENING STEEL 59
4.3.1. Work Hardening *59*
4.3.2. Interstitial Solid Solution Strengthening *60*
4.3.3. Substitutional Solid Solution Strengthening *62*
4.3.4. Effect of Grain Size on Strength *62*
4.3.5. Dispersion Strengthening *64*
4.4. THE CARBON-STEEL PHASE DIAGRAM 64
5. KINETICS OF THERMAL PHASE TRANSFORMATIONS 72
5.1. HEAT TREATMENTS 72
5.2. TIME-TEMPERATURE-TRANSFORMATION DIAGRAMS 72
6. AUSTENITE – MARTENSITE TRANSFORMATIONS OF TRIP STEELS 75
6.1. THE MARTENSITIC REACTION 75
6.2. THE TRANSFORMATION OF METASTABLE AUSTENITE 75
6.2.1. Temperature-Controlled Martensitic Reaction *75*
6.2.2. Strain-Controlled Martensitic Reaction *76*
6.2.3. TRIP Steels *77*
6.2.4. Nucleation and Growth of Martensite *78*
6.2.5. Effect of Carbon *80*
7. ALLOYING ELEMENTS TO DEVELOP METASTABLE AUSTENITE 81
7.1. AUSTENITE-STABILISING ALLOYING ELEMENTS (OPEN γ-FIELD) 81
7.2. ELEMENTS THAT ASSIST HEAT TREATMENT (EXPANDED γ-FIELD) 83

7.3. ELEMENTS THAT CLOSE THE γ-FIELD 85
7.4. ELEMENTS THAT CONTRACT THE γ-FIELD 86
7.5. COMBINATIONS OF ELEMENTS 86
7.6. PROCESSING METASTABLE AUSTENITIC STEELS 89
7.6.1. Time-Temperature Transformations and Work-Hardenability 90
7.6.2. Material Processing to Improve Strength 91
7.6.3. Effects of Composition and Thermal Process Temperatures on TRIP Steels 93
8. TRIP STEELS FOR SMART SHM MINING BOLTS 111
8.1. TRIP STEEL FULFILMENT OF SHM REQUIREMENTS 111
8.1.1. Criteria for Diagnostic Data 111
8.1.2. Integration Parameters 111
8.1.3. The Proposal 112
8.2. TRANSFORMATION-INDUCED PLASTICITY 112
8.2.1. Early Research Areas 112
8.2.2. Magnetic Properties of TRIP Steels as the Basis for a Smart SHM System 114
8.2.3. Physical Properties of TRIP Steels 114
8.3. TRIP STEELS AS SMART MATERIALS 121
8.4. FEASIBILITY OF METASTABLE ALLOYS AS DEFORMATION-SENSING MATERIALS 121
8.4.1. Characteristics of Metastable Alloys 121
8.5. FACTORS INFLUENCING THE FERROMAGNETIC RESPONSE 126
8.5.1. Inhomogeneous Deformation During Initial Phase Transformation 126
8.5.2. Temperature Dependence of the Transformation Process 126
8.5.3. The Effect of Strain Rate on Transformation Behaviour 129
8.5.4. Error Related to TRIP SHM Systems 130
8.6. ECONOMIC CONSIDERATIONS 130
8.7. CONCLUSIONS 131
9. FEA OF STRUCTURAL DEFORMATION AND MAGNETOSTATIC FIELDS 133
9.1. BROAD OVERVIEW OF THE FEA PROCESS 133
9.2. BENEFITS OF FEA 133
9.3. VARIATIONAL TECHNIQUES FOR SOLVING FUNCTIONALS 134
9.3.1. Analytical Formulations 134
9.3.2. Variational Techniques for Solving Numerical Equations 139
9.4. SOLVING FOR PHYSICAL LOAD CONDITIONS 140
9.4.1. Force Energy Potential 140
9.4.2. Strain Energy 142
9.4.3. Total Potential Energy 147
9.4.4. Generalised FEA Workflow for Physical Load Conditions 147
9.5. FEA FOR MAGNETOSTATIC MODELLING 169

9.5.1. Magnetic Field Equations *169*
9.5.2. FEA of Magnetostatic Fields *187*
10. DISCUSSION **192**
11. CONCLUSION **197**
12. APPENDIX A: SMART MATERIAL SURVEY **198**
12.1. PIEZOELECTRICITY 198
12.1.1. Piezoelectric Behaviour *198*
12.1.2. History *199*
12.1.3. Piezoceramics *200*
12.1.4. Manufacture For Specific Response *200*
12.1.5. Common Piezoelectric Materials *214*
12.1.6. Piezoelectricity and Smart SHM *214*
12.2. ELECTROSTRICTION 215
12.2.1. Electrostrictive Materials *215*
12.2.2. Electrostrictive vs. Piezoelectric Devices *216*
12.2.3. Measurement Techniques *220*
12.2.4. Lead Magnesium Niobate, PMN *220*
12.2.5. Conclusion *220*
12.3. MAGNETOSTRICTION 221
12.3.1. Definition *221*
12.3.2. Terfenol *221*
12.3.3. Applications *223*
12.3.4. Conclusion *223*
12.4. ELECTRORHEOLOGICAL AND MAGNETORHEOLOGICAL FLUIDS 224
12.4.1. History *224*
12.4.2. Electrorheology (ER) *224*
12.4.3. Magnetorheology (MR) *227*
12.4.4. Applications *227*
12.4.5. Measurement/Monitoring Systems *227*
12.4.6. Conclusion *228*
12.5. SHAPE MEMORY ALLOYS 228
12.5.1. Shape Memory and Crystallography *228*
12.5.2. Materials That Exhibit Shape Memory *229*
12.5.3. Commercial Availability *231*
12.5.4. Load - Deformation Diagrams *232*
12.5.5. Superelasticity *232*
12.5.6. Applications *233*
12.5.7. Measuring/Monitoring Systems *234*

12.5.8. Nitinol 237
12.6. ACTIVE DAMAGE CONTROL OF HYBRID MATERIAL SYSTEMS 240
13. APPENDIX B: HEAT TREATMENT PROCESSES FOR CARBON STEELS 243
13.2. QUENCHING 245
13.3. ANNEALING 245
13.4. NORMALISING 246
13.5. STRESS RELIEVING 246
13.6. TEMPERING 247
13.7. THROUGH HARDENING 247
13.8. CASEHARDENING 249
13.9. SURFACE HARDENING 250
13.10. FORGING 251
13.11. PRE-HEATING FOR WELDING 251
13.12. MELTING 251
14. REFERENCES 252

LIST OF TABLES

Table 2.1: Parameters for rockmass classification according to the Q system. 18

Table 2.2: Recommended ESR values for excavation safety level 21

Table 2.3: Estimations of support pressure, length and spacing of primary support systems based on the Q system. 22

Table 2.4: Support recommendation for good ground conditions (RCF < 0.7) 23

Table 2.5: Support recommendation for average ground conditions (0.7 < RCF < 1.4). 23

Table 2.6: Support recommendation for poor ground conditions (RCF > 1.4) 24

Table 2.7: Geomechanics Classification system 25

Table 2.8: Grading of joint orientations applicable to tunnels 25

Table 2.9: Adjustments to RMR for varying joint orientations. 26

Table 2.10: Rock mass class and description as determined from RMR. 26

Table 2.11: Rockmass parameters based on RMR. 26

Table 2.12: Excavation technique and support selection based on RMR. 27

Table 2.13: Percentage adjustment to joint spacing rating for joint number and number of non-vertical excavation rockwalls 29

Table 2.14: Support system guide based on Modified Geomechanics Classification. 30

Table 2.15: Tunnel support guidelines from the US Corps of Engineers. 32

Table 2.16: Empirical guidelines for excavations <15m span. 33

Table 3.1: Examples of passive sensors. 43

Table 3.2: Examples of active sensors 44

Table 3.3: Comparison of passive and active system characteristics. 45

Table 4.1: Crystal systems. 51

Table 4.2: Atomic sizes of non-metallic elements in iron. 55

Table 4.3: Solubility limits of carbon and nitrogen in γ- and α-iron. 55

Table 4.4: Constants for diffusivity calculations. 57

Table 7.1: Alloys used in the nickel-with-nitrogen partial replacement experiment. 84

Table 7.2: Percentage austenite retained at room temperature for the nickel-with-nitrogen partial replacement experiment. 84

Table 7.3: Comparison of thermal strengthening processes. 93

Table 7.4: Chemical compositions of examined steels in weight %. 93

Table 8.1: Energy Absorption Characteristics for TRIP Steels and HSLA 4340 Steel; $V_n = 0.1*V_0$. 119

Table 9.1: Maxwell's equations. 186

Table 9.2: Maxwell's equations for magnetostatic fields. 186

Table 12.1: Material properties 210

Table 12.2: Nitinol properties 239

LIST OF FIGURES

Figure 2.1: Schematic of the principal features of a typical deep level South African gold mine 6

Figure 2.2: Schematic of typical components of a tunnel and its reinforcement/support system 6

Figure 2.3: Simple support structure for a falling block (Stillborg, 1986) 9

Figure 2.4: Support of a sliding block (Stillborg, 1986). 10

Figure 2.5: Suspension of laminated rockmass based on tributary area loading (Stillborg, 1986). 10

Figure 2.6: Formation of unstable zone and natural stable arch within moderately jointed rockmass structures (Stillborg, 1986) 12

Figure 2.7: Creation of a reinforced beam within a stratified mass. 14

Figure 2.8: Creation of a reinforced rock mass arch within a highly discontinuous rock mass structure. 15

Figure 2.9: Semi-empirical instability extent after Stimpson (1989) [25] and based on the RMR guidelines. 28

Figure 2.10: Approach to support design under rockburst conditions after Kaiser et al. (1996) [29]. 34

Figure 2.11: Rockburst damage mechanisms as defined by Kaiser et al. (1996) 35

Figure 2.12: Primary functions of support elements. 35

Figure 4.1: Unit shell shapes 51

Figure 4.2: Thermal expansion of iron 52

Figure 4.3: Typical stress-strain curves for mild steel at elevated temperatures. 61

Figure 4.4: Dependence of the lower yield stress of mild steel on grain size. 63

Figure 4.5: Iron-carbon equilibrium phase diagram. 65

Figure 4.6: Invariant Reactions 66

Figure 5.1: The transformation process. The stable phases of the phase diagrams (right) are not obtained immediately, but require the times shown on the left. 74

Figure 7.1: Open γ-field phase diagram characteristics 82

Figure 7.2: Effect of alloying elements on the austenitic phase boundaries of the Fe-C phase diagram. 87

Figure 7.3: The Schaeffler diagram for prediction of phases present in complex alloying elements at room temperature. 88

Figure 7.4: Commercially available steel alloys. 88

Figure 7.5: The region of chromium and nickel equivalence most likely to produce strain memory alloys 89

Figure 7.6: An example of a metastable austenitic bay engineered into the TTT curve through the use of suitable alloying chemistry. 91

Figure 7.7: Hardness effects of thermal processing 92

Figure 7.8: Mechanical properties as a function of annealing temperature TA1. Annealing performed for 51 seconds, followed by austempering at 450°C for 300 seconds. 95

Figure 7.9: Transmission electron micrograph showing niobium carbide precipitation in steel S-Nb. 96

Figure 7.10: Scanning Electron Micrographs of cold rolled steels (a) S, (b) S-Mn, (c) S-Nb, (d) S-Si annealed at 830°C for 51 seconds, followed by austempering at 450°C for 300 seconds. 96

Figure 7.11: Mechanical properties as a function of austempering start temperature. Annealing performed at 830°C for 51 seconds, followed by austempering at various temperatures for 300 seconds. 97

Figure 7.12: Volume percent of austenite as a function of annealing time at various intercritical annealing temperatures for sample S-Mn. 98

Figure 7.13: Scanning electron micrograph of a cold rolled sheet of steel S-Mn. 98

Figure 7.14: Scanning electron micrographs of the cold rolled sheet of steel S-Mn heated at a rate of 6°C per second and held at 740°C for (a) 0 s, (b) 10 s and (c) 100s 99

Figure 7.15: Volume fraction of bainite, ΔV_B, transformed during austempering for steel S-Mn as a function of the austempering time at various temperatures. 100

Figure 7.16: Volume fraction of bainite, ΔV_B, transformed during austempering, ΔV_B, for the sample steels as a function of the austempering temperature for 300 seconds. 101

Figure 7.17: Martensite start temperature, Ms, as a function of the volume fraction of bainite transformed, ΔV_B, at various austempering temperatures (Sample S-Mn). 102

Figure 7.18: Volume percentage of retained austenite for sample S-Mn as a function of the volume fraction of bainite transformed during austempering, ΔV_B, at various temperatures (400°C, 450°C and 500°C). 102

Figure 7.19: Martensite start temperature for different steel compositions, M_s, as a function of the volume fraction of bainite transformed during austempering, ΔV_B, at 450°C. 103

Figure 7.20: Martensite start temperature, Ms, for different steel compositions as a function of the austempering time at 450°C. 104

Figure 7.21: Volume percentage of retained austenite as a function of the martensite start temperature, Ms, for different samples austempered at 450°C for 0 to 900 seconds. 105

Figure 7.22: Martensite start temperature, M_s, for the steels heated to form 28% and 100% austenite, as a function of the cooling rate. 106

Figure 7.23: Scanning electron micrographs of the steels heated to form 100% austenite, followed by cooling at various cooling rates. 107

Figure 7.24: Scanning electron micrographs of the steels heated to form 100% austenite, followed by cooling at various cooling rates. 108

Figure 8.1: A Comparison of the Tensile Properties of Structural Steels. 115

Figure 8.2: Engineering Stress-Strain Curves for Structural Materials. 116

Figure 8.3: Stress-Strain Curves for a Structural Grade Steel and a Low-Strength TRIP Steel. 116

Figure 8.4: A True Stress - Strain Curve Divided into Pre and Post-Necking Regions. 117

Figure 8.5: True Stress-Strain Curves for a HSLA Steel (AISI 4340) and a Type I TRIP Steel (different curves represent different processing conditions for the materials). 118

Figure 8.6: High cycle fatigue properties of a TRIP steel and high performance structural steels. 120

Figure 8.7: Generic strain-induced martensitic (α') phase transformation curve. 122

Figure 8.8: Stress-strain curves (top) and Hall output-strain curves (bottom) for Fe-8.4 Cr-8.4 Ni-2.1 Mn-0.262 C TRIP sensor material 124

Figure 8.9: Schematic of a TRIP element /transducer assembly and a structurally-attachable tensile strain monitor 125

Figure 8.10: Hall unit output as a function of displacement for a candidate sensor material. 127

Figure 8.11: Idealised behaviour of an installed sensor element. 127

Figure 8.12: Hall Unit Output as a Function of Displacement for Different Strain Rates in Carbon-Free Alloys. 130

Figure 9.1: Variation of extremising function $\phi(x)$ within prescribed limits, and the effect on $I(\phi)$ 134

Figure 9.2: Dependence of function $\phi(x)$ on parameter $\in$ 135

Figure 9.3: Variation of I with $\in$ 137

Figure 9.4: Uniaxial stress and strain system 144

Figure 9.5: Variation of force F with time 145

Figure 9.6: Variation of force F with displacement u 146

Figure 9.7: Global and natural local coordinate systems for triangular axisymmetric elements. 149

Figure 9.8: Calculating the area of a triangular element. 156

Figure 9.9: Element nodal force components on an 8-node brick element. 159

Figure 9.10: Distributed element edge load applied to an 8-node brick element. 160

Figure 9.11: Distributed surface load applied to an 8-node brick element. 162

Figure 9.12: Body force direction and magnitude for an 8-node brick element. 165

Figure 9.13: Magnetic field lines $\overline{B}$ surrounding a permanent magnet in the N-S direction 170

Figure 9.14: Calculating $\overline{E}$ due to Q_1 at location $\overline{r}$ in a global coordinate system 172

Figure 9.15: Charge Q in a spherical coordinate system 173

Figure 9.16: Cubic volume Δv in a vector field $\overline{D}$ 176

Figure 9.17: Magnetic field vectors due to a moving positive charge q. 181

Figure 9.18: Magnetic field vectors due to a current element. 182

Figure 9.19: Line integral of $\overline{B}$ around a closed path surrounding a long, straight conductor carrying current I 183

Figure 12.1: The piezoelectric effect in a cylinder of PZT material 198

Figure 12.2: The orthogonal co-ordinate system. 201

Figure 12.3: Shear deformation 202

Figure 12.4: Motor Transducer Relationships. 207
Figure 12.5: Generator transducer relationships. 208
Figure 12.6: Typical Thermal Properties 209
Figure 12.7: Expansion vs. Electric field 216
Figure 12.8: Expansion vs. temperature of PZT and PMN actuators. 217
Figure 12.9: Hysteresis vs. temperature of PZT and PMN actuators. 218
Figure 12.10: Actuator creep of PZT and PMN actuators 219
Figure 12.11: Electrostriction. 219
Figure 12.12: Viscosity changes 225
Figure 12.13: ER mechanism 226
Figure 12.14: Shape memory 229
Figure 12.15: Deformation-Temperature diagram 230
Figure 12.16: Crystallography 231
Figure 12.17: Load-Deformation diagrams 232
Figure 12.18: Superelasticity. 233
Figure 12.19: Braces make use of pseudoelasticity. 234
Figure 12.20: DSC test results 235
Figure 12.21: Constant load diagram 236
Figure 12.22: Active A_f test 237
Figure 13.1: Iron-carbon phase diagram indicating heat treatment zones. 244
Figure 13.2: TTT diagram showing 4 typical through-hardening treatments. 248

LIST OF SYMBOLS AND ABBREVIATIONS

α	Low temperature ferrite
δ	High temperature ferrite
δI	Variation of functional $I(\phi)$
$\in$	Parameter controlling the magnitude of variation
$\eta(x)$	Any continuously differentiable function of x with specific endpoints
σ	Conductivity
σ_0	Friction stress; constant used in Hall-Petch relationship
σ_1	Maximum principal stress
σ_3	Minimum principal stress
σ_c	Uniaxial compressive strength
σ_Y	Yield strength
$\delta\phi$	Variation of a function $\phi(x)$
φ	Angle
ξ	Angle
θ	Angle
ϕ	Angle of rockmass internal friction
ρ	Mass density
ρ	Charge density
ε	Martensite
ε	Permittivity
ε	Strain
γ	Austenite
γ	Shear strain
τ	Shear stress
μ	Permeability
ω	Frequency
Ω	Volume
Φ	Magnetic flux
Π	Total potential energy
Γ	Flux density vector
ψ	Electric flux
$\overline{\nabla}$	Differential operator
Δ	Variational field variable
$\{\Delta\}$	Displacement
a	Local system coordinate
a	Empirical rockbolt design methodology factor

a	Acceleration
A	Area
A_f	Austenitic transformation finish temperature
A_s	Austenitic transformation start temperature
b	Local system coordinate
$\bar{B}$	Body force
$\bar{B}$	Magnetic flux density
[B]	Transition matrix
BCC	Body-centred cubic
BCT	Body-centred tetragonal
c	Apparent rockmass cohesion
c	Local system coordinate
C	Carbon or carbide
C	Capacitance
[C]	Stress-strain constitutive matrix
d_{ij}	Piezoelectric strain constants
D	Height of rockmass de-stressed zone
D	Number of rockbolts per square metre
D	Atomic diffusivity
$\bar{D}$	Electric flux density
D_0	Frequency factor (relates to atomic diffusivity)
DSC	Differential scanning calorimetry
E	Rockbolt energy absorption requirement per unit area
E	Activation energy for atomic diffusion
$\bar{E}$	Electric field
ESR	Excavation Support Ratio
F	Force
F(b)	Blocked force of a piezoelectric actuator
F_c	Conservative force
FCC	Face-centred cubic
g	Gravitational acceleration
g_{ij}	Piezoelectric voltage constants
$\bar{H}$	Magnetic field strength
HCP	Hexagonal close-packed
HSLA	High Strength Low Alloy
I	Current
$I(\phi)$	Functional

$I(\tilde{\phi})$	Stationary functional
I-T	Isothermal transformation
$\|J\|$	Jacobian determinant
$\bar{J}$	Current density
Ja	Joint alteration number
Jn	Joint set number
Jr	Joint roughness number
Jw	Joint water reduction factor
[K]	Stiffness matrix
k	Boltzmann constant
[k]	Element stiffness matrix
k_{ij}, k_p	Electromechanical coupling coefficients
k_Y	Constant used in Hall-Petch relationship
K_{ii}	Relative diectric constants
l	Length
L	Length
$L(\phi)$	General functional
L_M	Mean chord length
m	Mass
M	Martensite
M_D	The temperature above which the formation of martensite due to strain ceases to occur
M_f	Martensitic transformation finish time
M_s	Martensitic transformation start time
MRMR	Modified Rock Mass Rating
n	Number of dislocations in a grain
[N]	Shape function
NDT	Non-destructive testing
PDA	Prior deformation of austenite
p	shear perimeter of reinforced rock unit
P_L	The reciprocal of the number of intersection points of a unit-length line with grain boundaries
$P_n(x,y,z)$	n-th order polynomial
$\bar{P}$	Magnetic flux density potential
q	Electric charge
{q}	Element nodal displacements
Q	Rating based on the NGI Q system classification

Q	Electric charge
Q	Factor of amplitude increase at resonance
{Q}	FEA load vector
Q_m	Equivalent circuit ratio of reactance to resistance in a mechanical vibrating piezoelectric system
r	Distance
R	Distance
R	Shear radius of the reinforced rock unit
R	Resistance
RCF	Rockwall condition factor
RMR	Rock mass rating
RQD	Rock quality designation
s	Rockbolt spacing
S	Driving energy for a magnetic field
SC	Simple cubic
SHM	Structural health monitoring
SRF	Stress Reduction Factor
S_V	Number of grain boundaries per unit volume
TRIP	Transformation-induced plasticity
t	Time
t	Thickness
t_f	Transformation finish time
t_s	Transformation start time
T	Temperature
$\overline{T}$	Traction vector
TTR	Transformation temperature range
TTT	Time-temperature-transformation
u	Displacement
U	Strain energy
UTS	Ultimate tensile strength
v	Velocity
v	Displacement
v	Volume
V	Force energy potential
V	Electric potential difference
w	Excavation span
w	Displacement
w_m	Maximum span of an unsupported opening

W	Rockmass load acting on a rockbolted area
W	Work field
W	Stored magnetic field energy
W_c	Work due to a conservative force
X(f)	Free deflection of a piezoelectric actuator
Y^D_{ii}	Young's modulus for an open-circuited piezoelectric material
Y^E_{ii}	Young's modulus for a short-circuited piezoelectric material

1. INTRODUCTION

The advent of smart structural health monitoring (SHM) technology has the potential to impact enormously on economic and safety concerns in mechanical and structural engineering industries. The integrity of previously inaccessible structural members can now be monitored, so that structural components approaching the end of their working life can be identified before potentially catastrophic failure occurs. Lifespan can also be extended through accurate damage accumulation assessment. Costs associated with structural failure and removal of components from service for inspection (including risk of damage) can be minimised though the use of integrated smart damage sensing and assessing systems.

The aims of this thesis are:

1. to identify a suitable monitoring system capable of providing information, when required, as to the deformed/stressed state of mine tunnel rockbolts, without having to remove them from service.

2. to identify suitable numerical techniques to simulate and predict the behaviour of such a system under various loading conditions. In-situ measurements may be compared to simulated scenarios to better establish the local rockmass movements/conditions.

1.1. OVERVIEW

At this time, the design of support systems within the South African mining industry is generally empirically or semi-empirically based [1]. In relatively deep, highly stressed rockmass environments, reference is generally made to empirical tables in an industry guide. In these cases the Rockwall Condition Factor (RCF) is a parameter that represents the condition of the local rockmass. Alternatively, support system design may be based on energy absorption criteria experienced during rockbursting. Test rigs [2] and numerical models defined using e.g. finite element methods [3] are other means of simulating the support system under prescribed loading conditions.

A suitable structural health monitoring system should have the inherent capacity to provide information as to the deformed state of a rockbolt unit at any given time in order for the responsible authority to determine whether the operational performance of the support system has been compromised. The monitoring

TRIP steels are strain-measuring alloys that have been identified in this study as candidates for smart mining bolt SHM. This very broad range of materials potentially provides the basis for a monitoring system for numerous applications across a wide spectrum of industries.

1.2. PARAMETERS OF THE INVESTIGATION

The objective of this investigation is not to develop new design methodologies, but rather to identify a potential structural health monitoring system to supplement the established design methodologies. This will assist inspectors to monitor and quantify the extent of any damage to tunnel support units, and thereby determine the integrity of the support system. It is left to future research to either integrate the results of any SHM systems into the existing design methodologies, or to develop new ones.

A further objective is to identify suitable methods for numerical simulation of the SHM system for test purposes, and to describe any mathematical formulations. Once again it is left to future research to produce a physical specimen, measure and calibrate any smart material properties, and incorporate these into a representative numerical model for simulation.

instrumentation at an underground site and be guaranteed of capturing the response of the support system. Detailed observations of numerous rock movement and rockburst incidents have therefore been used to form a database of the *in-situ* performance of different support systems within different rock mass environments. Because of the inherent design complexities associated with working within a highly variable medium such as rock mass, this observational approach currently forms the basis of the majority of practical support design guidelines within the civil tunnelling and mining industries [1].

In general, a limited variety of support units are currently available at individual mines. This means that a few standard support systems cater for a potentially wide range of environmental conditions. Unless the excavation is of a special nature, or problematic ground conditions are foreseen well in advance, only these standard support systems are used. The support design engineer is therefore currently limited to existing semi-empirically-designed support systems that must cater for a multitude of variations in the rock mass environment.

Where rockmass conditions differ substantially from those for which the support system was generally developed (e.g. changes in loading conditions), failure of the support system may occur. Here the load distribution offered by interaction of the components of the support system may be insufficient to control the associated deformations of the rockmass. In 1995 a survey of tunnel conditions and support practices by Haile, Jager and Wojno [1] found that a significant cause of problems, particularly in abnormal ground conditions, is the lack of a rational design methodology for support systems. This is particularly true in highly stressed and highly fractured gold mine environments.

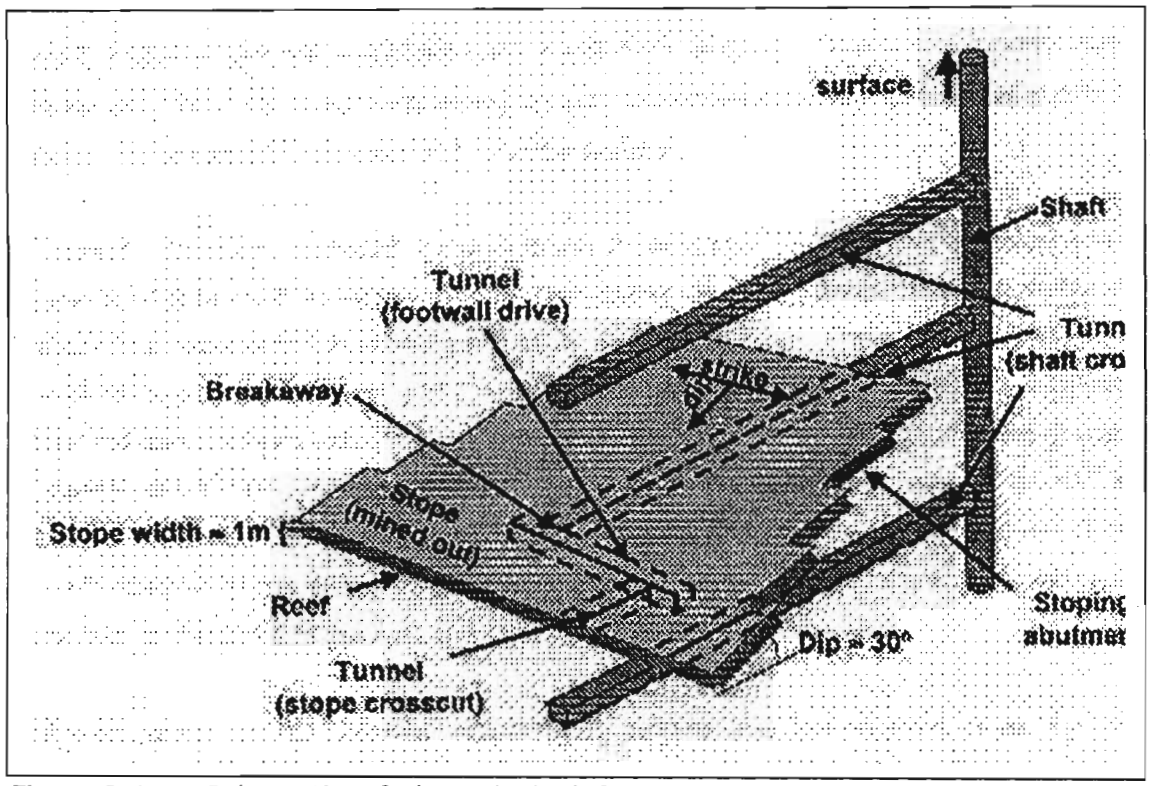


Figure 2.1: Schematic of the principal features of a typical deep level South African gold mine

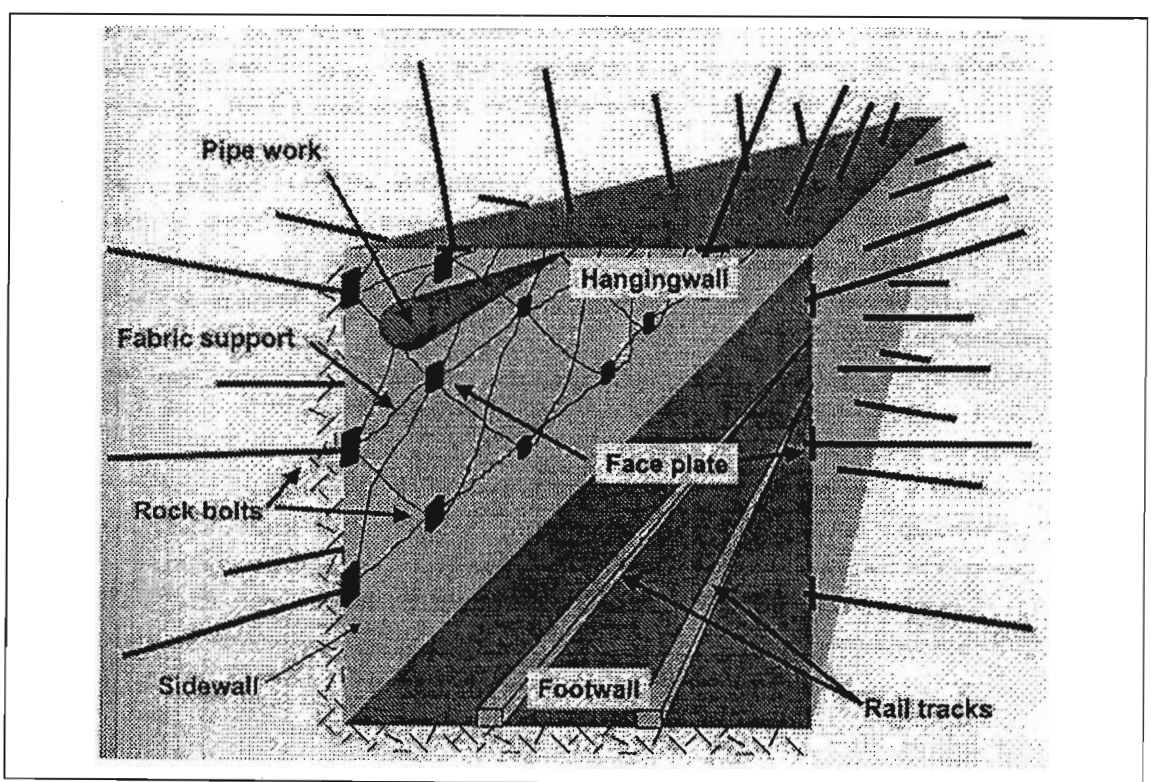


Figure 2.2: Schematic of typical components of a tunnel and its reinforcement/support system

2.2. TYPICAL ENVIRONMENTAL CONDITIONS

South African mines (gold mines in particular) comprise of tunnels that can reach great depths, in rockmasses that range from sedimentary to igneous (i.e. significant variation in the rockmass condition of the tunnels). Tunnels are subjected to loading conditions that are capable of causing fracturing of the rockwall. In these cases the enormous pressures cause the rock to break up into discontinuous blocks so that the rockmass dilates, forcing the rockwall into the tunnel.

A seismic event may notably occur on a fault due to changes in the stress environment in close proximity to a stope. Tunnels that access these areas of a mine experience violent shaking that may be sufficient to result in the rapid deformation and potential expulsion of fractured rock into the tunnel (called rockbursting). If this movement cannot be controlled by the tunnel support system, then total failure of the tunnel may result with potentially serious implications for the safety of personnel and the operational status of the tunnel.

The net result of these disruptions and damage has a huge economic and social impact; hence there is a strong incentive to supplement with methods of monitoring the state of the support system components at any time. The information gathered from a suitable health monitoring system would also offer additional insight into both the operational behaviour of support systems in general, as well as localised rockmass behaviour.

2.3. CONTROLLING HIGHLY STRESSED TUNNEL BEHAVIOUR

In 1966, Barrow [10] noted that supports cannot be practically designed to *prevent* the failure of an excavation wall, but that yielding bolt supports (rockbolts) can be designed to *control* the failed rockmass. His studies were based on the development of failure around an unsupported square tunnel (3m x 3m - typical of current excavation dimensions) that was sited in quartzite at a depth of approximately 1600m below surface. The tunnel was originally in a near virgin stress condition and was subsequently distressed by adjacent stoping operations (overstoping). To this day, rockbolts are still the common method of tunnel support, although a consistent and reliable means of quantifying the damage of installed units is still at conceptual level.

In the worst-case loading scenario, namely seismicity, the resultant stress at the excavation periphery is a function of the size and proximity of the seismic event to

the excavation. In 1983, Ortlepp [11] observed that around a given gold mine (deep) tunnel, fracturing is closely spaced at the surface of the excavation, but the spacing increases with depth into the rock wall. Wagner (1983) [12] found that the intensity of the fracturing was a function of the brittleness of the rock type and the magnitude of the local stress regime. Bedding planes also allow for increased rockmass movement. However, he concluded that there was very little understanding of the mechanism of rock fracture for the design of excavations. Any smart structural health monitoring system must therefore be independent of the mode of deformation.

2.4. DESIGN OF TUNNEL SUPPORT SYSTEMS

The mechanical behaviour and properties of tunnel support units are simple enough to analyse in detail using mechanical engineering methods, but the interaction between the rockmass and the rockbolt system is still poorly understood.

Fully grouted (frictional) rockbolts interact with the rock mass along the full barrel of the borehole in which they are installed. These bolts primarily influence the blocks with which they make contact. The stiff nature of this interaction make this type of support system generally applicable to the control of discrete structural blocks that must be retained in-place to maintain excavation stability. The number of required rockbolts in a given site is a direct function of the unstable volume to be supported. Alternative point-anchored support systems, with subsequent tensioning, result in an interaction of individual rock blocks to create a reinforced rock arch structure. The capacity and loading of the support unit structure is small compared to the loads acting within the rockmass.

The basic design principle is to create the most effective structure to allow the rockmass to be self-supporting. The current methodologies for the design of tunnel support systems are separated into three categories: analysis of structural stability (mechanistic evaluation); empirical design; and numerical modelling (finite element analysis, etc.).

As the excavation depth increases, so the understanding of the interaction of the rockbolt with the rockmass and applicable design methodologies decreases. The design of tunnel support systems is dependent primarily on the rockmass environment in which excavations are sited. The design engineer is presently restricted to using either experience or a rock mass classification system to assess the potential excavation instability and determine the support system requirements. The rockbolt support system should be designed to minimise the

instability of the excavation peripheral rockmass and thus excessive deformations of blocks within the rockmass. The length of the rockbolt units must be sufficient to anchor into the competent rockmass.

2.5. ANALYSIS OF STRUCTURAL STABILITY

1. For stabilisation, the simplest-case structural analysis considers the ability of a block to simply fall from the rockmass. The characteristics of the discontinuities that define the block are generally neglected. Static design considerations are based on the load capacity and length of the rockbolt system in relation to the size and mass of the defined unstable block. Increasing the number of rockbolts for a given unstable block improves the factor of safety. However, complex failure modes may result in sequential loading of the rockbolt units. Consequently, even with this simplest of support design analyses, an elevated factor of safety is required. See Figure 2.3.

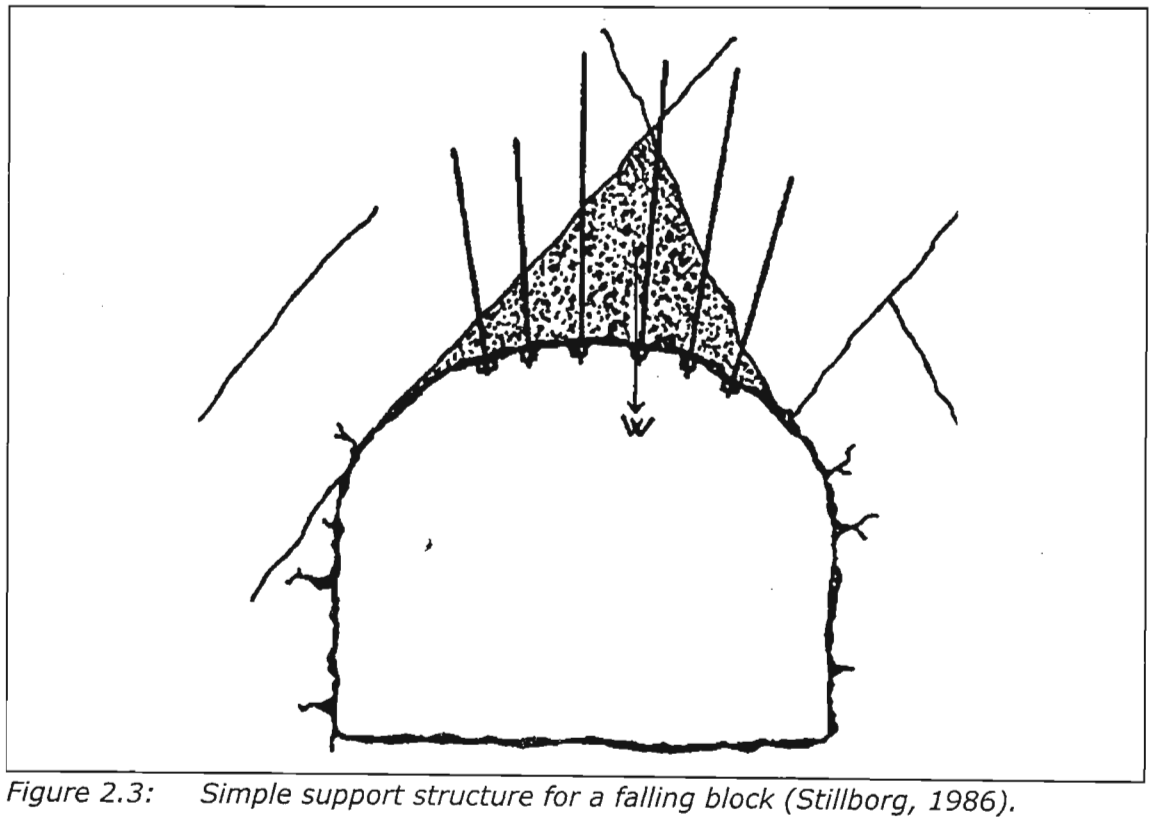


Figure 2.3: Simple support structure for a falling block (Stillborg, 1986).

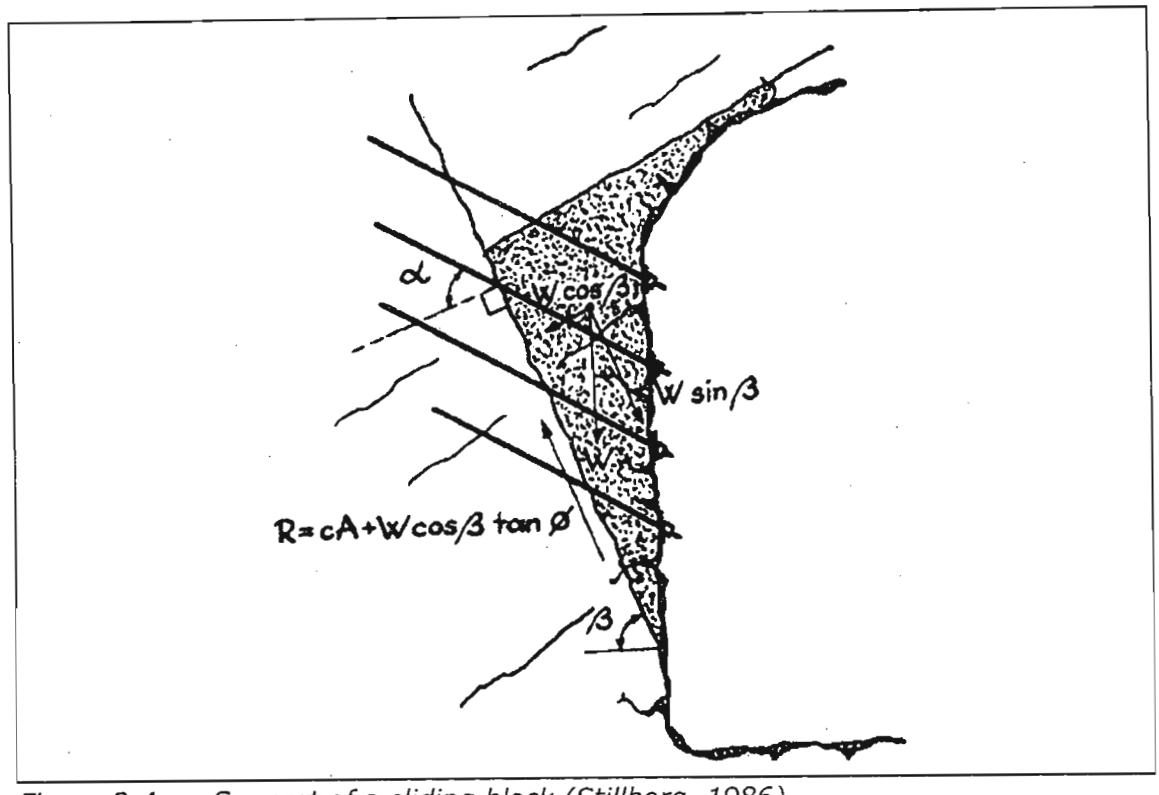


Figure 2.4: Support of a sliding block (Stillborg, 1986).

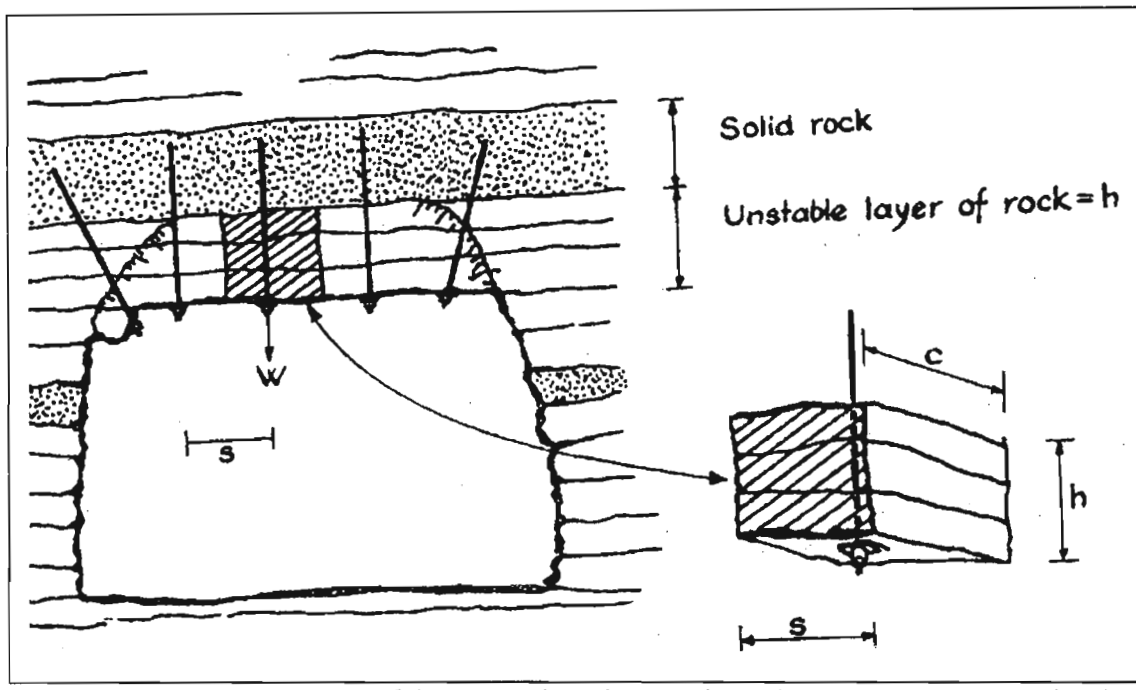


Figure 2.5: Suspension of laminated rockmass based on tributary area loading (Stillborg, 1986).

2. In conditions where sliding of the block may occur, the influence of the discontinuity characteristics are also considered. The development of tension within the rockbolt units increases the normal forces on the sliding plane of the block, and thus increases its frictional resistance (Figure 2.4). As the frequency of discontinuities within the rockmass increases, so the

rockbolting requirement changes from the stabilisation of individual, well defined blocks to more systematic rockbolt patterns in order to reinforce and/or support the potentially unstable rockmass volume.

3. In instances where the limits of rockmass instability can be defined, and are within a practical depth from the excavation boundary, then the rockbolt system is designed on the basis of suspension or containment. This is possible when a more competent horizon occurs immediately above the unstable strata (Figure 2.5). Here the length of the rockbolt unit must be sufficient to anchor beyond the limit of rockmass instability. The spacing of the rockbolts depends on the load capacity of the rockbolt unit and the volume of the unstable rockmass (s * c * h in the figure).

4. A similar analysis applies to cases where the discontinuous rockmass structure is competent yet moderately jointed. The creation of an excavation results in the formation of an unstable tensile zone in the immediate roof, above which a natural arch is formed where the rockmass loading condition is primarily compressive. The size of the unstable zone is a function of the rockmass structure, the loading environment, the span of the excavation and the stability of the sidewalls.

   Un-tensioned grouted support units are typically utilised in these conditions where the lower boundary of the natural arch is close to the boundary of the excavation. The unstable rockmass can be pinned to the competent natural arch (Figure 2.6) provided the bolt length is sufficient to anchor into the competent rock. A universally accepted estimation of the required length is:

$$L = 1.40 + 0.184 * w \text{ metres} \qquad (2\text{-}1)$$

   where w is the span of the excavation and L the required rockbolt length (Norwegian Institute for Rock Blasting Technique, after Stillborg, 1986 [17]).

   This design philosophy currently forms the basis of the mechanistic methodology for the design of rockbolt support systems within the South African mining industry [1]. The determination of actual depth instability should be based on accident data captured from an applicable geotechnical environment, and may consider both gravity-induced ground falls and seismically-induced rockbursts.

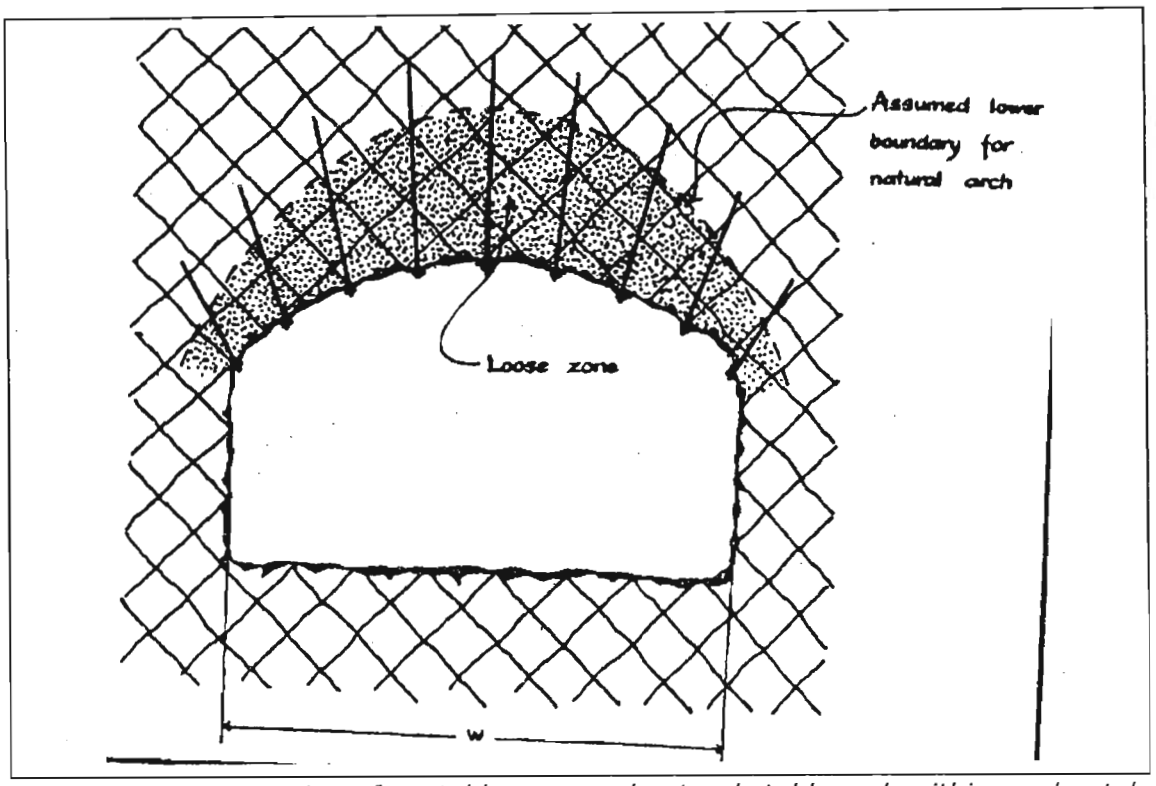


Figure 2.6: Formation of unstable zone and natural stable arch within moderately jointed rockmass structures (Stillborg, 1986).

Under gravitational loading, the number of required rockbolts is calculated from the load capacity of a single unit where the load W acting on the bolt/s is:

$$W = \rho * g * (s * c * h) \qquad (2\text{-}2)$$

where ρ is the rockmass density, g is the gravitational acceleration, and s, c and h (Figure 2.5) define the unstable volume. If highly friable ground conditions are anticipated, then additional use of blanket coverage support systems (mesh or shotcrete) is made to maintain the integrity of the rockmass between the bolt units.

Under seismic loading conditions the same design philosophy is adopted, except that consideration is also given to the energy demand on, and absorption capability of, the rockbolt unit.

Hangingwall:

$$E = \frac{1}{2}mv^2 + mgh \qquad (2\text{-}3)$$

Sidewall:

$$E = \frac{1}{2}mv^2 \qquad (2\text{-}4)$$

E energy absorption requirement of the system per unit area

m mass of ejected rock based on the defined unstable rockmass depth and density

v anticipated peak ground velocity

h yield capacity of the support system at any given time

g gravitational acceleration

Since the stress state may vary over the life of the excavation, consideration should also be given to the progressive deformation of the rockmass. A smart structural health monitoring system could help to determine the remaining yield capacity of the rockbolt units at a given time in the life of the excavation, and thus the support system energy absorption capacity.

5. In high stress environments, where significant fracturing of the rockmass may occur, or under low stress conditions but within more highly discontinuous rock mass structures, the depth of instability may exceed the practical depth of rockbolt anchorage. Under these conditions excavation stability may be enforced by the creation of reinforced beam (Figure 2.7) or arch (Figure 2.6) structures within the discontinuous rockmass.

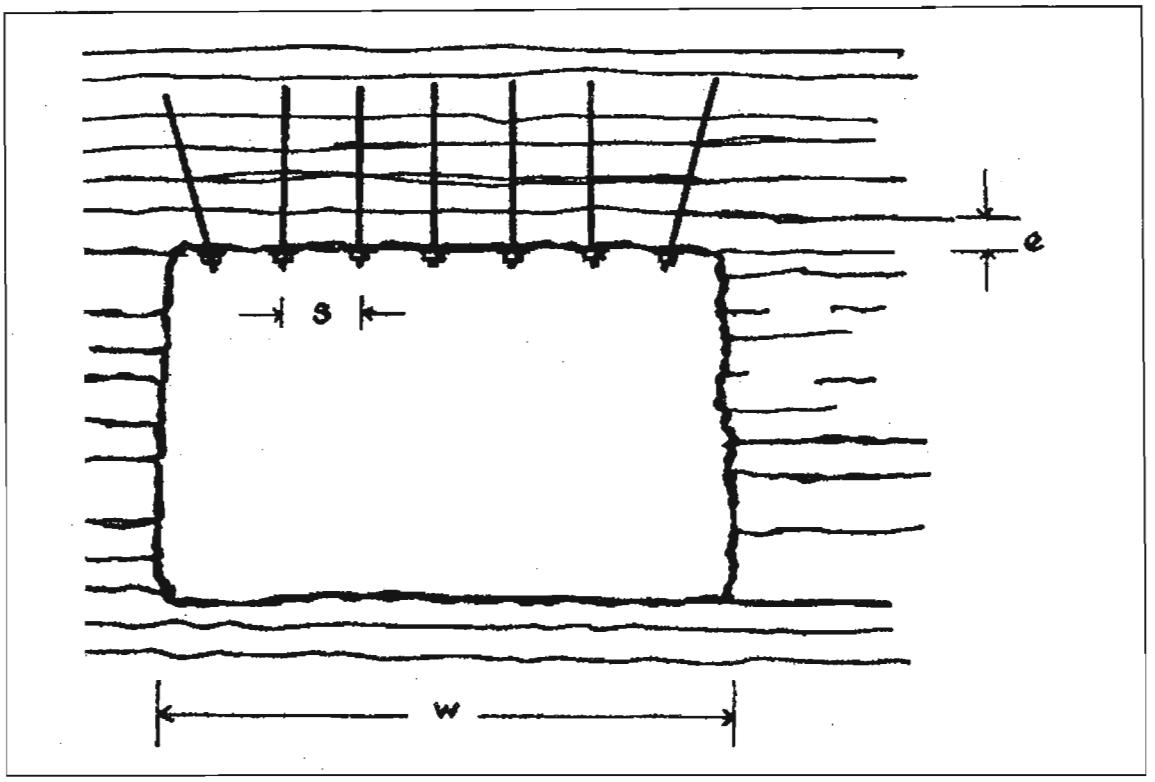


Figure 2.7: Creation of a reinforced beam within a stratified mass.

Under these conditions the rockbolts are tensioned to increase the friction between slabs or blocks within the defined zone, which in turn improves the resistance of the rockmass to shear and deformation. The design of a rockbolt system under these conditions is principally based on empirical rules (described later) and the guidelines discussed below. These techniques may be adapted into design charts for specific geotechnical applications.

One guideline for the design of a rockbolt system that creates a reinforced beam structure within a stratified rock mass (Figure 2.8) (Lang and Bischoff, 1982 [18]) considers the increase of shear resistance between layers within the rock mass:

$$\text{Bolt Length} = w\ 2/3 \tag{2-5}$$

$$\text{Bolt tension} = \left(\frac{(a\gamma AR)}{(\tan\phi)k}\right)\left(1-\frac{c}{\gamma R}\right)\left(\frac{1-\exp\left(-\frac{(\tan\phi)kD}{r}\right)}{1-\exp\left(-\frac{(\tan\phi)kL}{R}\right)}\right) \tag{2-6}$$

p shear perimeter of reinforced rock unit (p = 4 * s)

ϕ angle of internal friction for the rock mass.

k	ratio of average horizontal to average vertical stress
R	shear radius of the reinforced rock unit (A/P = s/4)
w	excavation span
A	area of roof carried by one bolt (s * s)
L	bolt length
s	bolt spacing
c	apparent cohesion of the rockmass
D	height of de-stressed zone
a	factor depending on time of installation of bolts
γ	unit weight of the rock

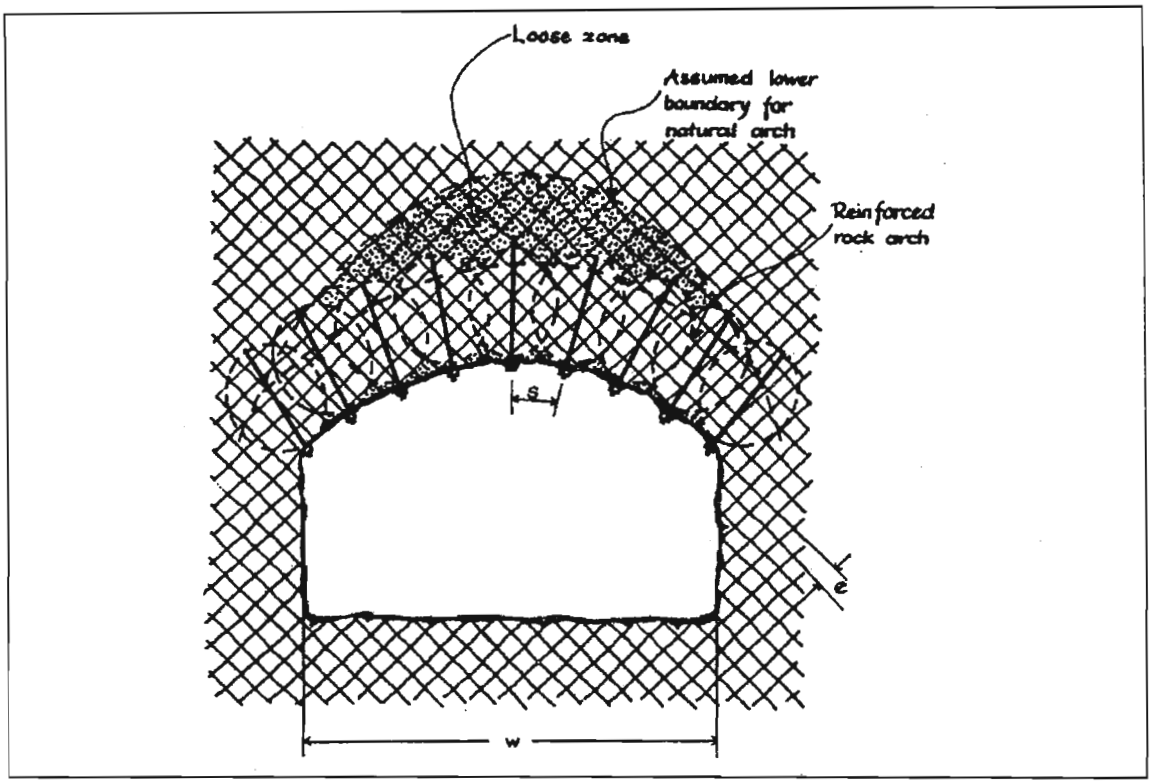


Figure 2.8: Creation of a reinforced rock mass arch within a highly discontinuous rock mass structure.

If the rock reinforcement is installed prior to the occurrence of significant deformation, then it is considered to make an active contribution to the excavation stability (a = 0.5). If passive (more conservative) reinforcement is assumed, then a=1. Cohesion should be taken as zero for initial design.

From Schach *et al.* (1979) [19] and Stillborg (1986) [17], a guideline for a support system design within a highly jointed rockmass is to create a reinforced rockmass arch (Figure 2-6) is:

$$\text{Bolt length} = 1.60 + \sqrt{1.0 + 0.012w^2} \qquad (2\text{-}7)$$

$$\text{Bolt spacing} = 3 * \text{joint spacing (e)} \qquad (2\text{-}8)$$

$$\text{Bolt tension} = (0.5 - 0.8) * \text{capacity of bolt} \qquad (2\text{-}9)$$

Blanket coverage support systems (e.g. mesh or shotcrete) must also be considered as a means to maintain the integrity of the rockmass between the rockbolt reinforcement.

These guidelines may find application within the South African mining environment although they are not commonly used [1]. Consideration is also not given to specific characteristics of the rock mass and thus tend to be generally conservative, although on occasion the requirements are under-estimated.

The concept of the creation of a structurally competent reinforced rock mass arch (Stillborg, 1986 [17]) is the basis of most of the empirical rockbolt system design methods.

2.6. EMPIRICAL DESIGN METHODOLOGIES

Because of complexities within the rockmass structure, or the lack of detailed design parameters, it may not be possible to design a rockbolt system based on a specific structural analysis. Under these conditions it is convenient to make use of empirical design methodologies that are based on analyses of the geotechnical environment. Design confidence is a function of the maturity of the design process and the availability of suitable geotechnical data. The classification systems reviewed in this section point to a specific support category for a given rockmass condition.

The majority of formal empirical design methodologies such as the Q system and the Rock Mass Rating (RMR) system (also referred to as the Geomechanics Classification system) are based on experience gained within shallow, low stress, civil engineering applications (not mining). They therefore include little consideration for an environment of stress change due to mining and the potential for rockbursting, and find limited application within the general South African

mining industry (Piper, 1985) [13]. Exceptions can be found, particularly in the platinum and diamond mining industries that operate at shallower depths. The Rockwall Condition Factor method (Wiseman, 1979) [14] is better suited to conditions of potential stress change.

As the rockmass changes, due to a change in stratigraphy or stress environment, so the support recommendations may dramatically alter with the transition to a different rockmass class. Within a mining environment, with numerous areas of simultaneous tunnel excavation, the logistics of supply and control of the support system components make this potential scenario problematic. Thus there is a tendency to install support systems that cater for more adverse ground conditions, resulting in inefficient and costly over-supporting of the more favourable ground conditions.

2.6.1. Local Estimations for Highly Stressed Environments

Wojno, Jager and Roberts (1987) [15] and Henderson (1990) [16] investigated tunnels that experience high stress conditions in an attempt to determine support performance requirements in these adverse conditions. The support guidelines applicable to this rockmass environment indicate a need for yielding tendons. The proposed specifications for a suitable support system are based on the mechanistic evaluations discussed in Section 2.4:

Work done > 25kJ

Static yield >100kN

Dynamic yield > 50 kN

Maximum yield < 500mm

Tendon strength > 25% static yield

These simple rules allow initial estimations of support requirements, but give limited or no consideration to the rockmass characteristics. They are therefore conservative in nature.

2.6.2. The NGI Q System

The widely used and internationally recognised Q system classification (Barton, lien and Lunde, 1974 [20]) is generally applicable to shallow tunnel excavations.

Recent experience with dynamic failure (rockbursting) has also been incorporated into the system (Barton, 1997 [21]). In South Africa, the Q system has found application in tunnel excavations of shallow mining orebodies and in a limited number of deep level mining environments with associated local adjustments. The Q system uses rockmass parameters to determine a Q value for which there is a recommended support system based on excavation dimensions and utilisation.

Table 2.1: Parameters for rockmass classification according to the Q system.

1. Rock Quality Designation (RQD)		RQD
A	Very poor	0 – 25
B	Poor	25 – 50
C	Fair	50 – 75
D	Good	75 – 90
E	Excellent	90 – 100
Notes:	i) Where RQD is reported or measured as ≤ 10 (including 0), a nominal value of 10 is used to evaluate Q. ii) RQD intervals of 5, i.e. 100, 95, 90 etc. are sufficiently accurate.	

2. Joint Set Number (Jn)		Jn
A	Massive, no or few joints	0.5 – 1.0
B	One joint set	2
C	One joint set plus random joints	3
D	Two joint sets	4
E	Two joint sets plus random joints	6
F	Three joint sets	9
G	Three joint sets plus random joints	12
H	Four or more joint sets, random, heavily-jointed, "sugar cube", etc.	15
J	Crushed rock, earth-like.	20
Notes:	i) For intersections, use 3 x Jn. ii) For portals, use 2 x Jn.	

3. Joint Roughness Number (Jr)		Jr
a) Rock wall contact and b) rock wall contact before 10 cm shear.		
A	Discontinuous joints	4
B	Rough or irregular, undulating	3
C	Smooth, undulating	2
D	Slickensided, undulating	1.5
E	Rough or irregular, planar	1.5
F	Smooth, planar	1
G	Slickensided, planar	0.5
Note:	Descriptions refer to small-scale features and intermediate-scale features in that order.	
c) No rock wall contact when sheared.		
H	Zone containing clay minerals thick enough to prevent rock wall contact	1
J	Sandy, gravelly or crushed zone thick enough to prevent rock wall contact	1
Notes:	Add 1.0 if the mean spacing of the relevant joint set is > 3 m. Jr = 0.5 can be used for planar slickensided joints having lineations, provided the lineations are orientated for minimum strength.	

4. Joint Alteration Number (Ja)		φr approx.	Ja
a) Rock wall contact (no mineral filling, only coatings).			
A	Tightly healed, hard, non-softening, impermeable filling, i.e. quartz or epidote	-	0.75
B	Unaltered joint walls, surface staining only.	25 - 35°	1.0
C	Slightly altered joint walls. Non-softening mineral coatings, sandy particles, clay free disintegrated rock, etc.	25 - 35°	2.0
D	Silty- or sandy-clay coatings, small clay fraction (non-softening).	20 - 25°	3.0
E	Softening or low friction clay mineral coatings (kaolinite or mica), chlorite, talc, gypsum, graphite, small amounts of swelling clays, etc.	8 - 16°	4.0
b) Rock wall contact before 10 cm shear (thin mineral fillings).			
F	Sandy, clay-free disintegrated rock, etc.	25 - 30°	4.0
G	Strongly over consolidated non-softening clay mineral fillings (continuous but < 5 mm thickness).	16 - 24°	6.0
H	Medium or low over consolidation, softening, clay mineral fillings (continuous but < 5 mm thickness).	12 - 16°	8.0
J	Swelling clay fillings, i.e. montmorillonite (continuous but < 5 mm thickness). Value of Ja depends on percent of swelling clay-size particles, access to water, etc.	6 - 12°	8 – 12
c) No rock wall contact when sheared (thick mineral fillings).			
K, L, M	Zones or bands of disintegrated or crushed rock and clay (See G, H, J for description of clay condition).	6 - 24°	6, 8 or 8 - 12
N	Zones or bands of silty or sandy clay, small clay fraction (non-softening).	-	5.0
O, P, R	Thick, continuous zones or bands of clay (See G, H, J for description of clay condition).	6 - 24°	10, 13 or 13 - 20

5. Joint Water Reduction Factor (Jw)		~Water Pressure (kg/cm^2)	Jw
a) Weakness zones intersecting excavation, which may cause loosening of rockmass when tunnel is excavated.			
A	Dry excavations or minor inflow, i.e. <5 l/min locally.	< 1	1.0
B	Medium inflow or pressure, occasional outwash of joint fillings.	1 – 2.5	0.66
C	Large inflow or high pressure in competent rock with unfilled joints.	2.5 – 10	0.5
D	Large inflow or high pressure, considerable outwash of joint fillings.	2.5 – 10	0.33
E	Exceptionally high inflow or water pressure at blasting, decaying with time.	> 10	0.2 – 0.1
F	Exceptionally high inflow or water pressure continuing without noticeable decay.	> 10	0.1 – 0.05
Notes:	i) Factors C to F are crude estimates; increase Jw if drainage measures are installed. ii) Special problems caused by ice formation are not considered.		

6. Stress Reduction Factor (SRF)				SRF
a) Weak zones intersecting excavation, which may cause loosening of rock mass when tunnel is excavated.				
A	Multiple weakness zones containing clay or chemically disintegrated rock, loose surrounding rock (any depth).			10
B	Single weakness zones containing clay or chemically disintegrated rock (depth of excavation ≤ 50 m).			5
C	Single weakness zones containing clay or chemically disintegrated rock (depth of excavation > 50 m).			2.5
D	Multiple shear zones in competent rock (clay free), loose surrounding rock (any depth).			7.5
E	Single shear zones in competent rock (clay free) (depth of excavation ≤ 50 m).			5.0
F	Single shear zones in competent rock (clay free) (depth of excavation > 50 m).			2.5
G	Loose, open joints, heavily jointed or "sugar cube", etc. (any depth).			5.0
Note:	i) Reduce these values of SRF by 25 – 50% if the relevant shear zones only influence but do not intersect the excavation.			
b) Competent rock, rock stress problems.		σ_c / σ_1	σ_θ / σ_c	SRF
H	Low stress, near surface, open joints	> 200	< 0.01	2.5
J	Medium stress, favourable stress condition	200 - 10	0.01 – 0.3	1
K	High stress, very tight structure. Usually favourable to stability, may be unstable for wall stability.	10 – 5	0.3 – 0.4	0.5 - 2
L	Moderate slabbing after > 1 hour in massive rock	5 – 3	0.5 – 0.65	5 - 50
M	Slabbing and rockburst after a few minutes in massive rock	3 – 2	0.65 – 1	50 - 200
N	Heavy rockburst (strain-burst) and immediate dynamic deformations in massive rock.	< 2	> 1	200 - 400
Note:	ii) For strongly anisotropic virgin stress field (if measured): When $5 \le \sigma_1/\sigma_3 \le 10$, reduce σ_c to $0.75\sigma_c$. When $\sigma_1/\sigma_3 > 10$, reduce σ_c to $0.5\sigma_c$. iii) Few case records are available where depth of crown below surface is less than span width. Suggest SRF increase from 2.5 to 5 for such cases (see H).			
σ_c is the unconfined compression strength, σ_1 and σ_3 are the major and minor principle stresses, and σ_θ is the maximum tangential stress (estimated from elastic theory).				
c) Squeezing rock: plastic flow of incompetent rock under the influence of high rock pressure			σ_θ / σ_c	SRF
O	Mild squeezing rock pressure		1.5	5 – 10
P	Heavy squeezing rock pressure		> 5	10 - 20
Note:	iv) Cases of squeezing rock may occur for depth $H > 350*Q^{1/3}$ (Singh et al., 1992). Rockmass compression strength can be estimated from $q = 7\gamma Q^{1/3}$ MPa where γ is the rock density in g/cm^3 (Singh, 1993).			
d) Swelling rock: chemical swelling activity depending on pressure of water				SRF
R	Mild swelling rock pressure			5 – 10
S	Heavy swelling rock pressure			10 - 15

Jr and Ja classification is applied to the joint set or discontinuity that is least favourable for stability both from the point of view of orientation and shear resistance (where $\tau = \sigma_n * \tan\left(\frac{Jr}{Ja}\right)$).

Finally:

$$Q = \frac{RQD}{Jn} * \frac{Jr}{Ja} * \frac{Jw}{SRF} \qquad (2\text{-}10)$$

Thus, by identifying the applicable rockmass characteristics (within the defined parameter fields) and their associated ratings, an overall classification of the rockmass, denoted as Q, is obtained. This Q value may now be used to assess the stability (that is also dependent on the size and utilisation of the proposed excavation) of the proposed excavation and recommend suitable support systems based on case studies. To quantify the relative safety of an excavation, the Excavation Support Ratio (ESR) is used. See Table 2.2.

Table 2.2: Recommended ESR values for excavation safety level

Type of excavation		ESR
A	Temporary mine openings, etc.	2 – 5
B	Permanent mine openings, water tunnels for hydropower (excluding high pressure penstocks), pilot tunnels, drifts and headings for large openings, surge chambers.	1.6 – 2.0
C	Storage caverns, water treatment plants, minor road and railway tunnels, access tunnels.	1.2 – 1.3
D	Power stations, major road and railway tunnels, civil defence chambers, portals, intersections.	0.9 – 1.1
E	Underground nuclear power stations, railway stations, sports and public facilities, factories, major gas pipeline tunnels.	0.5 – 0.8

Using the Q value and the ESR value, the maximum span of an unsupported opening is:

$$w_m = 2 * ESR * Q^{0.4} \qquad (2\text{-}11)$$

Even for self-supporting excavation spans, consideration must still be given to the support of potentially unstable jointed blocks.

For spans exceeding self-supporting requirements, support recommendations are calculated using a chart. Here the length of the rockbolts is purely a function of the relative size of the excavation. The spacing of the rockbolts and the necessity for (and type of) fabric support is a function of the rockmass.

Guidelines for the design of primary support systems based on the Q system are shown in Table 2.3.

Table 2.3: Estimations of support pressure, length and spacing of primary support systems based on the Q system.

Rockmass jointing	Support pressure (MPa)	Length and spacing of reinforcement
If the number of discontinuity sets > 2	$P_{roof} = \frac{0.2 * Q^{-1/3}}{Jr}$	Bolts: $L_{roof} = \frac{2 + 0.15 * B}{ESR}$ Anchors: $L_{roof} = \frac{0.4 * B}{ESR}$
If the number of discontinuity sets ≤ 2	$P_{roof} = \frac{0.2 * Jn^{-1/2} * Q^{-1/3}}{3 * Jr}$	$s_{roof} = \left(\frac{C * 10^{-3}}{P_{roof}}\right)^{1/2}$
	P_{wall} is calculated with the same formulae as P_{roof}, but replacing Q with Q': Q' = 5*Q if Q > 10 Q' = 2.5*Q if 0.1 ≤ Q ≤ 10 Q' = Q if Q < 0.1	Bolts: $L_{wall} = \frac{2 + 0.15 * H}{ESR}$ Anchors: $L_{wall} = \frac{0.35 * H}{ESR}$ $s_{wall} = \left(\frac{C * 10^{-3}}{P_{wall}}\right)^{1/2}$
B and H are the span and height of the excavation respectively (metres) L is the length of reinforcement (metres) s is the spacing of the reinforcement (metres) C is the load exceeding the yield strength of the bolt (kN)		

The support pressure, or distributed rockbolt load, can also be defined as a function of the Q value and the roughness of the potentially unstable discontinuity, and is also extracted from a chart.

2.6.3. The Rockwall Condition Factor Classification System

A limited number of support design guidelines are applicable to tunnels developed under high and often variable stress environments where they are susceptible to rockbursting. This is because of the unique nature of this environment. In the South African mining industry, the recommended and commonly used guidelines (Rockwall Condition Factor system) are based on the work of Wiseman (1979) [14] and are extended in the Chamber of Mines Research Organisation industry design guidelines (Anon, 1988) [22]. The Rockwall Condition Factor (RCF) is defined as:

$$RCF = \frac{3\sigma_1 - \sigma_3}{F * \sigma_c} \qquad (2\text{-}12)$$

where σ_1 and σ_3 are the maximum and minimum principal stresses within the plane of the excavation cross-section, and F is a factor representing the downgrading of

σ_c (the uniaxial compressive strength), for the rockmass condition and excavation size. In a highly discontinuous rockmass it is recommended that F is approximately 0.5; in large excavations (> 6m x 6m) F may be further reduced by 10%.

Wiseman [14] examined the RCF criterion in South African Witwatersrand goldmine tunnels (typically 3m x 3m). He found that for RCF < 0.7, good conditions prevailed with minimum support (Table 2.4); for 0.7 < RCF < 1.4 average conditions prevailed with typical support systems (Table 2.5); and for RCF > 1.4 poor conditions prevailed with specialist support requirements (Table 2.6).

Table 2.4: Support recommendation for good ground conditions (RCF < 0.7).

Case	Primary support (typical requirements)	Secondary support (typical requirements)
Static stress conditions	Spot support where necessary (split sets, rock studs, etc.).	Spot support where necessary (rock studs, etc.).
Stress changes anticipated	Spot support where necessary (split sets, rockbolts, etc.).	Fully grouted tendons > 1.5m long, installed on basic 2m pattern. Support resistance 30 – 50kPa. Rope lacing (sidewalls only).
Seismic activity anticipated	Split sets or tendons > 1.2m in length, installed as close to the face as possible on a basic 2m or 1.5m pattern. Support resistance 30 – 50kPa.	Fully grouted (preferably yielding) tendons > 1.5m in length, installed on a basic 2m pattern. Support resistance 30 – 50kPa. Rope lacing (sidewalls only).

Table 2.5: Support recommendation for average ground conditions (0.7 < RCF < 1.4).

Case	Primary support (typical requirements)	Secondary support (typical requirements)
Static stress conditions	Rock studs or tendons > 1.5m long, installed as close to the face as possible on a basic 2m pattern. Support resistance 30 – 50kPa.	Steel strapping or rope lacing integrated with primary support tendons; or gunite.
Stress changes anticipated	Fully grouted tendons or steel ropes > 1.8m long, installed as close to the face as possible on a basic 2m or 1.5m pattern. Support resistance 40 – 60kPa.	Rope lacing and wire mesh integrated with primary support tendons.
Seismic activity anticipated	Fully grouted (preferably yielding) tendons or steel ropes > 1.8m long, installed as close to the face as possible on a 1.5m or double 2m pattern. Support resistance 80 – 110kPa.	Rope lacing and wire mesh integrated with primary support tendons, plus optional gunite.

Table 2.6: Support recommendation for poor ground conditions (RCF > 1.4)

Case	Primary support (Typical requirements)	Secondary support (Typical requirements)
Static stress conditions	(If necessary) gunite to face, then fully grouted tendons or steel ropes, length > 1.8m on a basic 1.5m pattern, as close to the face as possible. Support resistance 80 – 110kPa.	Steel wire mesh integrated with primary support; optional gunite/shotcrete.
Stress changes anticipated	(If necessary) gunite to face, then fully grouted steel ropes or yielding tendons, length > 1.8m on a basic 1m or double 2m pattern, as close to the face as possible. Support resistance 120–230kPa.	Rope lacing and wire mesh integrated with primary support. Add integral gunite in long-life tunnels. If necessary, additional hangingwall support comprising grouted steel ropes.
Seismic activity anticipated	(If necessary) gunite to face, then fully grouted steel ropes or yielding tendons, length > 2.3m on a basic 1m pattern, as close to the face as possible. Support resistance 220–290kPa.	Rope lacing and wire mesh integrated with primary support. Add integral gunite in long-life tunnels. If necessary, additional hangingwall support comprising grouted steel ropes.

This empirical methodology primarily relates the influence of the stress environment on the excavation rockmass condition to the typical support systems utilised. The guidelines cater well for bedded deep level rockmass environments, but may be less effective in other environments.

2.6.4. The Geomechanics Classification System

This system (Bieniawski, 1973 and 1979) [23], [24] is also referred to as the RMR (Rockmass Rating) system and is based on case studies of civil constructions within South Africa. The basic RMR value considers critical rockmass parameters, including the intact rock strength, discontinuity frequency, discontinuity condition, and the presence of water within the rockmass. See Table 2.7.

Table 2.7: Geomechanics Classification system

Parameter			Range of Values						
1	Strength of intact rock material	Point-load strength index (MPa)	>10	4–10	2–4	1–2	For this low range, uniaxial compressive test is preferred		
		Uniaxial compressive strength (MPa)	>250	100–250	50–100	25–50	5–25	1–5	<1
	Rating		15	12	7	4	2	1	0
2	Drill core quality RQD		90–100%	75–90%	50–75%	25–50%	<25%		
	Rating		20	17	13	8	3		
3	Spacing of discontinuities		>2m	0.6–2m	200–600mm	60–200mm	<60mm		
	Rating		20	15	10	8	5		
4	Condition of discontinuities		Very rough surface. Not continuous. No separation. Un-weathered walls.	Slightly rough surface. Separation <1mm. Slightly weathered walls.	Slightly rough surface. Separation <1mm. Highly weathered walls.	Slick surfaces, or gouge <5mm thick, or separation 1 – 5 mm continuous.	Soft gouge >5mm thick, or separation >5mm continuous.		
	Rating		30	25	20	10	0		
5	Ground water	Inflow per 10m tunnel	None or	<10 litres/min or	10–25 litres/min or	25–125 litres/min or	>125 litres/min or		
		Ratio (joint water pressure/major principal stress)	0 or	0.0–0.1 or	0.1–0.2 or	0.2–0.5 or	>0.5 or		
		General conditions	Completely dry	Damp	Wet	Dripping	Flowing		
	Rating		15	10	7	4	0		

Further adjustments may also be made in relation to the orientation of the joints relative to the excavation orientation and type to obtain an overall rock mass description (Tables 2.8, 2.9 and 2.10).

Table 2.8: Grading of joint orientations applicable to tunnels.

Strike perpendicular to tunnel axis				Strike parallel to tunnel axis		Dip 0-20° irrespective of strike
Drive with dip		Drive against dip				
Dip 45-90°	Dip 20-45°	Dip 45-90°	Dip 20-45°	Dip 45-90°	Dip 20-45°	
Very favourable	Favourable	Fair	Unfavourable	Very unfavourable	Fair	Unfavourable

Table 2.9: Adjustments to RMR for varying joint orientations.

Strike and dip orientations of the joints		Very favourable	Favourable	Fair	Unfavourable	Very unfavourable
Rating adjustment	Tunnels (Table 2.8)	0	-2	-5	-10	-12
	Foundations	0	-2	-7	-15	-25
	Slopes	0	-5	-25	-50	-60

Table 2.10: Rock mass class and description as determined from RMR.

Rating	100 ← 81	80 ← 61	60 ← 41	40 ← 21	< 20
Class No.	I	II	III	IV	V
Description	Very good rock	Good rock	Fair rock	Poor rock	Very poor rock

The description of the rockmass may be used to define overall rock mass parameters (Table 2.11) that may be used to analyse the large-scale response of the rock mass.

Table 2.11: Rockmass parameters based on RMR.

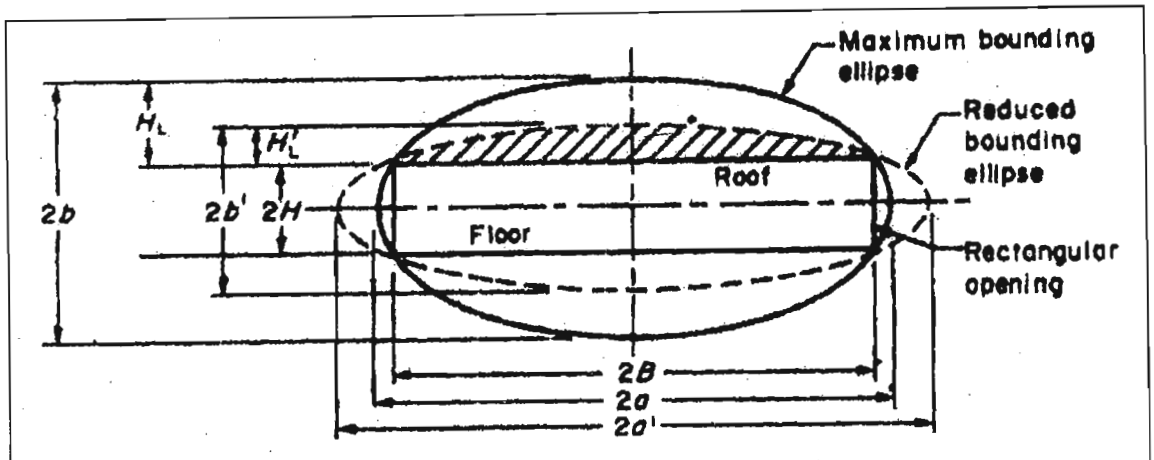
Class No.	I	II	III	IV	V
Average stand-up time	10 years for 15m span	6 months for 8m span	1 week for 5m span	10 hours for 2.5m span	30 mins for 1m span
Cohesion of the rock mass (kPa)	> 400	300 – 400	200 – 300	100 – 200	< 100
Friction angle of the rock mass	> 45°	35 - 45°	25 - 35°	15 - 25°	< 15°

Using the RMR system, support recommendations were made for the specific civil construction shown in Table 2.12. Because of the empirical nature of the design methodology, this is based on the experience and geotechnology of the specific sites incorporated in the analysis.

Table 2.12: Excavation technique and support selection based on RMR.

Shape: horseshoe; **Width**: 10m; **Vertical stress**: below 25MPa; **Construction**: Drill and Blast				
Rock mass class	Excavation	Support		
		Rockbolts (20mm diameter fully bonded)	Shotcrete	Steel Sets
Very good rock Class I RMR 81 - 100	Full face. 3m advance.		Generally no support required except for occasional spot bolting.	
Good rock Class II RMR 61 - 80	Full face. 1 – 1.5m advance. Complete support 20m from face	Local bolts in crown 3m long spaced 2.5m with occasional wire mesh.	50mm in crown where required	None
Fair rock Class III RMR 41 - 60	Top heading and bench. 1.5 – 3m advance in top heading. Commence support after each blast. Complete support 10 m from face.	Systematic bolts 4m long spaced 1.5 – 2m in crown and walls with wire mesh in crown.	50 – 100mm in crown and 30 mm in sides.	None
Poor rock Class IV RMR 21 - 40	Top heading and bench. 1 – 1.5m advance in top heading. Install support concurrently with excavation 10m from face.	Systematic bolts 4 – 5m long spaced 1 – 1.5m in crown and walls with wire mesh.	100 – 150mm in crown and 100mm in sides.	Light to medium ribs spaced 1.5m where required.
Very poor rock Class V RMR < 20	Multiple drifts. 0.5 – 1.5m advance in top heading. Install support concurrently with excavation, shotcrete as soon as possible after blasting.	Systematic bolts 5 – 6m long, spaced 1 – 1.5m in crown and walls with wire mesh. Bolt invert.	150 – 200 mm in crown,150 mm in sides and 50 mm on face.	Medium to heavy ribs spaced 0.75m with steel lagging and fore poling if required. Close invert.

Applications of the RMR system include a semi-empirical method for the estimation of the depth of instability around an excavation (Stimpson, 1989) [25]. This system defines an elliptical boundary around the excavation wherein the rock mass is considered to require reinforcement.



Maximum principal stress is horizontal	Maximum principal stress is vertical
$a = k * b$	$a = \sqrt{H^2 + k^2 * B^2}$
$b = \frac{\sqrt{B^2 + k^2 * H^2}}{k}$	$b = \frac{a}{k}$
$a' = \frac{B * b'}{\sqrt{b'^2 - H^2}}$	$a' = a * (1 - 0.01 * RMR) + H * 0.01 * RMR$
$b' = b * (1 - 0.01 * RMR) + H * 0.01 * RMR$	$b' = \frac{B * a'}{\sqrt{a'^2 - H^2}}$
$H_L = b - H$	$H_L = a - H$
$H_L' = H_L * (1 - 0.01 * RMR)$	$H_L' = H_L * (1 - 0.01 * RMR)$

Figure 2.9: Semi-empirical instability extent after Stimpson (1989) and based on the RMR guidelines.

2.6.5. The Modified Rock Mass Rating System

The Modified Rock Mass Rating (MRMR) system (Laubscher and Taylor, 1976) [26] caters for the differences in support requirements between the original civil environment and a mining environment. This is applied via a percentage alteration to the individual parameters within the RMR as a function of weathering, stress environment, excavation orientation relative to the rock mass structure, and the excavation method.

The influence of weathering is applied to the intact rock strength (decreases towards 96%), RQD (decreases towards 95%) and the condition of the joints (decreases towards 82%), dependent on the degree of weathering.

The influence of the stress environment is a function of the initial field stress level and any potential stress changes, and is considered to influence the relative joint condition. Where the stress environment is, or becomes, compressive relative to the rockmass structure and excavation periphery, then this will be favourable to joint conditions (increases towards 120%). Where shear deformation may be

anticipated on the joints in the periphery of the excavation, or a low stress results in the opening of joints, then this will be unfavourable to rock mass stability (decreases towards 90% or 76% respectively). Where subsequent stress changes occur, this may also cause adverse shear movement or joint opening within the excavation peripheral rock mass (decreases towards 60%).

Laubscher and Taylor [26] considered the influence of the excavation orientation relative to the rockmass structure as accounted for by the RMR system to be insufficient. They allowed for further adjustments to the joint spacing rating based on the number of non-vertical rockwalls of the excavation and the number of joints (Table 2.13).

Table 2.13: Percentage adjustment to joint spacing rating for joint number and number of non-vertical excavation rockwalls

Number of joints		70%	75%	80%	85%	90%
3	Number of non-vertical excavation rockwalls	3		2		
4		4	3		2	
5		5	4	3	2	1
6		6		4	3	2 - 1

A further adjustment to the joint spacing rating is also considered necessary for the presence and orientation of shear zones (decreases dependant on the angle of incidence: for 0 - 15° towards 76%, for 15 – 45° towards 84%, and for 45 - 75° towards 92%).

Finally, an adjustment may be made to the RQD and condition of joints rating for the influence of the method of excavation. The degree of the reduction of the parameters is on the consideration of the adverse effects of blasting (boring 100%, smooth wall blasting 97%, good conventional blasting 94%, poor conventional blasting 80%) on the immediate excavation peripheral rock mass stability.

Laubscher and Taylor noted that the application of these adjustments to the original parameters should not reduce the modified RMR by more than 50% of the original value. These modified Geomechanics Classifications were compared to typical mining support practices in order to define a relationship between the MRMR and support systems - see Table 2.14.

Table 2.14: Support system guide based on Modified Geomechanics Classification.

Adjusted ratings	Original rock mass ratings, RMR									
	90 - 100	80 - 90	70 - 80	60 - 70	50 - 60	40 - 50	30 - 40	20 - 30	10 - 20	0 - 10
60 - 100										
50 - 60		a	a	a	a					
40 - 50			b	b	b	b				
30 - 40				c,d	c,d	c,d,e	d,e			
20 - 30					g	f,g	f,g,j	f,h,j		
10 - 20						i	i	h,i, j	h,j	
0 - 10							k	k	l	l

a Generally no support; locally the joint intersections may require bolting.

b Patterned grouted bolts at 1m collar spacing.

c Patterned grouted bolts at 0.75m collar spacing.

d Patterned grouted bolts at 1m collar spacing and shotcrete 100mm thick.

e Patterned grouted bolts at 1m collar spacing and massive concrete 300mm thick and only used if stress changes are not excessive.

f Patterned grouted bolts at 0.75m collar spacing and shotcrete 100mm thick.

g Patterned grouted bolts at 0.75m collar spacing with mesh-reinforced concrete 100mm thick.

h Massive concrete 450mm thick with patterned grouted bolts at 1m spacing if stress changes are not excessive.

i Grouted bolts at 0.75m collar spacing if reinforcement potential is present, and 100mm reinforced shotcrete, and then yielding steel arches as a repair technique if stress changes are excessive.

j Stabilise with rope cover support and massive concrete 450mm thick if stress changes are not excessive.

k Stabilise with rope cover support followed by shotcrete to and including face if necessary, and then closely spaced yielding arches as a repair technique if stress changes are excessive.

l Avoid development in this ground otherwise use system j or k.

Additional notes:

1. The original Geomechanics Classification as well as the adjusted ratings must be taken into account in assessing the support requirements.

2. Bolts serve little purpose in highly jointed ground and should not be used as the sole support where the joint spacing rating is less than 6.

3. The recommendations in the above table are applicable to mining operations with stress levels less than 30MPa.

4. Large chambers should only be excavated in rock with adjusted total classification ratings of 50 or better.

These support design guidelines have found some application within the shallow mining environments (e.g. platinum and diamond) of the South African mining industry. Although comparable support systems are utilised at depth, the design is generally not based on these guidelines.

Applications of the MRMR have included the determination (from case studies) of the minimum rockbolt density for hard rock mines in Quebec (Choquet and Charette, 1988) [27]. The minimum required support density is given by:

$$D = -0.0214 * MRMR + 1.68$$

$$(\text{or } D = -0.227 * \ln(Q) + 0.839) \qquad (2\text{-}13)$$

$$s = \frac{1}{\sqrt{D}}$$

where:

D number of bolts per square metre

Q rating based on the NGI Q system classification

s bolt spacing on a square pattern

2.6.6. U.S. Corps of Engineers

The US Corps of Engineers have developed simple empirical design rules from evaluations of case studies. These rules have found widespread application, particularly as an initial estimate of support requirements, and are applied as the basis for rockbolt spacing calculations within many South African mines. See Table 2.15.

Table 2.15: Tunnel support guidelines from the US Corps of Engineers.

Parameter	Empirical Rule
Minimum length	Greatest of: a) 2 x bolt spacing b) 3 x thickness of critical and potentially unstable rock blocks c) For elements above the springline: spans < 6m: 0.5*span spans between 6m and 18m: interpolate 3 to 4.5m spans between 18m and 30m: 0.25*span d) For elements below the springline: height < 18m: as c) above height > 18m: 0.2*height
Minimum spacing	Least of: a) 0.5*bolt length b) 1.5*width of critical and potentially unstable blocks c) 2.0m (greater spacing difficult to attach fabric support)
Minimum spacing	0.9 to 1.2m
Minimum average confining pressure	Greatest of: a) For elements above the springline: either pressure = vertical rock load of 0.2 x opening width or 40 kPa b) For elements below the springline: either pressure = vertical rock load of 0.1 x opening height or 40 kPa c) Intersections: 2*confining pressure determined above

*The point of deviation from vertical in the upper sidewall.

2.6.7. Farmer and Shelton

Farmer and Shelton proposed alternative design rules for rockbolting within jointed rockmass structures in 1980 [28]. These are similar to those proposed by the US Corps of Engineers, but consideration is also given to the number of joint sets and their orientation relative to the excavation and rockbolt installation. See Table 2.16.

Table 2.16: Empirical guidelines for excavations <15m span.

Number of Discontinuity sets	Rockbolt design	Comments
≤ 2 inclined at 0 - 45° to horizontal	L = 0.3*B s = 0.5*L (depending on thickness and strength of strata). Install bolts perpendicular to lamination where possible with wire mesh to prevent flaking.	The purpose of bolting is to create a load-carrying beam over the span. Fully bonded bolts create greater discontinuity shear stiffness. Tensioned bolts should be used in weak rock, sub-horizontal tensioned bolts where vertical discontinuities occur.
≤ 2 inclined at 45 - 90° to horizontal	For side bolts: L > h.sinψ if installed perpendicular to the discontinuity. L > h.tanψ if installed horizontally. h is the distance of installation from the point of daylight of laminations in the sidewall that defines the largest unstable wedge. ψ is the inclination of laminations from vertical.	Roof bolting as above. Side bolts designed to prevent sliding along planar discontinuities. Spacing should be such that anchorage capacity is greater than sliding or toppling weight. Bolts should be tensioned to prevent sliding.
≤ 3 with clean, tight interfaces	L = 2 x s s = 3 – 4 x block dimension. Install bolts perpendicular to excavation periphery with wire mesh to prevent flaking.	Bolts should be installed quickly after excavation to prevent loosening and retain tangential stresses. Pre-stresses should be applied to create a zone of radial confinement. Sidewall bolting where toe of wedge daylights in sidewall.

2.6.8. Guidelines for Deep Mining

Recent work based on case studies in deep Canadian mines (Kaiser, McCreath and Tannant, 1996) [29] is another that examines the influence of seismicity on the design of tunnel support. The basis of the design recommendations is an evaluation of the potential for rockburst damage, and the mechanism of damage. This may vary from rockfall to violent ejection of the rockmass. The demand on a single support unit is evaluated, which thus allows an evaluation of the capacity of the support system as a whole. The design methodology for support selection is shown in Figure 2.10.

The demand on the support is a function of both the rockmass environment and the potential for seismic events. These indicate the mechanisms of rockburst damage as shown in Figure 2.11.

The support system requirements are based on envisaged damage mechanisms. The support capacity and means of interaction with the rockmass are thus defined as shown in Figure 2.12.

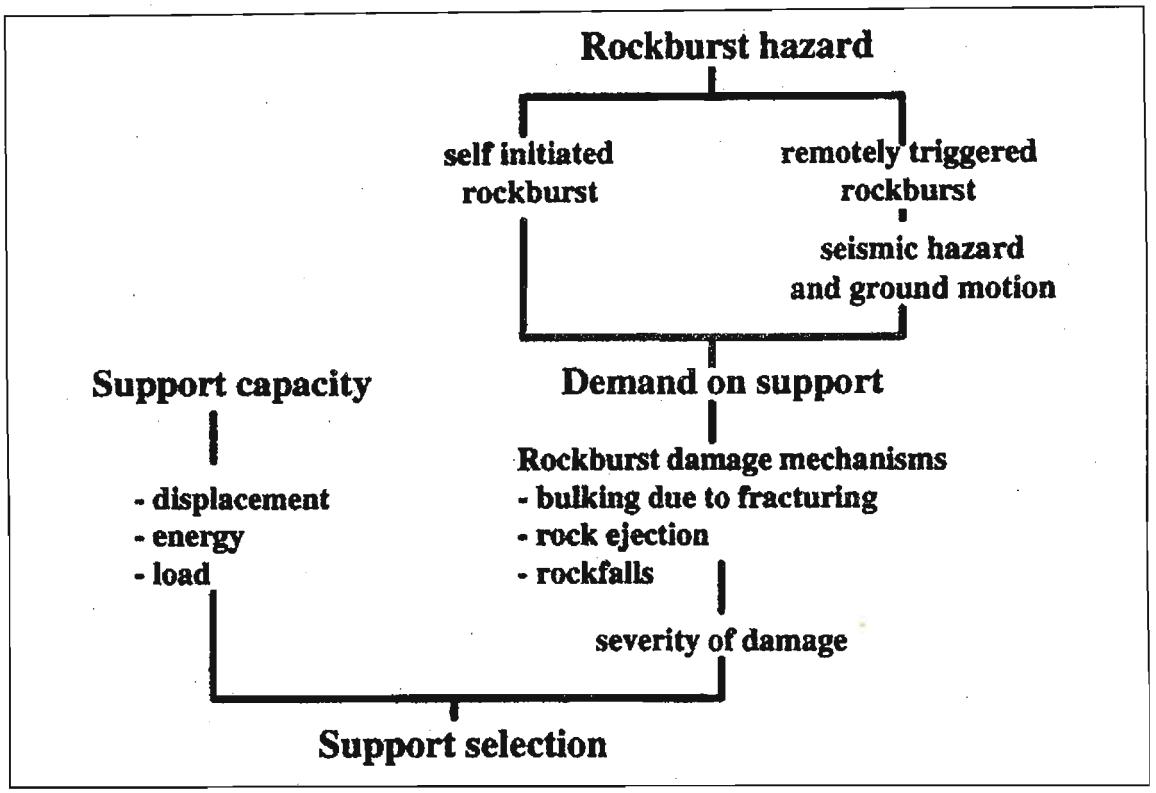


Figure 2.10: Approach to support design under rockburst conditions after Kaiser et al. (1996).

The capacity of typical support systems (rockbolts, mesh and shortcrete) was evaluated to determine the energy absorption capability of the systems compared to the envisaged demand. This knowledge can be utilised by the design engineer to assist in selecting suitable support systems. The methodology offers good insight into the demands on the support systems. However, it is assumed that the support is loaded as a system and thus a tributary area loading and average support resistance of the support system is considered. This assumption is generally valid under the more massive rockmass conditions of the excavation periphery in the North American mining environment, but differs significantly from cases where the interaction of the components of the support system within a highly discontinuous rockmass structure may result in significant differential loading.

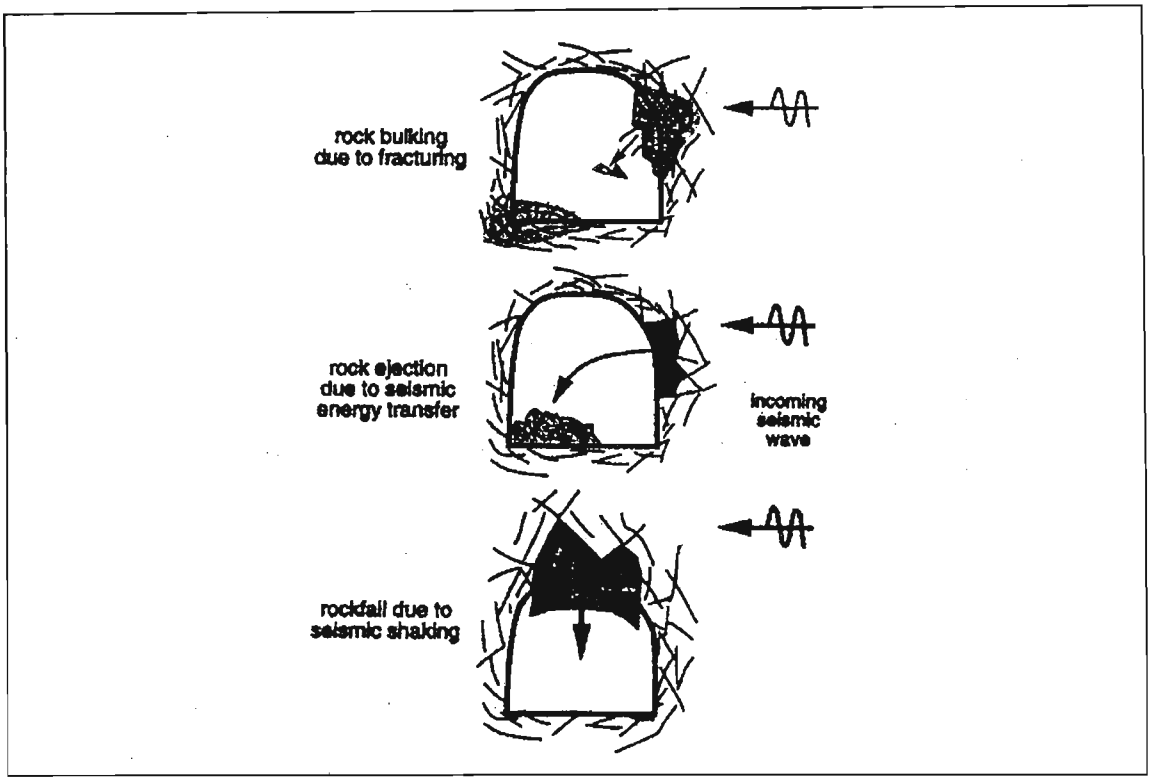


Figure 2.11: Rockburst damage mechanisms as defined by Kaiser et al. (1996).

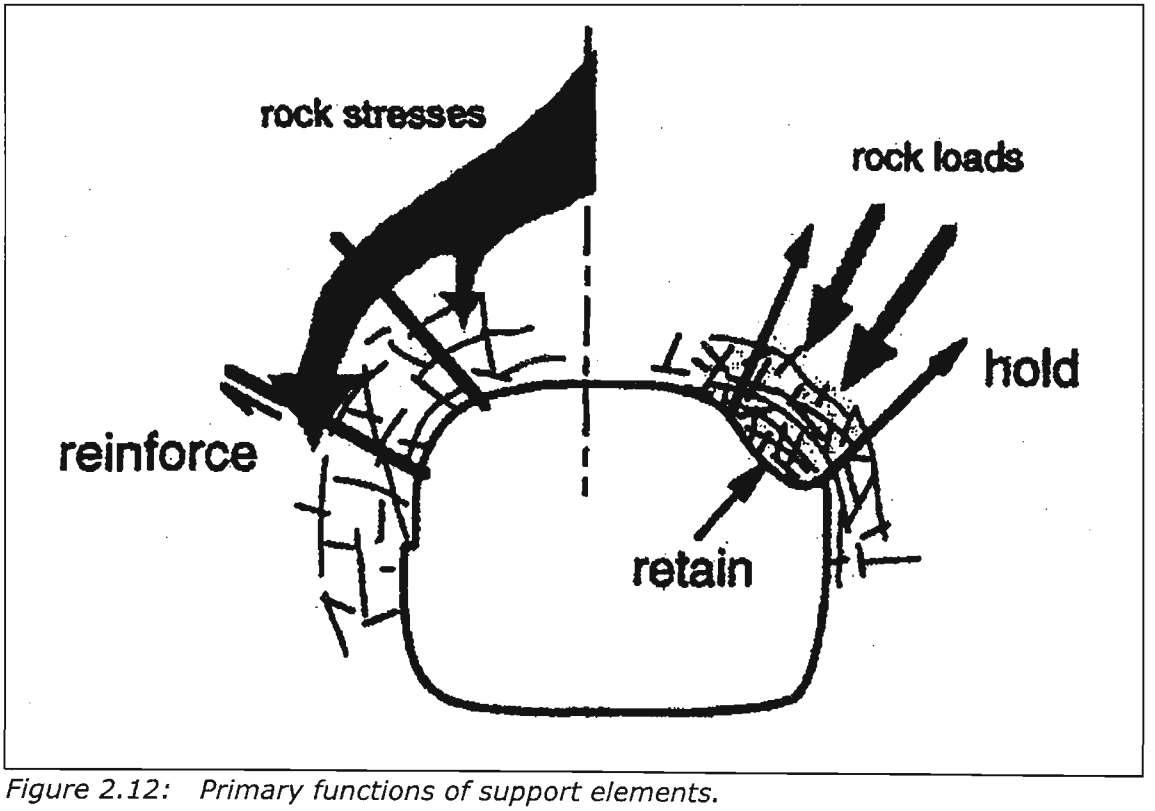


Figure 2.12: Primary functions of support elements.

2.7. SUMMARY

The designs of tunnel support systems are based principally on either mechanistic evaluation or empirical design rules.

Mechanistic evaluation allows the estimation of the demand on and characteristics of the support system based on an understanding of the interaction between the support and the rockmass. This allows an estimation of the load and deformation requirements of the specific components of the support system. However, these design considerations are currently based on relatively simple mechanisms and are generally only applicable to simple rockmass structures.

As the complexity of the rockmass environment increases, design procedures tend to be based more on empirical design rules derived from experience in similar rockmass environments. This design procedure requires some form of assessment, or classification, of the rock mass in order to define the comparable rockmass environments.

Most empirical classification systems currently in use have been derived from civil engineering experience. They therefore reflect a shallow, generally low, static stress environment and support systems that will ensure levels of excavation stability suitable for public use over long time frames. The applicability of these design procedures to the mining environment, and particularly deep-level mining, is limited due to the significant differences in the rockmass condition, the load characteristics, and the required level of support. Empirical design procedures that have been developed specifically for the South African mining environment have proved successful in the majority of rockmass conditions encountered. However, in more adverse rock mass conditions, where significant failure of the rockmass around a tunnel excavation has occurred, and where the tunnel is subjected to dynamic loading, then failure of the support systems is encountered. The large variation of rockmass environments within mining operations also leads to complex rockmass behaviour that is often not captured within the empirical design rules.

3. STRUCTURAL HEALTH MONITORING

3.1. OVERVIEW

The structural integrity, or health, of in-service equipment is essential to ensuring safe and reliable operating conditions, and is directly related to structural performance [32]. The timely detection and quantification of damage plays a crucial role in the assessment of structural health, as well as determining residual life based on usage. A typical health monitoring system should therefore have the capability of identifying the existence, size, location and severity of any damage. Information pertaining to the current condition of a structure allows personnel to confidently make optimal use of equipment and minimise downtime whilst avoiding catastrophic failure. Structural health monitoring (SHM) therefore has important safety and cost implications.

Fail-safe conditions require a substantial amount of monitoring. Since sensing and data processing technology, reliability and cost have improved tremendously over the past few decades, there has been an associated reduction in monitoring cost [31]. Research directed at smart SHM systems indicates that the traditional approach to SHM stands to be significantly enhanced by the implementation of smart systems for damage sensing and assessment.

3.2. TRADITIONAL SHM SYSTEMS

SHM systems typically acquire raw data from strategically located sensors placed within or on the structure. The data may be either automatically monitored and evaluated by a local system or fed to a remote location for processing. The processing of sensor data generally involves data conditioning and analysis with the aim of generating a quantifiable measure of the damage and possibly a prediction of the resultant degradation in structural performance.

A popular method of monitoring loads/stresses in a given structural entity is by bonding *strain gauges* to well-selected areas and measuring strain sequences. These sequences are then converted to stresses and/or loads, thereby facilitating fatigue evaluation, etc. Strain gauges are not suited to certain applications such as fully grouted rockbolts that potentially experience massive deformation at unpredictable locations.

Visual inspection is sometimes one of the quickest monitoring techniques for identifying potential problems (and psychologically important in critical situations)

in cases where e.g. ice, snow or fire might be affecting system performance, or where equipment might be incorrectly configured. This is also the case in mine tunnels [34] where a visual inspection of a particular area of a tunnel, based on experience, can help to identify potential problems. Human error, inspector fatigue and inability to examine inaccessible regions can lead to overly optimistic reports on structural health.

3.3. NON-DESTRUCTIVE TESTING

The traditional techniques described in Section 3.2 are limited by the fact that damage can only be predicted in this way and not accurately quantified. One means of quantifying damage is by performing non-destructive testing. Some common non-destructive techniques are described below.

1. *Acoustic emission* (AE) testing systems compare known acoustic signals to measured signals to determine structural health. Acoustic emission is the elastic energy suddenly released when materials undergo deformation. By applying stress to a structure/system, sources of failure can be located. AE can successfully monitor discontinuities, fatigue failures, materials behaviour, welds and stress corrosion cracking [31]. Handling, processing and interpretation of data has been improved through better computer technology.

2. *Acousto-ultrasonic* testing is a superior technique that can be used for characterising voids, cracks, inclusions and de-lamination in metals, ceramics, plastics and composite materials, although it is better suited to homogeneous-type structures [30]. Two probes are required, namely an ultrasonic transmitter (the actuator that emits a pulsed excitation) and an acoustic emission receiver (sensor). When the damaged area lies between the probes, the shape of the received acoustic signal changes because of the change in material damping characteristics.

3. The *Lamb Wave* technique also uses the actuator/sensor principle and is based on Lamb wave theory. A longitudinal and transverse plate wave is emitted into the structure. The ability of these waves to propagate over long distances is highly advantageous. However, much care is required to find the right angle for inducing the Lamb waves, especially when the structure to be monitored is of a geometric shape significantly different to that of a plate [31].

4. Vibration modes change due to a change in structural integrity. *Modal analysis* is suitable for monitoring damage such as displaced fasteners.

5. Advanced infrared thermal imaging cameras and data acquisition equipment have made *thermography* a viable source of damage assessment and health monitoring. Thermographic residual strength predictions have been found to be in agreement with finite element analysis predictions [30]. Thermal imaging is more effective with thin-wall structures.

6. A range of additional non-destructive techniques are available, including eddy currents (internal flaws); die penetrants (surface cracks); chemical sensing (corrosion); radiography (internal flaws); holography; shearography; Barkhausen noise; fibre optic sensors (strain); piezoelectrics (strain or acoustic signal monitoring); etc.

The term *non-destructive testing* generally refers to methods that require the item to be removed from service. In the case of mining rockbolts, this is not feasible since the bolt cannot be easily removed from its embedded location or simply re-installed after inspection.

3.4. SMART STRUCTURAL HEALTH MONITORING VS. NON-DESTRUCTIVE TESTING

Although considerable overlap may exist between traditional non-destructive testing (NDT) and smart SHM technologies, the two approaches differ significantly with respect to data acquisition and interpretation. Most importantly, smart structural health monitoring is a term used to describe *in-service* damage monitoring. Integrating SHM into a material or part has become an increasingly viable concept due to the significant progress achieved in sensor performance, miniaturisation and cost. The implementation of such a system requires development and demonstration by dedicated and qualified multidisciplinary teams, acceptance by designers, manufacturers and operators, and approval by the relevant authorities [32].

Damage assessment using NDT techniques generally requires the equipment undergoing inspection to be removed from service whilst sensors are placed for data acquisition. Direct measurement is usually employed to determine the physical condition of the structure, and in some cases no history data is required. The accuracy of the diagnosis depends on measurement resolution, which in turn relies on the accuracy of the technique and equipment used. In addition, the NDT

approach tends to be labour-intensive, requiring experienced and well-qualified inspection personnel. Both direct and indirect costs apply: direct costs refer to the cost of the NDT procedure; indirect costs relate to interrupted service resulting in a loss of production, or costs for leasing a temporary substitute, plus costs associated with the risk of dismantling, transporting and re-fitting the structure.

The primary motivation for developing smart SHM systems is that the monitored item is not removed from service until it is deemed to have reached the end of its working life. Costs are therefore reduced and productivity increased. Another strong motivation is that data acquisition and analysis can be automated and continuous, and does not interfere with other routines.

Damage assessment is usually based on the change in data measured at the same location at different times. Data is often automatically processed, either locally or remotely, to yield a quantifiable measure of the damage experienced by the structure. In some cases, a prediction of residual life may be computed against the damage assessment. The use of data history is therefore characteristic of the smart SHM approach. The accuracy of the structural health assessment depends on the sensitivity of the sensors and on the interpretation methods used. The potential direct benefits from such smart SHM systems as opposed to NDT include:

1. In-service, real-time monitoring and reporting thereby reducing direct and indirect costs associated with scheduled inspections.

2. Minimal human involvement, which reduces the dependency on specialised labour and the possibility of human error.

3. Automatic systems for data acquisition, reporting and analysis.

An integration of NDT techniques into smart SHM systems may also be beneficial, due to the technological maturity of the various NDT methods. The significant progress achieved in sensor performance, instrumentation miniaturisation, and equipment cost has made the integration of the NDT technologies within a smart SHM system entirely feasible for a multitude of applications.

3.5. SMART MATERIALS AND STRUCTURES FOR SHM

Smart materials encompass a very broad range of materials that react in a consistent manner to changing environmental (input) conditions. They can be compared to transducers, i.e. something that takes input energy in one form and converts it to output energy in another form. For this investigation, a smart

material that exhibits the inherent capacity to indicate the amount of deformation or damage that occurs in a structural item, specifically a mining rockbolt, is sought.

A *smart structure* is one that contains a class of materials known as *smart materials* and associated electronics and power supply [4]. In general, a smart structure can be broadly described as a non-biological physical structure having the following attributes [33]:

1. A definite purpose.
2. Means and imperative to achieve that purpose.
3. A biological pattern of functioning (an alternative definition might be: the consistent pattern of functioning of a Turing machine).

A Turing machine receives input, compares that input to its algorithm and internal state, and then either (i) produces an output, (ii) doesn't produce an output, (iii) produces an output and changes its internal state or (iv) doesn't produce an output and changes its internal state. A smart structure therefore senses the environment and decides whether or not to act upon itself and/or the environment based upon some control algorithm.

The smart structures concept is a multidisciplinary research area covering such fields as the mechanics of deformable bodies, nonlinear dynamics, control theories and materials science [4]. There are many different types of smart/sensor materials available that may be employed in the development of feasible smart SHM systems. Some are listed in Chapter 12 (Appendix A).

Adaptive computing, adaptive control systems and adaptive materials are at the forefront of smart structures research. Some smart systems currently under development are designed to react to given loading conditions and thereby increase operational performance, e.g. by flexing an aircraft wing in order to produce more lift. This investigation is primarily concerned with the use of smart materials for recording changes to key environment variables such as deformation (i.e. for measuring/monitoring purposes), specifically in mining bolt applications.

Measurement/monitoring materials fall into two classes:

1. Class I materials exhibit an inherent property that permanently changes with an increase in the environmental variable (e.g. peak deformation). Class I materials are used as the basis for *passive* monitoring systems.

2. Class II smart material behaviour is a function of the environment variable and can typically be reset. Class II materials are used as the basis for *active* monitoring systems).

Some active and passive systems can act as both structure and sensor simultaneously without any need for additional sensors. An advantage of a passive system is that information is retained regardless of power loss, etc., which is extremely useful for structural damage assessment after e.g. earthquakes.

In this study special attention is given to a particular passive smart material referred to as TRIP steel. TRIP steel exhibits the characteristics required of a candidate smart material for the mining bolt application.

3.6. SMART SHM TECHNOLOGY

In general, a smart SHM system must address five criteria:

1. *The sensor.* The nature of the application and the type of information required (i.e. the type of disturbance) determines which sensor is selected. For large structures an array of sensors may be required, measuring a single parameter or a multitude of parameters such as temperature and damage.

2. *Diagnostic signal generation.* Appropriate instrumentation must be connected to the sensor in some way in order to produce an output that can be interpreted.

3. *Signal processing.* Information needs to be compared to e.g. some control curve.

4. *Identification and interpretation.* Based on the history, the data must be interpreted for an assessment (either remotely or on location) of the damage to be made based on structural conditions (e.g. deformation and temperature).

5. *Integration.* The SHM must be integrated into the structure (or an array of structures), in some cases enhancing the structural integrity.

These criteria are described in greater detail in Sections 3.6.5 to 3.6.8. Selecting a suitable smart SHM system depends to a large extent on the type of structure that will be monitored and the data that will be required in order to assess structural

integrity. In general terms, a typical smart SHM installation consists of at least two major components:

1. The system hardware - usually comprising the sensors and instrumentation required for signal acquisition, conditioning and storage.

2. The software that carries out the analyses of the acquired data and thereby provides an assessment of the condition of the structure.

Smart SHM systems are generally designed to provide diagnostic data using one of two techniques: passive sensing with sensors only, or active sensing with actuators and sensors. An active sensor refers to one requiring electrical power and data acquisition/storage on a full-time basis, whereas a passive sensor collects information without requiring real-time electrical power and data acquisition/storage systems.

3.6.1. Passive Systems

Passive sensors act as "silent witnesses" to structural behaviour for interrogation at a later stage [34]. Damage is usually assessed by comparing sensor measurements (the peak value is inherently recorded) to a set of reference data. The reference data normally corresponds either to that of a healthy structure, or to various stages of damage/deformation in a similar structure. Because this approach relies entirely on data comparison, the provision of accurate reference data and comparison techniques is vital to the accuracy and reliability of the SHM system.

Table 3.1: Examples of passive sensors.

Sensor	Data provided	Operating characteristics
TRIP gauges (Chapter 8)	Peak deflection or strain	Metastable ferromagnetic sensor element signature correlates with peak deflection. Can be made active if warranted. Reset by replacing the sensor element. Adjustable range. ± 5% accuracy.
Mechatronic gauges	Peak deflection or strain	LVDT sensor output provides peak deflection. Can be configured to be reset and/or activated if warranted. Range limited by LVDT size and design. Accurate to within ± 2μm deflection.
Fibre-optic sensors		Highly accurate. Most based to signal transit time/reflected signals.
Note: Passive sensors are typically suitable for peak measurements. The SHM must be powered to interrogate the independently recorded sensor data.		

Passive sensors are invaluable for making damage assessments after instances where electrical power is lost (e.g. in blackouts or in mining tunnels where rockbursts may damage power cables).

3.6.2. Active Systems

With active sensing systems, known external mechanical or non-mechanical loads are applied to the structure through integrated actuators. The methodology is based on the premise that, since the inputs are known, any difference in the local sensor measurement based on the same input is directly related to a physical change in the condition of the structure. This therefore implies a requirement for consistent actuator action and sensor response in order to retain reliability and accuracy. There are a limited number of materials that can serve as built-in actuators for the purposes of structural health monitoring. The actuators must be small enough to be incorporated into the structure, but must be powerful enough to generate diagnostic signals that can be recorded by neighbouring active sensors (an active sensor requires electrical power and data acquisition/storage facilities [34]). The power transmission between the built-in actuators and the hosting structure must be fully understood and the diagnostic signals generated by the actuators must be controllable, repeatable, reliable, and must be sensitive to damage, defects, or any anomaly of particular concern. Research in this area has generally concentrated on the use of integrated piezoelectric films as both sensors and actuators in composite structures, although there is also significant interest being shown in the use of shape memory alloys in a similar manner (refer to Appendix A).

Table 3.2: Examples of active sensors.

Sensor	Data provided	Operating characteristics
Resistive strain gauges	Voltage fluctuation	Voltage correlates with strain/deflection. Highly accurate.
Vibrating wire gauges	Resonant frequency response	Frequency correlates with deflection.
Piezoelectric transducers	Resonant frequency response	Frequency response spectrum correlates with structural damage state.
Note: Active sensors typically require real-time power and data acquisition/storage systems.		

3.6.3. Passive vs. Active SHM

Table 3.3: Comparison of passive and active system characteristics.

Passive Systems	Active Systems
Power is only required during sensor interrogation. In the event of power failure, the sensor itself retains all information.	Full time power is required. In the event of power failure, data is lost.
Data is usually only interrogated after significant events, or at relevant or pre-determined time periods, but may be set up for continuous monitoring.	Data is constantly monitored and therefore instantaneously available. Data is stored in an extensive storage system.
Only successively higher levels of damage are recorded, i.e. a "peak" system.	Interpretation requires a full time history for data comparison.
Minimal data reduction is required to analyse and subsequently diagnose health.	Extensive data reduction is usually required before analysis and health diagnosis.
Less expensive. Capable of detecting minute, gradual changes within a structure.	Costly but highly accurate.
Robust and suitable for outdoor conditions.	Suitable for "friendly" environmental conditions.
Some systems provide additional stiffening. Cannot actively control damage levels.	Some sensors may be used as actuators to actively rectify damage.

Combining passive with active technologies is feasible if peak load times can be predicted, e.g. trains crossing a monitored bridge. This allows analysts to interpret current information as well observe any accumulated damage.

3.6.4. Selecting a Suitable Smart Sensor/Material

The primary requirement for a suitable sensor material is a reliable and consistent response over the entire measurement range, and insensitivity to variation in anticipated operational conditions such as heat, moisture, magnetic fields, etc. The particular choice of sensor technology will be dictated to a large extent by the structure itself and the characteristics of the data that will be recorded. Key factors influencing the choice of material include cost, accuracy, ruggedness, information and control requirements. Where damage control is needed (e.g. for fatigue control), an active sensor is required (such as piezoelectric, eloctrostrictive and magnetostrictive materials, and shape memory alloys (Chapter 12, Appendix A).

3.6.5. Sensor Arrays

Typically, smart SHM systems rely on a number of sensors (that may be of varying type) placed in strategic locations to record the required data against which structural assessments can be made. In most cases the responses of the individual sensors are recorded by centralised monitoring instrumentation that may be local or remote. The design and implementation of this installation would need to be such that it minimises the added complexity to the structure and does not interfere with the normal operation of the structure.

3.6.6. Signal Processing

In typical applications, the data retrieved directly from sensors contains substantial amounts of information, most of which is unusable and irrelevant. Furthermore, the data can be highly corrupted by the environment and noise. For these reasons, raw sensor measurements need to be processed before they can be used for interpretation.

Signal processing is usually carried out with dedicated hardware and software. It is crucial that the processing is consistent and reliable, since the processed data will be used to identify the condition of the structure. Signal uniformity is also essential. For example, measurements can be adversely affected by the properties of the interface between the host structure and the sensors, and it is also possible that in a network of distributed sensors, each sensor may produce a different output for the same given input. These issues need to be catered for and may involve fine-tuning and calibration of the system during installation and commissioning.

3.6.7. Damage Identification and Interpretation

The competence of a SHM system is dependent on the accuracy and reliability of the damage detection and identification. Sensor measurements record information at pre-determined points in continuous structures, and it is likely that damage or an abnormal condition may not appear at the exact location where the sensor is located. Therefore, sensor information needs to be extrapolated for damage that occurs at a distance away from the sensor locations. Furthermore, there are many other factors that may influence the sensor measurements other than defects, and it can become difficult to interpret the measurements in terms of the actual physical condition of the structure.

Mathematically, the determination of the physical condition of a structure based on sensor measurements is a nonlinear inverse problem. Several numerical and analytical techniques have been proposed or adopted for different applications. Most are limited to controlled laboratory environments and to simple structural configurations. In practice, an extensive amount of history data of the structure is required, which may be difficult to obtain. Research in this area has concentrated on the development of novel identification or interpretation algorithms that relate the sensor measurements to the physical condition of the structure in terms of damage and defects. Such quantifiable relationships provide the necessary foundation for the development of effective identification techniques for different applications. However, in many applications, the relationship between damage or defects and the measurable physical quantities is still to be established.

3.6.8. System Integration

Integration of the system into the structure requires both hardware and software. Ultimately the system must be reliable. It is a general concern that inclusion of sensors and actuators as part of the structure may influence or compromise the performance of the structure. Such influence should be minimised as much as possible if it cannot be avoided. In addition the system must be easy to use. It would be preferable if the proposed system could display through its initial interrogation the condition of the structure in a simple format, for example damage severity may be judged on one of the three coloured lights: green - safe to use; yellow- use with caution, needs inspection; and red - unsafe, do not use. More detailed information regarding the structural condition could be obtained should further interrogation be needed.

3.7. INTRODUCING SMART SHM SYSTEMS INTO NEW ENVIRONMENTS

Regulatory requirements, cost, reliability, ease of fabrication and implementation remain key issues in introducing smart SHM systems into any industry. The integration of such a system requires development and demonstration by dedicated and qualified multidisciplinary teams; acceptance by designers, manufacturers and operators; and approval by the authorities.

As an initial step to justifying the implementation of a smart SHM system, it is important to quantitatively determine the amount of manpower-related inspection effort and cost that can be replaced, as well as the cost for purchasing, operating,

maintaining and disposing of that system. In the case of mining excavations, warnings of impending failures have both an economical and a safety impact.

Reliability is a key factor when making a decision to implement a smart SHM system, in that any false information can have enormous economic and safety consequences. The SHM system is therefore required to be at least as reliable as manpower-related inspection. Human factors such as loss of alertness in cases of damage occurrences, and inspector fatigue in cases of long and tedious inspections, are important reasons to consider a smarter solution to inspection as an element of structural health monitoring.

In a mine tunnel support system, the support units are concealed within the rockmass. The current reliability of inspection and maintenance procedures is therefore directly influenced by poor accessibility as well as varying (unknown) modes of deformation. Where a stamp of approval for public/worker safety must be provided, in some cases this must be done without complete confidence. Hence implementing smart inspection systems as a means of maintaining structural integrity and safety conditions can be motivated for in terms of greater reliability and flexibility of inspection, and because quantitative data is required.

A smart monitoring system may also allow for a less conservative system to be installed in a given region, where those areas that undergo unfavourable rockmass changes can easily and quickly be identified and "beefed up".

3.8. MINIMISING STRUCTURAL INSPECTION AND REPAIR COSTS

In general, the cost of pre-smart SHM structural repair may be broadly divided into two groups. The first involves the direct costs associated with the implementation of the required repair work. The second involves the indirect costs associated with the temporary removal of the equipment from operation (e.g. the closure of a mineshaft). These direct and indirect costs may be minimised by adopting a suitable repair strategy. For example, direct costs may be minimised by repair at an early stage of damage development. Postponing the repair until a scheduled overhaul may reduce the indirect costs, and in this case it may become necessary to adapt operational usage to limit or even stop damage growth (not applicable to mining excavations where damage is determined by rockmass behaviour). Cost-effective decisions could be made based on an analysis of the quantitative data

relating to the severity and propagation rate of the damage, as supplied by a smart SHM system.

3.9. SUMMARY

The choice of system, sensor, instrumentation and software is heavily dependent on the type of structure that will be monitored and the type of assessment that is required. In the case of the smart mining bolt application, the focus lies on the provision of a strain sensing capability within the confines of a specific geometry and at specific locations within the structure. This implies that either the bolt itself should act as the sensor, or its design should be modified to cater for the implementation of a smart SHM system – in which case the design will be driven concurrently by both structural considerations as well as by the choice of sensor technology and type of system (active or passive). The associated instrumentation for data acquisition, conditioning, storage and analysis will depend on the monitoring requirements, i.e. continuous or scheduled, and will also be influenced by sensor technology and system type. In addition, any physical design changes of the smart bolt SHM system will need to consider installation limitations, allowable weight and size of the equipment and operational constraints.

It is likely that no total replacement SHM technology will ever be developed because of the inherent human satisfaction with conducting visual examinations. Smart SHM systems would rather augment the effectiveness of visual inspections by providing quantitative data from key locations.

4. STEEL MICROSTRUCTURES

In this study, attention is focused on specially processed materials called strain memory alloys (and specifically TRIP steels) that can be used simultaneously as structural component and sensor element in a structural health monitoring system. These alloys require controlled thermo-mechanical processing to develop and enhance the smart capabilities.

This Chapter introduces the effect of alloy composition on steel microstructures; and the relationship between microstructure and material properties.

4.1. BASIC STEEL PROPERTIES

Plain-carbon and alloy steels are the iron-based metals most commonly used by industry. The response of plain-carbon types to hardening is not good on heavier sections since they do not harden deeply [35]. As a result, uniform material properties throughout the piece are not obtained. In contrast, the more expensive alloy steels were developed partly because of their better response to heat treatment, and should therefore be used in a heat-treated condition. Where a component is not subjected to unusual stresses in service (or is not required to act simultaneously as a structural member and smart health monitoring sensor), a less expensive plain-carbon steel could probably suffice. Alloy content and thermal processing requirements are generally the principal factors governing the price of steel, an important factor to consider if expensive TRIP steels are to be considered as a replacement for structural components such as rockbolts. As the hardness of a particular alloy increases, so does the tensile and yield strength. Elongation and reduction of area, the mathematically unrelated measures of ductility, both decrease. Toughness, a measure of the energy required for mechanical failure due to impact (or the energy absorbed prior to fracture), also decreases. If the steel is to be machined, most grades can be done so economically at hardnesses up to about Rockwell C 32.

The most effective elements in the hardening of steels are the interstitial solutes, carbon and nitrogen. Other factors being equal, the higher the interstitial element content, the higher the hardness that can be developed. Carbon and nitrogen also play a vital role in interacting with dislocations, and in combining preferentially with some of the metallic alloying elements.

4.2. CRYSTALLINE PHASES AND PHASE EQUILIBRIA

A phase is that part of a material that is distinct from other parts in structure or composition. The high-strength and sensing properties of TRIP steels are developed by thermo-mechanically tailoring the phases.

4.2.1. Overview of Unit Cell Structures and Terminology

All metals, a major fraction of ceramics, and certain polymers crystallise when they solidify. Each phase is characterised by a specific arrangement of atoms (or ions) of a *unit cell* that is repeated at uniform intervals within each crystal. Figure 4.1 illustrates the 7 different unit cell arrangements. Differences between crystal systems (and thus material properties) relate to the arrangement, axial lengths and angles of the unit cells, as presented in Table 4.1.

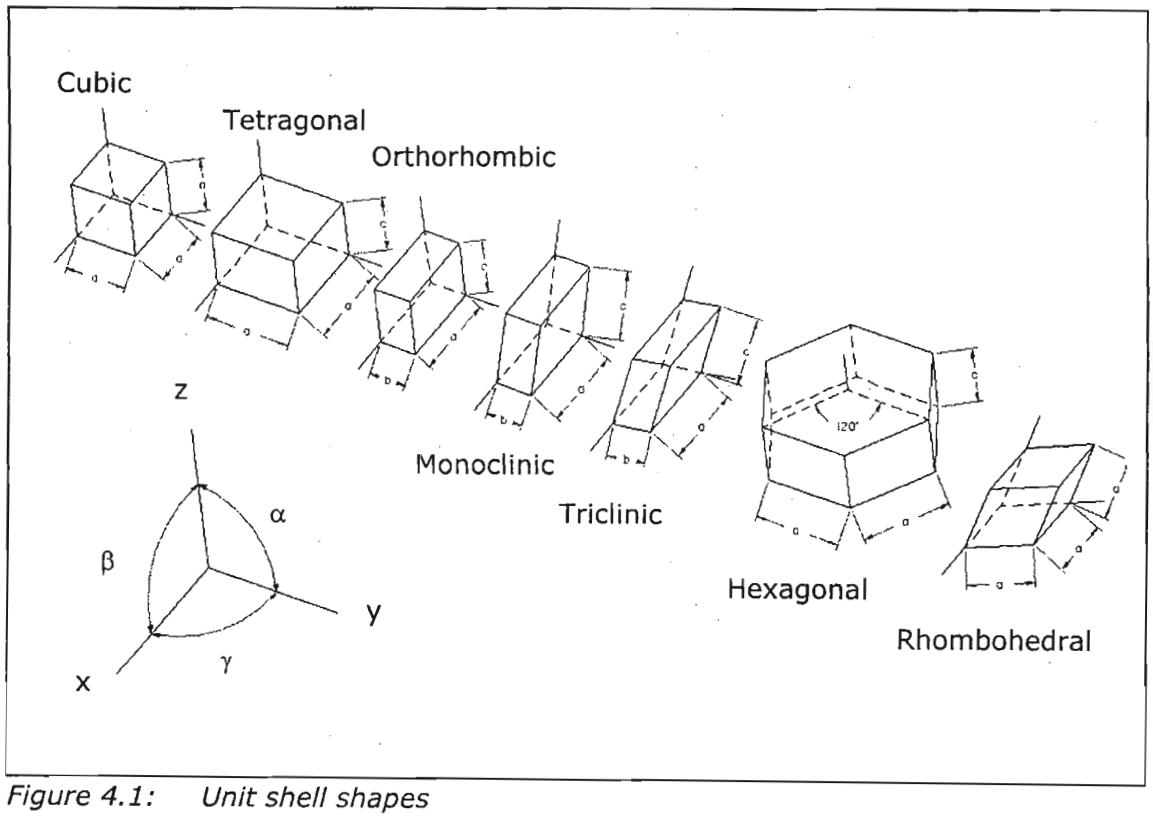


Figure 4.1: Unit shell shapes

Table 4.1: Crystal systems.

System	Axes	Axial Angles
Cubic	a = b = c	$\alpha = \beta = \gamma = 90°$
Tetragonal	a = b ≠ c	$\alpha = \beta = \gamma = 90°$
Orthorhombic	a ≠ b ≠ c	$\alpha = \beta = \gamma = 90°$
Monoclinic	a ≠ b ≠ c	$\alpha = \gamma = 90° \neq \beta$
Triclinic	a ≠ b ≠ c	$\alpha \neq \beta \neq \gamma \neq 90°$
Hexagonal	a = a ≠ c	$\alpha = \beta = 90°; \gamma = 120°$
Rhombohedral	a = b = c	$\alpha = \beta = \gamma \neq 90°$

Body-Centred Cubic Structures

Pure iron (Fe) crystallises with a BCC microstructure at temperatures up to 910°C. There is an iron atom at each corner of the unit cell, and another iron atom at the body centre.

Face-Centred Cubic Structures

In the FCC structure, in addition to the atom at each corner of the unit cell, there is an identical one at the centre of each face and none at the centre.

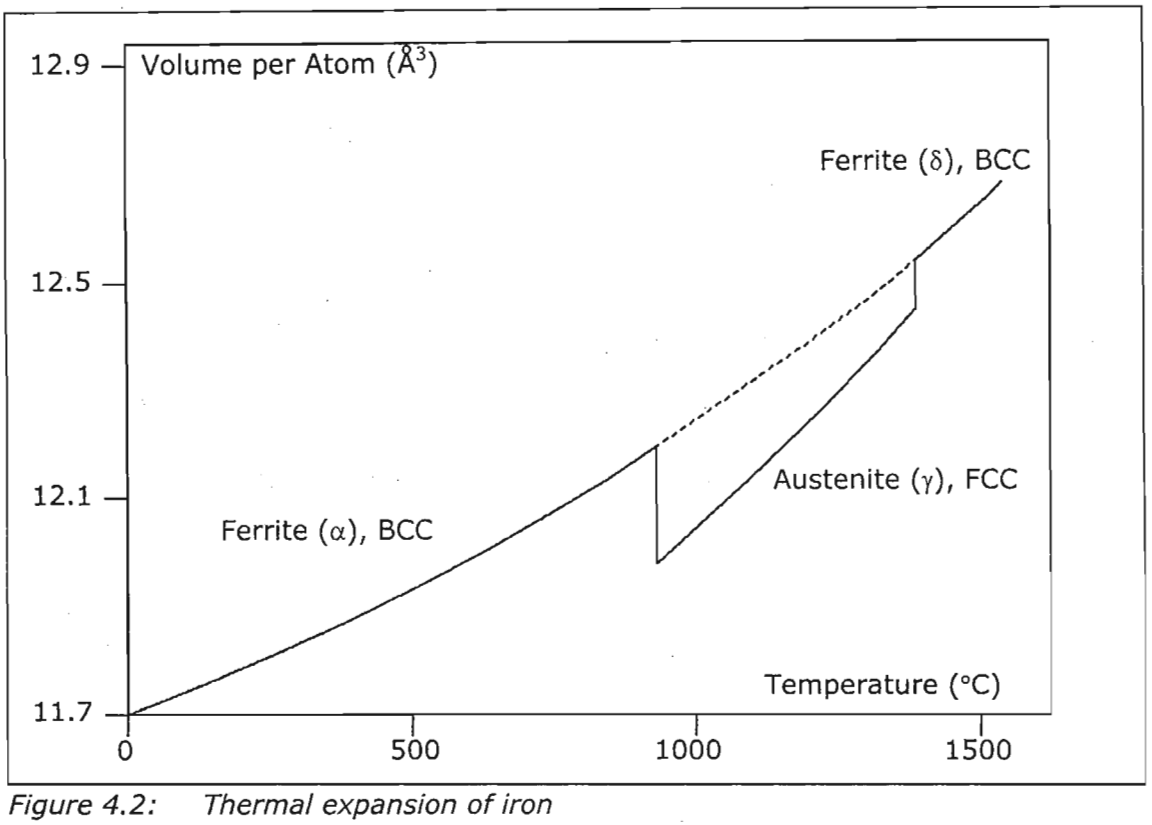


Figure 4.2: Thermal expansion of iron

Above 910°C, iron transforms from a BCC (ferrite, α) phase to a FCC (austenite, γ) phase. This transformation relates directly to the ability to heat-treat steel and modify its properties. FCC iron is stable up to 1390°C where it reverts to BCC before melting at 1536°C. Volume changes accompany the reactions (Figure 4.2) primarily because the FCC microstructure has a higher *packing factor* than BCC (the radius of the atoms is in fact slightly smaller in BCC than in FCC microstructures). The BCC→FCC transformation is accompanied by an atomic volume reduction of ~1% [36].

The α↔γ transformation of pure iron occurs very rapidly, so it is impossible to retain the FCC form at room temperature without the addition of other elements (notably including carbon), even though rapid quenching can substantially alter the

morphology [37]. If necessary, any detailed study of high temperature phases can be performed by X-ray diffraction. The transformation to ferrite on cooling can best be detected by precision dilatometry, which shows the phase transformation as a result of the volume change.

Simple Cubic Structures

BCC and FCC structures are replicated at the centre and face centres respectively. SC crystal structures are repeated at only full unit-cell increments. β'-brass is an alloy of 50% copper and 50% zinc. Although a Zn atom exists at the centre of 8 Cu atoms, the structure is SC and not BCC because the centre atom is different to the corner atoms (a Cu atom also exists at the centre of 8 Zn atoms).

Hexagonal Close-Packed Structures

In HCP structures, each atom in one layer is located directly above or below interstices among 3 atoms in the adjacent layers. Consequently, each atom "touches" 3 atoms in the layer below its plane, 6 atoms in its own plane, and 3 atoms in the layer above.

Other Non-Cubic Structures

The non-cubic crystals far outnumber all the BCC, FCC and SC materials combined. The distortion is due to inter-atomic bonds having different strengths and orientations.

Cementite, Fe_3C, is orthorhombic, with 12 iron atoms and 4 carbon atoms in each unit cell. The small carbon atoms are in the interstices among the larger iron atoms. This compound is found as a minor phase in almost all steels [37]. It greatly affects the mechanical properties of steel because Fe_3C is much harder and more rigid than are the other phases.

Crystal Directions and Anisotropy

Many material properties are anisotropic with respect to specific crystal directions. For example, the elastic modulus of BCC iron is greater parallel to the body diagonal than it is to the cube edge [36]. Conversely, the magnetic permeability of iron is greatest in a direction parallel to the edge of the unit cell.

4.2.2. Grains (Crystals) and Phases

The steel microstructure comprises polycrystalline phases. Microstructures vary with respect to size, shape and orientation of the grains. These variables are not wholly independent, because the shape and size of the grains are both consequences of grain growth. Likewise, grain shape is commonly dependent on the crystalline orientation of grains during growth.

Grains are separated by *grain boundaries*. The *number of grain boundaries per unit volume* (varies inversely with grain size) affects the behaviour (e.g. diffusion and nucleation) and characteristics (e.g. strength and corrosion) of a polycrystalline material.

The *orientation* of grains within a metal is typically quite random, but specific processes can create a preferred orientation (e.g. the [100] directions of iron have a higher magnetic permeability than do the other directions).

The driving force for *grain growth* is the energy released as an atom moves across the boundary from the grain with the convex surface to the adjacent grain with the concave surface. The boundary therefore moves in the opposite direction towards the centre of curvature of the surface. Since small grains tend to have surfaces of sharper convexity than do large grains, they disappear because they feed the larger grains. The net effect is grain growth. All crystalline materials – metal and non-metal – exhibit this characteristic of grain growth. Grain growth is heavily dependent on temperature. A temperature increase increases the thermal vibrational energy, which in turn accelerates the net diffusion of atoms across the boundary from small to large grains. A subsequent decrease in temperature slows down the boundary movement, but does not reverse it. The only way to refine (reduce) the grain size is to deform the grains plastically (work/strain-hardening) and to start new grains by recrystallisation through annealing.

4.2.3. Solutions and Mixtures

Different components can be combined into a single material as a solution or a mixture. A *solution* is a single phase consisting of more than one component (e.g. the austenite phase of plain-carbon steel comprises FCC Fe and interstitial C atoms). A *mixture* is a material with more than one phase, e.g. cementite comprises BCC ferrite crystals (iron with limited interstitial C atoms) interspersed amongst iron carbide crystals.

Interstitial Solid Solutions

A sufficiently small atom may be located in the interstices (gaps) between larger atoms. For iron in FCC form a relatively large hole exists at the centre of the unit cell. Carbon, being an extremely small atom, can move into this hole to produce a solid solution of iron and carbon, albeit with a certain amount of lattice distortion. The presence of even a small concentration of carbon (e.g. 0.1 to 0.2wt%) has a great strengthening effect on iron [37]. Nitrogen also enter the iron lattice as *interstitial* solute atoms.

Table 4.2: Atomic sizes of non-metallic elements in iron.

Element	Atomic Radius, r (Å)	$\frac{r}{r_{Fe}}$
α-Fe	1.28	1.00
B	0.94	0.73
C	0.77	0.60
N	0.72	0.57
O	0.60	0.47
H	0.46	0.36

The interstices between BCC iron atoms are much smaller, limiting the solubility. Separate phases consisting of ferrite and iron carbide, are therefore formed on cooling an FCC solution of carbon in iron below 910°C. The marked differences in solubility of the main interstitial solutes in FCC austenite (γ) and BCC ferrite (α) are of profound significance in the heat treatment of steels, and are fully exploited to increase strength and other physical properties.

Table 4.3: Solubility limits of carbon and nitrogen in γ- and α-iron.

	Temperature (°C)	Solubility (wt%)
C in γ-Iron	1150 723	2.04 0.80
C in α-Iron	723 20	0.02 <0.00005
N in γ-Iron	650 590	2.8 2.35
N in α-Iron	590 20	0.10 <0.0001

Substitutional Solid Solutions

In contrast to carbon and nitrogen, the metallic alloying elements such as manganese, nickel and chromium have much larger atoms, i.e. nearer in size to those of iron, and consequently they enter into *substitutional* solid solution.

Substitutional solid solutions form most readily when the solvent and solute atoms have similar sizes and comparable electron structures. As the difference in atomic size increases, less substitution can occur. Adding alloying elements in excess of the *solid solubility* must result in another phase, so that the resulting alloy is a *mixture* of distinct phases.

Diffusion

The atoms in a solid are not static. The average kinetic energy of the atomic vibrations increases roughly in proportion to the absolute temperature. Adding thermal energy increases the mean atomic distances, leading to thermal expansion. Atoms in a lattice diffuse to surrounding areas by either squeezing between adjacent elements (interstitially) or by breaking existing bonds and forming new bonds with adjacent neighbours (substitutionally). The energy required for an atom to change position is the *activation* energy. Carbon diffuses interstitially in FCC γ-iron at 1000°C; nickel diffuses substitutionally.

At a given temperature, there is a statistical distribution of atomic energies in a particular material. As the temperature increases, the average energy increases, as does the number of atoms exceeding some elevated specified energy value. The atoms with higher energies can introduce changes within solids by breaking old bonds and joining new neighbours, so that a small fraction of high-energy atoms relocate themselves in the lattice (e.g. *self-diffusion* of iron in iron). The fraction also depends on how tightly the atoms are bonded in position. For comparable material types (e.g. steels), there is an inverse relationship between the strength of the atomic bonds (indicated by the melting temperature, T_m) and the thermal expansion coefficient, α.

Diffusivity, D, of a given solute in a given solvent can be calculated from the following exponential relationship [36]:

$$D = D_0 e^{-E/kT} \qquad (4\text{-}1)$$

Where E is the activation energy, D_0 is the frequency factor (independent of temperature), k is the Boltzmann constant (13.8e-24 J/atom.K), and T is the temperature (K).

Table 4.4: Constants for diffusivity calculations.

Solvent	Solute	D_0 (m^2/sec)	E (J/atom)
γ-iron	C	0.2e-4	0.236e-18
	Fe	0.18e-4	0.445e-18
	Ni	0.77e-4	0.465e-18
	Mn	0.35e-4	0.469e-18
α-iron	C	2.2e-4	0.204e-18
	Fe	2.0e-4	0.400e-18

Using equation (4-1) it is readily shown that substitutional atoms move several orders of magnitude slower than interstitial ones. Alloys comprising substitutional elements therefore require homogenisation treatments to be prolonged and performed at higher temperatures.

It should also be noted that at a given temperature, diffusion occurs much more readily in BCC α-iron than in FCC γ-iron (in spite of BCC iron having significantly lower solubility limits). This is because α-iron is more loosely packed than γ-iron.

Phase Distribution

Microstructures with 2 or more phases (i.e. a mixture) possess a variety of geometries characterised by:

1. Relative amounts of the various phases.
2. Phase distributions (the minor phase can be dispersed or present as a grain-boundary network).
3. Various sizes and shapes of the different phases in the mixture.

A *supercooled* phase is a phase that has been cooled below its range of full equilibrium, so that changes to the microstructure (separation) can be expected. *Precipitation* describes the separation of a minor phase from a supercooled solid solution.

γ-iron containing 0.02wt% C is substantially supersaturated with carbon if quenched to 700°C. This supersaturated solid solution is not stable because of the ease with which carbon can diffuse in α-iron. Carbon begins to precipitate as iron carbide, Fe_3C. Because of the high hardness of Fe_3C, this process is referred to as precipitation hardening (or age-hardening).

Solid solutions can be cooled sufficiently rapidly that a *supersaturated* alloy can be retained as a single phase at ambient temperatures. This retention is possible

because for a new phase to be nucleated, and for atoms to diffuse to the grains of such a new phase, *time* is required. Solubility limits (or phase transformations) with respect to time and temperature can be plotted as Time-Temperature-Transformation (TTT) curves (described in Section 5.2).

4.2.4. Imperfections in Crystalline Solids

Inherently, disorder is present in all materials. In metals, disorder manifests as structural imperfections. These irregularities can be used and manipulated to develop useful and desired properties.

1. Point defects. Gaps/vacancies arising in a crystal due to imperfect packing or thermal vibrations.

2. Line defects (dislocations). An *edge dislocation* is a plane of extra atoms "wedged" into a crystal structure creating zones of tension and compression. A spiral *screw dislocation* has an imperfection line down its axis, with shear stresses associated with the atoms adjacent to the screw dislocation. The stresses result in a net energy increase in these regions. Shear deformation introduces both edge and related screw dislocations.

3. Surfaces. Atomic co-ordination at a surface is not fully comparable to the atoms within a crystal. Having neighbours on only 1 side, they have higher energy and are less firmly bonded than the internal atoms.

4. Grain boundaries. These are the disordered transition zones (mismatches) between two adjacent grains each comprising uniformly oriented unit cells. The less efficiently packed atoms along the boundary have a higher energy than those within the grains, promoting nucleation and diffusion. Grain boundaries also modify the progression of dislocation movements, and thereby the plastic strain of a material.

5. Compositional imperfections (impurities) may involve substitutional or interstitial atoms.

4.2.5. Nucleation

A new grain or phase does not automatically form in a supersaturated solution; it must be nucleated [36]. That is, a "seed" must exist before growth can start. Nucleation requires external energy to form a phase boundary, but can occur

unaided - albeit with delays – if an imperfection, an impurity, a grain boundary or an inoculant serve as a seed for reaction initiation.

At low temperatures, *homogeneous* nucleation can be triggered by a few atoms statistically clustering together non-spontaneously as a nucleus with initial radius r. Only once this radius exceeds a critical value r_c does the growth proceed automatically. With added growth, the new phase attains stability. Atom movements are sluggish at low temperatures, so growth is generally slow.

Irregularities in crystal structure, such as point defects and dislocations, possess strain energy that facilitates *heterogeneous nucleation* by reducing the external energy requirements and the critical radius value. So-called *inoculants* that do not participate in the reaction (e.g. container surfaces) can also facilitate nucleation.

4.2.6. Homogenisation

In order to obtain uniform material properties in a given batch or sample, *homogenisation processes* are performed. The alloy is heated to a temperature just below melting point and *soaked* (or *extended annealed*) so that the atoms can move away from regions of higher concentration to regions with deficiencies.

4.3. STRENGTHENING STEEL

Pure iron, a relatively weak material, can be strengthened by several basic mechanisms, the most important of which are:

1. Work hardening.
2. Solid solution strengthening by interstitial atoms.
3. Solid solution strengthening by substitutional atoms.
4. Refinement of grain size.
5. Dispersion strengthening, including lamellar and random dispersed structures.

4.3.1. Work Hardening

Work- or strain-hardening provides opportunities for the engineer to modify and control the properties of ductile metals. The material is stressed beyond its yield strength (i.e. plastic deformation) and the applied load then removed. Even plain-

carbon steels can have their strength levels significantly raised (greater effect with higher carbon content) simply through work hardening at temperatures below the *recrystallisation* temperatures. With cold working, the number and length of dislocations increases. The dislocation entanglements increase the critical shear stress and therefore the strength of the material [37]. The yield strength is also increases (heavily dependent on temperature), but ductility is reduced.

During deformation, slip occurs on several planes of the unit cell of BCC α-iron ({110}, {112} and {123}), but always in the close-packed [111] direction that is common to each of these planes. The diversity of slip planes leads to irregular wavy slip bands in deformed crystals.

Recrystallisation is the process of growing new crystals from previously deformed crystals. Materials that have been plastically deformed have more energy than do unstrained materials because they are loaded with dislocations and point imperfections. Given the opportunity, the atoms will move to form a more perfect, unstrained array. This occurs during e.g. annealing, whereby the material is heated and then cooled in a furnace so that a re-ordering of the atoms into less distorted, softer grains occurs. The number of dislocations is significantly reduced and the hardness drops. The *recrystallisation temperature* is that temperature where the microstructure and strength change drastically. The recrystallisation temperature is dependent on the amount of strain hardening, dropping as the amount of strain hardening is increased because of the higher stored energy available.

Cold work reduces the amount of plastic deformation a metal can undergo during a subsequent shaping operation. The hardened, less ductile metal requires more power for further working and is subject to cracking.

4.3.2. Interstitial Solid Solution Strengthening

Interstitial atoms already have associated strain fields around them. When a test specimen is loaded, the interstitial atoms (carbon) move within the dislocation strain fields (i.e. saturate the dislocations). In extreme cases this can lead to lines of the interstitial atoms along the cores of the dislocations, which locks the dislocations in position. The stress required to cause subsequent dislocation movement is substantially raised and there is an overall reduction in the total strain energy.

Only a very small concentration of interstitial atoms is needed to produce locking in annealed iron. Hence even small amounts of carbon and/or nitrogen in iron raises the yield strength and leads to a sharp transition between elastic and plastic deformation. An applied load raises the stresses in the material to an upper yield point, where the load then drops dramatically (where the dislocation movement resumes) with an accompanying decrease in the stress to a lower yield point. See Figure 4.3(b). There is a subsequent horizontal section on the stress-strain curve where the plastic deformation propagates at a front (the *Luders band*). The Luders band moves uniformly along a tensile specimen, and in complex arrays (called *stretcher strains*) in components where stresses are not uniaxial. The presence of the yield point in a steel can be detrimental to its behaviour, for example in pressings where complex patterns of Luders bands can produce rough surfaces and lead to poor workability. When the entire gauge length of a tensile specimen has yielded, general work hardening commences.

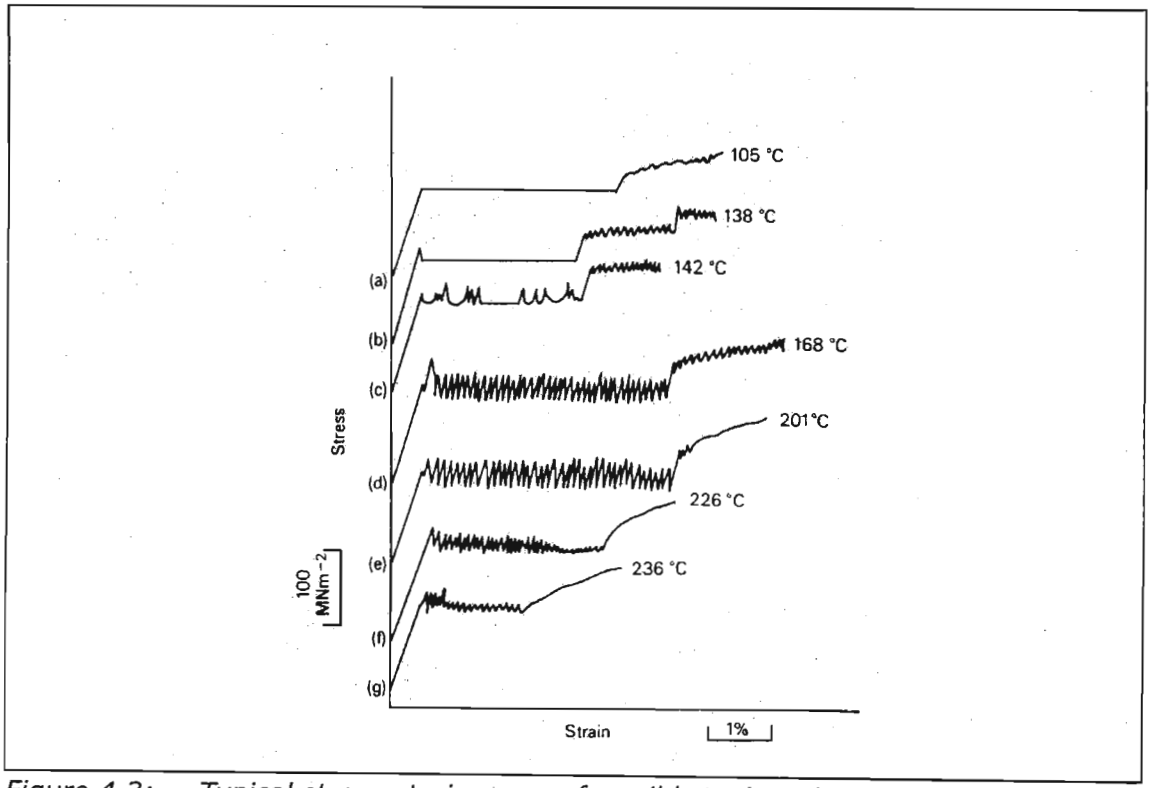


Figure 4.3: Typical stress-strain curves for mild steel at elevated temperatures.

At elevated temperatures, the horizontal zone beyond the yield point on the stress-strain curve becomes increasingly serrated (Figure 4.3). This is referred to as *dynamic strain ageing*, in which the serrations represent the replacement of the primary yield point with numerous localised yield points. These arise because the temperature is high enough to allow interstitial atoms to diffuse during deformation and cluster around new dislocations. Steels in these conditions exhibit low ductility

partly because of (i) the high dislocation density, and (ii) the nucleation of carbide particles on those dislocations where the carbon concentration is high. The phenomenon is often referred to as *blue brittleness* (item 9 in Figure 5.2; the dislocation density increases substantially here, resulting in increased work-hardening rates).

The *rapid* quenching of austenite with a high carbon content up to ~10% results in a phase transformation not to ferrite but to a body-centred tetragonal (BCT) phase called *martensite*. This transformation forms via a diffusionless shear transformation. If the quench is sufficiently rapid, the martensite is essentially a supersaturated solid solution of carbon in a BCT iron matrix. Because the carbon content can greatly exceed the equilibrium concentration in ferrite, the strength is raised substantially. High-carbon martensites are normally very hard but brittle. The increase in strength (including yield strength) can be attributed partly to (i) increased interstitial solid solution hardening, and also to (ii) the high dislocation density that is characteristic of martensitic transformations in iron-carbon alloys. The martensite phase can be *tempered* (described later) to retain a proportion of the strength and significantly increase the toughness and ductility of the steel.

4.3.3. Substitutional Solid Solution Strengthening

Many metallic elements form solid solutions in austenitic and ferritic iron. These are invariably substitutional solid solutions, with large variations in strength. The relative strengthening of a given alloying element may alter with varying test temperatures, as well as with concentrations of interstitial solute and the presence of other phases.

Alloying elements are usually added for reasons other than to increase strength (e.g. Si to achieve deoxidisation, Mn to combine with sulphur, etc.); therefore the solid solution hardening effect can be viewed as a useful bonus.

4.3.4. Effect of Grain Size on Strength

The refinement of the grain size (increasing the number of grain boundaries) provides one of the most important strengthening routes in the heat treatment of steels. Although grain size is often referred to in terms of diameter, very few grains are spherical. The mean chord length L_M is an index to represent grain size [36]. Determined by placing a random line of known length across a polished and etched microstructure, the mean chord length, L_M, is the reciprocal of the number of intersection points of a unit-length line with grain boundaries, P_L.

$$L_M = \frac{1}{P_L} \tag{4-2}$$

The Hall-Petch relationship describes the effect of grain "diameter", L_M, on yield stress, σ_Y, for a wide variety of steels, as well as for many non-ferrous metals and alloys:

$$\sigma_Y = \sigma_0 + \frac{k_Y}{\sqrt{L_M}} \tag{4-3}$$

where σ_0 and k_Y are constants. σ_0 can be regarded as the yield strength of a single crystal and is particularly sensitive to temperature and composition. Figure 4.4 shows the linear relationship between σ_Y and $\frac{1}{\sqrt{L_M}}$ for 3 test temperatures. The k_Y term represents the slope of the plot and is insensitive to temperature, composition and strain rate.

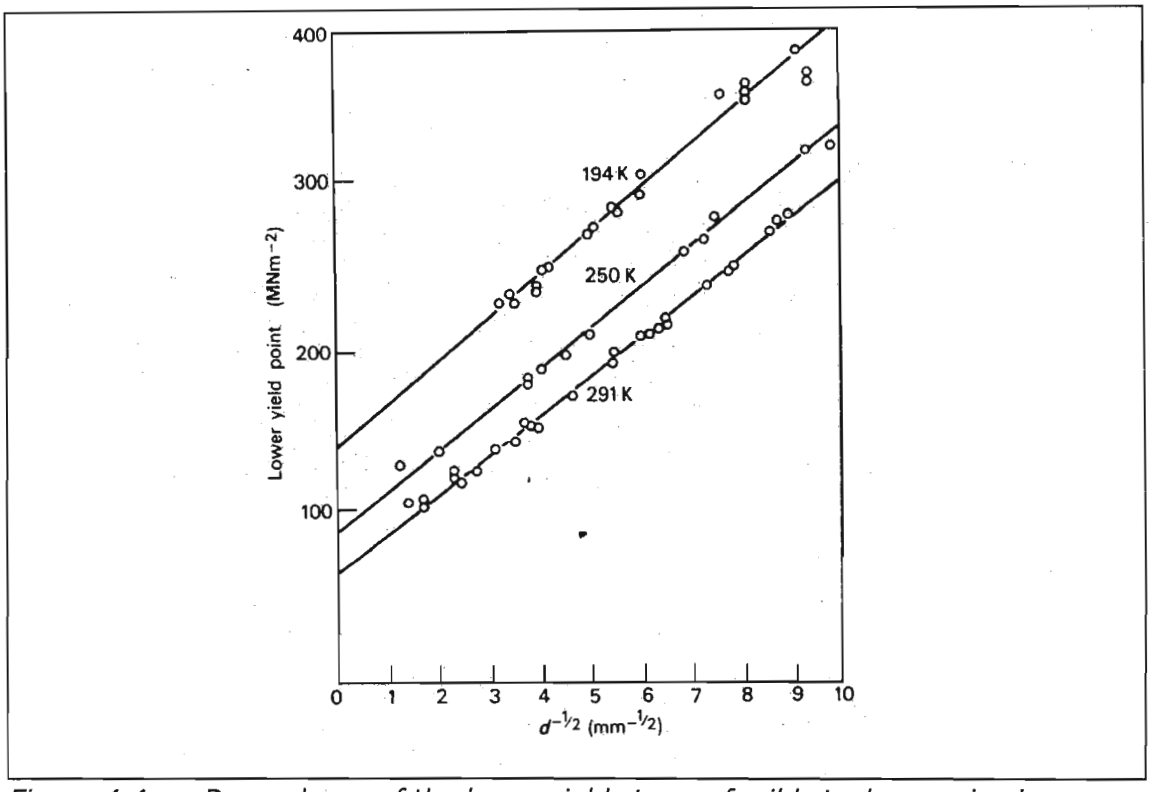


Figure 4.4: Dependence of the lower yield stress of mild steel on grain size.

The grain size effect on the yield stress can be explained by assuming that where a dislocation source (load) operates within a crystal, dislocations move and eventually pile up at the grain boundary. The pile-up causes a stress to be generated in the adjacent grain which, when it reaches a critical value, creates a new source in that grain. This can be observed macroscopically by the passage of

a Luders band. The grain size determines the distance dislocations have to move to form grain boundary pile-ups, and hence the number of dislocations involved. With larger grain sizes, the pile-ups will contain larger numbers of dislocations that will in turn cause higher stress concentrations in neighbouring grains. The shear stress τ_i at the head of a dislocation pile-up is equal to $n*\tau$, where n is the number of dislocations involved and τ is the shear stress on the slip plane. This means that the coarser the grain size, the easier it will be to propagate the yielding process. In other words, the finer the grain size, the higher the resulting yield stress. While a coarse grain size L_M of 0.25mm gives a yield stress of ~100MPa in mild steels, grain refinement of L_M to 0.0025mm raises the yield stress to over 500MPa.

4.3.5. Dispersion Strengthening

Often several phases can be recognised in a certain steel microstructure. The matrix can be strengthened to a considerable degree by controlling the dispersions of additional phases in the ferrite (BCC) or austenite (FCC) matrix.

The most common other phases are the carbides that form as a result of the low solubility of carbon in α-iron. In plain carbon steels this carbide is normally Fe_3C (cementite), which can occur in a wide range of forms ranging from coarse lamellar (e.g. in pearlite) to fine rod or spheroidal precipitates (as in tempered steels, described later). In alloy steels, the same wide range of structures is also encountered, except that in many cases iron carbide is replaced by other carbides that are thermodynamically more stable.

Most dispersions lead to strengthening, but often they can have adverse effects on ductility and toughness. The yield strength increases as the spacing between the particles decreases (e.g. as the lamellar layers of ferrite and cementite in pearlite are refined). Further heat treatment coarsens the dispersion, and the strength of the alloy falls.

4.4. THE CARBON-STEEL PHASE DIAGRAM

Phase diagrams are a collection of curves illustrating the unique phase equilibria (solubility limits) of a given alloy as a function of composition and temperature [35].

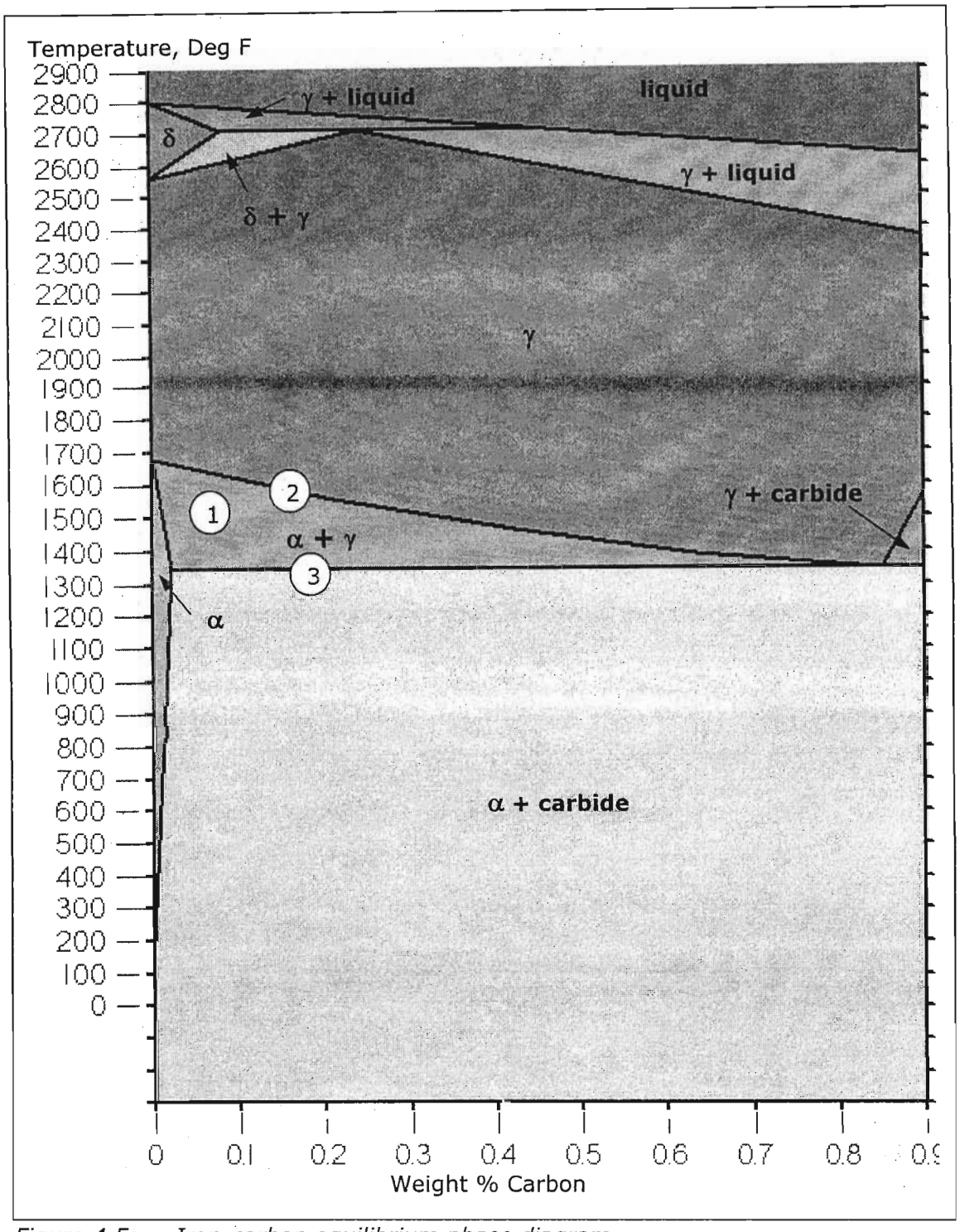


Figure 4.5: Iron-carbon equilibrium phase diagram.

Because TRIP steels have been identified as strong candidate materials for a smart mining bolt, the plain-carbon steel phase diagram is used as a starting basis to describe how the mechanical properties can be optimised through suitable heat treatments and additional alloying elements. The solubility limits illustrated in Figure 4.5 (material at equilibrium, i.e. no metastable phases) vary as the weight percent of carbon increases from zero. Additional alloying elements alter the form of the plain-carbon steel phase diagram.

Invariant Reactions

Eutectoid, eutectic and monotectic reactions are invariant heating/cooling reactions that appear frequently in phase diagrams at fixed temperatures. They share a common feature - 2 new phases form simultaneously during *cooling*. Peritectic, peritectoid and syntectic reactions are characterised by the simultaneous formation of 2 new phases during *heating*.

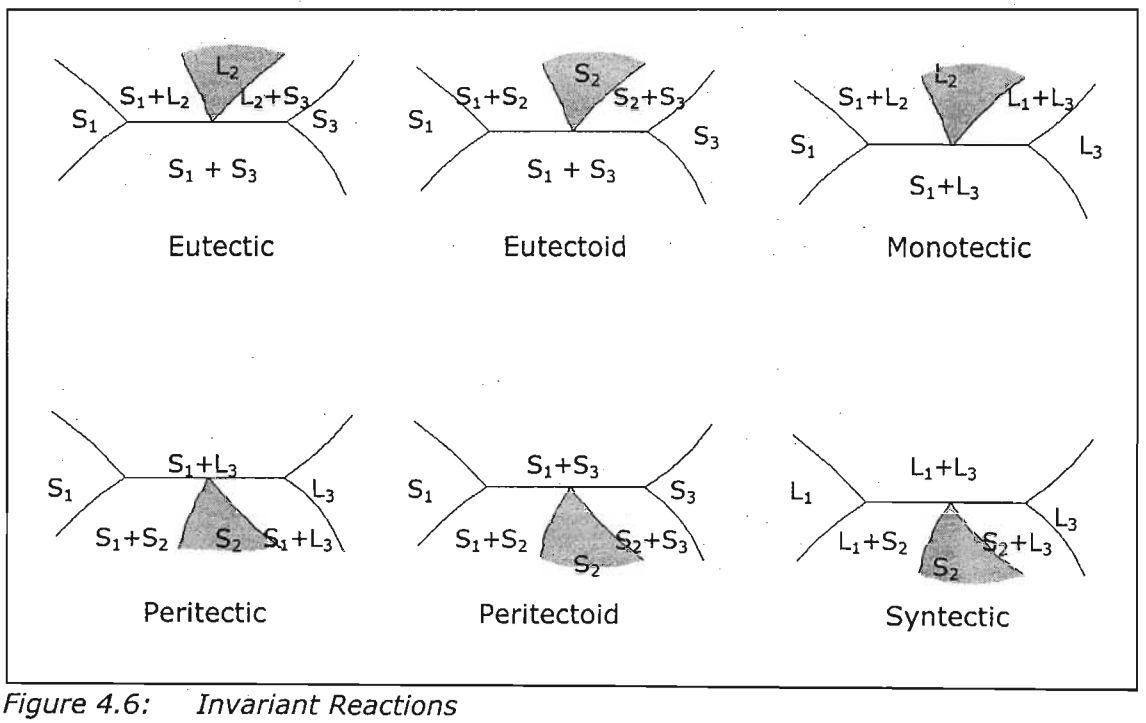


Figure 4.6: Invariant Reactions

Eutectic: $L_2 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_1 + S_3$

Eutectoid: $S_2 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_1 + S_3$

Peritectic: $S_1 + L_3 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_2$

Peritectoid: $S_1 + S_3 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_2$

Monotectic: $L_2 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_1 + L_3$

Syntectic: $L_1 + L_3 \underset{\text{heating}}{\overset{\text{cooling}}{\rightleftharpoons}} S_2$

Critical Range

The critical range (item 1 in Figure 4.5) is the region bounded by A_1 and A_3, described below. The upper and lower limits are easily detected by thermal analysis or dilatometry during cooling or heating cycles, and some hysteresis is observed. Steels undergoing temperature changes through this region experience changes to their internal atomic structures that radically affect the material properties.

Within the critical range is the A_2 Curie Point (item 14 in Figure 5.2), the temperature where iron changes from the ferro- to the paramagnetic condition. This temperature is 769°C for pure iron.

Lower Critical Temperature

In the plain-carbon steel phase diagram, the lower critical temperature (item 2 in Figure 4.5) is a horizontal 723°C curve termed A_{c1} on heating (chauffage) and A_{r1} on cooling (refroidissement).

Below A_{c1} the structure ordinarily consists of ferrite (α) and pearlite (a lamellar mixture of α and cementite, Fe_3C), or pearlite and cementite. On heating through A_{c1} the ferrite, pearlite and cementite begin to dissolve in each other to form austenite (γ). This dissolving action continues on heating through the critical range until the solid solution is complete at the upper critical temperature A_{c3}.

Upper Critical Temperature

Item 3 in Figure 4.5, and termed A_{c3} on heating, A_{r3} on cooling. Above A_3 and at equilibrium, pearlite and cementite have completely transformed to austenite, which coarsens with increasing time and temperature. For plain-carbon steel, A_3 decreases as the carbon content increases to 0.85% (at the *eutectoid point*, A_3 meets A_1) before increasing again for higher carbon-content steel.

Ferrite

Ferrite is practically pure Fe (very slight solid solubility for carbon; maximum 0.02wt% C at 723°C) designated as α below the lower critical temperature, and as δ at elevated temperatures. A BCC phase, it is ferromagnetic.

Austenite

Austenite is the soft, non-magnetic (essential property for TRIP steels) FCC form of iron that has the ability to dissolve carbon and alloying elements interstitially. The high solubility of carbon in γ-iron is of extreme importance in heat treatment, when solution treatment in the γ-region followed by rapid quenching to room temperature allows a supersaturated solution of carbon in iron to be formed.

Cementite

Cementite (Fe_3C) is a chemical compound of Fe and C. It is a hard phase and less readily deformable than either ferrite or austenite.

Pearlite

On slowly cooling austenite containing less than 0.85% C below A_{r3}, *hypo-eutectoid* ferrite begins to form, "ejecting" the carbon atoms. This continues through the critical region until at 723°C (A_{r1}) the remaining austenite, now containing 0.85% carbon, transforms to a lamellar (layered) mechanical mixture of α plus Fe_3C called pearlite. In other words, pearlite is formed by decomposing austenite of *eutectoid* composition. Ferrite and carbide growth usually begins at the austenite grain boundaries, and lamellae of the 2 phases grow simultaneously inward.

On cooling austenite with 0.85% to 2.06% C, *hyper-eutectoid* cementite first forms, progressively depleting the austenite of carbon. At 723°C (A_1) the austenite contains 0.85% C and transforms to pearlite.

Therefore, below A_3, carbon steel containing less than 0.85% C consists of ferrite plus pearlite; carbon steel containing more than 0.85% C consists of cementite plus pearlite. This is provided the material has been subjected to relatively slow cooling rates to avoid the formation of metastable phases. The pearlite nucleates at preferred sites in the austenite, and the nuclei then grow until they impinge on eachother. The process is both time- and temperature-dependant, and is controlled by the diffusivity of the relevant atoms. If the cooling rate is slow, carbon atoms can diffuse greater distances and *coarse pearlite* (thicker layers) is formed, as opposed to *fine pearlite* that is associated with accelerated cooling rates and increased strength. In eutectoid (i.e. fully pearlitic) steels:

1. the yield strength and UTS are linearly related to the reciprocal of the square root of the inter-lamellar spacing;

2. the inter-lamellar spacing is inversely proportional to the degree of undercooling;

3. the yield strength and UTS are linearly related to the square root of the degree of undercooling

The degree of undercooling is the temperature difference between the cooling temperature and the temperature of A_{r1}.

Pearlite has, however, an adverse effect on the ductility and toughness of plain carbon steels. The impact transition temperature is that temperature where the mode of fracture due to impact for a given material changes from brittle to ductile. Studies have shown that a 1% by volume increase in pearlite (accompanies a higher carbon content steel) raises the transition temperature by about 2°C.

Eutectoid Microstructures

Eutectoid steel contains approximately 0.85% carbon on the plain-carbon steel phase diagram, and on the Fe-C phase diagram is vertically in-line with the eutectoid point. This steel is fully pearlitic at room temperature.

Hypo- and Hypereutectoid Microstructures

Since pearlite comes from austenite of eutectoid composition, the amount of pearlite in a material should be equal to the amount of eutectoid austenite that was decomposed.

A *hypoeutectoid* steel has less carbon than the eutectoid composition, and some ferrite (called *proeutectoid ferrite;* "massive" in comparison to the lamellar *eutectoid ferrite* in pearlite) separates from the austenite before the austenite begins to transform to pearlite. Coarse pearlite forms in clusters within large expanses of proeutectoid ferrite. On further cooling, the final product (bainite) has many fine carbide particles dispersed within the ferrite matrix.

Hypereutectoid steels have more carbon than the eutectoid composition. On cooling, some carbon diffuses from the austenite to the austenite grain boundaries, where it precipitates as *proeutectoid carbide*, Fe_3C. Pearlite then forms within these proeutectoid carbide boundaries.

Martensite

The body-centred tetragonal (BCT) martensitic phase arises spontaneously from a diffusionless, shear-like phase transformation due to the rapid cooling of austenite, whose FCC structure becomes unstable below the critical temperature. The rate of cooling must be sufficient to suppress the higher temperature diffusion-controlled ferrite and pearlite reactions, as well as other intermediate reactions such as the formation of bainite. All the atoms shift in concert, and no individual atom departs more than a fractional nanometer from its previous neighbours. The critical rate of cooling required is very sensitive to the alloying elements present in the steel.

Martensite is a very hard structure in which the carbon, formerly in solid solution in the austenite, remains in solution in the new BCT α'-phase. In plain-carbon steels this phase is not stable at room temperature and, given an opportunity, martensite will proceed to form α + carbide. It is therefore referred to as *metastable*. For this reason it is not shown on the plain-carbon steel phase diagram. By manipulating the alloy geometry, martensite can become stable at room temperature.

The martensitic reaction in relation to TRIP steel development is discussed in Chapter 6.

Bainite

The pearlite reaction is essentially a high temperature one characterised by slow cooling from austenite, while the martensite reaction is characterised by rapid quenching to low temperatures. There is a wide intermediate temperature range where neither pearlite nor martensite forms. Rather, lath-shaped fine aggregates of ferrite and cementite are formed either during isothermal decomposition of austenite, or by athermal treatment with cooling rates that are too fast to form pearlite yet not rapid enough to form martensite [37]. This intermediate structure is termed bainite, and is discussed further in Chapter 13. The nature of bainite changes as the transformation temperature is lowered. Two main forms can be identified: *upper* and *lower* bainite.

Upper bainite is composed of long ferrite laths in the austenite expanse. These laths have a fine structure comprising smaller laths that are only slightly disoriented from each other. The upper bainitic ferrite has a much lower carbon concentration than the austenite from which it forms; consequently as the laths grow, the remaining austenite is enriched in carbon. This is an essential feature of upper bainite which forms in temperature ranges where the diffusivity of carbon is

still high enough to allow partition between ferrite and austenite. As a result, carbide precipitation does not form within the laths, but rather in the austenite at the lath boundaries when a critical carbon concentration is reached. The morphology of the cementite at the lath boundaries is dependent on the carbon content of the steel.

Lower bainite has more clearly defined individual plates that are broader than in upper bainite. While these plates nucleate at austenitic grain boundaries, there is also much nucleation within the grains (i.e. intragranular nucleation), and secondary plates form from primary plates away from the grain boundaries. The plates exhibit a higher dislocation density than upper bainite, but not as dense as martensites of similar composition. A striking microscopic characteristic of lower bainite is the growth of carbide rods within the ferrite plates. This is either cementite or ε-iron carbide depending both on the transformation temperature and on the composition of the steel. The carbide nucleates at the γ/α interface and grows as the interface moves forward. The lower bainite reaction is basically an interface-controlled process leading to cementite precipitation, which then decreases the carbon content of the austenite and enhances the driving force for further transformation.

5. KINETICS OF THERMAL PHASE TRANSFORMATIONS

This Chapter introduces the effect of heat treatments on steel microstructures and material properties. Of primary importance is the kinetics of the transformation, represented on a Time-Temperature-Transformation curve, as the material is cooled from an elevated temperature in a controlled manner.

5.1. HEAT TREATMENTS

Different heat treatments are performed by cooling a material from an elevated temperature through specific temperature ranges at different rates. Typical product phases of carbon steel include pearlite, ferrite, bainite and cementite. Metastable phases such as martensite also appear. Exactly what product or combination of products is formed depends on the rate of cooling and the alloy content. Alloy content and heat treatment processes are selected based on the required material properties of the finished product.

Chapter 13 (Appendix B) is an extension of the topics introduced in this Chapter and describes the heat treatments used to manipulate the austenite-ferrite (γ-α) and austenite-martensite (γ-α') transformations. Heat treatments are primarily used to control hardness. The discussion of these treatments is included as an appendix because the specialised processes used to develop strain memory alloys are thermo-mechanical rather than thermal. TRIP steels depend on the formation of a metastable austenitic parent phase at room temperature through the control of the alloy composition. Chapters 6 through 8 discuss the methods of developing suitable TRIP steel compositions for smart structural health monitoring applications.

5.2. TIME-TEMPERATURE-TRANSFORMATION DIAGRAMS

During heat treatment, the temperature of the material is raised to a prescribed level and then cooled in a controlled manner to manipulate the microstructure (and in turn the material properties). The phase composition at any time (other than metastable phases) is described by that particular steel's phase equilibrium diagram, but the distribution of phases throughout the solid solution is dependent on the heat treatment process and cooling rates.

The time-temperature-transformation (TTT) diagram for a specific alloy corresponds to that alloy's phase diagram and is used to describe the effect of

cooling sequences and rates on the microstructure. In their simplest form, these curves exhibit a well-defined "C"-shape, where the nose of the curve represents the temperature at which the transformation proceeds most rapidly, slowing down at both higher and lower temperatures. This can be explained with reference to Figure 5.1 which is representative of plain-carbon steel. Near the A_{r1} temperature, although the atoms move readily in the solid solution, the degree of undercooling ($\Delta T = T - T_{A_{r1}}$) is low, so that the driving force for the transformation is low. The only location for easy nucleation is the grain boundaries, where the atoms are already disordered. Precipitation of α in γ is primarily along the grain boundaries; this is *intergranular precipitation*. As ΔT increases the driving force also increases until the maximum rate at the nose of the curve. Below this temperature, the driving force for the reaction continues to increase, but the reaction is now impeded by the lack of diffusivity of the rate-controlling element (C in carbon steels). The *supersaturation* is great, therefore nucleation will occur wherever possible (at point imperfections such as vacancies or interstitials, along dislocation lines, or adjacent to impurities) within the interior of the supersaturated grains; this is *intragranular precipitation*, where the microstructure has a *dispersion* of the minor phase. Two different samples of the same material having opposing microstructures (intergranular vs. intragranular precipitation) will have significantly different properties.

As illustrated in Figure 5.1, if the temperature is held at a fixed cooling value, the corresponding isothermal transformation (I-T) curve is horizontal from left to right; at specific times nucleation and or diffusion begins, followed by precipitation. Curves illustrating some common controlled cooling (annealing, direct quenching, interrupted quenching and austempering) and re-heating (tempering) processes are included in the diagram. These topics are covered in greater detail in Chapter 13 (Appendix B).

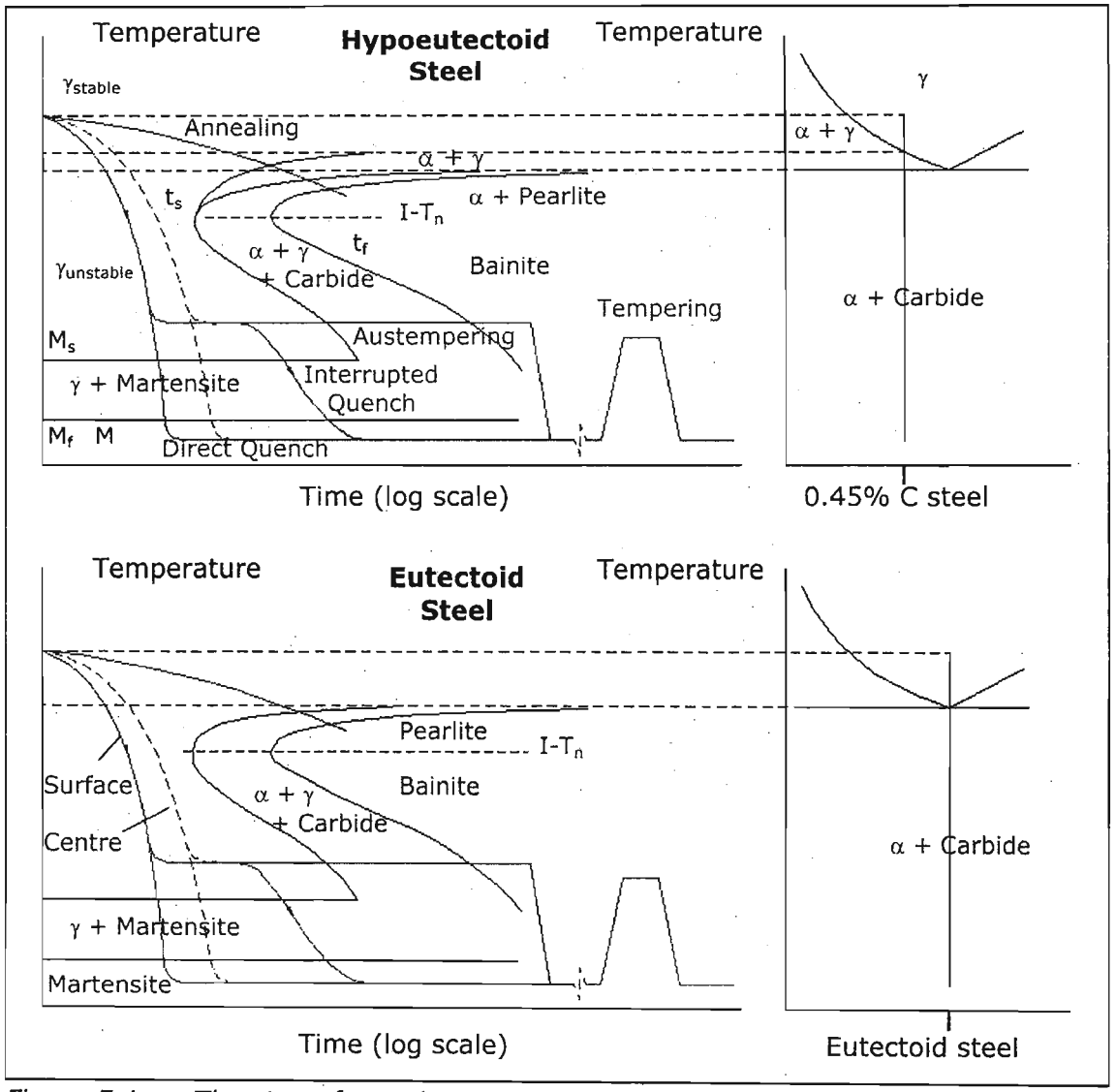


Figure 5.1: The transformation process. The stable phases of the phase diagrams (right) are not obtained immediately, but require the times shown on the left.

6. AUSTENITE – MARTENSITE TRANSFORMATIONS OF TRIP STEELS

6.1. THE MARTENSITIC REACTION

The high hardness and brittleness of rapidly quenched steels is the result of the formation of martensite. The martensite reaction in steels normally occurs athermally, i.e. during cooling in a temperature range that can be precisely defined for a particular steel. The rate of cooling must be sufficient to suppress the higher temperature diffusion-controlled ferrite and pearlite transformations, as well as the other intermediate transformations such as the formation of bainite.

During cooling, martensitic transformations commence at the M_s temperature and continue until the material is cooled below the M_f temperature. The M_s temperature can be as high as 500°C or well below room temperature, depending on the concentration of γ-stabilising elements in the alloy (refer to Chapter 7). Below M_f, a small proportion of *retained austenite* usually remains.

There are martensitic reactions that can proceed at constant temperature in so-called metastable austenitic steels. It is this phenomenon, coupled with the simultaneous change from material paramagnetism to ferromagnetism, that forms the basis of smart TRIP steel technology.

6.2. THE TRANSFORMATION OF METASTABLE AUSTENITE

6.2.1. Temperature-Controlled Martensitic Reaction

Some austenitic steels are often close to transformation, in that the M_s temperature is just below room temperature. This is true for example for low-carbon 18%Cr-8%Ni austenitic steel, which can undergo a martensitic transformation merely by refrigeration.

Defining M_s as the martensitic start temperature on cooling and A_s as the austenitic start temperature on heating, the driving force for the start of the martensitic transformation can be expressed as T_0-M_s. T_0 is the temperature between M_s and A_s at which austenite and martensite possess the same free energy. Similarly, The driving force for the start of the austenitic transformation can be expressed as A_s-T_0. The martensitic reaction requires a degree of *supercooling*; the austenitic reaction requires a degree of *superheating*.

In alloys where the transformation from austenite to martensite results in a large shape change, the driving force T_0-M_s is large and the temperature range M_s-M_f is also large; with small shape changes the reverse is true. For iron-based martensites the shape change is large and the M_s-M_f range is often several hundred degrees. Hence the strain energy that arises when a small martensite plate is formed plays a significant role in nucleation.

6.2.2. Strain-Controlled Martensitic Reaction

In some metastable austenitic steels, applying a plastic deformation to a specimen at room temperature can also lead to the formation of martensite at deformed locations. In general, an M_D temperature exists below which cold work (plastic deformation) increases the M_s temperature, so that transformation to martensite occurs. In other words, if a specimen is plastically deformed at a temperature T that lies within the M_D – M_s temperature range, the cold work raises the M_s temperature of the strained region to T, and transformation to martensite is initiated here.

In general, the higher the alloying element content, the lower the M_s and M_D temperatures. In the absence of cold working, the M_s temperature can be approximated using empirical equations:

$$M_s(^\circ C) = 539 - 423(\%C) - 30.4(\%Mn) - 17.7(\%Ni) - 12.1(\%Cr) - 7.5(\%Mo) \quad (6\text{-}1)$$

In austenitic steels, 2 factors influence the extent of work-hardening:

1. The stacking fault energy of the matrix, determined by the composition.

   Manganese can be substituted for nickel in Cr-Ni austenitic steels, but the Cr-Mn solid solution then has a much lower stacking fault energy. This means that the FCC austenite is energetically closer to an alternative close-packed hexagonal structure, and that the dislocations will tend to dissociate to form broader stacking faults than is the case with Cr-Ni austenites. Under these circumstances, the martensite that forms first is hexagonal in structure (ε-martensite) and paramagnetic, and nucleates on stacking faults. This type of martensite forms as parallel-sided plates. Frequently, α′-martensite eventually forms, nucleating at the interface between ε and the austenite matrix.

Ni is in fact one of the elements that raises the stacking fault energy of austenite, resulting in low work hardening. The elements Cr, Mn, Co, Si, C and N tend to lower the stacking fault energy of austenite. By severe cold working, e.g. up to 80% reduction in wire drawing, the relatively modest yield strengths of ordinary austenitic steels can be raised to over 1200MPa.

Manganese on its own can stabilise austenite at room temperature provided sufficient carbon is in solid solution. An example of such an alloy is Hadfield's manganese steel with 12%Mn-1.2%C. The alloy exists as austenite at room temperature and even after extensive deformation does not form martensite. However, if the carbon content is lowered to 0.8%, then M_D is above room temperature and transformation due to deformation is possible.

2. The stability of the matrix.

   The transformation to martensite has undoubtedly the largest effect on work hardening rates.

6.2.3. TRIP Steels

TRIP steels have been developed to exploit the properties obtained when the martensite reaction occurs during low temperature plastic deformation. They exhibit increases in work hardening rate and a marked increase in uniform ductility prior to necking (if necking occurs at all before failure). The basis for TRIP steel development is the same as for other controlled transformation steels, but a broader range of thermo-mechanical treatments are utilised.

Effect of Thermo-Mechanical Processes

Two common thermo-mechanical treatments utilise the metastable bay:

1. One process used to develop TRIP steels involves initially balancing the composition of the austenitic steel to produce an M_D temperature above room temperature. The steel is then heavily deformed at a temperature above M_D (usually in the range 250°C to 500°C) in a process known as the *prior deformation of austenite*.

   An example of such a steel is Fe-0.3%C-2%Mn-2%Si-9%Cr-8.5%Ni-4%Mo. After an 80% reduction at 475°C, the alloy exhibits the following properties at room temperature:

0.2% proof stress 1430MPa

Tensile strength 1500MPa

Elongation 50%

2. Higher strength levels (proof stress ~2000MPa) with ductility between 20% - 25% can be obtained by adding strong carbide-forming elements such as vanadium and titanium, and by causing the M_D temperature to be below room temperature. As in the previous process, severe thermo-mechanical treatments in the range 250°C to 500°C are used to deform the austenite and dispersion strengthen it with fine alloy carbides in a process known as *ausforming*. The M_D temperature is, as a result, raised to above room temperature. The net effect is an excellent combination of strength and ductility, as well as substantial improvements in fracture toughness, although the strain memory effect is lower than that obtained by PDA because of the formation of phases other than martensite.

Applications of TRIP Steels

As with other controlled transformation steels, TRIP steels require extremely good metallurgical control and are very expensive to make. They are only used in applications where extremely high demands are made on mechanical properties.

Because the transformation from metastable austenite to martensite is accompanied by an increase in ferromagnetism, TRIP steels lend themselves to passive smart structural health monitoring systems.

6.2.4. Nucleation and Growth of Martensite

The overall free energy change, ΔG, when nucleation of martensite in the austenite matrix takes place, is a result of 3 components:

1. chemical free energy
2. strain energy
3. interfacial energy between martix and martensite

From classical theory of *homogeneous* nucleation, in order to form a nucleus of martensite with an oblate spheroid shape having radius r and semi-thickness c:

$$\Delta G = \frac{4}{3}\pi r^2.c.\Delta g + \frac{4}{3}\pi r.c^2 N + 2\pi r^2 \alpha \qquad (6\text{-}2)$$

K strain energy factor

α free energy per unit area of the γ/α' interface

Δg chemical free energy change per unit volume.

The critical nucleus size (σ) is determined by c^* and r^* at which the free energy change is ΔG^*, which defines a "saddle" on the free energy c-r curve; thus:

c^* strain energy factor

r^* free energy per unit area of the γ/α' interface

and:

$$\Delta G^* = \frac{32\pi K^2 \sigma^3}{3(\Delta g)^4} \qquad (6\text{-}3)$$

However, if reasonable values of Δg, K and σ are used, the value of ΔG^* is so high that the barrier to nucleation is orders of magnitude too large. It is therefore impossible for martensite nuclei to occur as a result of random fluctuations. This suggests that nucleation must take place *heterogeneously* on pre-existing embryos that are already beyond the saddle point on the free energy curve [37]. However, the search for such embryos has not been particularly successful, and they still remain a deduction from formal nucleation theory.

In some special cases, for example in high manganese steels, stacking faults readily occur as the austenite has a low stacking fault energy. On transformation to martensite, a hexagonal, paramagnetic ε-martensite is shown to nucleate at the stacking faults. Olson and Cohen [38] have developed a theory of nucleation in which the first step is faulting on the close-packed planes derived from existing groups of dislocations. The most likely sites for such nucleation are grain boundaries, incoherent twin boundaries, and inclusion particle interfaces. The Olson and Cohen model may be used to characterise the kinetics of the strain-induced transformation reaction in terms of the temperature and plastic strain.

6.2.5. *Effect of Carbon*

The interstitial solid solution of carbon that results in the tetragonality of BCT martensite is one reason for the strength of martensite [37]. Experiments have shown that increasing the carbon concentration, c, increases the yield strength by a factor of $c^{½}$.

The yield strength of martensite, like that of ferrite, is markedly temperature dependent, but this dependence is little affected by the presence or absence of precipitate or by the amount of carbon in solution.

7. ALLOYING ELEMENTS TO DEVELOP METASTABLE AUSTENITE

Commonly available iron alloys include compositions of carbon (C), nickel (Ni), chromium (Cr), manganese (Mn), silicon (Si), nitrogen (N), molybdenum (Mo), niobium (Nb), phosphorus (P), etc. These elements are included for their effects on the alloy microstructure in order to promote certain physical properties (strength, hardness, toughness, ductility, corrosion resistance, etc.) [37]. The magnetic properties of iron in smart sensing TRIP steels are required to record the strain history. Suitable alloying elements in precise quantities together with controlled thermal processing are essential to establishing the capability by promoting the formation of metastable austenite at room temperature.

Iron binary equilibrium systems exhibit either open or closed γ-fields. Alloying elements can influence the phase equilibrium diagram either by expanding the γ-field and encouraging the formation of austenite over wider compositional limits (*austenitic stabilisers*), or by contracting the γ-field and encouraging the formation of ferrite over wider compositional limits (*ferritic stabilisers*). The role of austenitic stabilisers in the development of TRIP steels is therefore extremely important.

The form of the phase diagram depends to some degree on the electronic structure of the alloying elements, which is reflected in their relative positions on the Periodic Table.

7.1. AUSTENITE-STABILISING ALLOYING ELEMENTS (OPEN γ-FIELD)

When added to iron, certain elements produce an open austenitic field on the phase diagram (as in Figure 7.1(a)) whereby FCC austenite is preserved at room temperature, either in a stable or metastable condition. This can be attributed to the suppression of the ferrite phase. Elements used to create this effect include nickel, manganese, cobalt and the inert materials ruthenium, rhodium, palladium, osmium, iridium and platinum [37].

When added in sufficient quantity, both nickel (Ni) and manganese (Mn) eliminate BCC ferrite entirely at room temperature and replace it with the FCC austenite. When added in sufficient quantities, nickel also improves the ductility and formability, as well as the corrosion resistance.

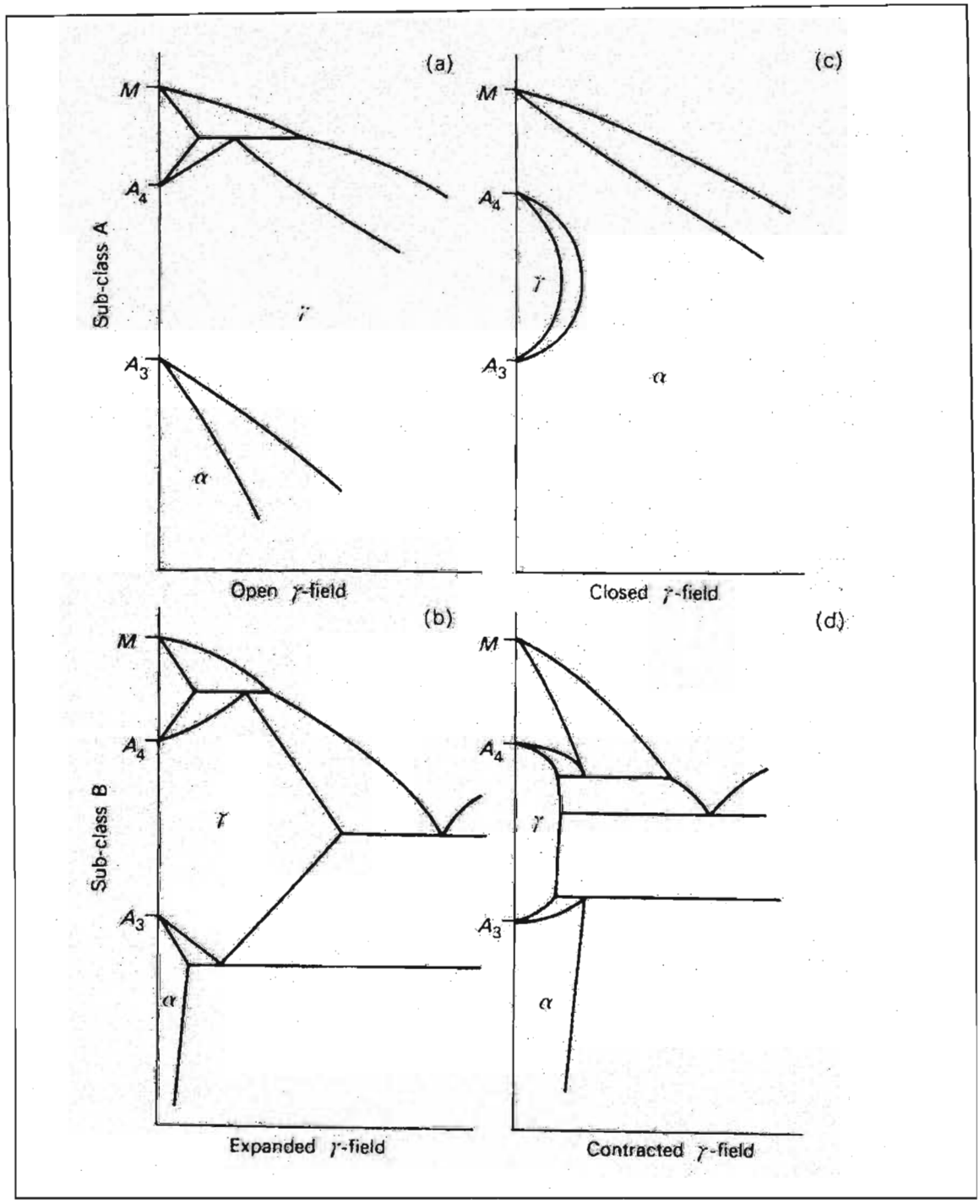


Figure 7.1: Open γ-field phase diagram characteristics

An important characteristic of TRIP steels is their excellent response to thermo-mechanical processing. In austenitic steels, 2 factors influence the extent of work-hardenability:

1. The stacking fault energy of the matrix, determined by the alloy composition. The elements Cr, Mn, Co, Si, C and N (not all of which are austenitic stabilisers) all tend to lower the stacking fault energy of austenite.

Manganese (Mn) is not as powerful an austenitic stabiliser as Nickel, but importantly serves to lower the stacking fault energy of the austenite (Ni is in fact one of the elements that raises the stacking fault energy of austenite, resulting in low work hardening). This means that the FCC austenite is energetically closer to an alternative HCP structure, and that the dislocations will tend to dissociate to form broader stacking faults than is the case with Cr-Ni austenites. In these circumstances the martensite that forms first (an intermediate phase) is hexagonal in structure (ε-martensite, introduced in Section 6.2.4), and nucleates on stacking faults. This type of martensite forms as parallel-sided plates. Alloys of Fe-Cr-Mn that exhibit the intermediate ε-martensite phase have a 10% or greater Manganese content. The presence of chromium in concentrations of 13% - 18% stabilises the BCC (BCT if carbon present) ferrite phase α sufficiently to facilitate the transformation from ε- to α'-martensite. α'-martensite nucleates at the interface between ε and the austenite matrix.

By severe cold working, yield strengths of ordinary austenitic steels can be significantly increased.

2. The stability of the matrix.

   The transformation to martensite has undoubtedly the largest effect on work hardening rates.

The Fe-Cr-Mn group of austenitic stainless steel are cheaper than the Fe-Cr-Ni group and have the advantage of reducing the sensitivity of the transformation to changes in the ambient temperature that is common amongst strain memory alloys. This temperature dependence is minimised by the formation of the ε-martensite (which like austenite is paramagnetic) at lower temperatures than the expected operating range. The nucleation of α'-martensite from within ε-martensite should be a function of the applied strain only.

7.2. ELEMENTS THAT ASSIST HEAT TREATMENT (EXPANDED γ-FIELD)

Some elements expand (enlarge) the austenite field on the phase diagram. Refer to Figure 7.1(b). The interstitial elements carbon and nitrogen are the most important in this group because of their role in the heat treatment of steels (they allow the formation of a homogeneous solid solution of austenite that contains up

to 2%C or 2.8%N. Other elements used to create the expanded γ-field effect include copper, zinc and gold [37].

The austenitic field is expanded, but not open, because the formation of compounds limits its existence. The above-mentioned elements are therefore austenite stabilisers at elevated temperatures.

Nitrogen is an extremely cheap austenitic stabiliser. Researchers at the University of Cape Town investigated whether nitrogen (expands the γ field) can be used to replace part of the nickel (opens the γ field) in strain memory alloys such as AISI 301 and 302 In order to lower the cost. The experimental compositions are shown below in Table 7.1.

Table 7.1: Alloys used in the nickel-with-nitrogen partial replacement experiment.

	Cr	Ni	N	C	Mn	Si	V	Cu	Co	Mo	Fe
301	17.7	7.48	0.074	0.029	1.5	0.5	0.07	0.08	0.02	0.03	Bal
A1	17.2	5.12	0.087	0.084							
A3	17.5	5.09	0.129	0.087							
A5	17.9	5.17	0.163	0.075							
A6	17.9	5.03	0.276	0.071							
D1	17.9	3.41	0.185	0.078							
D2	17.9	3.36	0.275	0.070							

The effectiveness of nitrogen as a nickel replacement can be seen in the amount of austenite retained at room temperature, as listed in Table 7.2. In all cases ferrite was absent, hence the remainder of the microstructure is composed of martensite.

Table 7.2: Percentage austenite retained at room temperature for the nickel-with-nitrogen partial replacement experiment.

Alloy	301	A1	A3	A5	A6	D1	D2
% Retained Austenite	98	70	83	89	91	43	67

It is evident that a careful balance is needed if nitrogen is to partially replace nickel (compare A6 to D2). The alloy designated A6 was the most successful of those tested for austenite retention. Although not as stabilised as the 301 alloys, A6 was found to exhibit comparable work-hardening rates. It should be noted that delayed cracking has been observed in alloys with high nitrogen levels after deep drawing operations – a common operation applied to 301 alloys. This cracking, which gets

worse over time, is believed to occur as a result of the embrittling effect the nitrogen has on the martensite.

Although carbon can be used as an austenitic stabiliser, care must sometimes be taken to avoid the formation of free carbides. This occurs when elements such as chromium are leached out of the solution to form chromium carbides within certain temperature ranges. In the case of chromium, this has a negative effect on corrosion resistance. Elements such as titanium (Ti) and niobium (Nb) are sometimes added to prevent chromium carbides from forming. As an interstitial component, carbon itself can cause a decrease in ductility. Carbon is known to depress the M_s temperature by approximately 10°C per 0.01%C added.

Although an austenitic stabiliser, copper is more commonly used to improve corrosion resistance, usually at the slight expense of ductility.

7.3. ELEMENTS THAT CLOSE THE γ-FIELD

Certain elements restrict the formation of austenite and encourage the formation of ferrite (ferritic stabilisers). The γ phase is forced to contract as shown in Figure 7.1(c) so that in some cases the ferrite α and δ phases become continuous [37]. These alloys are not amenable to normal heat treatments involving cooling through the γ/α phase transformation. Ferritic stabilising elements include silicon, aluminium, beryllium and phosphorus. Strong carbide formers such as chromium, molybdenum, titanium and vanadium also fall into this category, with Mo the most effective and Cr the least effective.

One reason a carbide-former might be selected in spite of its effect on the γ-field is that the formation of carbides enhances dislocation multiplication at elevated temperatures. When combined with thermo-mechanical processing, this substantially increases the yield strength at room temperature. Both Mo and V can produce work hardening that is independent of temperature and strain rate.

Chromium is usually added to improve corrosion resistance. As little as 5% improves the corrosion properties, but a minimum of 12% is required to form stainless steel. For chromium contents less than around 13% there is an austenite to ferrite transformation on cooling from within the γ loop of Figure 7.1(c). For Cr contents greater than 13% a solid solution of chromium in ferrite is formed.

Mo and Si have also been shown to improve corrosion resistance. Aluminium can be included to improve high temperature scaling and is commonly added (like Ti and Nb) interstitially to prevent sensitisation.

7.4. ELEMENTS THAT CONTRACT THE γ-FIELD

Certain elements promote a phase equilibrium diagram resembling Figure 7.1(d), where the γ loop is strongly contracted [37]. These include boron and the carbide formers tantalum, niobium and zirconium.

Niobium is added as a grain refiner, which may then improve strength. As little as 0.13 to 0.22% can result in marked γ grain refinement.

7.5. COMBINATIONS OF ELEMENTS

Although the effects of each of the most prominent alloying elements have been discussed in relation to iron, most alloys are not binary in nature. A combination of several elements is usually required depending on the properties they produce. For example, when selecting a suitable material to be used in smart mining bolts, corrosion resistance and a stable austenitic phase at room temperatures are just two of the requirements. This section examines the effects of the interaction of different compositions of alloying elements on the final product.

For a fixed carbon content, as the alloying element is added the γ-field is either expanded or contracted depending on the particular solute. For example, Si restricts the γ-field and there is a corresponding enlargement of the α-field; vanadium contracts the γ-field and there will be vanadium carbide in equilibrium with ferrite over much of the ferrite field; Ni does not form a carbide and expands the γ-field. Normally elements with opposing tendencies will cancel each other out, but in some cases anomalies occur. For example, chromium added to a steel containing Ni in concentrations around 18% (e.g. 18Cr 8% Ni) retards the kinetics of the γ→α transformation, thereby contributing to the retention of austenite at room temperature.

One convenient method of illustrating quantitatively the effect of a *particular* alloying element on the austenitic field of the Fe-C system is to plot the austenitic solubility boundaries for increasing concentrations of the alloying element. Refer to Figure 7.2.

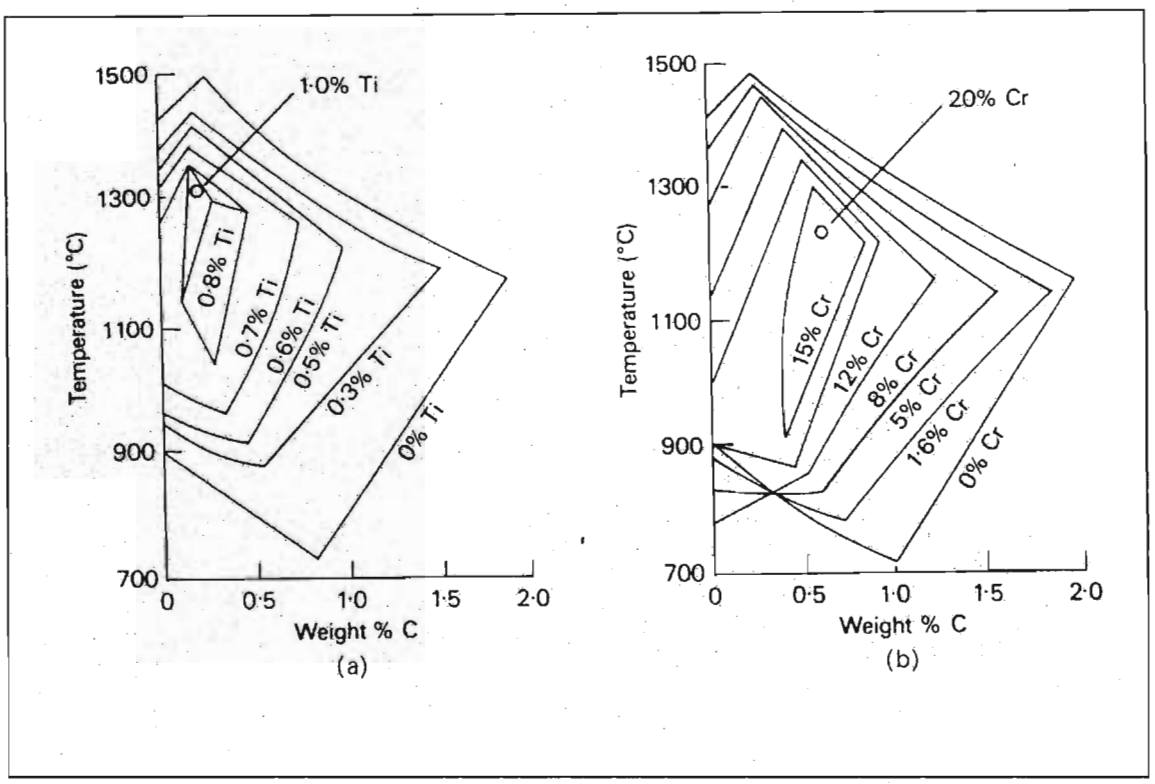


Figure 7.2: Effect of alloying elements on the austenitic phase boundaries of the Fe-C phase diagram.

The Schaeffler diagram (Figure 7.3) is a convenient method of representing the effect of *various* elements on the basic structure of chromium-nickel stainless steels [37]. It is a plot of the compositional limits at room temperature of γ-austenite, δ-ferrite and M-martensite in terms of Ni and Cr equivalents. The *chromium equivalent* has been empirically determined using the most common ferrite-forming elements:

$$C.E. = (\%Cr) + 1.5(\%Mo) + 2(\%Si) + 5(\%V) + 5.5(\%Al) + 1.75(\%Nb) + 1.5(\%Ti) + 0.75(\%W) \quad (7\text{-}1)$$

Similarly, the *nickel equivalent* has been determined with the familiar austenite-forming elements:

$$N.E. = (\%Ni) + 30(\%C) + 0.5(\%Mn) + 0.3(\%Cu) + 25(\%N) \quad (7\text{-}2)$$

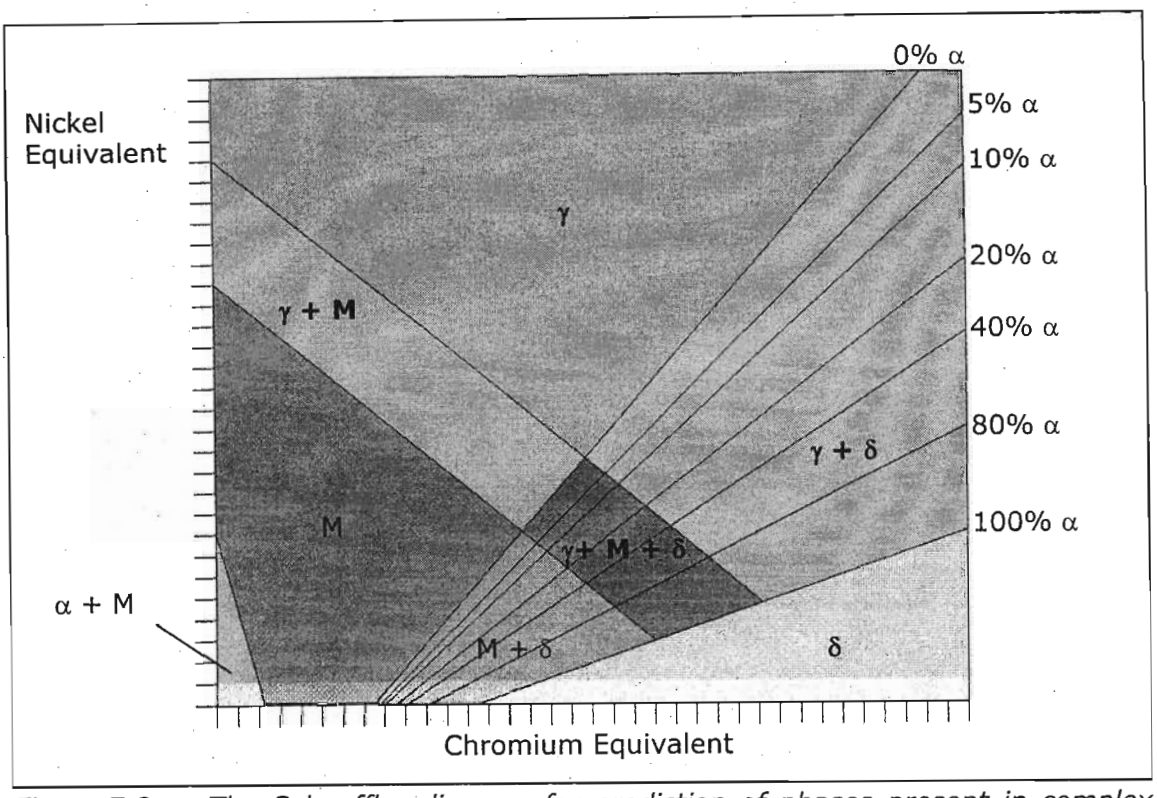


Figure 7.3: The Schaeffler diagram for prediction of phases present in complex alloying elements at room temperature.

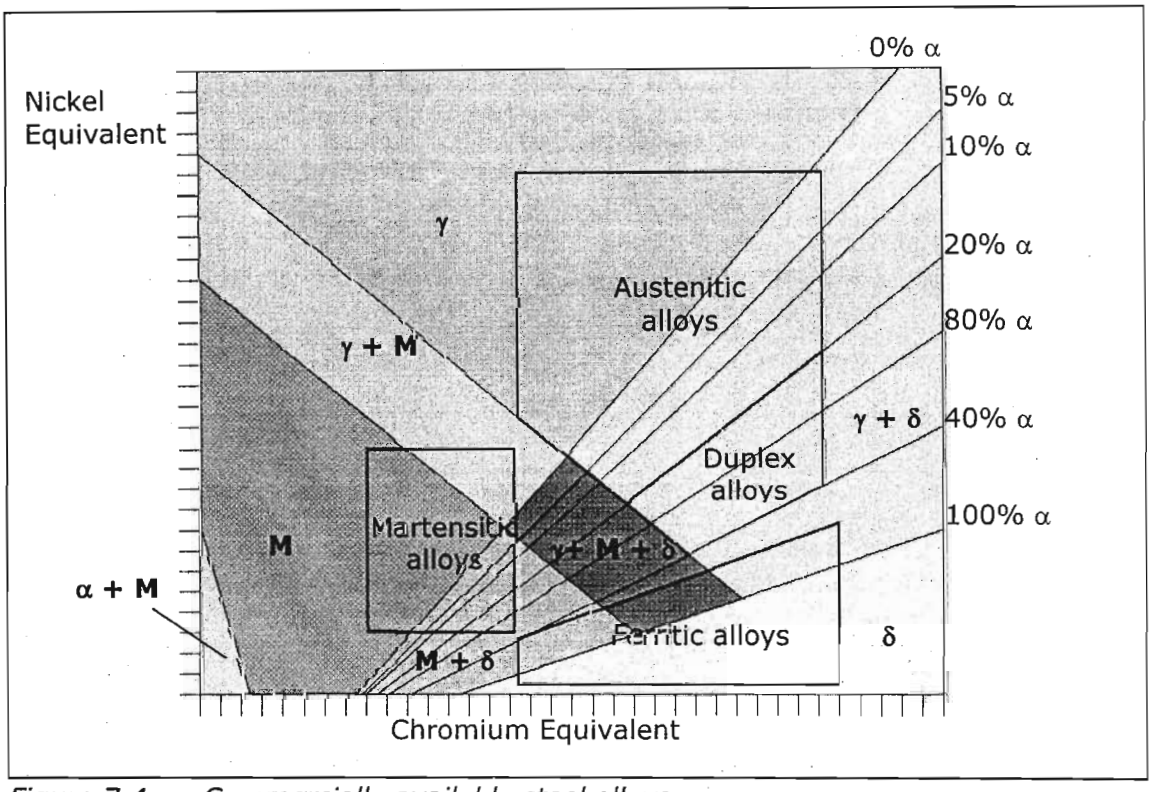


Figure 7.4: Commercially available steel alloys.

Figure 7.4 shows the commercially available grades of steels consisting of the above-mentioned alloying elements. Figure 7.5 indicates the region on the Schaeffler diagram most likely to facilitate the strain memory effect.

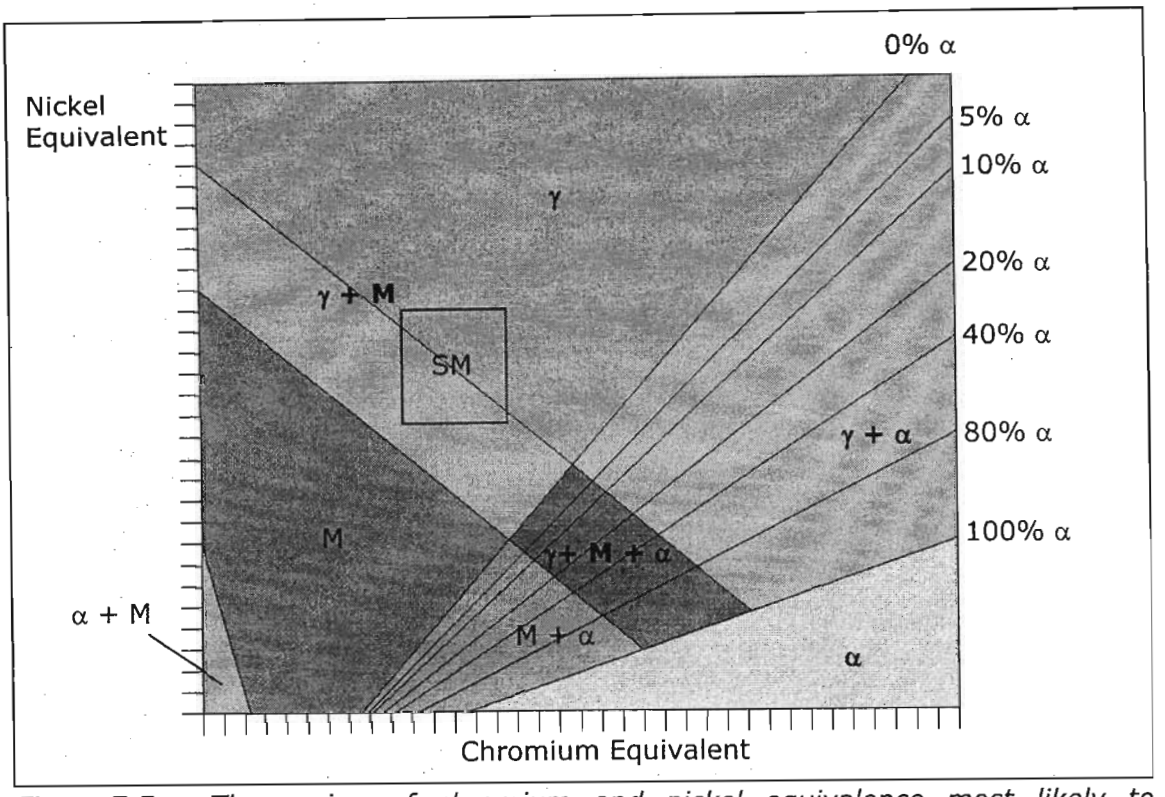


Figure 7.5: The region of chromium and nickel equivalence most likely to produce strain memory alloys.

7.6. PROCESSING METASTABLE AUSTENITIC STEELS

Usually steels of a particular composition are strengthened either through:

1. heat-treating (transformation from austenite to either martensite, or tempered martensite, or some form of ferrite plus carbides) or

2. cold working (thereby increasing the dislocation density and the yield strength).

Strain memory alloys cannot be heat treated in the normal fashion because the alloys have been engineered to remain austenitic down to room temperature. Cold working is also unsuitable; the resulting deformation would immediately trigger the strain-induced martensitic transformation, which is highly undesirable if advantage is to be taken of the smart capabilities of the material in a strain monitoring application. By using alloying elements to manipulate the TTT diagram, thermo-mechanical processes can be applied that increase the yield strength of the material without actually creating martensite. Section 6.2.2 introduced 2 suitable thermo-mechanical processing methods, further discussed in the Section 7.6.2.

7.6.1. Time-Temperature Transformations and Work-Hardenability

Refer to Figure 5.1 for the TTT curve of plain carbon steel. The transformation start and finish times are clearly marked as t_s and t_f. The lower horizontal curves indicate the start and finish time curves of austenite transformation to martensite (marked M_s and M_f respectively). The position of these start and finish curves in relation to the actual operating temperatures influences the strain-induced transformation of austenite to martensite. The term M_D was introduced in Chapter 6 and represents the temperature above which the formation of martensite due to strain ceases to occur. This temperature varies according to the alloy composition.

Alloying elements have three major effects on TTT curves:

1. The composition of an alloy determines the shape of that alloy's TTT diagram. For example, the horizontal width of the nose of the C-shaped section of the transformation is usually increased by the addition of most alloying elements to iron.

2. Those alloying elements that expand or open the austenite field of the material's phase diagram, also tend to depress the position of the curve with respect to temperature.

3. It is possible to create an austenitic bay in the TTT curve if certain alloying elements, notably chromium, are added. See Figure 7.6. If the M_D temperature is above this bay, then it is a *metastable austenitic bay* and representative of the optimal TTT curves for thermo-mechanically developing strain memory.

   The TTT diagram of Figure 5.1 shows a continuous t_f curve along which the ferrite, pearlite and bainite reactions occur, but it difficult to disentangle the reactions sufficiently to study their kinetics. However, the addition of certain alloying elements separates the reactions to the extent that they can be represented as individual curves on the TTT diagram, which then adopts a more complex form than that of the familiar C-curve. In Figure 7.6. the pearlite and bainite reactions are very clearly separated, and the relationship of the bainite transformation to the Ms temperature is revealed.

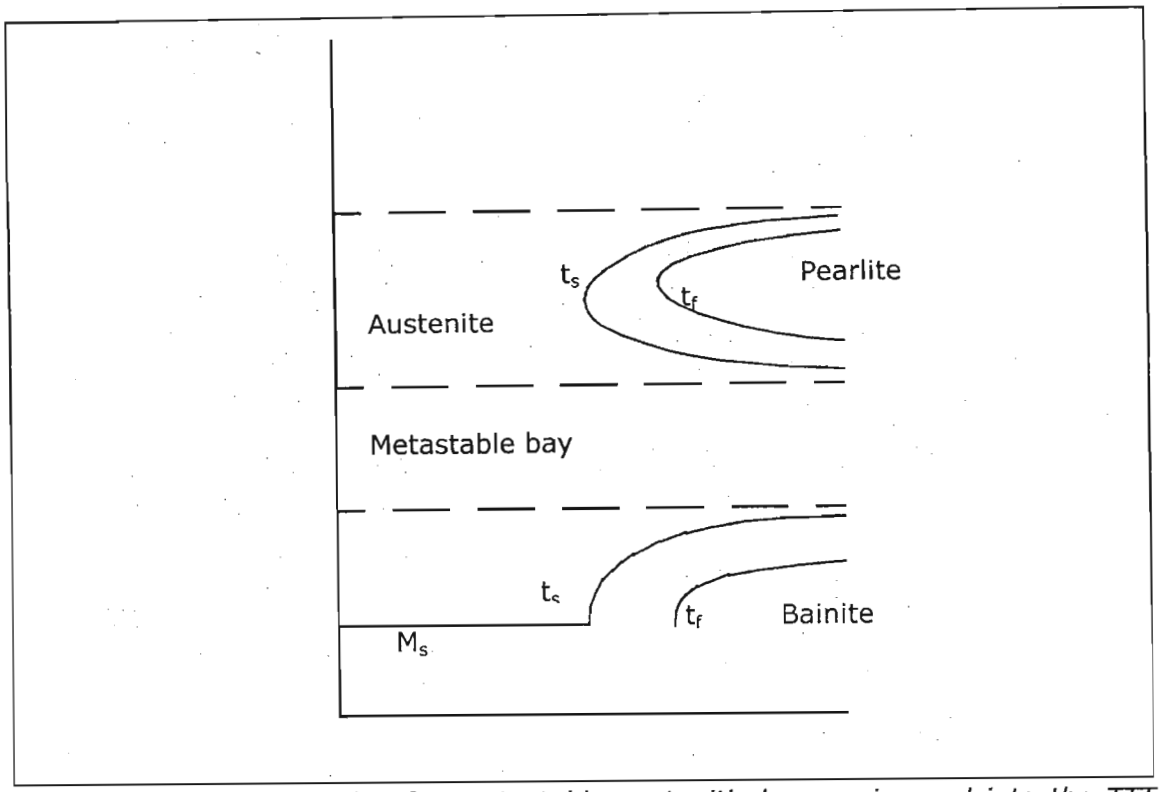


Figure 7.6: An example of a metastable austenitic bay engineered into the TTT curve through the use of suitable alloying chemistry.

7.6.2. Material Processing to Improve Strength

Prior Deformation of Austenite

The creation of the metastable austenitic bay facilitates heavy warm working in this temperature range (neither specifically hot nor cold) that substantially improves the strength and ductility of the material. This classical method of increasing strength in metastable alloys, introduced in Section 6.2.3, is referred to as *prior deformation of austenite* (PDA) [37].

The method with which PDA is applied determines the dislocation substructure. This in turn influences martensitic nucleation, which also influences the M_D temperature.

Ausforming

An alternative to warm rolling is thermo-mechanical processing (*ausforming*), which involves deforming the specimen in the metastable region and then cooling through α + C regions. The resulting multi-phase steel is TRIP assisted, since some austenite is retained for transformation to martensite. However, the bulk of the material at room temperature is bainite and other α + carbides. Carbides serve

to raise the M_D temperature and thereby broaden the temperature range within which martensite can be induced by strain.

Although the strength is higher than that obtained through PDA and ductility is high because of the partial transformation to martensite during deformation, ausformed material is not suitable for sensitive strain memory requirements.

The Koppenaal Alternative

By cycling repetitively between a martensitic and reverted austenite (reverse martensite shear transformation) structure, strengthening occurs [39] [40]. The effect is illustrated in Figure 7.7.

The Koppenaal method as applied to the alloy Fe-24.5%Ni-4.1%Mo-0.29%C can briefly be described as follows: Samples of the alloy are austenised and water quenched to a hardness of 80HR_B. They are then cooled to −175°C and returned to room temperature, where the hardness of the $\gamma + \alpha'$ structure measures between 31.5 and 33 HR_C. This is followed by immersing the samples in salt baths at temperatures of between 500 and 940°C, and then air-cooling to room temperature. The microstructure, now reverted austenite with its hardness of approximately 27.5 HR_C, is appreciably harder than the as-quenched austenite.

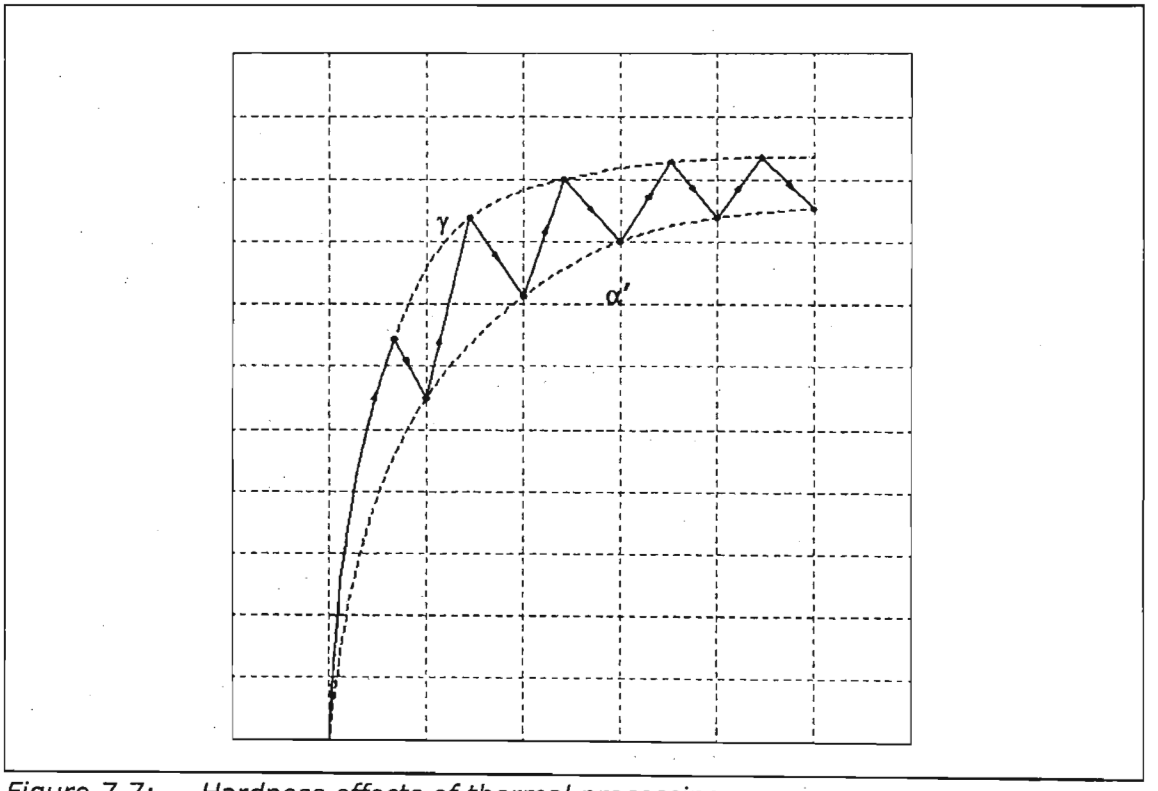


Figure 7.7: Hardness effects of thermal processing

Several requirements must be addressed if the Koppenaal process is to have the desired effect:

1. The amount of martensite that can be formed by cooling to liquid nitrogen temperature must be sufficient.

4. The austenite recrystallisation temperature must be above the austenite reversion temperature.

5. The M_D temperature must be raised by thermal processing if it is below room temperature when the material is in an annealed state.

Properties of alloys that have been thermally processed in this manner compare favourably to those that have been thermo-mechanically treated (Table 7.3).

Table 7.3: Comparison of thermal strengthening processes.

Alloy	Treatment	Yield strength (MPa)	UTS (MPa)
Fe-24%Ni-4%Mo-0.3%C	Thermo-mechanical reduction of 80% at 517°C	1130.78	1213.52
Fe-24.2%Ni-4.1%Mo-0.29%C	5 thermal cycles between -175°C and 722°C	1117	1330.74

7.6.3. Effects of Composition and Thermal Process Temperatures on TRIP Steels

To examine the effects of alloy composition on the volume change of transformed phases, Baik, Kim, Jin and Kwon (2001) [41] tested the compositions listed in Table 7.4 for the 800 MPa grades of cold-rolled TRIP steel sheets. Their intention was also to clarify the relationship between the phase transformation and the mechanical properties of TRIP steels.

Table 7.4: Chemical compositions of examined steels in weight %.

Reference	C	Si	Mn	P	S	Sol. Al	Nb
S	0.14	0.93	2.11	0.013	0.010	0.045	-
S-Mn	0.15	0.99	2.39	0.011	0.009	0.035	-
S-Si	0.14	0.48	2.20	0.010	0.009	0.038	0.018
S-Nb	0.14	0.98	2.10	0.010	0.009	0.038	0.018

Specimens were prepared by successive heating, hot rolling and cooling. The resulting hot rolled samples were then pickled and cold rolled before being

annealed in the metastable austenitic bay. The retained austenite fractions of the samples were calculated from data captured using X-ray diffraction analysis.

Mechanical Properties

Figure 7.8 shows the mechanical properties (elongation, yield strength, UTS and a combination of strength and ductility measured as the product of UTS and elongation) as a function of the annealing temperature TA1. The alloying elements Mn and Si were found to exist as solid solution atoms, while Nb was precipitated as a carbide (Figure 7.9). The volume fractions of second phases are strongly influenced by chemical composition as shown in Figure 7.10. The mechanical properties are dependent on the morphology and volume fractions of the second phases as well as on solid solution elements.

The volume fraction of bainite (carbide) was found to increase as the content of Mn was increased. This is because the volume of austenite increases during intercritical annealing as the amount of austenitic stabiliser (Mn) increases. The tensile strength of the alloy was also found to increase with Mn content.

The volume fraction of bainite was found to increase with decreasing Si, since silicon is a ferritic stabiliser. The tensile strength also increased with decreasing Si.

Hence the strength is more strongly influenced by the change of the bainite volume fraction than by the solid solution hardening effect.

The lower elongation of the steel S-Mn than that of the other steels can be attributed to the pearlite observed in Figure 7.10. It is well known that pearlite is a hard phase that reduces ductility.

Test specimens were then austempered. Figure 7.11 compares the tensile properties of the experimental steels for various austempering start temperatures. Tests were performed for start temperatures of 400°C, 450°C and 500°C. In all specimens, the highest value for the product of UTS and elongation (a measure for the optimal combination of strength and ductility) was observed for the 450°C austenitic start temperature.

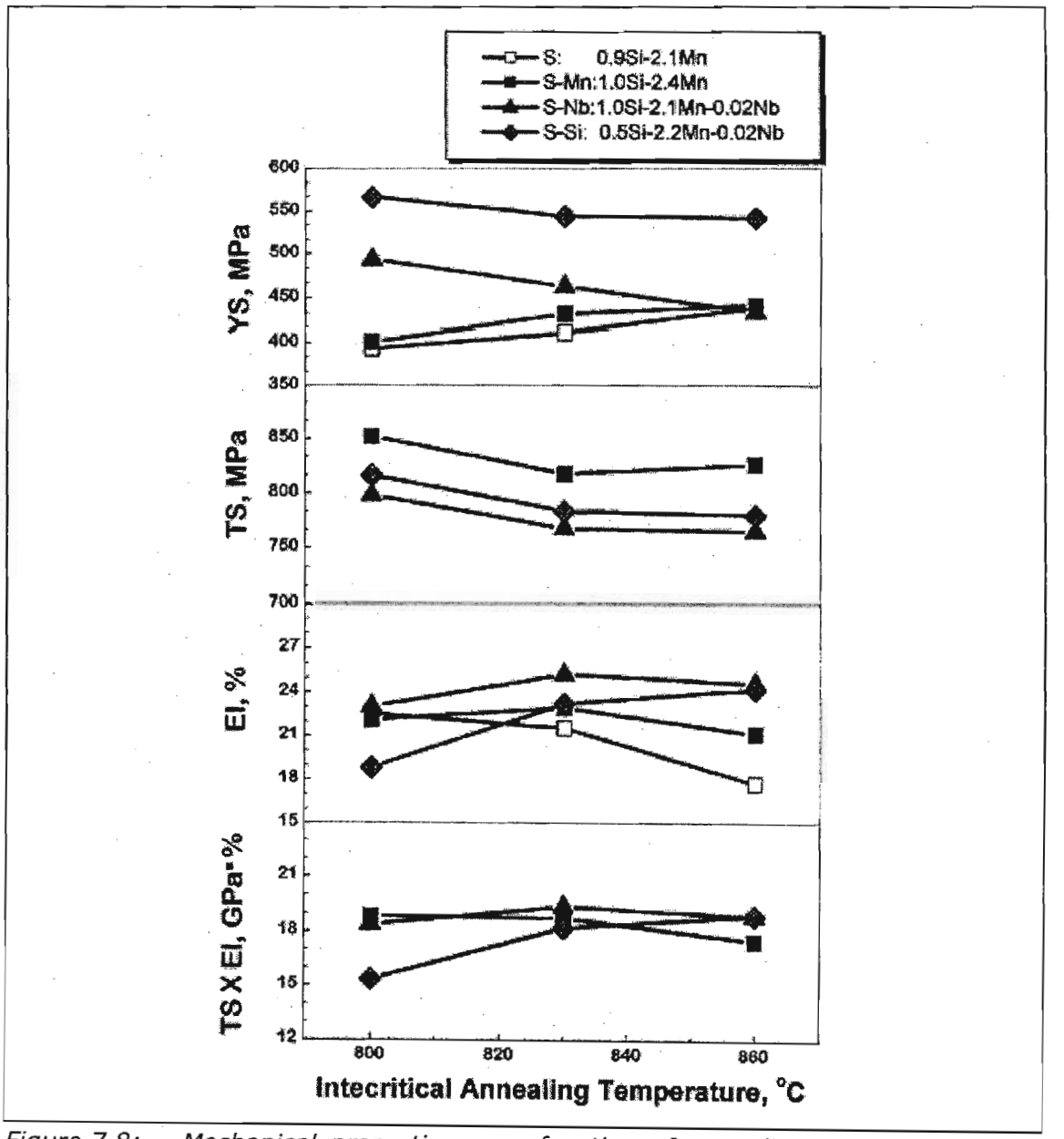


Figure 7.8: Mechanical properties as a function of annealing temperature TA1. Annealing performed for 51 seconds, followed by austempering at 450°C for 300 seconds.

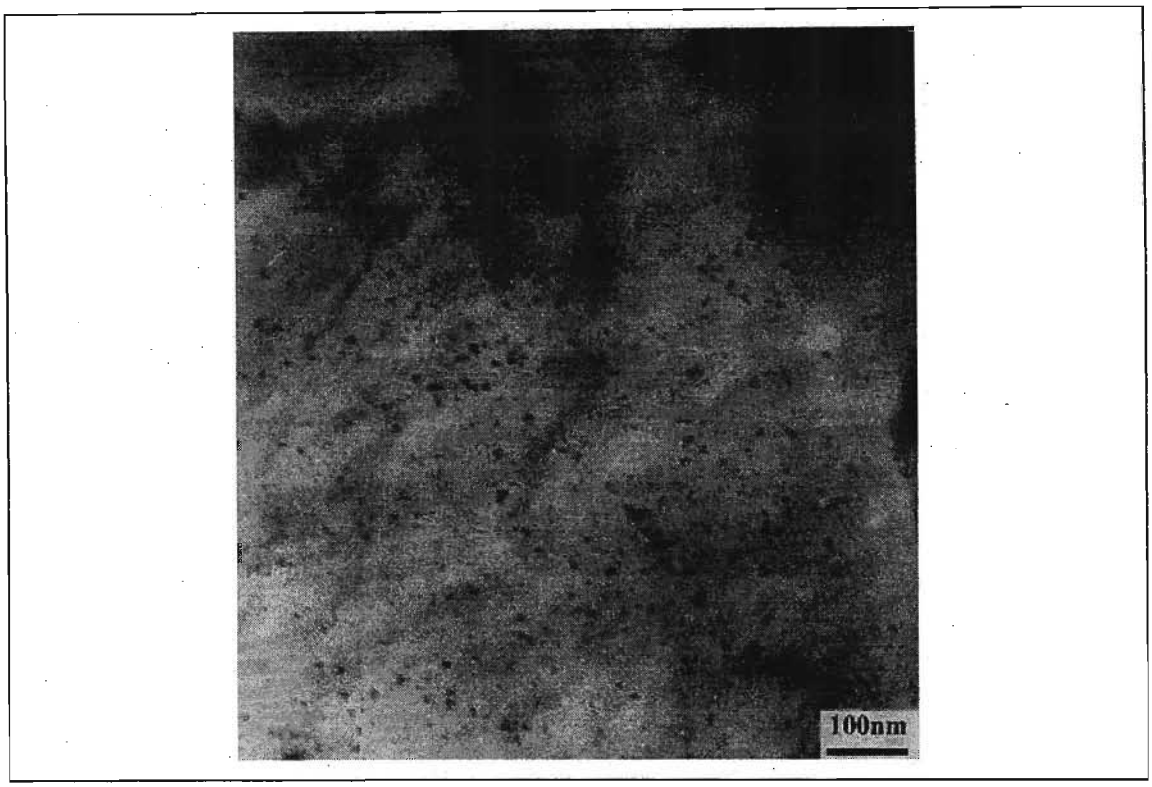


Figure 7.9: Transmission electron micrograph showing niobium carbide precipitation in steel S-Nb.

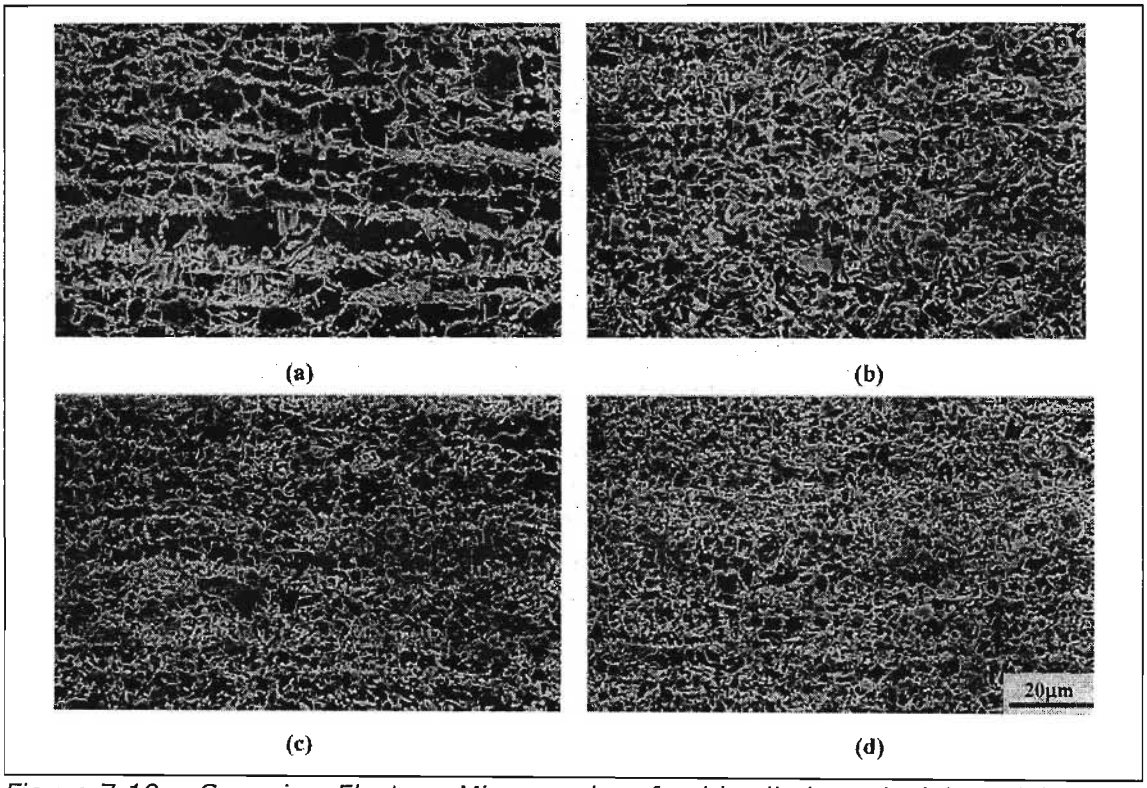


Figure 7.10: Scanning Electron Micrographs of cold rolled steels (a) S, (b) S-Mn, (c) S-Nb, (d) S-Si annealed at 830℃ for 51 seconds, followed by austempering at 450℃ for 300 seconds.

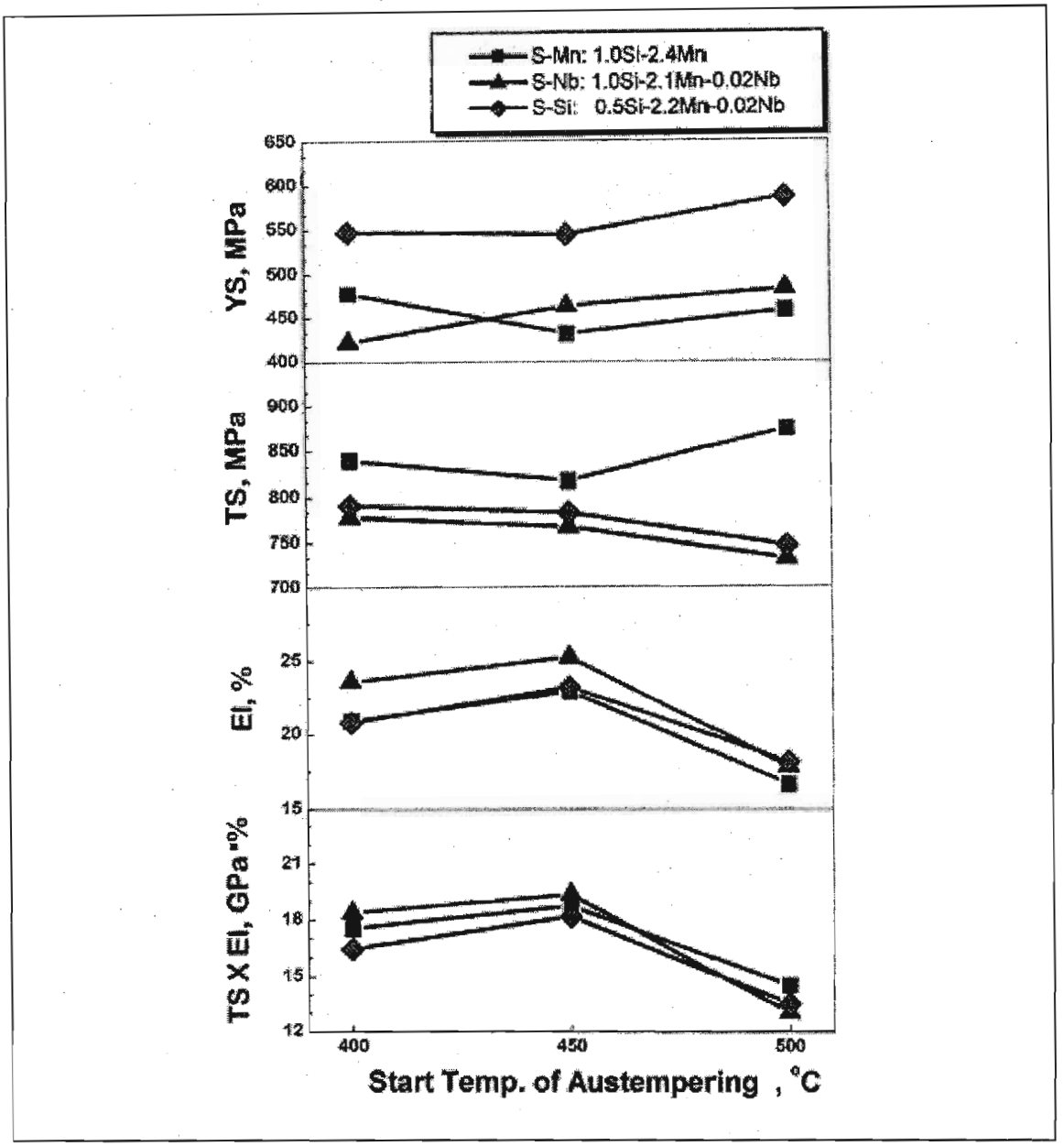


Figure 7.11: Mechanical properties as a function of austempering start temperature. Annealing performed at 830℃ for 51 seconds, followed by austempering at various temperatures for 300 seconds.

Phase Volume Fraction and Transformation Temperature

Figure 7.12 illustrates the changes of the austenite volume fraction of sample S-Mn as a function of holding time at various intercritical annealing temperatures.

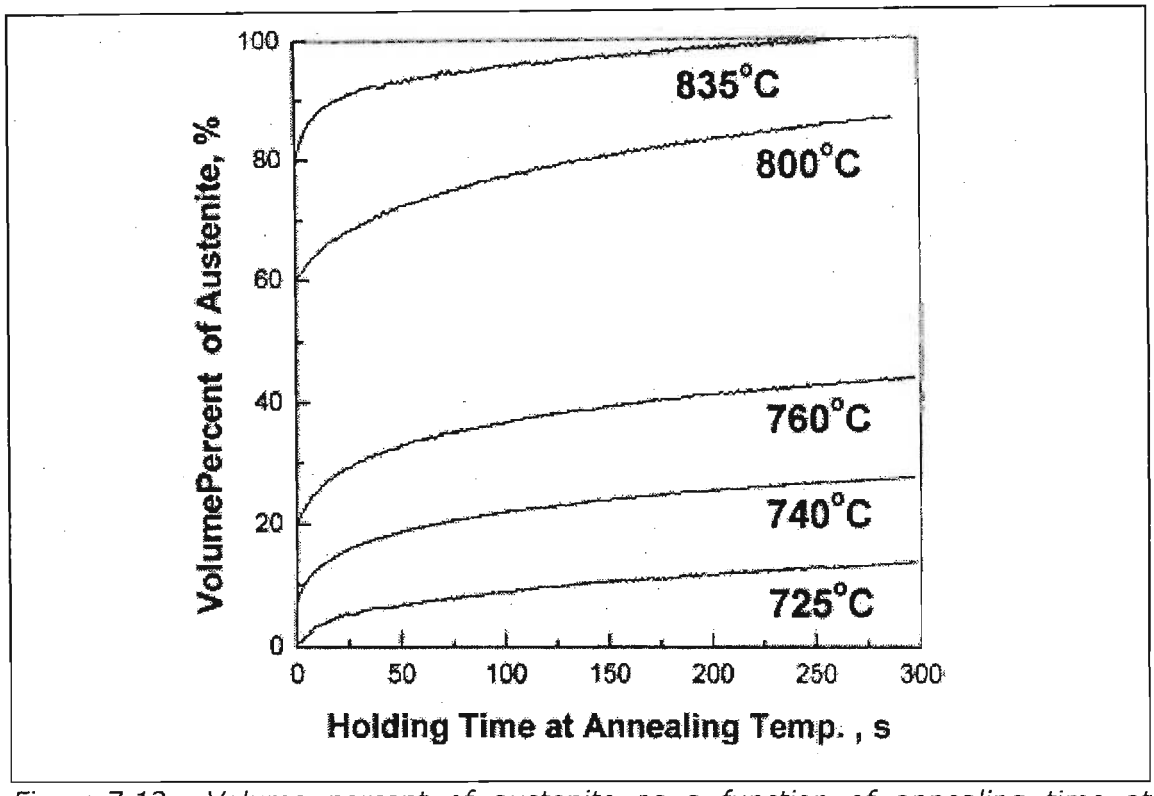


Figure 7.12: Volume percent of austenite as a function of annealing time at various intercritical annealing temperatures for sample S-Mn.

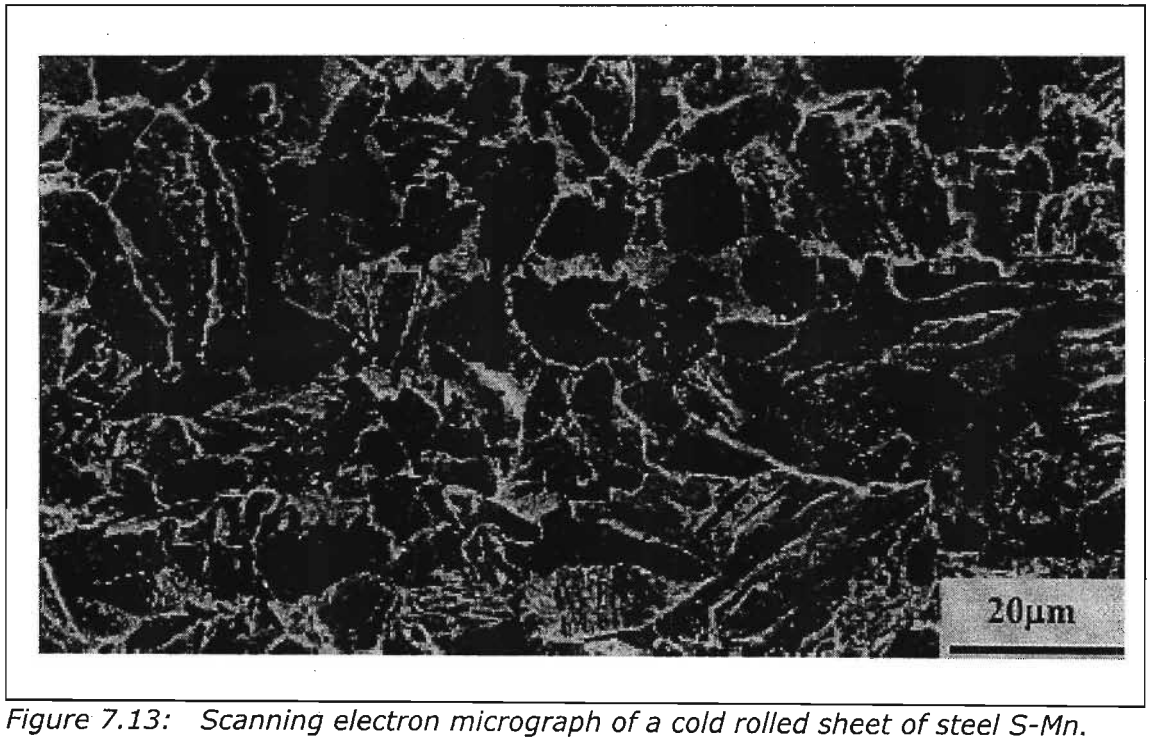


Figure 7.13: Scanning electron micrograph of a cold rolled sheet of steel S-Mn.

Figure 7.13 shows the cold rolled steel composition (pearlite and ferrite) of sample S-Mn before annealing. For annealing at 740°C, it was observed that pearlite transforms to austenite in under 10 seconds (Figure 7.14). This confirmed that pearlite is more rapidly transformed to austenite on heating than ferrite.

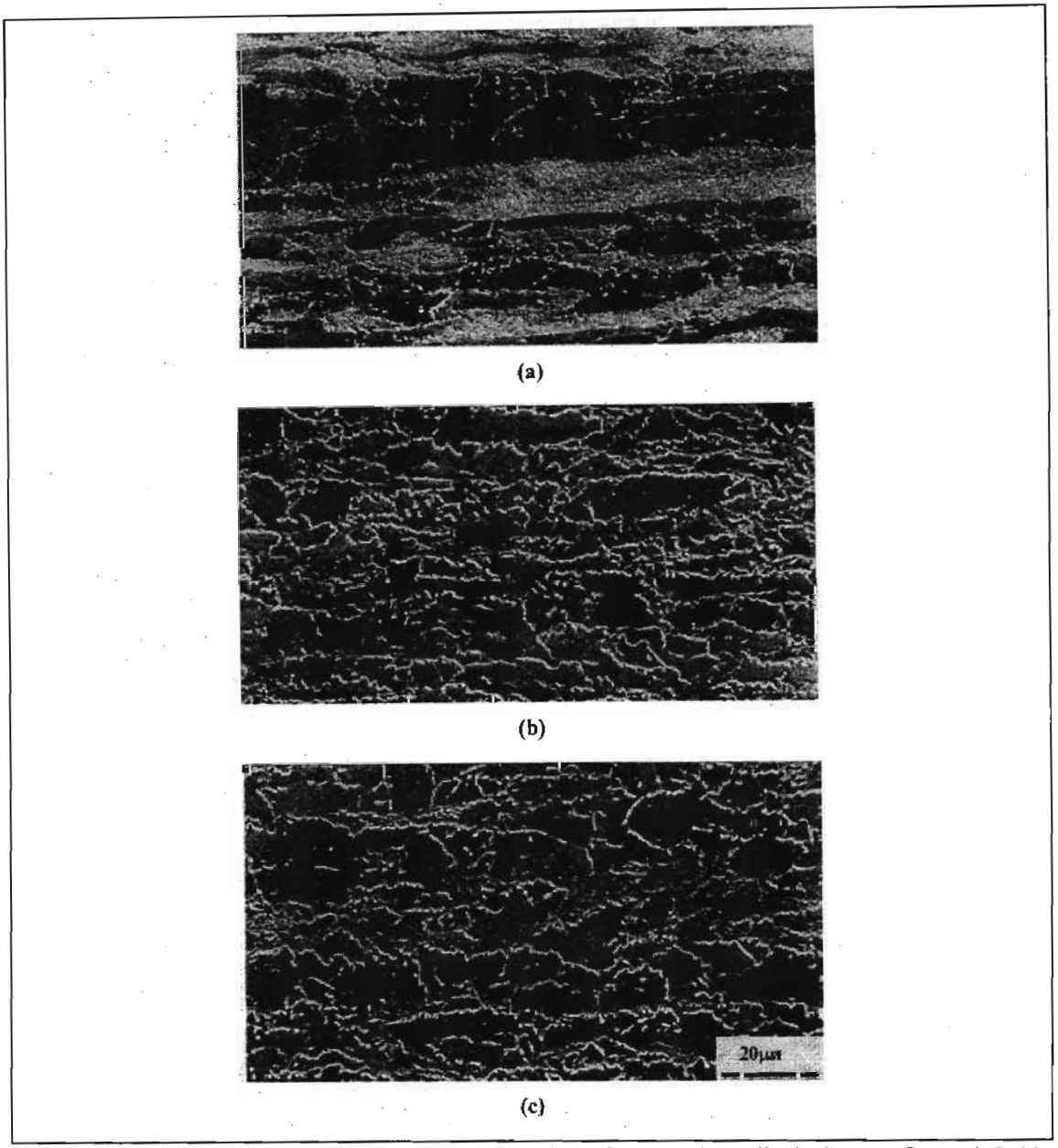


Figure 7.14: Scanning electron micrographs of the cold rolled sheet of steel S-Mn heated at a rate of 6℃ per second and held at 740℃ for (a) 0 s, (b) 10 s and (c) 100s.

The volume change of bainite, ΔV_B, for steel S-Mn is shown in Figure 7.15 as a function of austempering time at various austempering temperatures. ΔV_B in the various steels austempered for 300 seconds is shown in Figure 7.16 as a function of austempering temperature (temperatures of 400°C, 450°C and 500°C were used).

The following observations were made:

1. Austempering at 450°C resulted in the lowest bainite volume fraction for all compositions.

6. The addition of Nb had a slight effect on the bainite formation.

7. ΔV_B increased as the Mn content was increased. This is because Mn retards the transformation during cooling prior to austempering, hence a large amount of austenite is transformed to bainite during austempering.

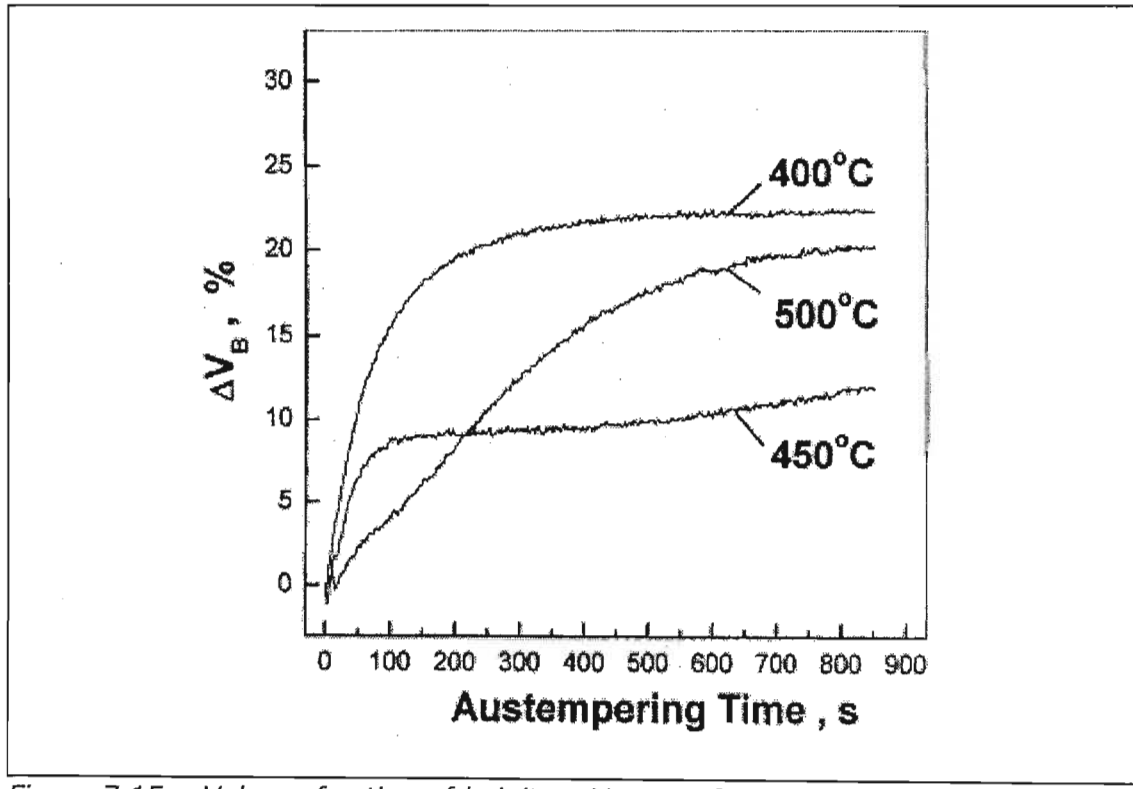


Figure 7.15: Volume fraction of bainite, ΔV_B, transformed during austempering for steel S-Mn as a function of the austempering time at various temperatures.

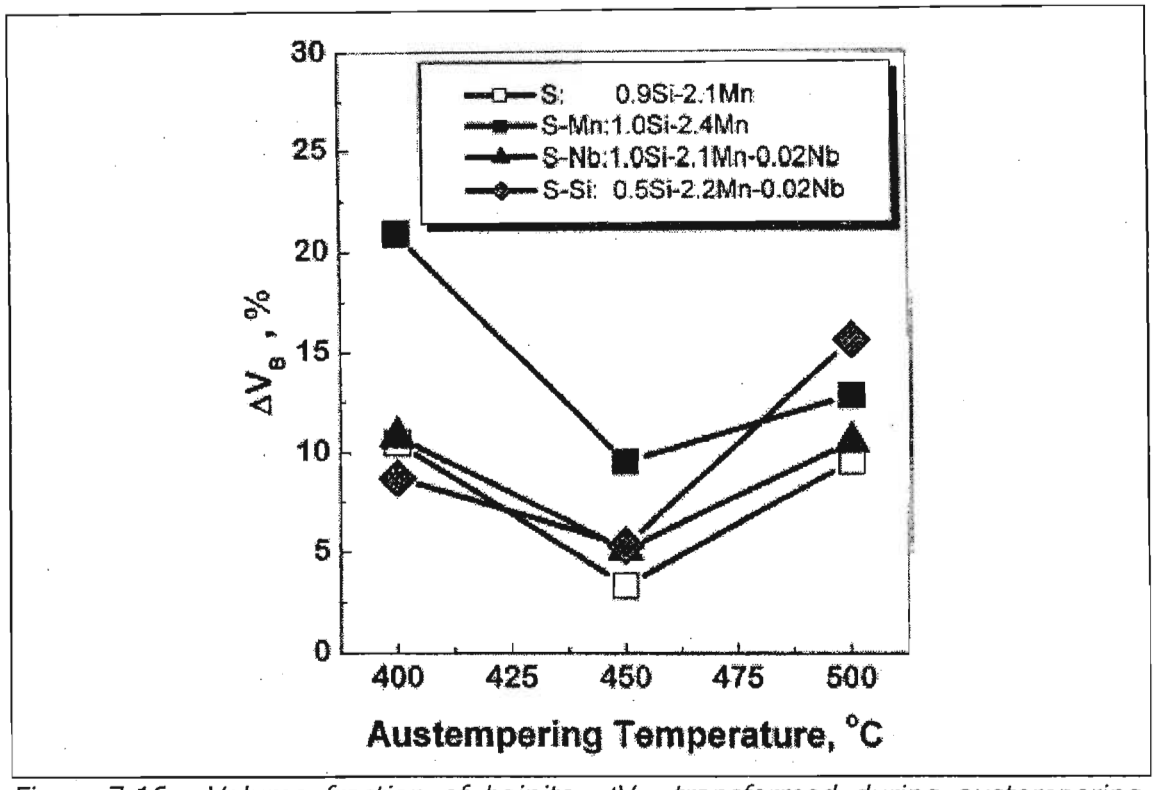


Figure 7.16: Volume fraction of bainite, ΔV_B, transformed during austempering, ΔV_B, for the sample steels as a function of the austempering temperature for 300 seconds.

To examine the stability of austenite, the martensite start temperature, M_s, was measured as a function of ΔV_B during austempering for 0 to 900 seconds at various temperatures (400°C, 450°C, 500°C). Refer to Figure 7.17.

The results indicate that M_s decreases with increasing ΔV_B. The rapid rate of decreasing M_s at 450°C can be explained in terms of the rapid diffusion of soluble C to austenite, hence stabilising the retained austenite. To confirm the effect of the austempering condition on retained austenite, the volume fraction of retained austenite was measured as a function of ΔV_B (Figure 7.18).

The fastest rate of decreasing M_s occurs at the austempering test temperature of 450°C. This explains why the highest elongation was obtained at the austempering temperature of 450°C in Figure 7.11.

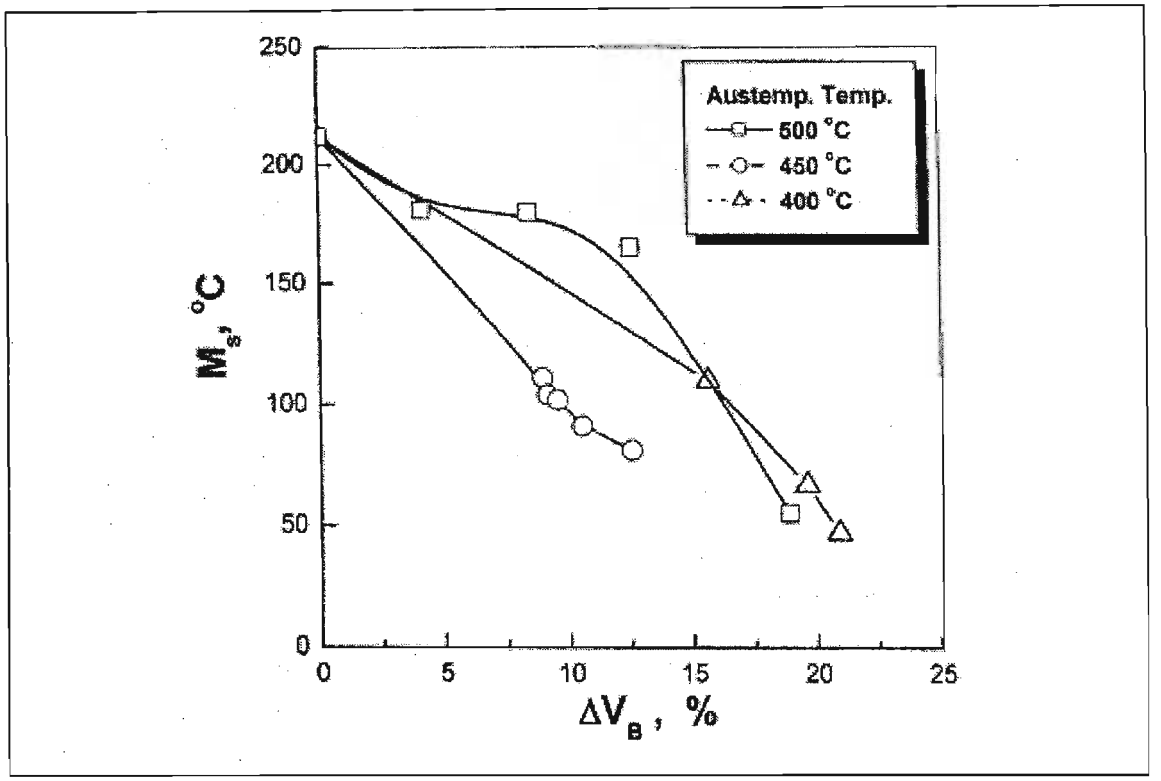


Figure 7.17: Martensite start temperature, Ms, as a function of the volume fraction of bainite transformed, ΔV_B, at various austempering temperatures (Sample S-Mn).

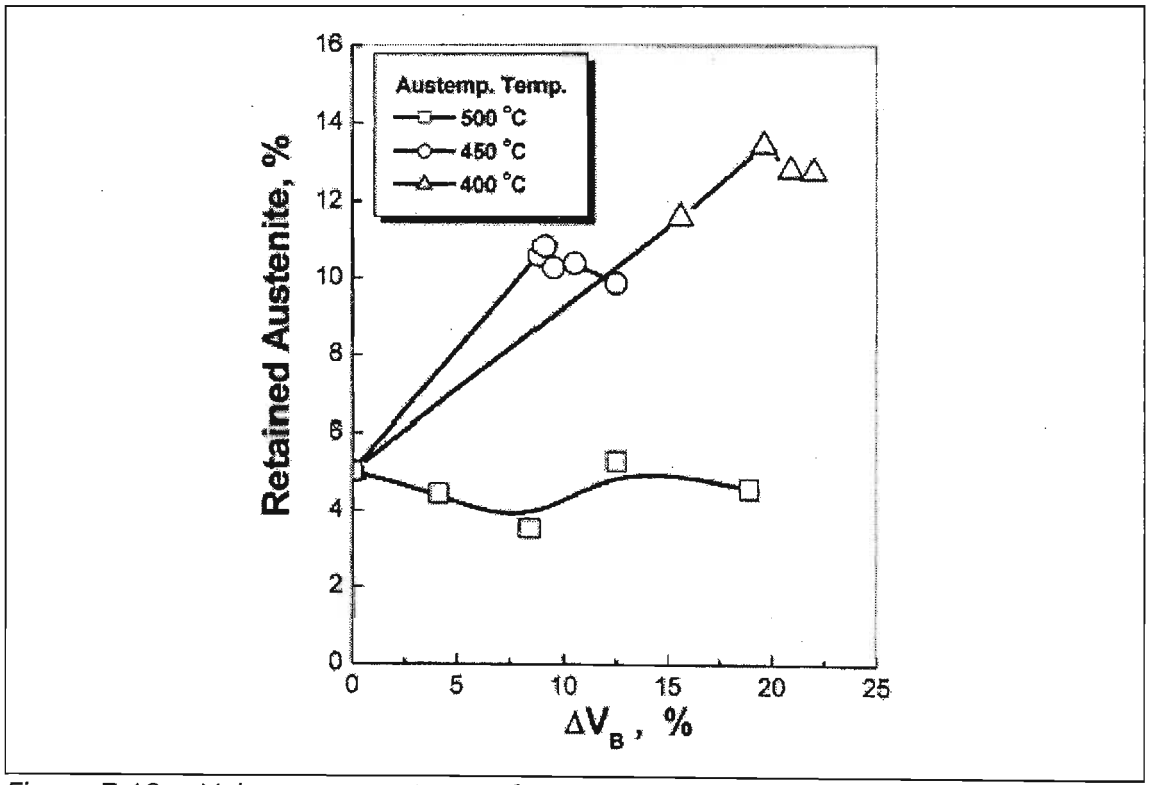


Figure 7.18: Volume percentage of retained austenite for sample S-Mn as a function of the volume fraction of bainite transformed during austempering, ΔV_B, at various temperatures (400℃, 450℃ and 500℃).

To examine the effects of the chemical composition on M_s, the relationship between M_s and ΔV_B was analysed for different samples austempered at 450°C as shown in Figure 7.19.

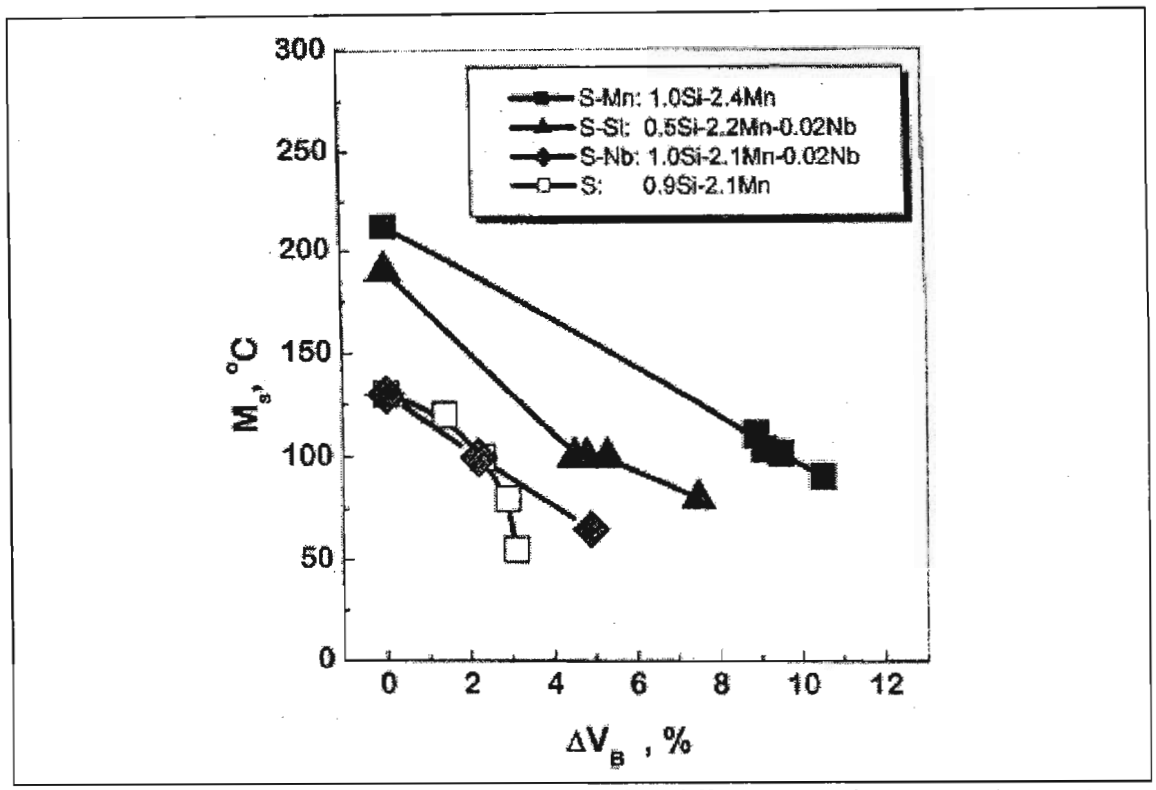


Figure 7.19: Martensite start temperature for different steel compositions, M_s, as a function of the volume fraction of bainite transformed during austempering, ΔV_B, at 450°C.

The following observations were made:

1. The martensitic start temperature is dependent on chemical composition.

8. M_s decreases with an increase of ΔV_B.

9. The addition of Nb has little effect on the variation of Ms.

10. Increasing the Si content decreases M_s by stabilising the retained austenite.

11. M_s is raised by increasing the Mn content. This is due to the reduction of volume fraction of ferrite formed during cooling to austempering temperature (as an austenitic stabiliser, Mn retards ferrite formation).

Stabilisation of Austenite vs. Austempering Time

Figure 7.20 shows how M_s is reduced with increased austempering time. The samples of S-Mn and S-Si exhibited a significant slowing in the reduction of M_s after 200 seconds of austempering.

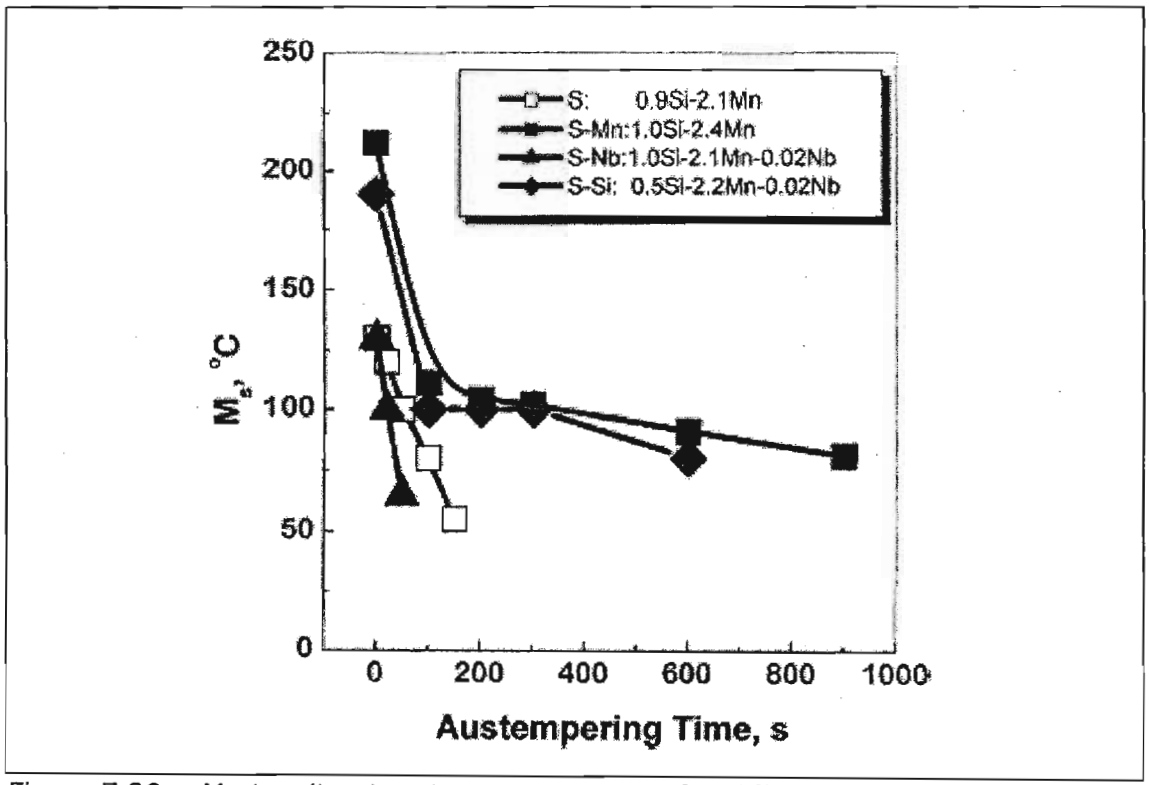


Figure 7.20: Martensite start temperature, Ms, for different steel compositions as a function of the austempering time at 450℃.

The retained austenite fraction in the various samples austempered at 450°C for 0 to 900 seconds was measured as shown in Figure 7.21, indicating an increased volume fraction of retained austenite with decreasing M_s. Baik et al. postulated that the retained austenite fraction can be predicted using the fit curve and the M_s obtained in the dilatometric curve.

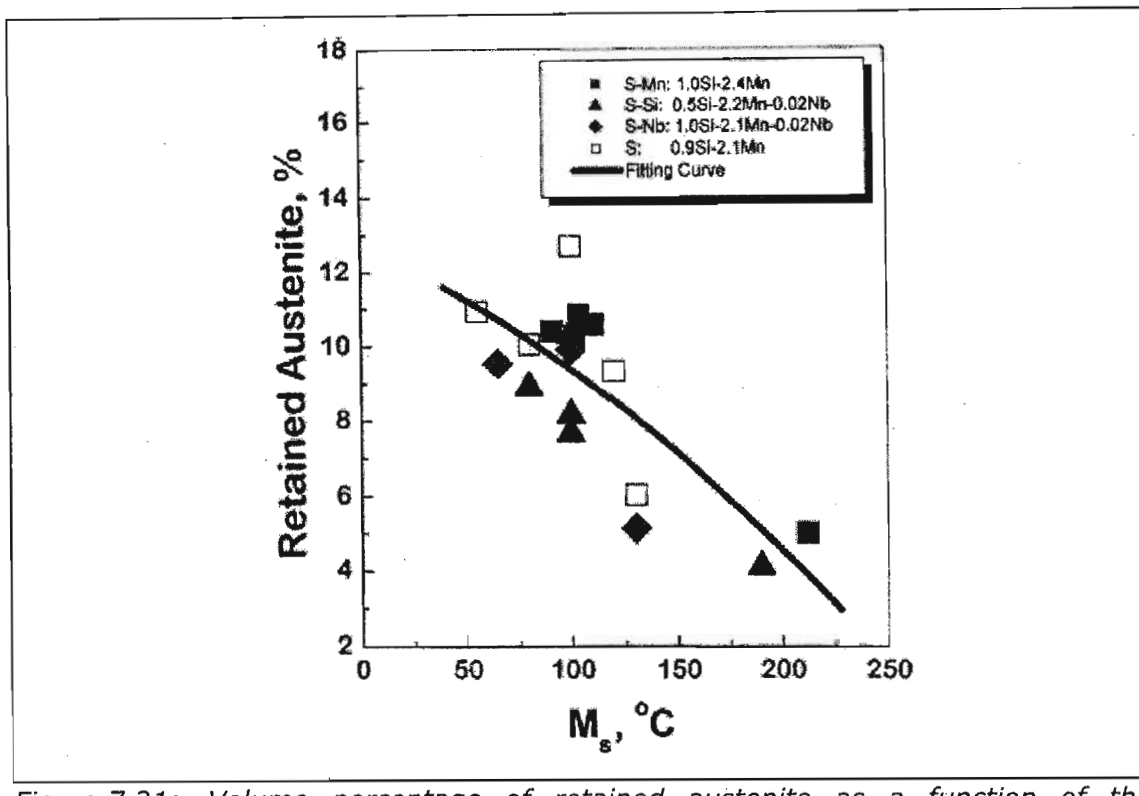


Figure 7.21: Volume percentage of retained austenite as a function of the martensite start temperature, Ms, for different samples austempered at 450℃ for 0 to 900 seconds.

Interpretion of the Test Results

Effect of Mn Content on Austenite Stability

Manganese has been identified in earlier sections as an austenitic stabiliser. In order to analyse the effect of the Mn content on M_s, samples of S and S-Mn were annealed to obtain the austenite volume fraction of 28% and 100%, followed by cooling to room temperature at various cooling rates.

Figure 7.22 shows the change in M_s as a function of cooling rate for the 2 alloys. The results indicate:

1. M_s is *lower* for the S-Mn steel than for the S steel for cooling from 100 % austenite.

2. This is reversed for cooling from 28% austenite, where M_s is *higher* for the S-Mn steel than for the S steel.

This can be explained by the volume fraction change of ferrite generated during cooling to the austempering temperature:

1. When the specimens were cooled at the rate of 100°C per second after forming 100% austenite, the ferrite formation was very little (as shown in the scanning electron micrograph images of Figure 15.27). In this case, M_s of the lower Mn steel was lower than that of the higher Mn steel.

2. When the austenite volume fraction was 28%, the higher carbide content in austenite lowered M_s for the lower Mn steel below that of the higher Mn steel.

The results also indicated that for all conditions, Ms decreased for slower cooling rates. This is because a decrease in the cooling rate results in an increase in the ferrite formation, as shown in Figure 7.23. The increased ferrite causes the increase in carbide content in austenite, resulting in a progressive decrease in M_s as the cooling rate is decreased.

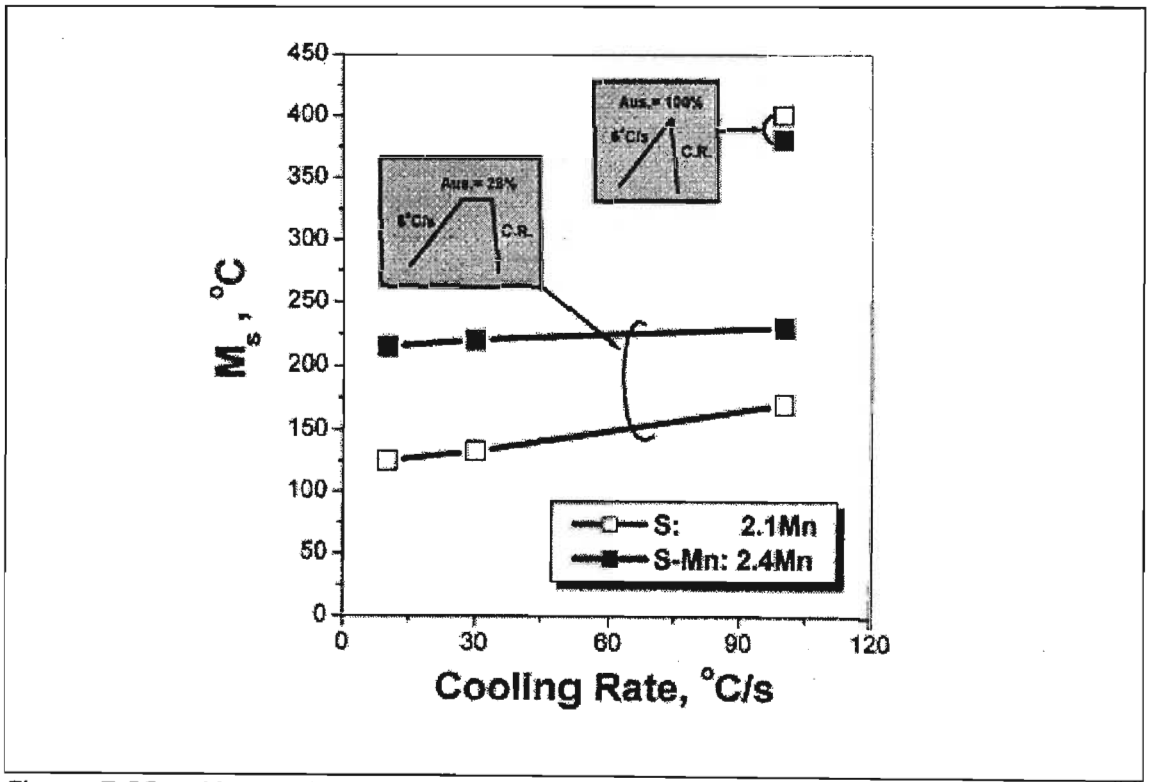


Figure 7.22: Martensite start temperature, M_s, for the steels heated to form 28% and 100% austenite, as a function of the cooling rate.

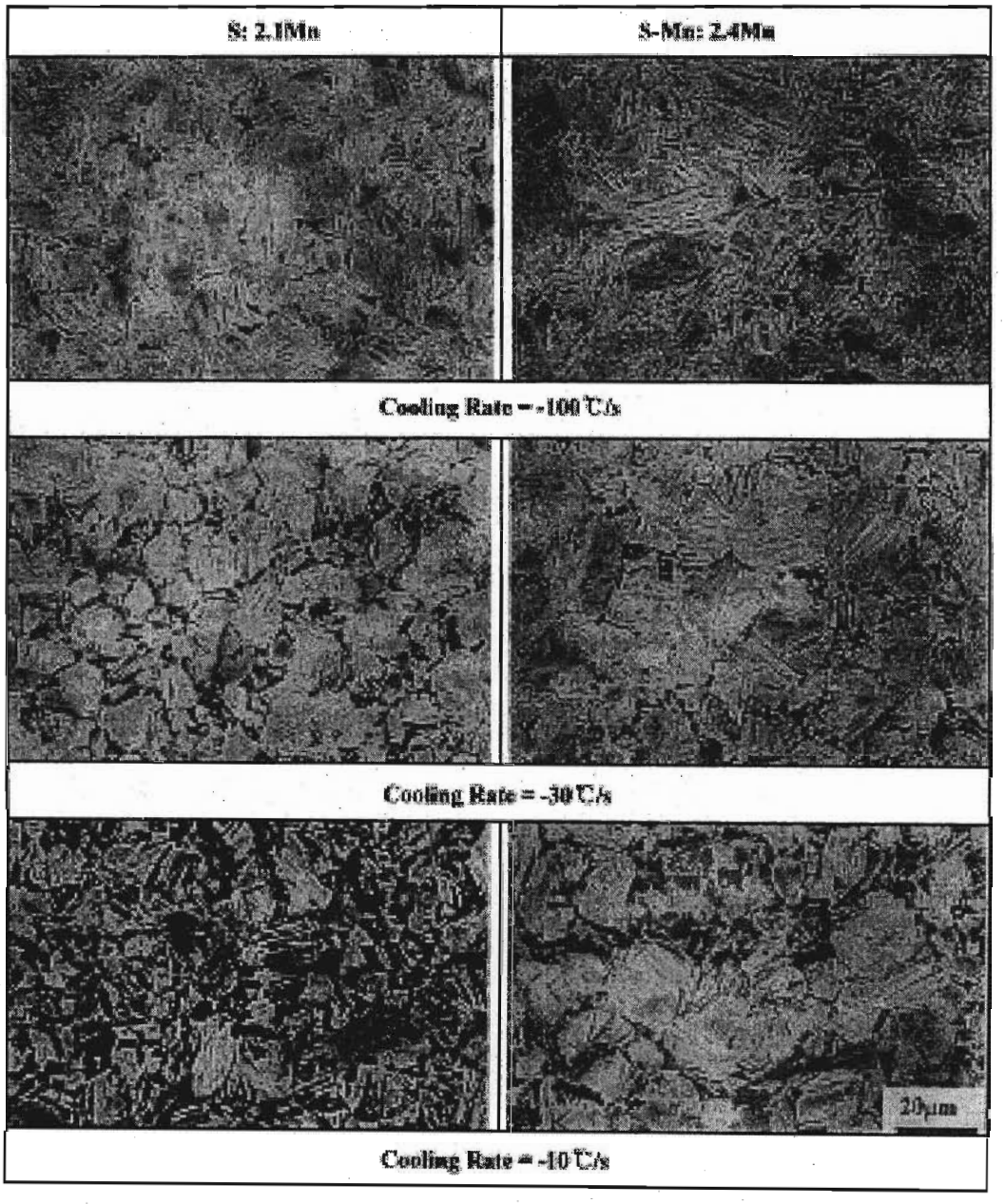


Figure 7.23: Scanning electron micrographs of the steels heated to form 100% austenite, followed by cooling at various cooling rates.

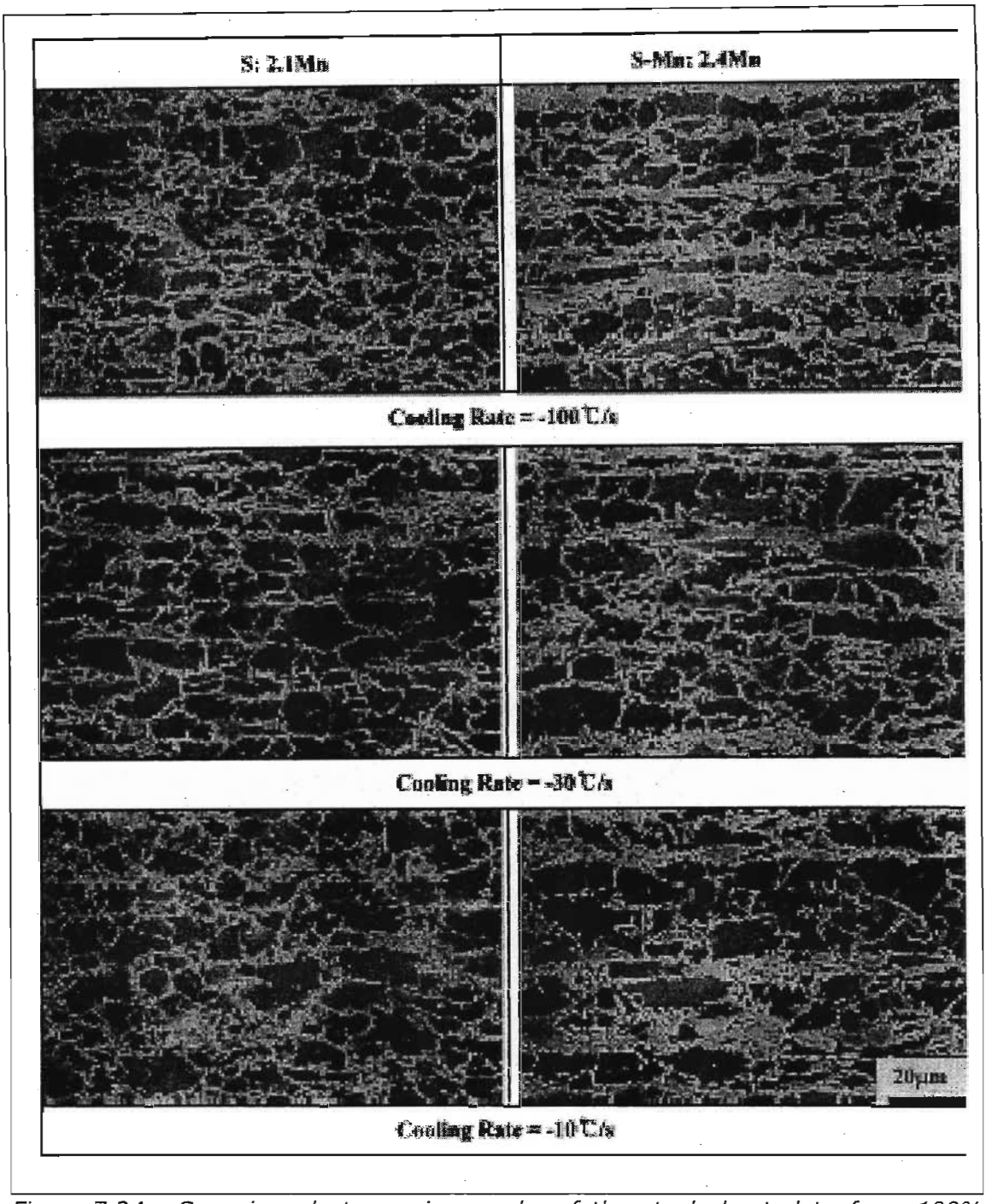


Figure 7.24: Scanning electron micrographs of the steels heated to form 100% austenite, followed by cooling at various cooling rates.

Phase Transformation Behaviour at Austempering Temperature

Bainite morphology is dependent on the austempering temperature. Variations to the bainite morphology:

1. causes variation of M_s

2. causes variations in the retained austenitic fraction for austempering temperatures between 400°C and 500°C (illustrated in Figures 7.17 and 7.18)

Maximum elongation is obtained at the austempering temperature when the rate of decrease of the martensitic start temperature as a function of the bainite volume fraction, M_s vs. ΔV_B, is greatest (Figures 7.8 and 7.16).

The soluble C diffuses to the retained austenite as the bainite volume fraction generated during austempering increases, as illustrated in Figure 7.19. This indicates that the retained austenite becomes more stable as the bainite volume fraction increases.

Transformation Behaviour vs. Steel Composition

Mn is an austenitic stabiliser, i.e. it stabilises the retained austenite. There is an optimal content of Mn for a given steel alloy. If an excess of Mn is added, ferrite transformation is retarded. This results in the decrease of soluble C and the destabilisation of the retained austenite.

An addition of Si retards precipitation of cementite and decreases the soluble C content in ferrite. This results in the increase of soluble C content in austenite. As shown in Figure 7.19, the decrease of Ms with increasing Si content indicates that the cementite formation can be effectively prevented by Si during cooling to the austempering temperature.

Neither the bainite volume fraction nor the martensitic start temperature, Ms, were significantly affected by the addition of Nb (Figure 7.16 and 7.19). This implies that the Nb precipitates have a negligible effect on the transformation.

Conclusions

In order to analyse the effects of alloying elements Mn, Si and Nb on the transformation behaviour and mechanical properties of 0.14%C-(2.1~2.4)%Mn-(0.5~1.0)%Si-(0.00~0.02)%Nb TRIP steels, tensile and dilatation tests were conducted. The following observations were made:

1. The steel austempered at 450°C exhibited the highest balance of UTS * elongation because Ms was most rapidly decreased with transforming to bainite at 450°C.

2. The volume fraction of retained austenite was found to increase as M_s decreases for the steels austempered at various conditions.

3. M_s was found to be dependent on the chemical composition and decreases with increasing bainite volume fraction. The declination rate of M_s vs. the bainite volume fraction was little influenced by the composition.

4. Ms increased as the amount of Mn was increased from 2.1% to 2.4%. This was due to the retarded ferrite transformation which resulted in a decreased soluble C content of the partially transformed austenite.

5. M_s was found to decrease with increased Si content just prior to austempering. This indicated that the formation of cementite was effectively prevented due to Si during cooling to the austempering temperature.

6. Both the bainite volume fraction and Ms were little influenced by the addition of Nb, implying that Nb precipitates had little effect on the bainite transformation behaviour.

8. TRIP STEELS FOR SMART SHM MINING BOLTS

8.1. TRIP STEEL FULFILMENT OF SHM REQUIREMENTS

8.1.1. Criteria for Diagnostic Data

Although seismic events cause fracturing of the surrounding rockmass, this study is solely interested in the rockbolt units themselves. Ultimate bolt failure is not fatigue-related, therefore the accumulated damage to the bolt during rockmass movement (slip) is deemed to be of greater significance than the seismic load characteristics (amplitude, angular frequency, etc.). Ultimate rockbolt failure (fracture) is generally attributed to massive quasi-static deformation. Rockmass slip is a direct consequence of seismic fracturing and other natural shifting.

Suitable diagnostic data for damage detection and interpretation should provide an indication as to the deformed state of a particular unit when interrogated. Transformation-induced plasticity (TRIP) steels produce a magnetic field that becomes stronger with increasing deformation. Because the magnetic field can be measured and processed, it is a suitable candidate for diagnostic data.

8.1.2. Integration Parameters

Holes are drilled into the rockmass for a suitable bolt to be installed. Often the bolt is bonded to the rockmass by a sleeve of grout. The design of a suitable SHM system would need to cater for these physical parameters in order to integrate the monitoring system into the support system.

In an ideal case, a smart material will form part of the structure itself. In addition to fulfilling any load-bearing requirements, the support structure would be capable of measuring its own level of damage. Few currently available smart materials have the capacity to fulfil the requirements of both roles. At this time, strain memory alloys appear to be the strongest contenders.

Transformation-Induced Plasticity (TRIP) steels are strain memory alloys that exhibit high strengths and are capable of withstanding enormous strains (deformations). From a physical perspective, a bolt made from such a material is superior to off-the-shelf units.

Because of the huge number of bolts required to provide adequate support to a particular mine tunnel, cost becomes an important consideration. TRIP steels are

expensive, partly because of their current limited use in industry. The cost of manufacture per unit would decrease significantly if the demand (nationally and internationally) increases. To keep costs down, not every unit placed need be made of TRIP steel; perhaps for example only 1 in 10.

8.1.3. The Proposal

TRIP steels exhibit the 5 criteria required of a smart SHM system for use in mining tunnel support applications.

Sensor: TRIP steel.

Diagnostic data: Magnetic field.

Signal Processing: Plot the magnetic field value onto a curve that measures magnetic field strength as a function of applied strain.

Interpretation: If the magnetic field is affected by e.g. temperature variations, these variations must be taken into account to improve the accuracy of the results.

Integration: The rockbolt unit itself would be made from strain memory alloy. Modifications to the bolt for the measurement equipment to be able to function correctly would be required.

8.2. TRANSFORMATION-INDUCED PLASTICITY

8.2.1. Early Research Areas

TRIP steel development initiated [42] from research in 1967 at U.C. Berkely into ultra-high strength structural steels exhibiting improved toughness (i.e. fracture resistance) over the high-strength, low-alloy steels commonly used in aircraft landing gear and similar applications [43].

At that time it was already known that some austenitic stainless steels are metastable with respect to deformation. As deformation of these materials progresses, the strained areas transform from austenite to the stronger martensite. The transformation is inhomogeneous because local weaker areas transform first. Necking in a test specimen is therefore delayed until the entire gauge length and beyond has transformed to the martensitic phase. Once the transformation progresses to a significant extent along the entire gauge length, final fracture

occurs when the UTS of the martensite is exceeded. Failure typically occurs with very little or no necking. At Berkeley, specimen elongations of 40 to 60% were observed for materials with yield strengths of over 1380MPa. These measured levels of ductility far exceeded those of high-strength steels, where typical elongations are around 10 to 15%.

With their combination of high strength and extraordinary ductility, the austenitic stainless steels were found to be ideal as replacements for ultra-high-strength low alloy steels and civil engineering structural steels since the phase transformation to martensite at micro-crack tips prevents further propagation of the cracks. This was the original reason for the classification as smart materials.

During this period, research was also being conducted into ausforming methods of thermo-mechanical fabrication. This warm rolling process increases the yield strength of metastable austenite by producing a high dislocation density [37]. The procedure is performed at a temperature above that necessary to initiate any strain-induced phase transformations.

It was found that it is possible to develop an austenitic microstructure in the Fe-Mn-Cr system, and attempts were made to develop austenitic stainless steels using this base alloy system. The combination of findings from subsequent research into ausforming and into metastable phase transformations in AISI 300 austenitic stainless steels and Fe-Mn-Cr-C alloys were key to further TRIP steel development [42]. Manganese (Mn) has the effect of reducing the stacking fault energy of Fe-based austenites, which results in a hexagonally close-packed (HCP) epsilon (ε) martensite phase. This phase is of intermediate thermodynamic stability with respect to the parent austenite and the alpha prime (α') martensite (still the most stable product phase) [42]. Fe-Mn-Cr phase transformations are triggered by lowering the *temperature* below the ε and α' martensite start temperatures just as in the austenitic stainless steels; however, they also exhibited the same transformation sequence when mechanically *strained* (the basis for TRIP steels), provided that the phase stabilities were adjusted properly through control of the alloy chemistry.

In spite of high levels of interest in the 1970's [47], TRIP steels were never fully commercialised primarily because of their high manufacturing costs. These costs can be attributed to the alloy chemistries, and to fabrication difficulties arising from their high toughness and energy absorption capabilities. Also, in their high-

strength condition that requires additional thermo-mechanical processing, the resulting microstructures are not weldable.

8.2.2. Magnetic Properties of TRIP Steels as the Basis for a Smart SHM System

During deformation, TRIP steels undergo an irreversible, solid-state, strain-dependent phase transformation from the metastable austenitic parent phase to the thermodynamically stable martensitic product phase. This strain-induced phase transformation is responsible for the large values of ductility and strain hardening that is observed in these steels. The phase transformation is triggered as the material yields and begins to plastically deform, and the martensitic fraction increases with the applied strain.

The transformation from an austenitic face-centred cubic (FCC) crystal structure to a martensitic body-centred cubic (BCC) structure results in changes to the magnetic properties of the steel. The austenitic phase is paramagnetic and the martensitic phase is ferromagnetic. The magnetic field of the steel is therefore dependent on the fraction of martensite present, and increases as the transformation progresses with increasing levels of strain.

Because the phase transformation is irreversible, and because the magnetic field strength increases with increased martensite formation, only successively higher strain readings are "registered" by the volume fraction of martensite. A means of establishing the peak strain (damage accumulation) experienced by the material is thus available. This provides the principle basis for a passive smart structural health monitoring system for use in mining bolt applications. Power is required only while a particular location's magnetic strength is being measured at some pre-determined or chosen interval.

8.2.3. Physical Properties of TRIP Steels

Figure 8.1 compares the tensile properties of high-strength (Type I) and low-strength (Type II: Fe-Mn-Cr alloys) variants of TRIP steels with those of structural steels and high strength low alloy (HSLA) steels [42]. Structural steels are found in civil engineering applications, while HSLA steels are suitable for high stress applications. The strength of Type I TRIP steels compares favourably with the HSLA steels, and the ductility and energy absorption capacities of Type I and II TRIP steels are higher than those offered by the HSLA steels. The corrosion

resistance of the Type II TRIP steels is superior to that of the Type I steels due to the higher levels of chromium present.

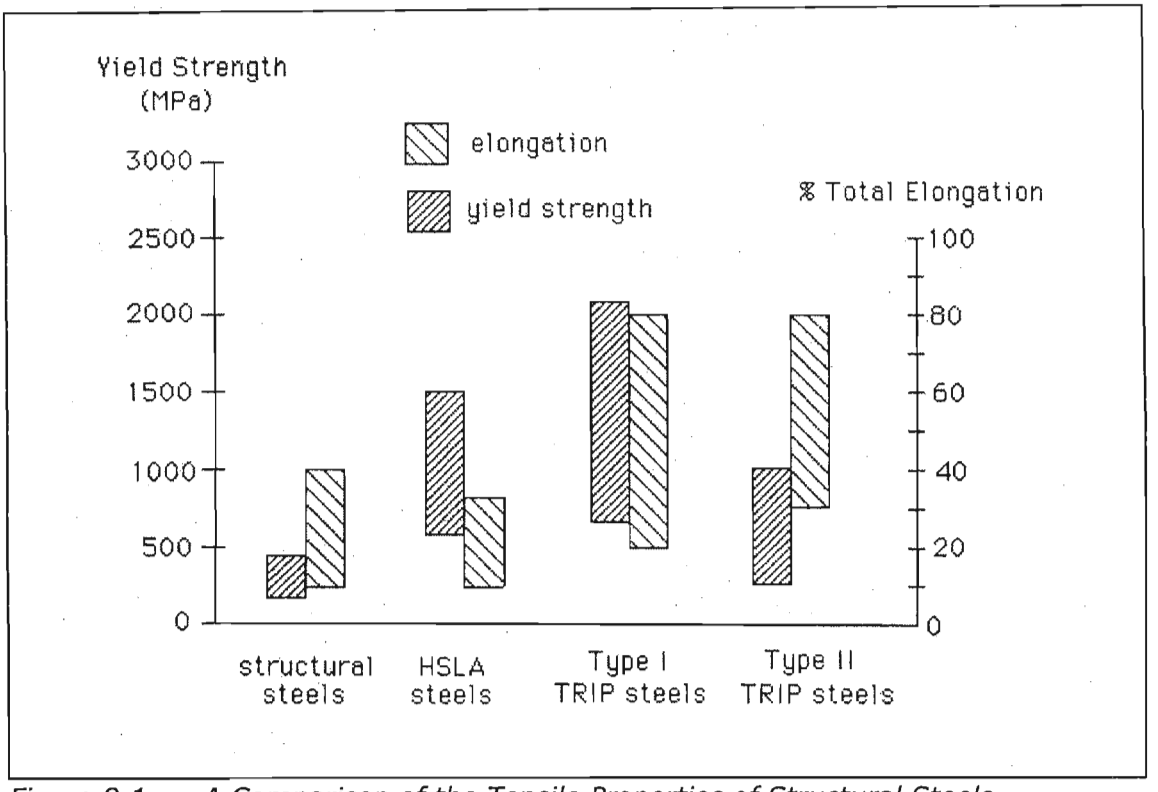


Figure 8.1: A Comparison of the Tensile Properties of Structural Steels.

Figure 8.2 compares stress-strain curves of TRIP steels, HSLA steel, and grade 60 structural steel. Although the Type II steels are not as strong as the Type I steels, it should be noted that they have not been evaluated other than in the as-austenitised condition using the simplest heat treatment schemes and either air-cooled or quenched to room temperature. To obtain the highest yield strengths depicted in Figure 8.1 for the Type I TRIP steels, it is necessary to thermo-mechanically process the material using complicated and expensive schemes.

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 condition are referred to as Low Strength Type I TRIP steels. The lower yield strengths of the Type II TRIP steels, more typical of those in standard structural grade steels, also 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 to yield strength ratios in the range of three to four being common [42].

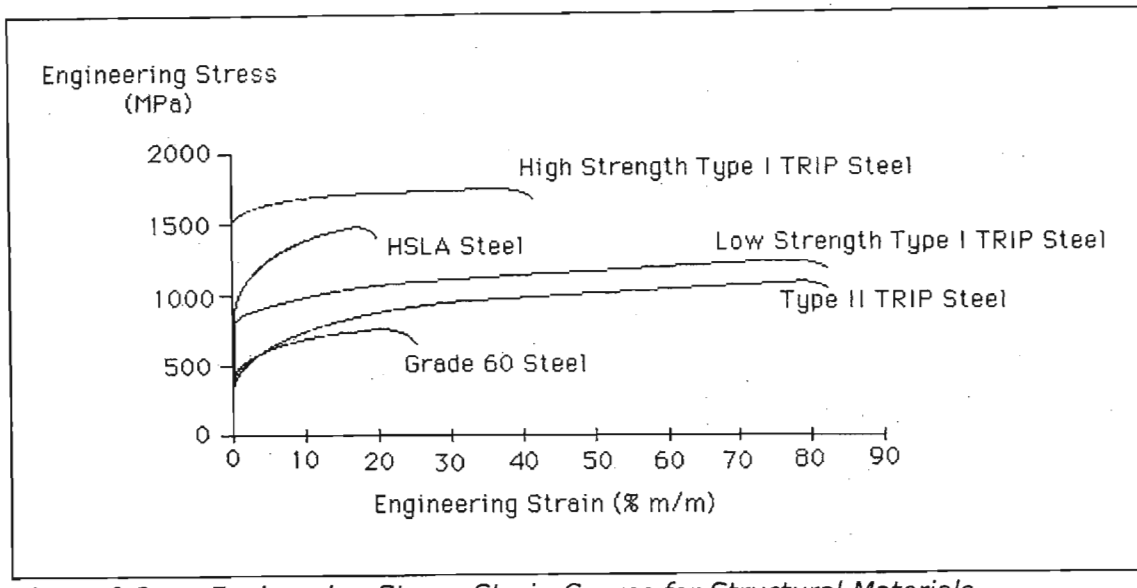


Figure 8.2: Engineering Stress-Strain Curves for Structural Materials.

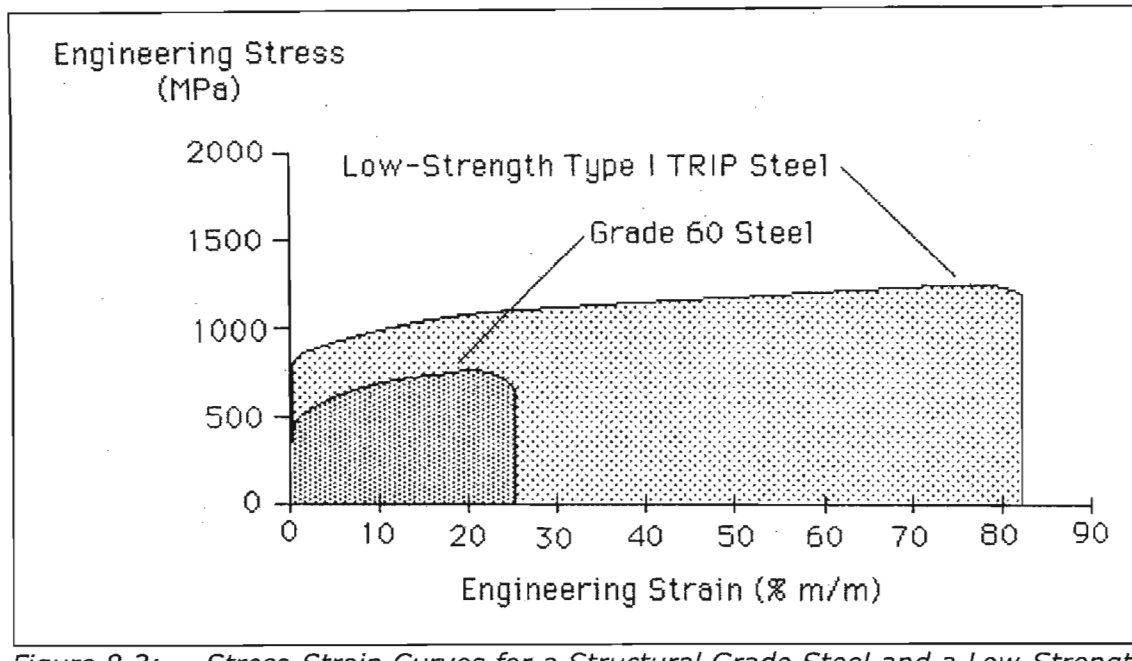


Figure 8.3: Stress-Strain Curves for a Structural Grade Steel and a Low-Strength TRIP Steel.

Figure 8.3 uses the information from Figure 8.2 to compare the stress-strain curves for a low-strength TRIP steel and a Grade 60 structural steel. The high levels of toughness exhibited by the TRIP steels (indicated by the large area under the engineering stress-strain curve) are a direct consequence of the phase transformation of the austenitic parent phase into the stronger martensitic phase.

Both the Type I and Type II TRIP steels experience the strain-induced phase transformations responsible for delaying mechanical instability and necking in uniaxial tensile testing. The increased uniform plastic deformation observed in these alloys corresponds to increased levels of energy absorption capacity. A direct

measurement of the fracture energy absorption capacity can be found using the true stress-strain curves, i.e.

$$E_V = \frac{\text{energy absorbed}}{\text{unit volume}} = \int_0^{\varepsilon_f} \sigma(\varepsilon) d\varepsilon \qquad (8\text{-}1)$$

where $\sigma(\varepsilon)$ represents the true stress expressed as a function of true strain, and E_V is the energy absorbed during fracture.

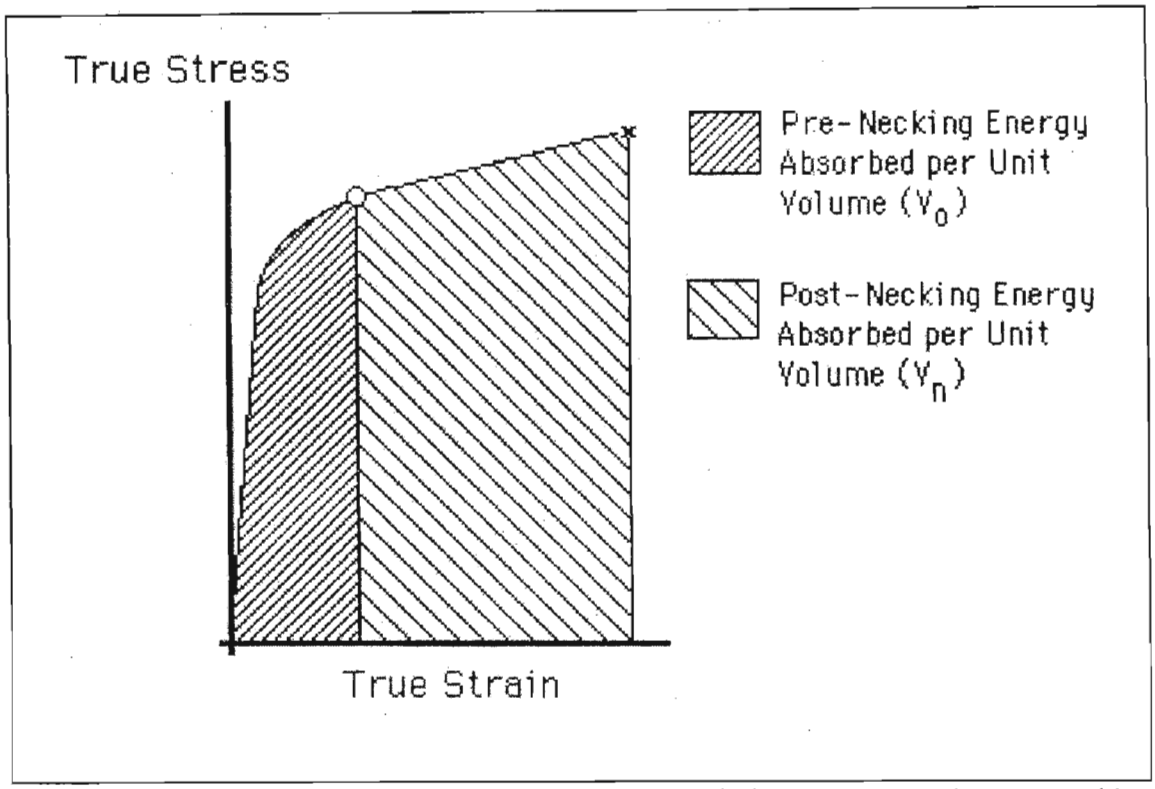


Figure 8.4: A True Stress - Strain Curve Divided into Pre and Post-Necking Regions.

Energy absorption of any test piece should be considered in terms of pre-necking and post-necking regions as shown in Figure 8.4. While the energy absorbed per unit volume is generally much greater after the neck forms, the actual volume of material participating in the deformation process plays an important role. TRIP steels display uniform elongations (i.e. non-localised deformation throughout the entire gauge length), therefore the entire gauge length contributes to energy absorption. Once a mechanical instability initiates during plastic deformation, any further straining is concentrated within the unstable region, leading to the formation of a true neck in conventional steels. In TRIP steels, the unstable region transforms from the weaker austenitic phase to the stronger martensitic phase, resulting in the region becoming locally stronger than the surrounding material so that the deformation continues without the formation of a true neck.

Figure 8.5 shows the true stress-strain curves for a TRIP steel processed using three different treatments, as well as the stress-strain curves for two different tempers of 4340 HSLA steel. A circle on each of the curves shows the necking point for each test. Any further deformation beyond the necking strain is 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 neck volume absorbs energy. The true strain at necking (sometimes referred to as the strain hardening exponent, n) is therefore a 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 high ductility, which accounts for their increased fracture resistance.

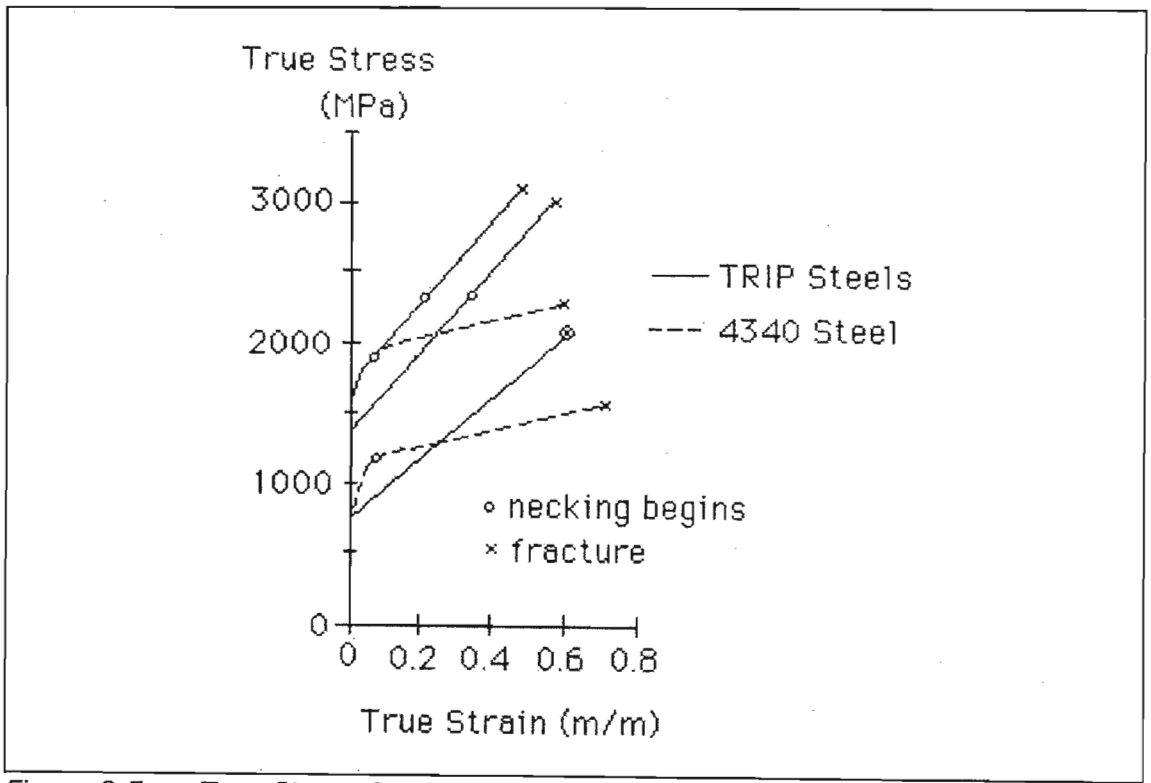


Figure 8.5: True Stress-Strain Curves for a HSLA Steel (AISI 4340) and a Type I TRIP Steel (different curves represent different processing conditions for the materials).

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

$$E_v = \int_0^{\varepsilon_{uts}} \sigma(\varepsilon)d\varepsilon + \int_{\varepsilon_{uts}}^{\varepsilon_f} \sigma(\varepsilon)d\varepsilon \qquad (8\text{-}2)$$

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 volume fraction participating in deformation before (V_0) and after (V_n) necking.

$$E_v = V_0 \int_0^{\varepsilon_{uts}} \sigma(\varepsilon)d\varepsilon + V_n \int_{\varepsilon_{uts}}^{\varepsilon_f} \sigma(\varepsilon)d\varepsilon \qquad (8\text{-}3)$$

Table 8.1 compares the energy absorption characteristics of the materials whose properties are displayed in Figure 8.5 [42]. The fraction of material volume associated with necking, V_n, was assumed to be one-tenth of the total gauge length volume, i.e. $0.1*V_0$. As mentioned previously it is assumed that the entire gauge volume absorbs energy during the uniform plastic deformation while the true strains are less than the strain hardening exponent n, whereas only the necking volume is absorbing energy at true strain levels greater than n. The table indicates that the three TRIP steel variants absorbed considerably more energy during fracture than the two variants of HSLA steels. Although both types of steel 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.

The data also shows that the lower strength TRIP steel variant, which has a true strain at necking equal to 0.60, absorbed considerably more energy than all the high-strength steels including the high strength TRIP steels.

These estimates of energy absorption indicate the enhanced structural safety that may be realized through selective utilisation of TRIP steels as structural components. The additional energy absorption would be particularly beneficial in structures subjected to cyclic deformation, delaying structural failure and enhancing structural safety.

Table 8.1: Energy Absorption Characteristics for TRIP Steels and HSLA 4340 Steel; $V_n = 0.1*V_0$.

	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

A true strain of 0.60 corresponds to an engineering strain of 0.82, i.e. a uniform elongation of 82%, which is exceptional for a material having a yield strength above 700MPa. This compares with an engineering strain at necking of 10.5% for the low strength 4340 and 8.3% for the high strength 4340 variant.

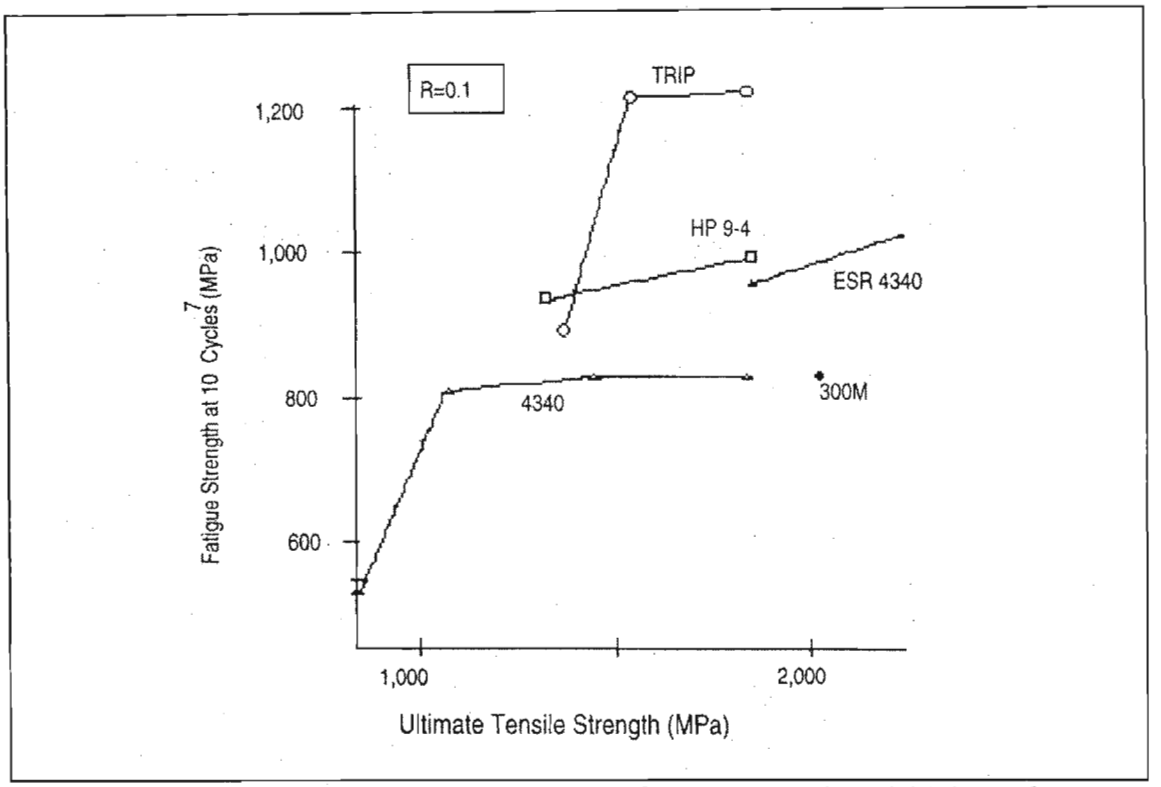


Figure 8.6: High cycle fatigue properties of a TRIP steel and high performance structural steels.

In addition to the studies mentioned above, low-cycle and high-cycle fatigue properties of TRIP steels have also been investigated [42]. The energy absorption capability of TRIP steels results in reduced rates of fatigue crack growth thereby delaying ultimate failure. Figure 8.6 shows the fatigue results (performed in the late 1970's) for TRIP steels as compared with those of conventionally melted 4340, electro-slag re-melted (ESR) 4340, high-strength alloy 300M (currently used in aircraft landing gear applications), and high performance alloy HP 9-4. The ultimate tensile strengths and fatigue endurance limits for 10^7 cycles are plotted for fatigue tests with a minimum to maximum load ratio, R, of 0.1. The TRIP steels displayed good fatigue properties, particularly in the higher strength materials where the combination of tensile strength and fatigue endurance limits are superior to those of the other commercially available materials.

8.3. TRIP STEELS AS SMART MATERIALS

TRIP steels fall into the Class I category of smart materials (i.e. passive monitoring systems), since they can act as both structure and sensor simultaneously. The material deformation state corresponds to the extent of martensitic transformation. The ferromagnetic response increases markedly as the amount or volume fraction of strain-induced martensite increases. Instrumentation is required to interrogate the material in order to obtain the ferromagnetic response, and ultimately correlate the response with the amount of strain required to produce the corresponding degree of phase transition.

The element geometry can also be adjusted to control the amount and location of material that actually deforms in response to structural loading. Magnification of the effect by increasing the attachment distance and decreasing the gauge length makes it possible to measure extremely small strains.

8.4. FEASIBILITY OF METASTABLE ALLOYS AS DEFORMATION-SENSING MATERIALS

8.4.1. Characteristics of Metastable Alloys

The strain-induced phase transformation exhibited by TRIP steels identifies them as candidate mechanisms for detecting and monitoring strain in mechanical components.

In a metastable sensor material such as TRIP steel, the product martensitic phase is ferromagnetic due to the BCC structure whereas the parent austenitic phase is FCC and exhibits no significant ferromagnetic response (is paramagnetic). It is therefore possible to monitor the extent of material deformation using a transformation calibration curve for the particular strain-sensing alloy.

The transformation behaviour depends on the chemical composition of the material. By tailoring the sensor alloy chemistry, one can produce different transformation rates and transformation strain ranges depending on the type of application being considered and the specific monitoring requirements.

Figure 8.7 shows a generic transformation curve for a metastable sensor alloy. The incubation strain corresponds to the initiation of the strain-induced transformation and is therefore the lowest measurable strain. It is possible to adjust or eliminate the incubation strain. The martensitic (α') transformation rate, which controls the

sensitivity and range of the sensor material response, can be increased or decreased depending on the monitoring strain range desired and the sensitivity required. A great advantage of TRIP steels is that the shape of the transformation curve can be controlled, and the amount of incubation strain can be modified. The sensor can therefore be adjusted to produce the optimal transformation behaviour for the application.

Steep transformation rates are desirable for many structural TRIP steel applications, e.g. aircraft fasteners such as bolts, where little deformation is generally acceptable before repairs are required. Incubation strains equal to or close to zero would also be required to identify an impending structural deformation problem as soon as possible. This approach, in which a deforming structural component is monitored over a certain time period, is possible only where some structural instability (albeit perhaps very small) is tolerable in the design.

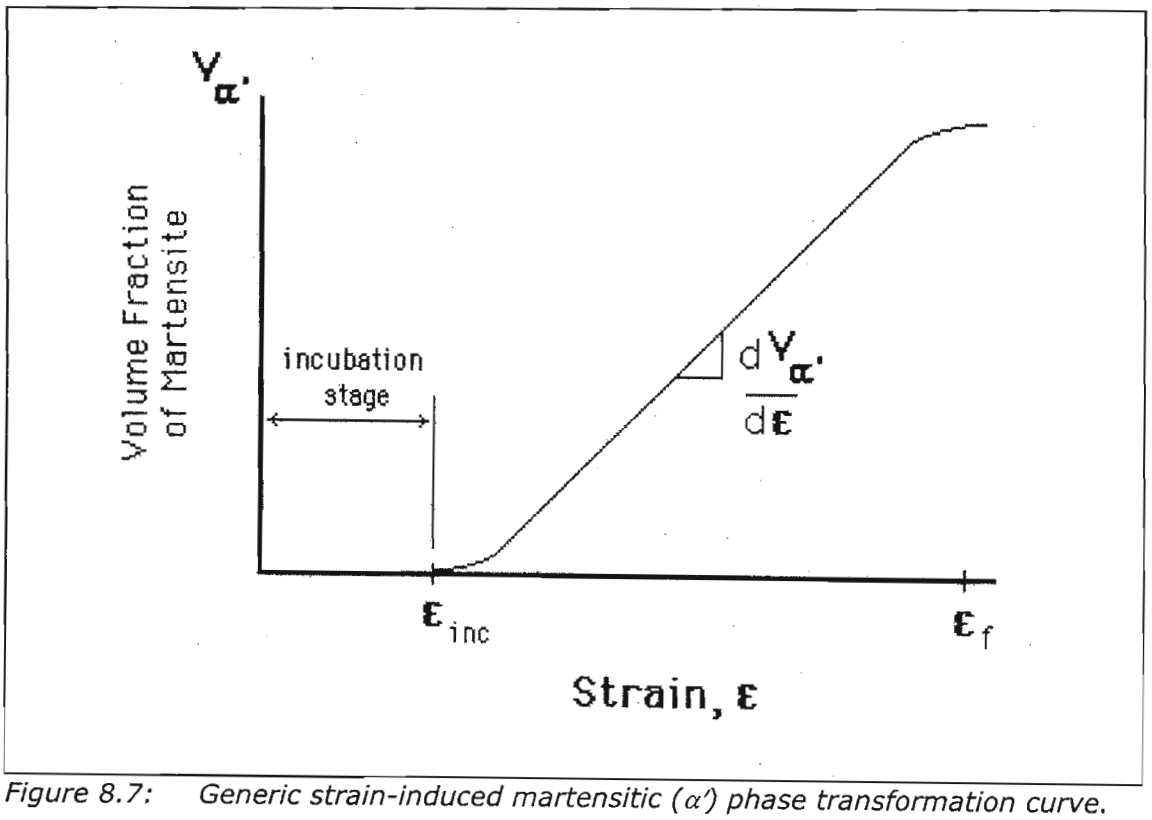


Figure 8.7: Generic strain-induced martensitic (α') phase transformation curve.

Deformation is confined to the gauge length of the sensor element. If the structural attachment distance is much greater than the gauge length of the sensor element, all of the structural deformation becomes concentrated in the gauge length, thus increasing the strain detection capabilities.

After being thermo-mechanically processed to produce the required metastable, austenitic microstructure, TRIP steels inherently possess the capacity for damage

detection and monitoring by producing the ferromagnetic martensitic product phase during deformation. The magnetic susceptibility changes by a factor of approximately five orders of magnitude between purely paramagnetic austenite and a 100% ferromagnetic martensite [42]. Thus, the detection of even small amounts of martensite (of the order of less than one volume percent) is relatively simple using off-the-shelf detectors based either on the Hall effect or on inductance readings.

Although altering the alloy chemistry can be used to change and control the transformation behaviour, it is also possible to manipulate the transformation by altering the thermo-mechanical processing history of the material. In Figure 8.8 a PDA (prior deformation of austenite) steel was warm rolled at 450° C [42] to introduce a 0, 20, 40, 60 or 80% (indicated as /0, /1, /2, /3, /4 respectively) PDA. The austenite does not transform at 450° C, but the dislocation density does increase with increasing strain, hence an increase in yield strength accompanies increased warm rolling. Tests A15/1 and A15/3 were not tested to failure whereas the remaining specimens were. The transformation curves become steeper indicating higher transformation rates with increasing prior warm rolling. This occurs because the prior deformation of austenite increases the martensite nucleation site density, thereby increasing the rate of martensite formation as a function of applied strain.

Engineering component materials can therefore be chosen based on the strength required or the specific transformation behaviour that would yield the optimum monitoring data for the intended application consistent with the strain sensitivity and range desired. A lower amount of initial austenite deformation would suit applications where relatively low strengths and high service strains are tolerable in the component design. Higher strengths are achieved with a corresponding decrease in the total strain range over which the transformation occurs (i.e. the range over which strains can be monitored once the initiation of structural instability has occurred). Although the warm rolling does increase the material fabrication costs, the additional cost may be offset by the amount of material required to support the design stresses considering the increased strength that is achieved.

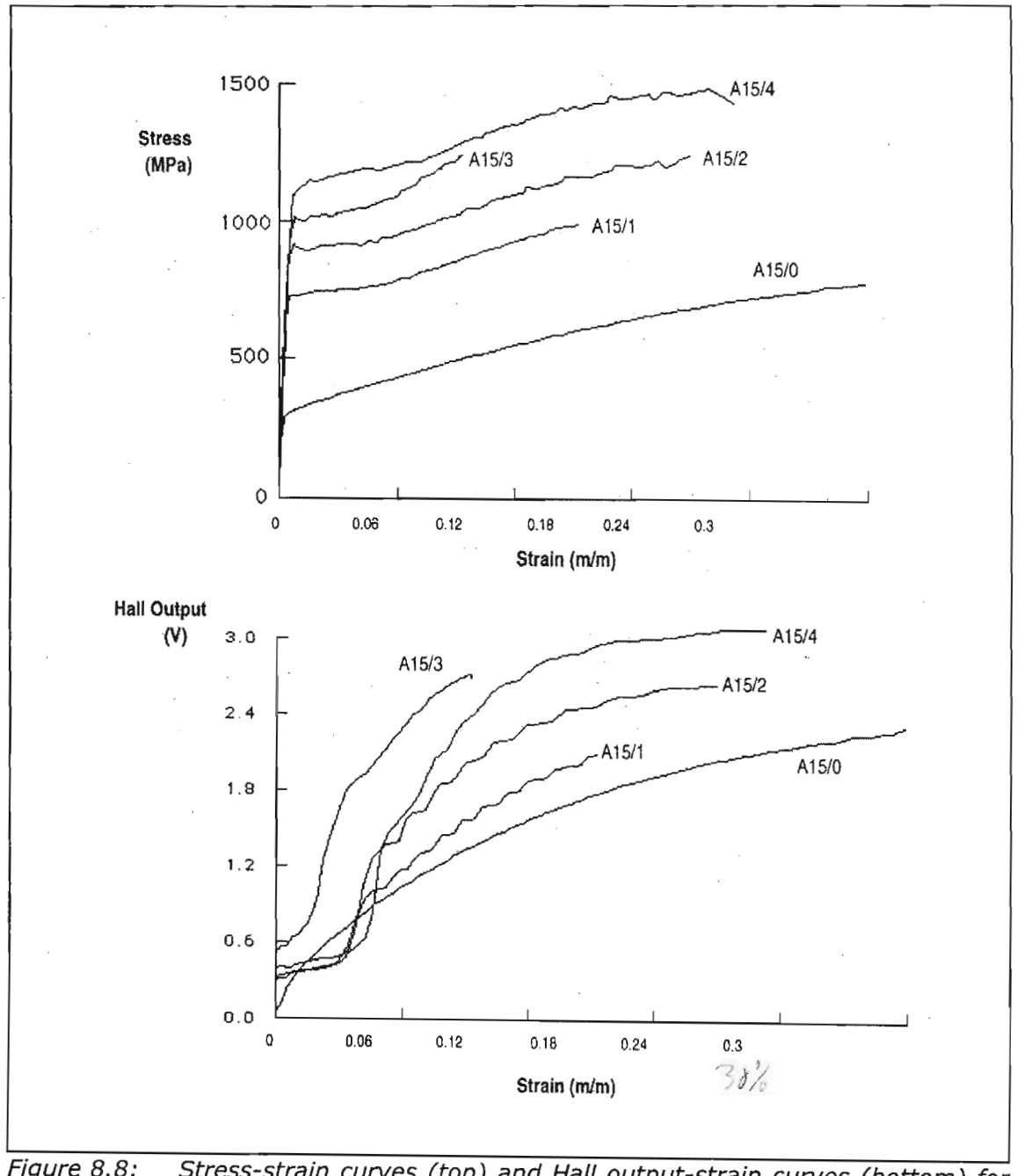


Figure 8.8: Stress-strain curves (top) and Hall output-strain curves (bottom) for Fe-8.4 Cr-8.4 Ni-2.1 Mn-0.262 C TRIP sensor material

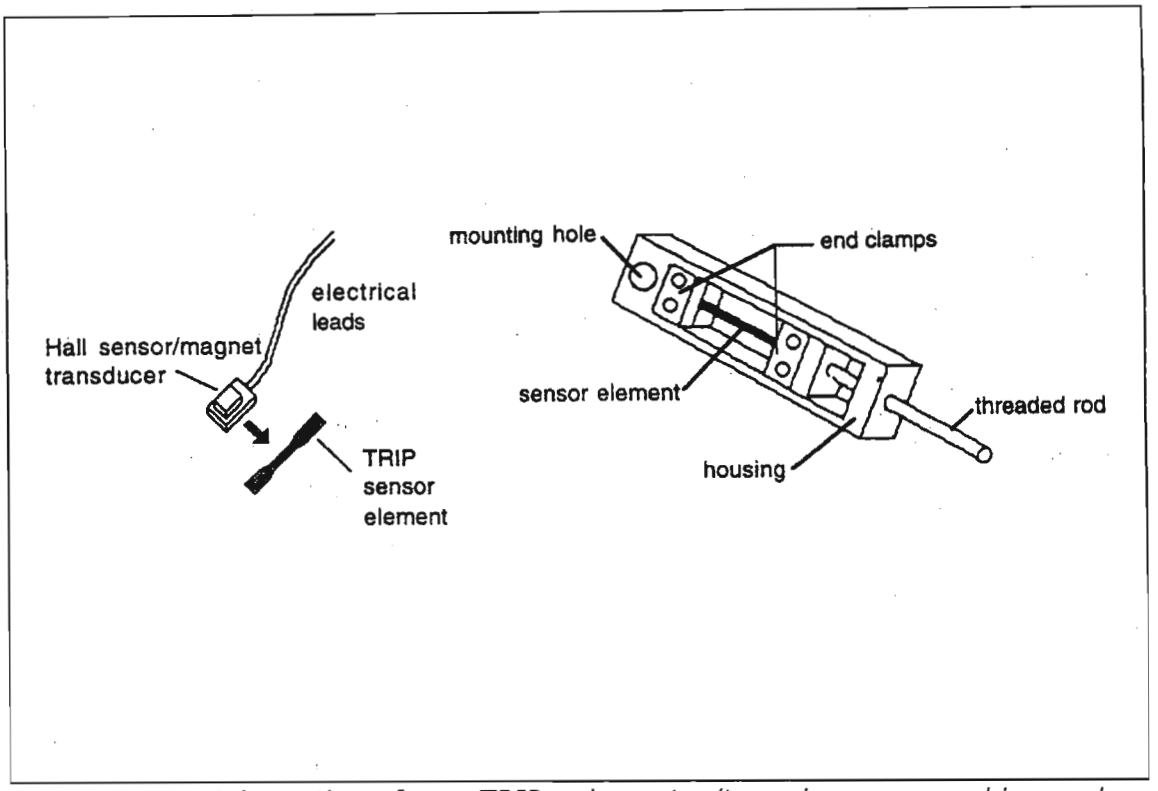


Figure 8.9: Schematic of a TRIP element /transducer assembly and a structurally-attachable tensile strain monitor

The transformation rates are highest in the early stages of martensite formation thereby making it possible to pre-strain the material to a level where any further deformation will cause a rapid increase in the rate of martensite formation and a corresponding marked increase in ferromagnetic response. The incubation strain is excessively large for the warm rolled materials, therefore in order to optimise the alloys for structural applications either a pre-straining (cold-working treatment) would be required to eliminate the incubation strain or the alloy chemistry could be adjusted to destabilise the austenitic phase with respect to applied strain. Pre-straining thus provides an additional variable in thermo-mechanical processing that can he utilised to decrease any incubation strain effect that is unwanted in a specific application.

TRIP steels that have been thermo-mechanically processed to achieve high strengths are not weldable, since welding would alter the microstructure and hence result in weak material in the heat-affected regions of the weld. Design considerations would have to anticipate this constraint, perhaps by employing bolted connections as opposed to welded connections. Some designs can easily be modified or reconfigured to accommodate this constraint. The lower strength TRIP steels, particularly the Type II materials, should be weldable based on their compositions and the fact that warm rolling is not employed. The Type II materials

can also be prepared as fully stainless materials with corrosion resistance levels typical of or better than the AISI 300 series austenitic stainless steels.

8.5. FACTORS INFLUENCING THE FERROMAGNETIC RESPONSE

Three influencing factors need to be addressed when selecting metastable alloy sensors for structural deformation sensing and monitoring, namely the inhomogeneity of the phase transformation, the temperature dependence of the transformation process, and the effect of strain rate on the transformation behaviour.

8.5.1. Inhomogeneous Deformation During Initial Phase Transformation

As previously mentioned, metastable alloys (including TRIP steels) deform inhomogeneously during the initial stages of plastic deformation. The ferromagnetic measurement technique determines how extensively any inhomogeneous deformation affects the measured value.

A *ferromagnetic sensor* (e.g. a Hall sensor) produces a localised measurement at the location where the sensor is positioned. When taking measurements from TRIP steels, the Hall sensor will register stepped measurements for small strains of generally below 2%.

An *induction coil* measurement arrangement produces a reading that reflects the overall behaviour of the deforming gauge length (i.e. an "averaged" response). This may assist in overcoming the initial inhomogeneous deformation limitation.

8.5.2. Temperature Dependence of the Transformation Process

The level of martensite that forms increases as the deformation temperature is decreased. Figure 8.10 shows the ferromagnetic response as a function of temperature for a carbon-free metastable sensor alloy version of AISI 301 austenitic stainless steel (known to transform similarly to TRIP steels).

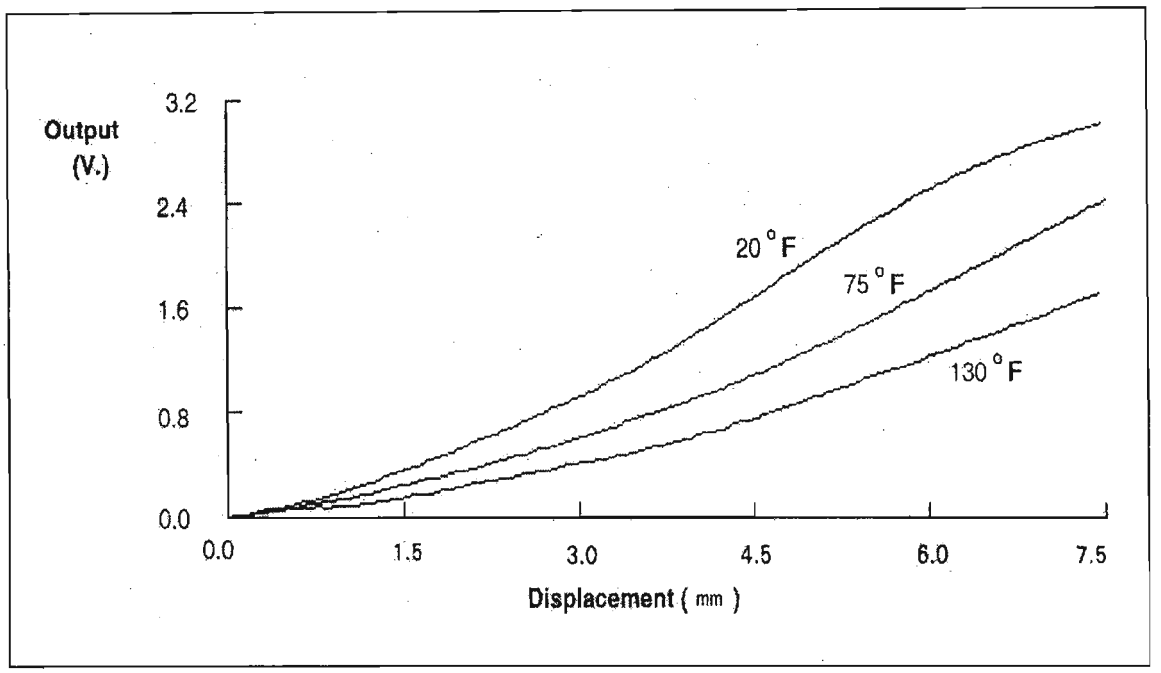


Figure 8.10: Hall unit output as a function of displacement for a candidate sensor material.

The temperature dependence of the response affects the overall performance of a passive monitoring system by increasing the error bands associated with a particular reading from a particular gauge. The error introduced by the temperature dependence increases as the amount of martensite formed increases. This is illustrated in Figure 8.11.

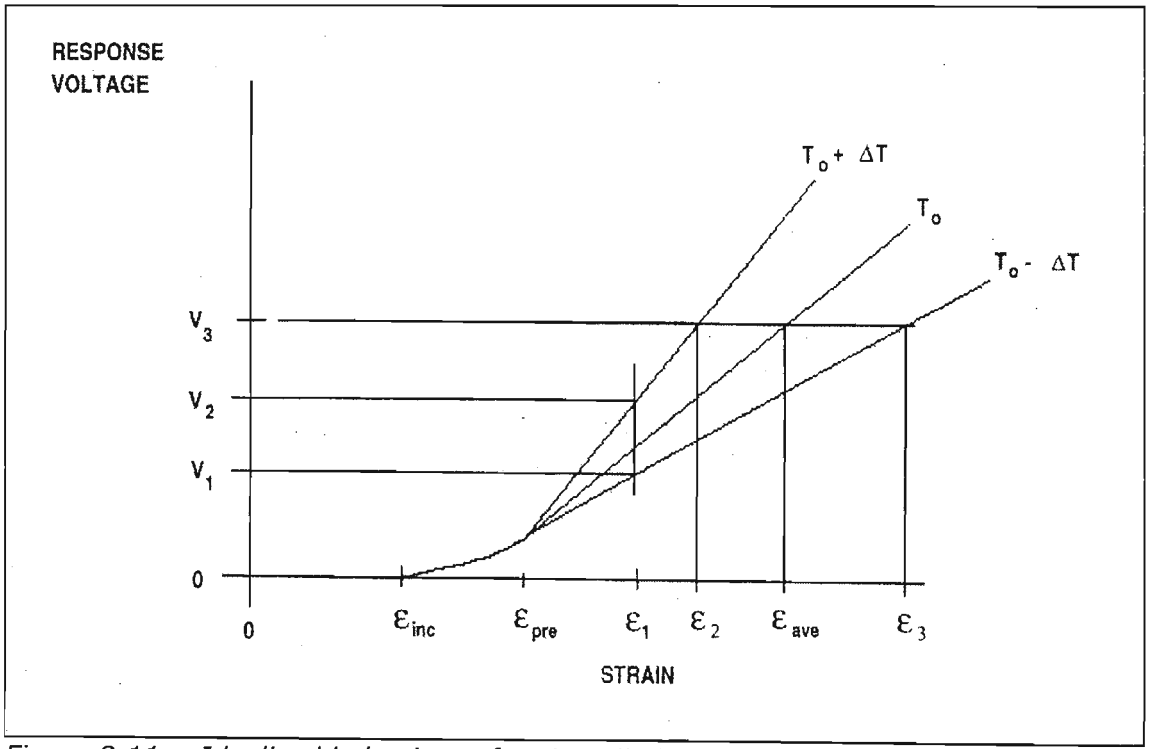


Figure 8.11: Idealised behaviour of an installed sensor element.

When the sensor is installed the element is pre-strained to produce the ε_{pre} shown on the curve (where $\varepsilon_{pre} > \varepsilon_{inc}$). A set point reading is then taken that is the reference reading for that sensor element. Any deformation (i.e. further plastic straining of the element) will increase the amount of martensite beyond the initially measured level. The amount of additional martensite depends on the deformation temperature at which it was formed. If the additional strain produces a total strain of ε_1 , then the corresponding ferromagnetic response reading, presented here in terms of measured voltage, would lie between V_1 and V_2. If we consider instead the sensor point of view when taking a measurement, a particular voltage - say V_3 - would correspond to an additional straining from the pre-strained ε_{pre} of between (ε_2 - ε_{pre}) and (ε_3 - ε_{pre}), depending on the temperature/s at which the additional straining occurred. One could estimate the actual strain by assuming the deformation occurred at the average temperature during the time between the last inspection and the current inspection. Here the average temperature is assumed to be the same as the sensor installation temperature, T_0. The error bars on the strain measurement would then extend between ε_2 and ε_3. For illustration purposes, V_3 corresponds to a considerable proportion of martensite formed. If instead the measured voltage is nearer the V_0 set point voltage, say at V_1 , then the error bars are far smaller. Should any strain changes be observed, it would be prudent to take readings far more frequently. Average temperature data would then reflect values fairly close to the actual data assuming that deformation occurs statistically above and below the average temperature.

In many instances, e.g. if the system was installed to make a post-earthquake damage assessment, the temperature at peak strain is readily available. There are also many applications where the system operates under fairly constant temperature conditions (e.g. the temperatures in a mine remain constant within about ± 2 °C). In certain applications it may be possible to change the operating mode from passive to semi-active, for example a thermistor and data acquisition chip may be installed in conjunction with a particular sensor or sensor array so that changes are stored as a function of time along with the temperature when the deformation occurred.

The ideal solution to overcoming temperature variation effects is to identify alloy compositions with little or no temperature dependence. This is a difficult problem considering that thermodynamics control the transformation rate. The problem is equivalent to finding alloys where there is little change in the martensite Gibbs free energy over the range of temperatures appropriate for the monitoring application. The Fe-Mn-Cr alloy systems are attractive in this case since the intermediate

phase, ε (epsilon) martensite, forms prior to the formation of the stable, ferromagnetic, α′ (alpha prime) martensite. The ε martensite forms as intersecting bands parallel to the {1 1 1} planes in the austenite. Considering that α′ martensite nucleates at ε martensite band intersections, it is possible to control the number of effective α′ martensite nucleation sites by controlling the ε band intersection density within the austenite. The amount of α′ martensite that can form is then constant regardless of temperature assuming that the lowest temperature of service is above the temperature where the precursor ε bands are formed.

Another approach to reducing the temperature sensitivity of the transformation process involves producing a dislocation substructure by deforming the strain memory alloy at a temperature below the lowest service temperature. The substructure will be consistent with the low temperature where the deformation was applied and assuming that the initial deformation strain is small the alloy will contain a density of α′ martensite nucleation sites that should be greater than would be produced at a higher temperature. Considering that the range of expected service temperatures would be higher than the pre-deformation temperature, the amount of α′ martensite that forms during service should be relatively constant reflecting the initial distribution of nucleation sites formed prior to installation of the sensor. In effect, the sensor material will behave as though it was being loaded at the pre-deformation temperature as long as the actual temperature is higher than the pre-deformation temperature.

8.5.3. The Effect of Strain Rate on Transformation Behaviour

The third and final concern/issue involves the effect of the strain rate on the phase transformation behaviour. Sensor alloys that are based on carbon-free chemical formulations do not display strain rate effects that are large enough to significantly affect the readings. For the range of strain rates considered typical for the accumulation of damage in conventional structures, carbon-free alloys display virtually no strain rate effects, as shown by the data in Figure 8.12.

Alloys that contain carbon as a principal alloying element (including the high strength TRIP steel formulations) are significantly affected by the strain rate, with more α′ martensite forming at lower strain rates.

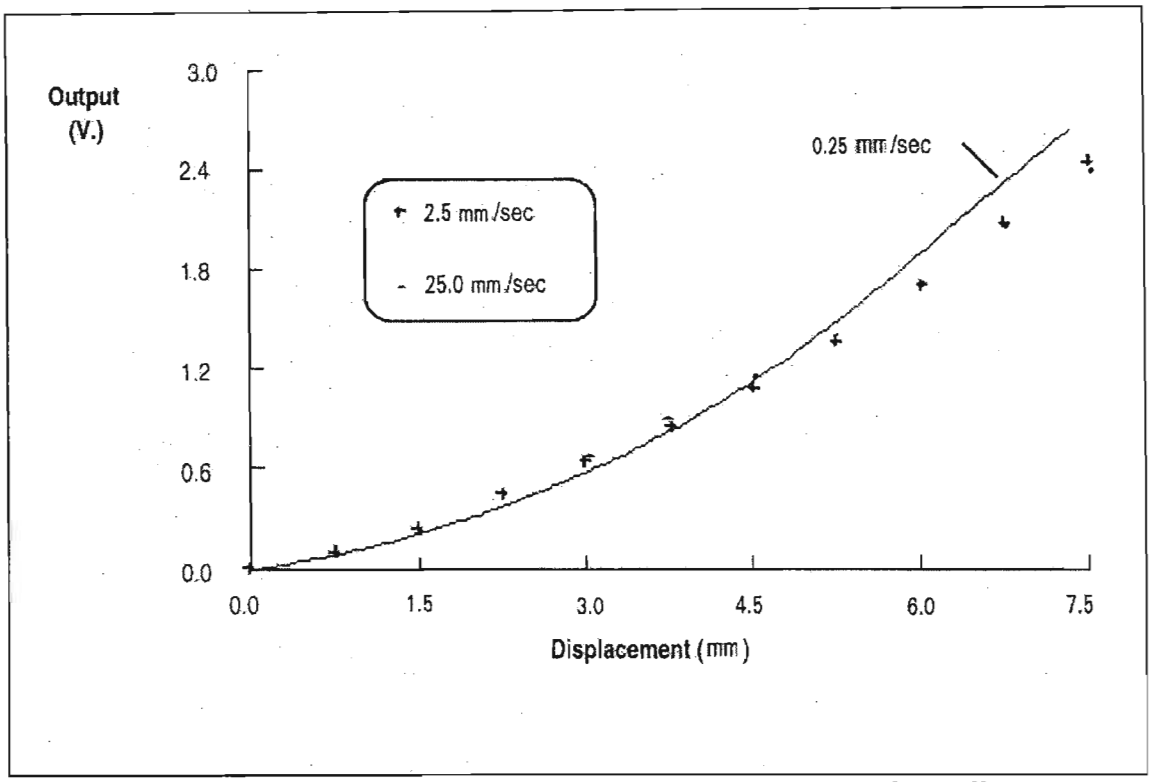


Figure 8.12: Hall Unit Output as a Function of Displacement for Different Strain Rates in Carbon-Free Alloys.

8.5.4. Error Related to TRIP SHM Systems

Variations in ferromagnetic response due to the instrumentation and variances that might exist in the sensor elements themselves are lumped together to produce what is termed the error related to the sensor system. The utilisation of inductance measurements rather than Hall sensor measurements will produce less variability in the output. Likewise, careful fabrication of sensor elements so as to produce consistent sensors will reduce the variability within the sensor elements themselves, thus quality control of the sensor element manufacturing becomes of extreme importance in reducing variability.

8.6. ECONOMIC CONSIDERATIONS

As well as the obvious safety benefits, consideration must also be given to the financial implications of incorporating TRIP steels selectively in structural applications [47]. Usage includes but is not limited to new engineering components, and to the repair or replacement of existing components to extend the remaining service life of existing structural systems. Economic impact assessments should include:

1. Material costs. These are high due to the exotic chemical compositions.

2. Fabrication costs. Additional expense is required to fabricate the materials to various useful structural forms since they absorb a relatively huge amount of energy during deformation. Also, the warm rolling process is an additional step to achieve the higher strength levels and must be performed at 450°C.

3. Tooling costs. TRIP steels are associated with increased machinery workloads, resulting in an anticipated corresponding increase in equipment costs. This includes repair or replacement of rolling mills, forging assemblies, stamping units, etc.

4. Reduced material requirement. Some expenses may be recovered solely because of the significantly higher yield strengths over structural grade steels, so that smaller units will suffice.

5. Enhanced structural integrity leads to extended lifespan. This relates to improved fracture resistance and higher strength, as well as to the ability to accurately determine the actual amount of accumulated damage and corresponding residual life remaining in a component. The current physical capabilities can also be assessed. Damaged components may therefore remain in-service longer than their traditional counterparts.

6. Design factors of safety might also be reduced, since any structural deformation is readily pinpointed. Costs associated with over-design would thereby be reduced.

7. Not all structural members would require replacement – only critical or (in the case of mining rockbolts) a percentage of a large array of units.

8. TRIP steels are passive sensors and thus less expensive than competitive active systems with their integrated electronic equipment requirements. Peak measurements are also retained in the event of power loss, so that any SHM benefits are not lost.

8.7. CONCLUSIONS

The TRIP steels are strain-sensing in that they are capable of providing an indication of the peak strain experienced by the material. This indication is manifested in an irreversible austenitic-to-martensitic phase transformation that is initiated by strain. The extent to which the transformation has taken place serves as the indicator of how much strain the material has been subjected to.

Considering that the martensite phase is magnetic whilst the austenitic phase is not, a useful and convenient method for the measurement of the extent to which the steel has transformed, and hence the peak strain, involves the measurement of the magnetic properties of the steel.

The advantages of TRIP steels lie in the fact that they can be used concurrently as both high strength structural and sensor elements. Provision does not therefore have to be made for the attachment of sensors to structural components that can feasibly be manufactured from TRIP steels. Measurement equipment does still need to be connected. The level of magnetism of the TRIP steels is also easily measured using standard components such as Hall sensors and inductance coils, thereby facilitating compact and relatively simple instrumentation. Disadvantages of TRIP steels involve their lack of commercial availability and the difficulties associated with the thermo-mechanical treatments required to produce the high strength varieties. Notwithstanding these issues, TRIP steels have been extensively investigated in the literature and have demonstrated their potential in several successful structural health monitoring systems.

9. FEA OF STRUCTURAL DEFORMATION AND MAGNETOSTATIC FIELDS

9.1. BROAD OVERVIEW OF THE FEA PROCESS

Finite element analysis (FEA) involves finding an approximate piecewise continuous solution for field variables in continuum problems (where space and time are the independent variables) that do not lend themselves to analytical solution techniques, usually because of complex geometry.

In the Eulerian approach to continuum problems, field quantities are defined at every point in space. Common field quantities investigated include displacement, stress, temperature, mass concentration, electromagnetic potentials, acoustic potentials, etc. The behaviour of the field quantities can be approximately represented by partial differential equations and their boundary conditions.

A region Ω where a field variable ϕ exists is divided into a number of elements each having a prescribed shape and a prescribed number of nodes. The nodal values of the field variable (e.g. displacement) are the unknowns. The degree of approximation depends on the size and number of elements as well as on the interpolation functions selected to describe the variation of the field variable through each element. The interpolation function selected should satisfy the *continuity* requirement, whereby the field variable and its derivatives (e.g. displacement and strain) are continuous across adjacent element boundaries. Polynomials are preferred since they are easily integrated and differentiated (e.g. Lagrange's interpolation formula). Expressions based on the approximated variation of the field variable are then formulated for individual elements using prescribed element material properties that may be anisotropic. These element expressions are then assembled to represent the entire problem, and solved using e.g. variational techniques described below.

9.2. BENEFITS OF FEA

Advantages of using FEA are that irregular geometries can be readily analysed, with refined element densities improving accuracy in areas of steep e.g. stress gradient. Problems involving non-homogeneous and anisotropic materials undergoing any type of static or dynamic external load are readily handled, and flexible boundary conditions can be defined. Nonlinear material and nonlinear geometric problems can also be solved.

9.3. VARIATIONAL TECHNIQUES FOR SOLVING FUNCTIONALS

9.3.1. Analytical Formulations

Functionals (e.g. the total potential energy Π of a system, discussed later) are functions of functions and their derivatives. Whereas a function produces a number as a result of giving values to one or more independent variables, a functional depends on the entire form of one or more functions between prescribed limits.

Functionals have the form [50], [51]:

$$L(\phi) = \int_a^b F\left(x, \phi(x), \frac{d\phi}{dx}, .., \frac{d^m\phi}{dx^m}\right)dx \qquad (9\text{-}1)$$

Consider the following functional:

$$I(\phi) = \int_a^b f\left(x, \phi(x), \frac{d\phi}{dx}\right)dx \qquad (9\text{-}2)$$

where the limits $\phi(a) = A$ and $\phi(b) = B$. Variational techniques are a method of finding a function $\phi(x) = \tilde{\phi}(x)$ for which $I(\phi) = I(\tilde{\phi})$ is stationary (is an extremum).

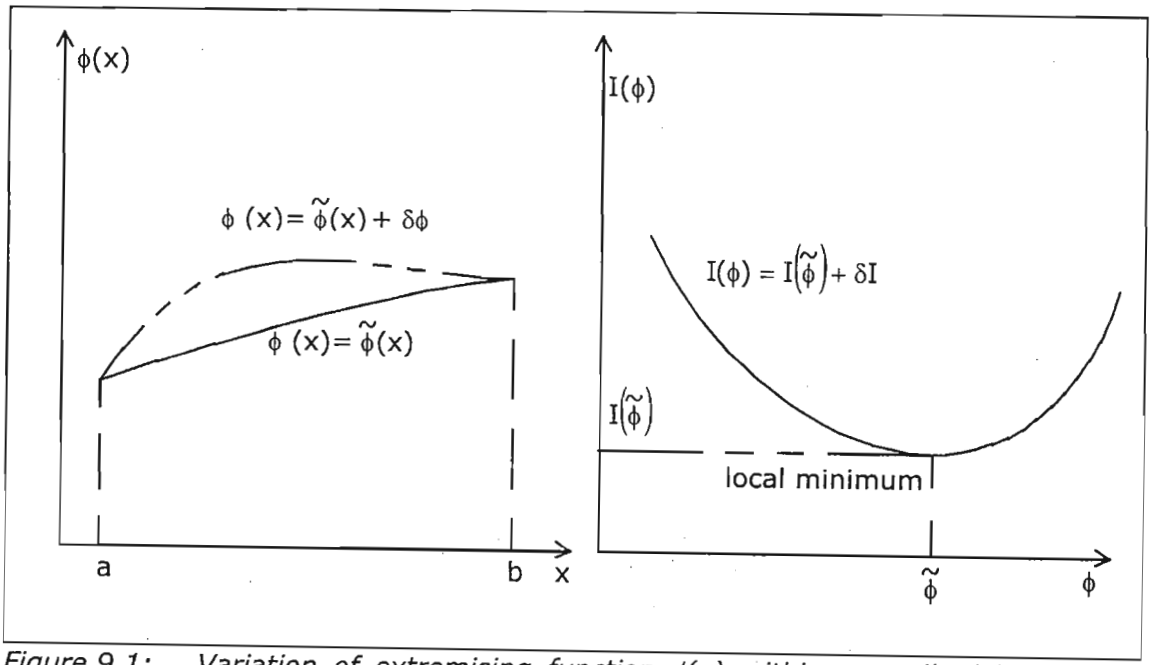


Figure 9.1: Variation of extremising function $\phi(x)$ within prescribed limits, and the effect on $I(\phi)$.

The operator δ is referred to as the *variational symbol*. The variation $\delta\phi$ of a function $\phi(x)$ is an infinitesimal change in ϕ for a fixed value of the independent variable x (i.e. $\delta x = 0$). $\delta\phi$ vanishes at those points where ϕ is prescribed and is

arbitrary elsewhere. A variation $\delta\phi$ results in a corresponding variation δI. See Figure 9.1. It is important to distinguish between (i) the differential of e.g. $\phi(x)$ (i.e. $d\phi$), which is the change of $\phi(x)$ corresponding to a differential change dx along the curve defined by $\phi(x)$, and (ii) the variation of $\phi(x)$ (i.e. δx), which is the change in this function when moving from one curve to another at a constant x.

The solution to equation (9-1) requires that:

$$\delta I = 0 \tag{9-3}$$

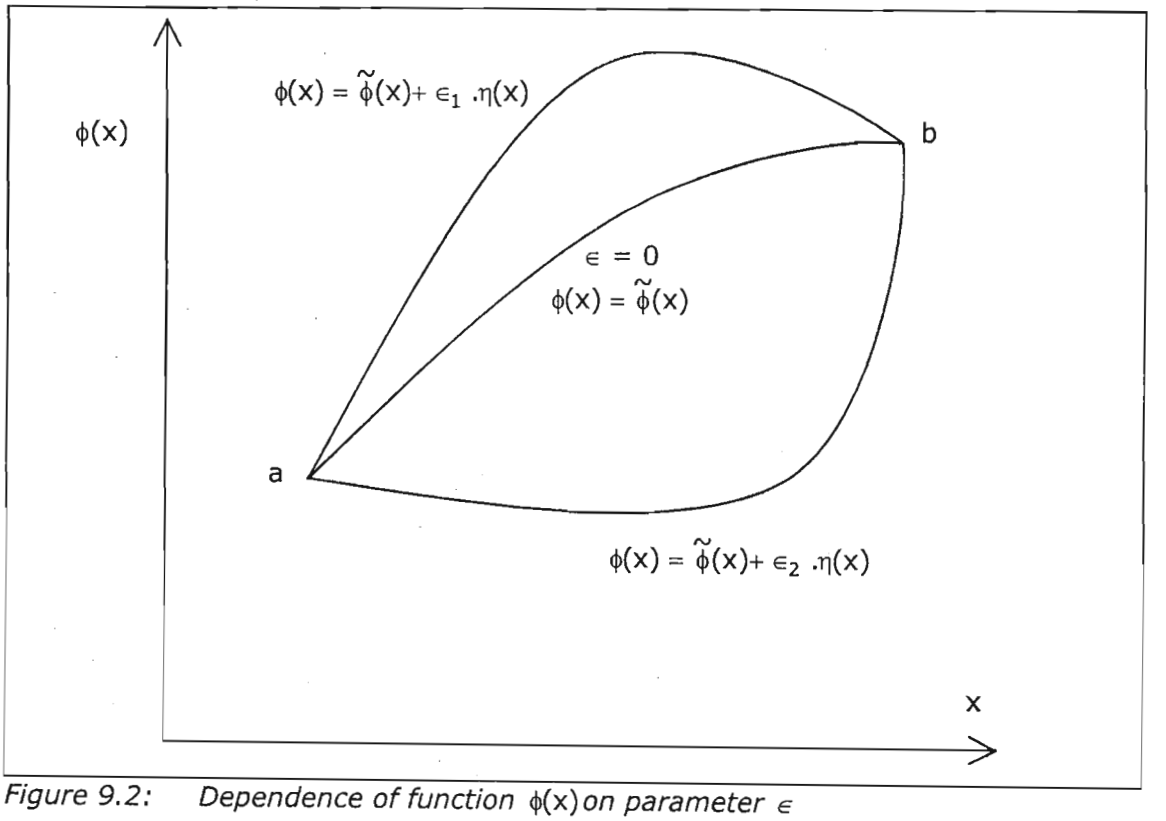


Figure 9.2: Dependence of function $\phi(x)$ on parameter ϵ

Now the variation of $\tilde{\phi}(x)$ is denoted by $\delta\phi$ and is equal to the difference between 2 curves $\tilde{\phi}(x)$ and $\phi(x)$ where, for example:

$$\phi(x) = \tilde{\phi}(x) + \epsilon .\eta(x) \tag{9-4}$$

and:

$$\begin{aligned} \phi'(x) &= \frac{d\phi}{dx} \\ &= \frac{d\tilde{\phi}}{dx} + \epsilon \frac{d\eta}{dx} \\ &= \tilde{\phi}' + \epsilon .\eta'(x) \end{aligned} \tag{9-5}$$

$\eta(x)$ may be any continuously differentiable function of x that, for defined end points on $\phi(x)$, is equal to zero at the limits of the integral, i.e.:

$$\eta(a) = \eta(b) = 0 \tag{9-6}$$

The "constant" ϵ is a parameter that controls the magnitude of the difference $\delta\phi$ between the functions $\tilde{\phi}(x)$ and $\phi(x)$, is independent of x, and can be varied in a continuous manner. The relationship between $\tilde{\phi}(x)$ and $\phi(x)$ can be visualised as sketched in Figure 9.2. Thus an infinite number of varied paths $\phi(x)$ can be generated for a given function $\eta(x)$ by adjusting the parameter ϵ. All these paths pass through points a and b.

From equation (9-4), at any constant x between a and b:

$$\delta\phi = \epsilon . \eta(x) \tag{9-7}$$

Recall that $\tilde{\phi}(x)$ is the function that minimises the functional I. By definition of $\tilde{\phi}(x)$, functional I attains an extreme value when $\phi(x) = \tilde{\phi}(x)$, i.e. when $\epsilon = 0$. Hence, with the agreement to denote $\tilde{\phi}(x)$ as the minimising function, then $\tilde{I}$ becomes the extreme value of the integral of equation (9-2):

$$I = \int_a^b f\left(x, \phi(x), \frac{d\phi}{dx}\right) dx$$

For the desired extremal $\tilde{\phi}(x)$, $\tilde{I}$ is now a function of the parameter ϵ. See Figure 9.3. I is extremised ($I = \tilde{I}$)when $\epsilon = 0$, thus:

$$\left(\frac{dI}{d\epsilon}\right)_{\epsilon=0} = 0 \tag{9-8}$$

From equation (9-2):

$$\frac{dI}{d\epsilon} = \frac{d}{d\epsilon}\left(\int_a^b f(x, \phi(x), \phi'(x)) dx\right) = \int_a^b \left(\frac{df(x, \phi(x), \phi'(x))}{d\epsilon}\right) dx \tag{9-9}$$

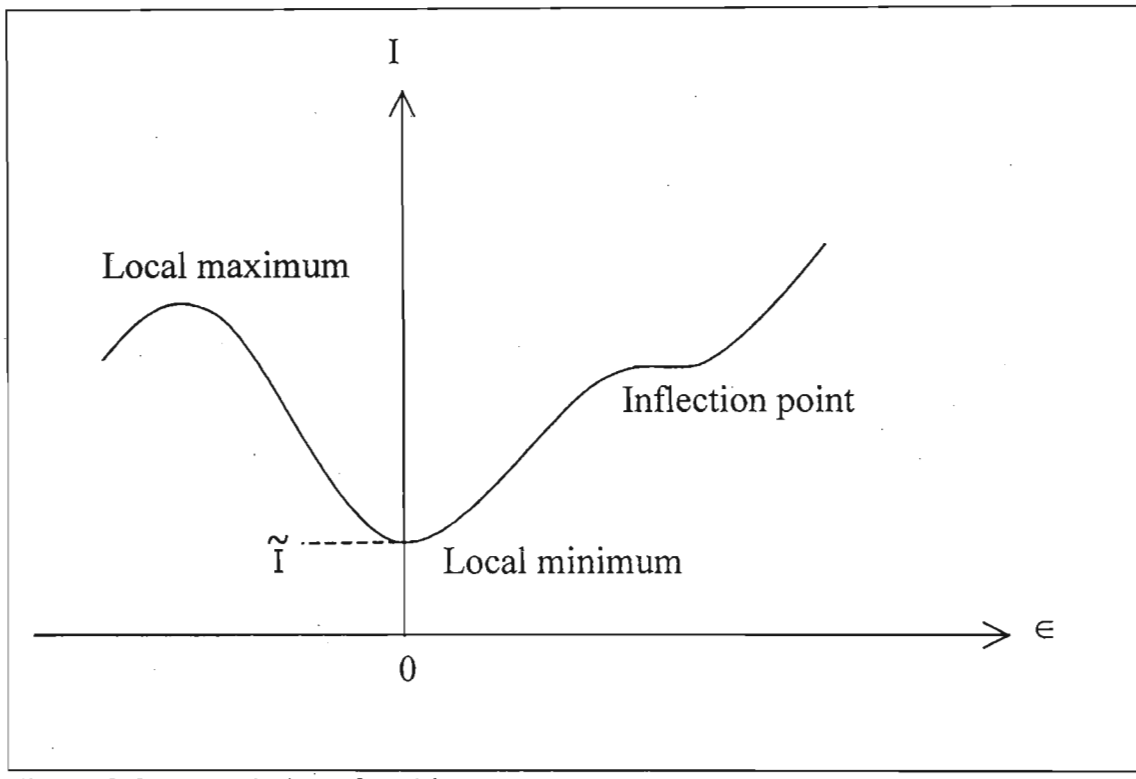


Figure 9.3: Variation of I with ϵ

Now:

$$\frac{df}{dx} = \frac{\partial f}{\partial \phi}\frac{\partial \phi}{\partial x} + \frac{\partial f}{\partial \phi'}\frac{\partial \phi'}{\partial x} \tag{9-10}$$

$$\begin{aligned}\frac{df}{d\epsilon} &= \frac{df}{dx}\frac{dx}{d\epsilon} \\ &= \frac{\partial f}{\partial \phi}\frac{\partial \phi}{\partial x}\frac{\partial x}{\partial \epsilon} + \frac{\partial f}{\partial \phi'}\frac{\partial \phi'}{\partial x}\frac{\partial x}{\partial \epsilon} \\ &= \frac{\partial f}{\partial \phi}\frac{\partial \phi}{\partial \epsilon} + \frac{\partial f}{\partial \phi'}\frac{\partial \phi'}{\partial \epsilon}\end{aligned} \tag{9-11}$$

From equations (9-4), (9-5) and (9-11):

$$\frac{df}{d\epsilon} = \frac{\partial f}{\partial \phi}\eta + \frac{\partial f}{\partial \phi'}\eta' \tag{9-12}$$

At $\epsilon = 0$, $I = \tilde{I}$, therefore:

$$\left(\frac{\partial f}{\partial \phi}\right)_{\epsilon=0} = \frac{\partial f}{\partial \tilde{\phi}} \tag{9-13}$$

$$\left(\frac{\partial f}{\partial \phi'}\right)_{\epsilon=0} = \frac{\partial f}{\partial \tilde{\phi}'} \tag{9-14}$$

so that:

$$\left(\frac{df}{d\epsilon}\right)_{\epsilon=0} = \frac{\partial f}{\partial \tilde{\phi}}\eta + \frac{\partial f}{\partial \tilde{\phi}'}\eta' \tag{9-15}$$

Substituting (9-15) into (9-9):

$$\left(\frac{dI}{d\epsilon}\right)_{\epsilon=0} = 0 = \int_a^b \left(\frac{\partial f}{\partial \tilde{\phi}}\eta + \frac{\partial f}{\partial \tilde{\phi}'}\eta'\right)dx \tag{9-16}$$

Expanding equation (9-16), integrating the right hand side by parts, and using equation (9-5):

$$\begin{aligned}
\int_a^b \left(\frac{\partial f}{\partial \tilde{\phi}}\eta + \frac{\partial f}{\partial \tilde{\phi}'}\eta'\right)dx &= \int_a^b \left(\frac{\partial f}{\partial \tilde{\phi}}\eta\right)dx + \left[\int_a^b \left(\frac{\partial f}{\partial \tilde{\phi}'}\eta'\right)dx\right] \\
&= \int_a^b \frac{\partial f}{\partial \tilde{\phi}}\eta dx + \left[\frac{\partial f}{\partial \tilde{\phi}'}\eta\Big|_a^b - \int_a^b \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\eta dx\right] \\
&= \int_a^b \frac{\partial f}{\partial \tilde{\phi}}\eta dx + 0 - \int_a^b \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\eta dx \\
&= \int_a^b \left[\frac{\partial f}{\partial \tilde{\phi}}\eta - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\eta\right]dx \\
&= \int_a^b \left[\frac{\partial f}{\partial \tilde{\phi}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\right]\eta dx \\
&= 0
\end{aligned}$$

Therefore:

$$\int_a^b \left(\frac{\partial f}{\partial \tilde{\phi}}\eta + \frac{\partial f}{\partial \tilde{\phi}'}\eta'\right)dx = \int_a^b \left[\frac{\partial f}{\partial \tilde{\phi}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right)\right]\eta dx = 0 \tag{9-17}$$

Since η is arbitrary, it follows that:

$$\frac{\partial f}{\partial \tilde{\phi}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right) = 0 \tag{9-18}$$

or:

$$\frac{d}{dx}\left(\frac{\partial f}{\partial \phi'}\right) - \frac{\partial f}{\partial \tilde{\phi}} = 0 \tag{9-19}$$

This is the Euler or Euler-Lagrange equation. Therefore, for a functional $I(\phi) = \int_a^b f\left(x, \phi(x), \frac{d\phi}{dx}\right)dx$ to have an extremum for a given function $\phi(x)$, a necessary condition is that $\phi(x)$ satisfies the Euler equation.

This result is easily extended to the generalised functional presented in equation (9-1), leading to Euler's equations:

$$\frac{d}{dx}\left(\frac{\partial f}{\partial \phi_\alpha'}\right) - \frac{\partial f}{\partial \phi_\alpha'} = 0 \text{ for } \alpha = 1,..,n \tag{9-20}$$

9.3.2. *Variational Techniques for Solving Numerical Equations*

When applying variational calculus to finite element analysis, the functional $I(\phi)$ through domain Ω is approximated as the sum of the individual element functionals $I^e(\phi^e)$ defined for m elements within Ω, i.e. [52]:

$$I(\phi) \approx \sum_{e=1}^{m} I^e\left(\phi^e\right) \tag{9-21}$$

The function $\phi(x)$ that makes the functional $I(\phi)$ stationary is expressed in terms of the nodal values of the functional's derivatives. That is, for a given domain with n discrete values of the field variable ϕ:

$$\frac{\partial I(\phi)}{\partial \phi} \approx \sum_{i=1}^{n} \frac{\partial I}{\partial \phi_i} = 0 \tag{9-22}$$

Because the terms ϕ_i are independent, equation (9-22) can be satisfied only if the system of equations:

$$\frac{\partial I}{\partial \phi_i} = 0 \quad (i = 1,..,n) \tag{9-23}$$

In FEA, a requirement is to find and solve the system of displacements $\{\Delta\}$ within a region Ω that corresponds to the *minimum* total potential energy Π within that region. This results in equation (9-23) leading to n equations having the following form:

$$\frac{\partial \Pi}{\partial \Delta_i} = K.\Delta_i - P = 0 \tag{9-24}$$

where K and P are known. This is known as the Rayleigh-Ritz method. The resulting solution for Δ_i and thus $\{\Delta\}$ is, by definition, an approximation.

Equation (9-24) is specific to the functional having form described in (9-2). Consider the variational solution to the general functional of equation (9-1):

$$L(\phi) = \int_a^b F\left(\phi, x, \frac{d\phi}{dx}, \ldots, \frac{d^m\phi}{dx^m}\right)dx$$

In FEA, if the highest derivative of the field variable ϕ that appears in F is *m*, then the field variable ϕ^e calculated for each element and its derivatives to order *m* should be able to assume a constant value in the element as the element size decreases to zero. It is the interpolation function selected that determines the value of *m*. This *completeness requirement* is necessary for convergence to the exact solution. For continuity (*continuity requirement* mentioned above), the field variable ϕ^e and its derivatives up to order *m*-1 should be continuous at the element boundaries. This continuity condition is not vital for convergence, provided the errors introduced by the approximation decrease fast enough as the element size decreases. If an element satisfies both the completeness and continuity requirements, it is said to be *conforming* or *compatible*, and convergence is guaranteed. Otherwise it is *non-conforming* or *incompatible*.

9.4. SOLVING FOR PHYSICAL LOAD CONDITIONS

The total potential energy of a region of material subjected to physical distortion comprises 2 components, namely the external force potential energy and the strain energy. For the purposes of this thesis, axisymmetric elements will be considered in the cylindrical (r,θ,z) coordinate system.

9.4.1. Force Energy Potential

A conservative force F_c has a magnitude independent of the displacement of the force. The work W_c performed by such a force in moving a body along path $\bar{R}$ is calculated as follows [52]:

$$W_c = \int_{R_1}^{R_2} F_c \cdot d\bar{R} = \phi_{c2} - \phi_{c1} \qquad (9\text{-}25)$$

Function ϕ_c is a work function that exists only when dealing with conservative forces. Because the forces are conservative, the integral depends on the endpoints

of the path but is independent of the path itself. Since ϕ_c is a function of position only, its derivative is an exact differential:

$$d\phi_c = \frac{\partial \phi_c}{\partial r} + \frac{\partial \phi_c}{\partial \theta} + \frac{\partial \phi_c}{\partial z} \tag{9-26}$$

Rewriting equation (9-25):

$$W_c = \int F_c \cdot d\bar{R} = \int_{\phi_{c1}}^{\phi_{c2}} d\phi_c = \int_{r_1}^{r_2} F_{cr} dr + \int_{\theta_1}^{\theta_2} F_{c\theta} d\theta + \int_{z_1}^{z_2} F_{cz} dz \tag{9-27}$$

Combining equations (9-26) and (9-27):

$$F_{cr} = \frac{\partial \phi_c}{\partial r}$$

$$F_{c\theta} = \frac{\partial \phi_c}{\partial \theta} \tag{9-28}$$

$$F_{cz} = \frac{\partial \phi_c}{\partial z}$$

In vector form:

$$\bar{F}_c = \bar{\nabla} \phi_c \tag{9-29}$$

The vector differential operator $\bar{\nabla}$ represents the following:

$$\bar{\nabla} = \hat{i} \frac{\partial}{\partial r} + \hat{j} \frac{\partial}{\partial \theta} + \hat{k} \frac{\partial}{\partial Z} z \tag{9-30}$$

$\hat{i}$, $\hat{j}$ and $\hat{k}$ are unit vectors in the r, θ and z directions respectively.

The work done by a conservative force plus its change in potential for doing work equals zero, i.e.:

$$W_c + (V_2 - V_1) = 0 \tag{9-31}$$

From equation (9-25):

$$(\phi_{c2} - \phi_{c1}) + (V_2 - V_1) = 0 \tag{9-32}$$

Considering differential changes:

$$d\phi_c + dV = 0 \tag{9-33}$$

$$\phi_c + V = C \tag{9-34}$$

This relationship exists at any position along the path $\bar{R}$. Setting constant C to zero (i.e. $\phi_c = 0$ when $W_c=0$ and $V = 0$ when W_c is a maximum):

$$\phi_c = -V \tag{9-35}$$

Combining this with equation (9-28) and expanding:

$$\bar{F}_c = -\bar{\nabla}V \tag{9-36}$$

$$\hat{i}.F_{cr} + \hat{j}.F_{c\theta} + \hat{k}.F_{cz} = -(\hat{i}\frac{\partial}{\partial r} + \hat{j}\frac{\partial}{\partial \theta} + \hat{k}\frac{\partial}{\partial z}) \tag{9-37}$$

Integrating both sides:

$$\int_0^u F_{cr}.dr + \int_0^v F_{c\theta}.d\theta + \int_0^w F_{cz}.dz = -V \tag{9-38}$$

Finally:

$$V = -(F_{cr}.u + F_{c\theta}.v + F_{cz}.w) \tag{9-39}$$

Thus the potential energy expression of any force that has a magnitude independent of the displacement of the force is simply the negative of the scalar dot product of the force and the displacement vector. Unless otherwise stated, all forces referred to throughout the remainder of this text are conservative.

9.4.2. Strain Energy

Longitudinal and shear strain components in the r, θ and z directions are defined according to the following expressions, where u, v and w represent the displacement of a point in the r, θ and z directions respectively:

$$\varepsilon_r = \frac{\partial u}{\partial r}$$

$$\varepsilon_\theta = \frac{\partial v}{\partial \theta}$$

$$\varepsilon_z = \frac{\partial w}{\partial z} \tag{9-40}$$

$$\gamma_{r\theta} = \frac{\partial u}{\partial \theta} + \frac{\partial v}{\partial r}$$

$$\gamma_{\theta z} = \frac{\partial v}{\partial z} + \frac{\partial w}{\partial \theta}$$

$$\gamma_{zr} = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial r}$$

In isotropic materials, elastic longitudinal and shear stresses are related as follows:

$$\sigma_r = E.\varepsilon_r - E.\nu.(\varepsilon_\theta + \varepsilon_z)$$

$$\sigma_\theta = E.\varepsilon_\theta - E.\nu.(\varepsilon_r + \varepsilon_z)$$

$$\sigma_z = E.\varepsilon_z - E.\nu.(\varepsilon_r + \varepsilon_\theta) \tag{9-41}$$

$$\tau_{r\theta} = G.\gamma_{r\theta}$$

$$\tau_{\theta z} = G.\gamma_{\theta z}$$

$$\tau_{zr} = G.\gamma_{zr}$$

These equations can be written in matrix form:

$$\begin{Bmatrix} \sigma_r \\ \sigma_\theta \\ \sigma_z \\ \tau_{r\theta} \\ \tau_{\theta z} \\ \tau_{rz} \end{Bmatrix} = \begin{bmatrix} E & -E.\nu & -E.\nu & 0 & 0 & 0 \\ -E.\nu & E & -E.\nu & 0 & 0 & 0 \\ -E.\nu & -E.\nu & E & 0 & 0 & 0 \\ 0 & 0 & 0 & G & 0 & 0 \\ 0 & 0 & 0 & 0 & G & 0 \\ 0 & 0 & 0 & 0 & 0 & G \end{bmatrix} \begin{Bmatrix} \varepsilon_r \\ \varepsilon_\theta \\ \varepsilon_z \\ \gamma_{r\theta} \\ \gamma_{\theta z} \\ \gamma_{zr} \end{Bmatrix} \tag{9-42}$$

Using symbolic notation, where [C] represents the stress-strain constitutive matrix:

$$\{\sigma\} = [C]\{\varepsilon\} \tag{9-43}$$

Refer to the simple system shown in Figure 9.1. A force F is applied to the end of a bar of length L and cross-sectional area A. The bar deforms linearly whereby the base of the bar remains fixed while the end of the bar is displaced an amount u. Now F varies from 0 to F_f as illustrated in Figure 9.2. The force-time relationship is shown as linear, but this is not a requirement. As the force varies from 0 to F_f, so

the displacement varies from 0 to u_f at the endpoint (Figure 1.3). Note that the force-displacement relationship is linear, i.e.:

$$F = k.u \tag{9-44}$$

where k is the slope of the curve. Since the curve is linear (i.e. force is proportional to displacement for an elastic body), the value of k is easily found [53]:

$$F = \frac{F_f}{u_f} u \tag{9-45}$$

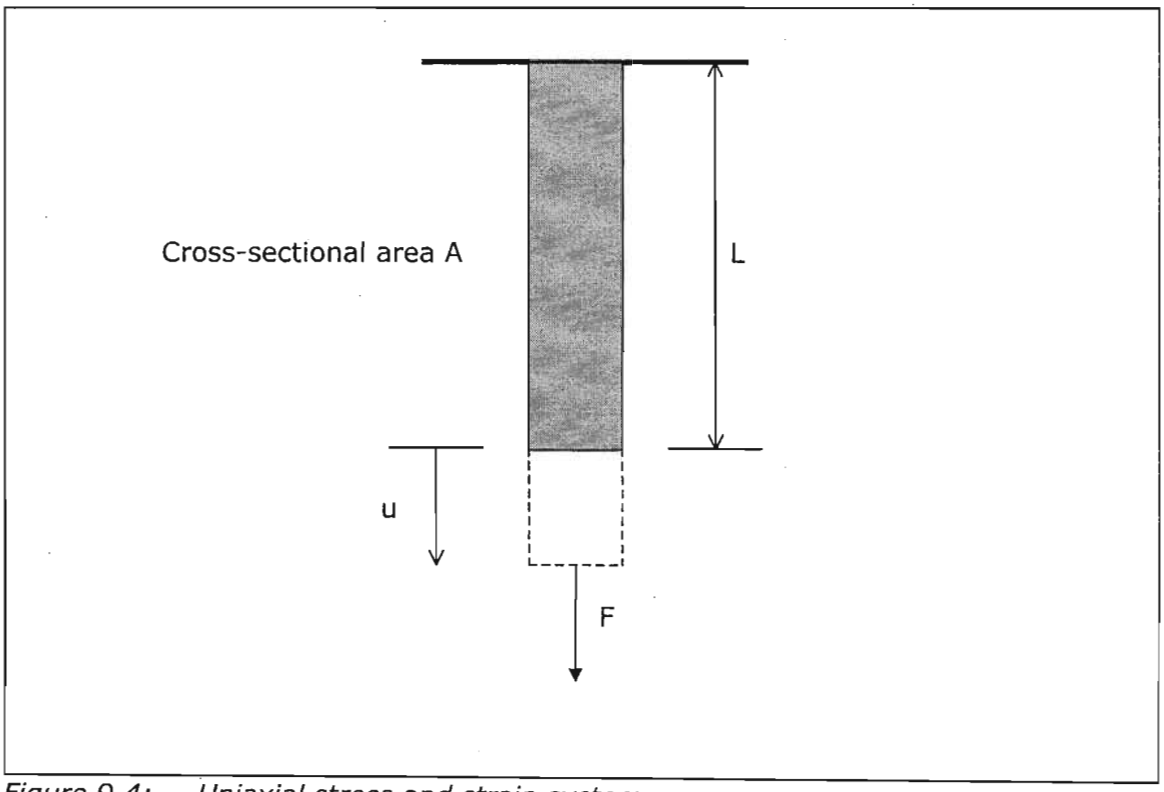


Figure 9.4: Uniaxial stress and strain system

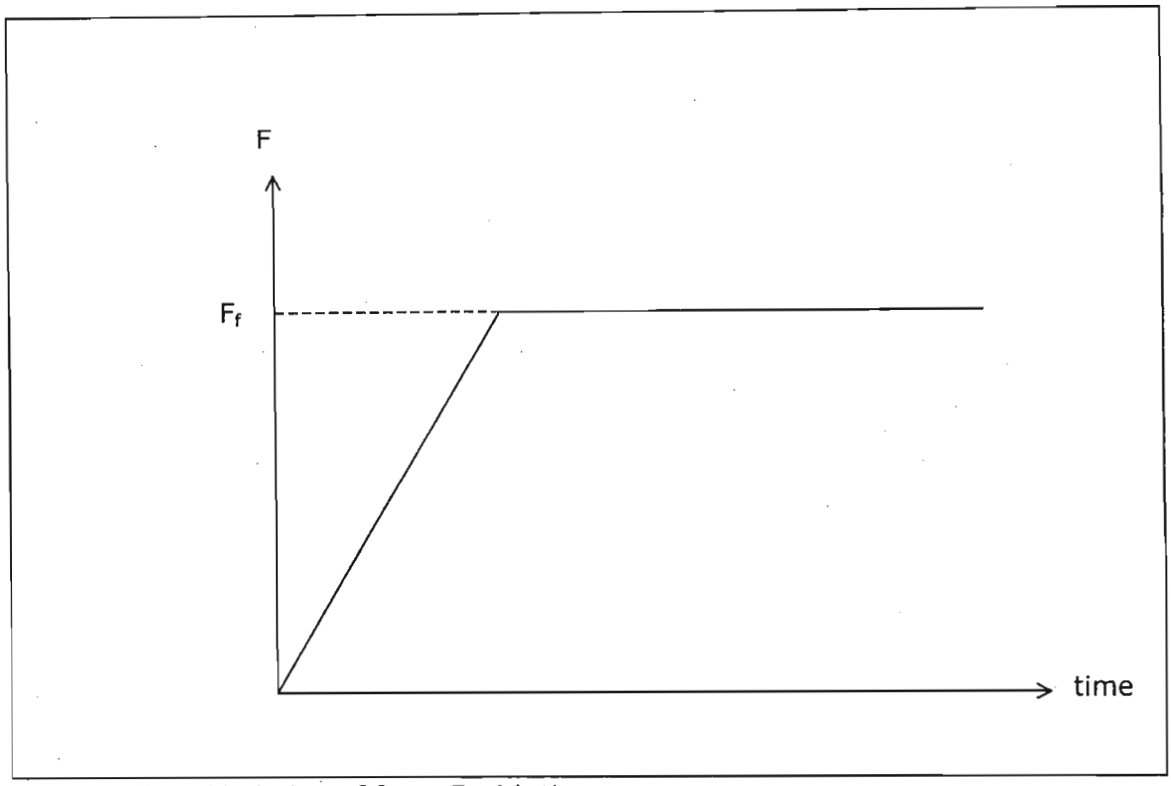


Figure 9.5: Variation of force F with time

The work done by the applied load can be defined as follows:

$$W_C = \int_{R_1}^{R_2} \bar{F} \cdot d\bar{R} \tag{9-46}$$

$$W_C = \int_0^{u_f} F.du \tag{9-47}$$

Note that because the force is applied to the endpoint of the bar, the integral boundaries are the boundaries of the endpoint displacement.

From equation (9-41), for the uniaxial stress state:

$$W_C = \int_0^{u_f} \left(\frac{F_f}{u_f} u \right) du = \frac{F_f}{u_f} \left[\frac{1}{2} u^2 \right]_0^{u_f} = \frac{1}{2} F_f . u_f \tag{9-48}$$

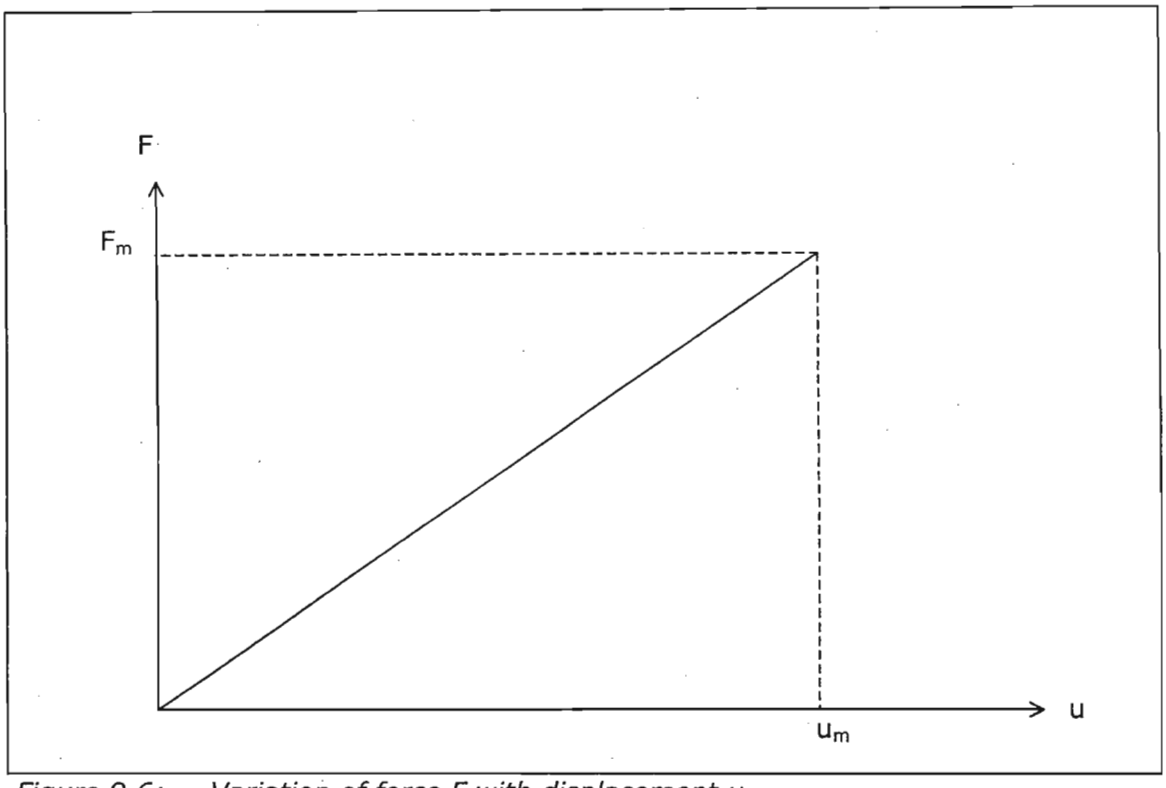


Figure 9.6: Variation of force F with displacement u

The internal energy of the system must increase an equal amount to this work performed by the applied force. A gradual reduction to zero of the force would result in the return of a perfectly elastic system to its original undeformed state. The internal energy of the system would also return to the zero. This returnable energy that is stored in a deformed elastic body is called strain energy, U.

Strain energy per unit volume (where Ω is the element volume):

$$\frac{U}{\Omega} = \frac{1}{2}\frac{F_f.u_f}{A.L} \tag{9-49}$$

In terms of stress and strain:

$$\frac{U}{\Omega} = \frac{1}{2}\sigma_f\varepsilon_f \tag{9-50}$$

$$dU = \frac{1}{2}\sigma.\varepsilon.d\Omega \tag{9-31}$$

For a 3-dimensional stress state undergoing normal and shear loads:

$$dU = \frac{1}{2}\left(\sigma_r\varepsilon_r + \sigma_\theta\varepsilon_\theta + \sigma_z\varepsilon_z + \tau_{r\theta}\gamma_{r\theta} + \tau_{\theta z}\gamma_{\theta z} + \tau_{zr}\gamma_{zr}\right)d\Omega \tag{9-51}$$

Or, using symbolic notation:

$$dU = \frac{1}{2}\{\varepsilon\}^T\{\sigma\}d\Omega \tag{9-52}$$

$$U = \frac{1}{2}\int_{\Omega}\{\varepsilon\}^T\{\sigma\}d\Omega \tag{9-53}$$

Ω is the volume of integration. Combining with equation (9-23) and re-writing:

$$U = \frac{1}{2}\int_{\Omega}\{\varepsilon\}^T[C]\{\varepsilon\}d\Omega \tag{9-54}$$

9.4.3. Total Potential Energy

The Principle of Minimum Potential Energy states that, of all the geometrically possible configurations that a body can assume, the true one, corresponding to the satisfaction of stable equilibrium, is identified by a minimum value for the total potential energy.

The total potential energy Π is defined as the sum of the strain energy and the force potential energy:

$$\Pi = U + V \tag{9-55}$$

Variational techniques are used to find the minimum value of the functional Π.

9.4.4. Generalised FEA Workflow for Physical Load Conditions

The following 5-step process is applicable to all physically loaded models:

1. Select a suitable element type and define a local natural (dimensionless) coordinate system.

2. Assume an interpolation function (displacement function) in terms of the local element coordinate system.

3. Derive the element strain and the element strain energy.

4. Derive the applied load energy functions.

5. Assemble the element equations for the total potential energy of the system (region Ω) in terms of the global coordinate system and solve using variational techniques.

Selecting the Element Types and the Local Natural Coordinate System

Natural coordinate systems are dimensionless and identify locations in an element without regard to the element size or shape.

The local coordinate system used in a particular analysis depends on the nature of the geometry, the load conditions and the constraints. For example, problems consisting of axisymmetric geometry constrained and loaded in a manner that can be defined in terms of cylindrical coordinates may be discretised into axisymmetric 2D elements. An example of an applicable local natural coordinate system (a,b,c) for such an element type [53] is shown in Figure 9.4, where:

$$a + b + c = 1 \tag{9-56}$$

a, b and c each vary from 0 to 1. The location of any point within the element can be established by any 2 local coordinates (i.e. there are 2 independent variables), since:

$$c = 1 - a - b \tag{9-57}$$

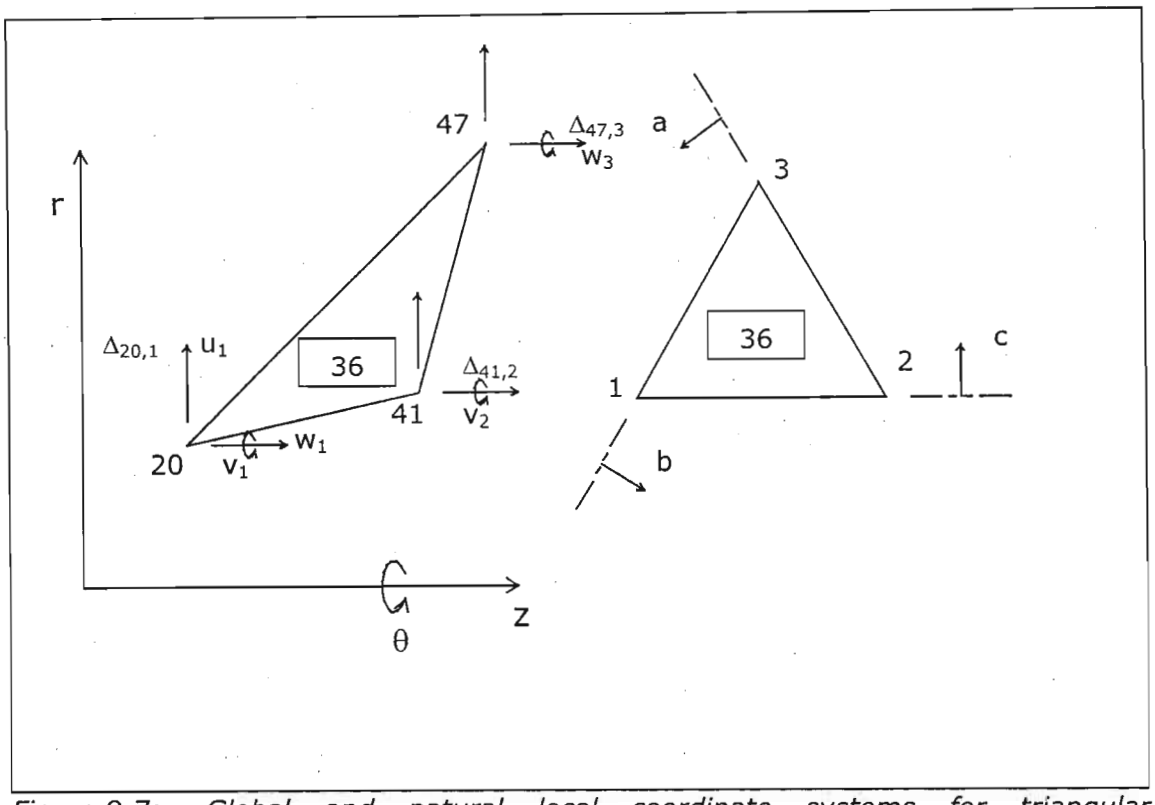


Figure 9.7: Global and natural local coordinate systems for triangular axisymmetric elements.

Assuming an Interpolation Formula to Describe the Displacement Behaviour of the Points Within Each Element

Utilising Lagrange's interpolation formula, the displacement u of any point p in the element in terms of a corresponding global coordinate system direction is calculated as follows:

$$u_p = N_1u_1 + N_2u_2 + .. + N_mu_m \tag{9-58}$$

where N_i is the shape function $N_i(a,b)$ described in terms of the local natural coordinate system (a,b). u_i is the displacement of node i in the element in the u direction. In simplified terms, the displacement is assumed to vary throughout each element according to the following equations:

$$u = \sum_{i=1}^{m} N_i(a,b).u_i$$

$$v = \sum_{i=1}^{m} N_i(a,b).v_i \tag{9-59}$$

$$w = \sum_{i=1}^{m} N_i(a,b).w_i$$

where u, v and w are the displacements in the 3 directions of the selected coordinate system. In Figure 9.4, a typical triangular axisymmetric element numbered 36 fills the region between nodes 20, 41 and 47. These are labelled 1, 2 and 3 respectively in the corresponding "unit" element 36. The displacement of node 47 in the z direction is designated $\Delta_{47,3}$ (discussed further later in this section), and is equivalent to w_3 of unit element 36.

For each function $N_i(a,b)$, it is imperative that at node *i*:

$$N_i(a,b) = 1 \text{ and } N_j(a,b) = 0 \text{ if } j \neq i \tag{9-60}$$

It is extremely simple to derive suitable shape functions $N_i(a,b)$ for 3-node triangular elements having the natural local coordinate system described above. Assuming that the influence of the "known" displacement of each node on the displacement of a given point within the element increases linearly with respect to that node's proximity to the point:

$$\begin{aligned} N_1(a,b) &= a \\ N_2(a,b) &= b \\ N_3(a,b) &= c = 1 - a - b \end{aligned} \tag{9-61}$$

Therefore, from equations (9-58) and (9-61):

$$\begin{aligned} u &= a.u_1 + b.u_2 + (1-a-b).u_3 \\ v &= a.v_1 + b.v_2 + (1-a-b).v_3 \\ w &= a.w_1 + b.w_2 + (1-a-b).w_3 \end{aligned} \tag{9-62}$$

If the interpolation coefficients $N_i(a,b)$ used in the displacement interpolation formula can also be used to define the spatial coordinates of each point within the element (as is the case here), then the elements are referred to as isoparametric:

$$\begin{aligned} r &= a.r_1 + b.r_2 + (1-a-b).r_3 \\ \theta &= a.\theta_1 + b.\theta_2 + (1-a-b).\theta_3 \\ z &= a.z_1 + b.z_2 + (1-a-b).z_3 \end{aligned} \tag{9-63}$$

A Note on Interpolation Functions for Higher Order Elements

Although many types of functions could serve as shape functions (assembled for a given element to form [N]), polynomials have received the most widespread use. This is primarily because of the ease of integration and differentiation. In the above example, the natural coordinates selected for the triangular element required a slightly different approach to the typical procedure for polynomials described below.

An n-th order polynomial defined in terms of the local coordinate system describes the shape function.

$$\begin{aligned} N(x) &= P_n(x) \\ N(x,y) &= P_n(x,y) \\ N(x,y,z) &= P_n(x,y,z) \end{aligned} \tag{9-64}$$

The number of nodes along an element's edges and in the middle determines the maximum order of the interpolation function that may be specified.

1. For a local coordinate system with a single independent variable (wire-frame element):

$$N(x) = P_n(x) = \sum_{i=0}^{T_n^{(1)}} \alpha_i x^i \tag{9-65}$$

   where $T_n^{(1)} = n+1$ is the number of terms in the polynomial. A linear variation of the field variable through the element (first order interpolation function) is represented as:

$$P_1(x) = \alpha_0 + \alpha_1 x$$

   A second order variation of the field variable through the element (parabolic interpolation function) is represented as:

$$P_2(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2$$

2. For a local coordinate system with two independent variables (shell element):

$$N(x,y) = P_n(x,y) = \sum_{k=0}^{T_n^{(2)}} \alpha_k x^i y^j, \quad i + j \le n \tag{9-66}$$

where $T_n^{(2)} = \frac{(n+1)(n+2)}{2}$ is the number of terms in the polynomial. A linear variation of the field variable through the element (first order interpolation function) is represented as:

$$P_1(x, y) = \alpha_0 + \alpha_1 x + \alpha_2 y$$

A second order variation of the field variable through the element (parabolic interpolation function) is represented as:

$$P_2(x, y) = \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 xy + \alpha_4 x^2 + \alpha_5 y^2$$

3. For a local coordinate system with three independent variables (volume element):

$$N(x, y, z) = P_n(x, y, z) = \sum_{l=0}^{T_n^{(3)}} \alpha_l x^i y^j z^k, \quad i + j + k \le n \qquad (9\text{-}67)$$

where $T_n^{(3)} = \frac{(n+1)(n+2)(n+3)}{6}$ is the number of terms in the polynomial. A linear variation of the field variable through the element (first order interpolation function) is represented as:

$$P_1(x, y, z) = \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 z$$

A second order variation of the field variable through the element (parabolic interpolation function) is represented as:

$$P_1(x, y, z) = \alpha_0 + \alpha_1 x + \alpha_2 y + \alpha_3 z + \alpha_4 xy + \alpha_5 yz + \alpha_6 zx + \alpha_7 x^2 + \alpha_8 y^2 + \alpha_9 z^2$$

Calculating the Element Strain and the Element Strain Energy

The calculation of the element strain requires the differentiation of the displacement functions relative to the local coordinate system. The resulting equations must then be converted into terms of the global coordinates for assembly into a system matrix. For elements with local natural coordinate system (a,b) and global cylindrical coordinate system (r,θ,z), this is achieved using partial derivatives as follows:

$$\frac{\partial u}{\partial a} = \frac{\partial u}{\partial r}\frac{\partial r}{\partial a} + \frac{\partial u}{\partial \theta}\frac{\partial \theta}{\partial a} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial a}$$

$$\frac{\partial u}{\partial b} = \frac{\partial u}{\partial r}\frac{\partial r}{\partial b} + \frac{\partial u}{\partial \theta}\frac{\partial \theta}{\partial b} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial b} \tag{9-68}$$

For axisymmetric 2D elements, the middle terms of the right hand side of equation (9-68) equal zero. Also:

$$\frac{\partial v}{\partial a} = \frac{\partial v}{\partial b} = 0 \tag{9-69}$$

The complete system of element equations for the axisymmetric example is as follows:

$$\begin{aligned}
\frac{\partial u}{\partial a} &= \frac{\partial u}{\partial r}\frac{\partial r}{\partial a} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial a} \\
\frac{\partial u}{\partial a} &= \frac{\partial u}{\partial r}\frac{\partial r}{\partial a} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial a} \\
\frac{\partial w}{\partial a} &= \frac{\partial w}{\partial r}\frac{\partial r}{\partial a} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial a} \\
\frac{\partial w}{\partial b} &= \frac{\partial w}{\partial r}\frac{\partial r}{\partial b} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial b}
\end{aligned} \tag{9-70}$$

Focusing on the strain due to displacement in the u direction, and writing in matrix form:

$$\begin{Bmatrix} \frac{\partial u}{\partial a} \\ \frac{\partial u}{\partial b} \end{Bmatrix} = \begin{bmatrix} \frac{\partial r}{\partial a} & \frac{\partial z}{\partial a} \\ \frac{\partial r}{\partial b} & \frac{\partial z}{\partial b} \end{bmatrix} \begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{Bmatrix} \tag{9-71}$$

From equations (9-66) and (9-67):

$$\begin{Bmatrix} u_1 - u_3 \\ u_2 - u_3 \end{Bmatrix} = \begin{bmatrix} r_1 - r_3 & z_1 - z_3 \\ r_2 - r_3 & z_2 - z_3 \end{bmatrix} \begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{Bmatrix} \tag{9-72}$$

Inverting:

$$\begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 \\ r_3 - r_2 & r_1 - r_3 \end{bmatrix} \begin{Bmatrix} u_1 - u_3 \\ u_2 - u_3 \end{Bmatrix} \tag{9-73}$$

Now:

$$\begin{Bmatrix} u_1 - u_3 \\ u_2 - u_3 \end{Bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} \tag{9-74}$$

From (9-73) and (9-74):

$$\begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 \\ r_3 - r_2 & r_1 - r_3 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} \tag{9-75}$$

Where:

$$|J| = (r_1 - r_3)(z_2 - z_3) - (z_1 - z_3)(r_2 - r_3) \tag{9-76}$$

Simplifying (and similarly for the strain due to displacement in the w direction):

$$\begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial u}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 & z_1 - z_2 \\ r_3 - r_2 & r_1 - r_3 & r_2 - r_1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix}$$

$$\begin{Bmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 & z_1 - z_2 \\ r_3 - r_2 & r_1 - r_3 & r_2 - r_1 \end{bmatrix} \begin{Bmatrix} w_1 \\ w_2 \\ w_3 \end{Bmatrix} \tag{9-77}$$

Equations (9-77) can now be used to calculate the strain variation through the element in terms of the global coordinates. The strain at a given point in the element is defined as follows:

$$\varepsilon_r = \frac{\partial u}{\partial r}$$

$$\varepsilon_\theta = \frac{\partial v}{\partial \theta} \text{ (= 0 for 2D axisymmetric elements)}$$

$$\varepsilon_z = \frac{\partial u}{\partial z}$$

$$\gamma_{r\theta} = \frac{\partial u}{\partial \theta} + \frac{\partial v}{\partial r} \text{ (= 0 for 2D axisymmetric elements)} \tag{9-78}$$

$$\gamma_{\theta z} = \frac{\partial w}{\partial \theta} + \frac{\partial v}{\partial z} \quad (= 0 \text{ for 2D axisymmetric elements})$$

$$\gamma zr = \frac{\partial u}{\partial z} + \frac{\partial w}{\partial r}$$

The strain matrix for each 2D triangular axisymmetric element is therefore from (9-77) and (9-78):

$$\begin{Bmatrix} \varepsilon_r \\ \varepsilon_z \\ \gamma_{zr} \end{Bmatrix} = \begin{Bmatrix} \frac{\partial u}{\partial r} \\ \frac{\partial w}{\partial z} \\ \frac{\partial u}{\partial z} + \frac{\partial w}{\partial r} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & 0 & z_3 - z_1 & 0 & z_1 - z_2 & 0 \\ 0 & r_3 - r_2 & 0 & r_1 - r_3 & 0 & r_2 - r_1 \\ r_3 - r_2 & z_2 - z_3 & r_1 - r_3 & z_3 - z_1 & r_2 - r_1 & z_1 - z_2 \end{bmatrix} \begin{Bmatrix} u_1 \\ w_1 \\ u_2 \\ w_2 \\ u_3 \\ w_3 \end{Bmatrix}$$

(9-79)

In symbolic notation:

$$\{\varepsilon\} = [B]\{q\} \tag{9-80}$$

where {q} is the matrix of element nodal displacements. Equations (9-59) can also be written as:

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = [N]\{q\} \tag{9-81}$$

It is clearly evident from equations (9-76) and (9-79) that for the simple case of 2D, 3-node, axisymmetric, triangular elements, the element strain is constant throughout the element.

From equation (9-54), the element strain energy is defined as follows:

$$U = \frac{1}{2} \int_{\Omega} \{\varepsilon\}^T [C] \{\varepsilon\} d\Omega$$

Expanding using equation (9-80), and for the axisymmetric example:

$$
\begin{aligned}
U &= \frac{1}{2}\int_{\Omega}\{q\}^T[B]^T[C][B]\{q\}d\Omega \\
&= \frac{1}{2}\{q\}^T[B]^T[C][B]\{q\}\int_{\Omega} d\Omega \\
&= \frac{1}{2}\{q\}^T[B]^T[C][B]\{q\}\int_{\theta=0}^{2\pi} A.r_c.d\theta \\
&= \frac{1}{2}\{q\}^T[B]^T[C][B]\{q\}(2\pi A.r_c)
\end{aligned}
\qquad (9\text{-}82)
$$

where A is the area of the 2D element, and r_c is the distance of the element centroid from the z-axis.

For a triangular element, the area A can be calculated as follows (refer to Figure 9.5):

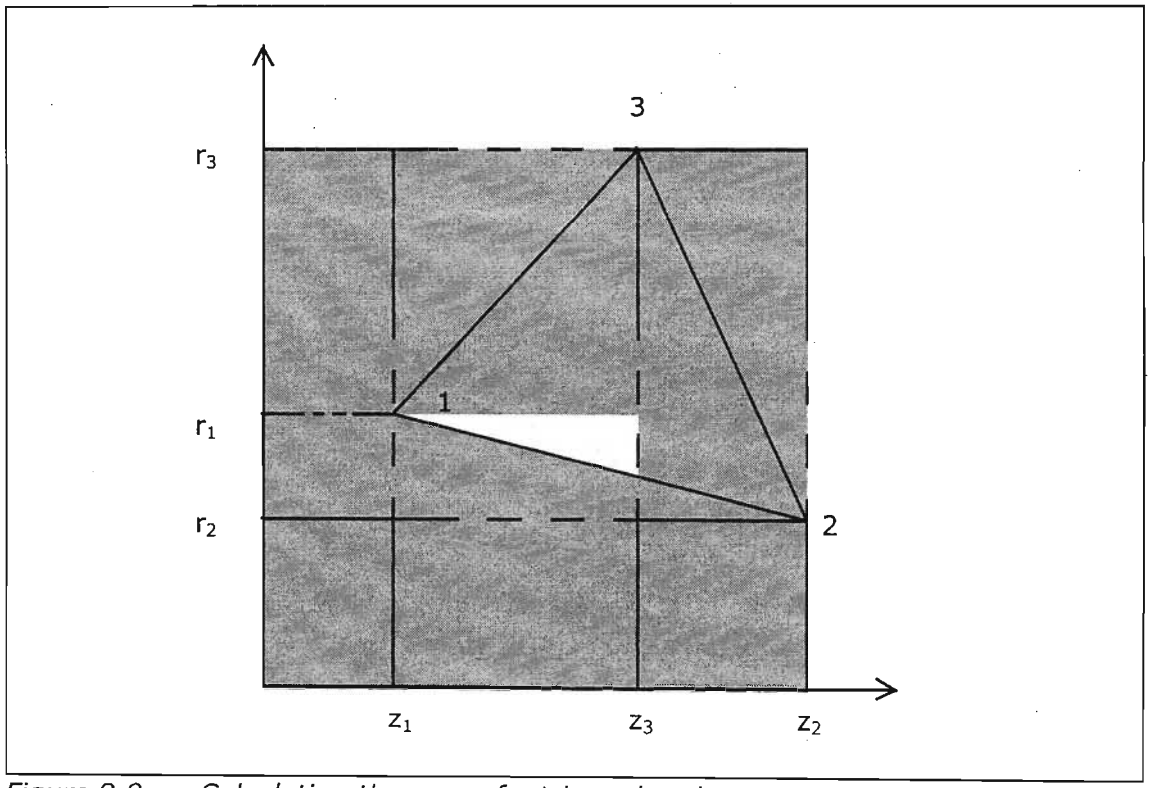


Figure 9.8: Calculating the area of a triangular element.

$$
\begin{aligned}
A &= r_3z_2 - r_3z_1 - r_2(z_2 - z_1) - \frac{1}{2}(r_1 - r_2)(z_2 - z_1) - \frac{1}{2}(r_3 - r_1)(z_3 - z_1) - \frac{1}{2}(r_3 - r_2)(z_2 - z_3) \\
&= \frac{1}{2}(r_1z_2 - r_1z_3 - r_2z_1 + r_2z_3 + r_3z_1 - r_3z_2)
\end{aligned}
\qquad (9\text{-}83)
$$

Now, from equation (9-76):

$$
\begin{aligned}
|J| &= (r_1 - r_3)(z_2 - z_3) - (z_1 - z_3)(r_2 - r_3) \\
&= r_1z_2 - r_1z_3 - r_2z_1 + r_2z_3 + r_3z_1 - r_3z_2
\end{aligned}
\qquad (9\text{-}84)
$$

Comparing equations (9-83) and (9-84), it is evident that:

$$A = \frac{1}{2}|J| \tag{9-85}$$

The equation of the element strain energy defined in equation (9-82) may now be written as:

$$U = \frac{1}{2}\{q\}^T[B]^T[C][B]\{q\}\pi.|J|.r_c \tag{9-86}$$

Referring once again to equation (9-58) and re-organising the terms:

$$U = \frac{1}{2}\{q\}^T\left(\int_\Omega [B]^T[C][B]d\Omega\right)\{q\} \tag{9-87}$$

Now, from equation (9-24), for a rod constrained at one end and a uniaxial load F applied to the other:

$$F = k.u$$

where k is the stiffness of the rod and u the deflection at the end. From equation (9-48), the work exerted by the applied load in stretching the rod by an amount u is:

$$W_c = \frac{1}{2}F.u = \frac{1}{2}k.u^2 \tag{9-88}$$

Comparing equation (9-88) to (9-87), it is apparent that the portion in brackets of equation (9-87) represents the stiffness of the element, i.e.:

$$U = \frac{1}{2}\{q\}^T[k]\{q\} \tag{9-89}$$

where:

$$[k] = \int_\Omega [B]^T[C][B]d\Omega \tag{9-90}$$

For the 3-node axisymmetric element, from equation (9-86):

$$[k] = [B]^T[C][B]\pi.|J|.r_c \tag{9-91}$$

and is easily calculated

A Note on Integration by Gauss Quadrature

The element strain energy has been shown to require integration over the element volume (i.e. integration over up to 3 variables). The triangular constant strain elements (including the 3-node axisymmetric elements) yield functions that are very easily integrated. This is not the case for higher order elements or those that possess midside and even internal nodes. In these instances the integral expressions are so complex that they must be evaluated numerically. There are many techniques of numerical integration, and the so-called Gauss method has found great acceptance in finite element work.

Integration over the natural coordinates has the following form:

$$I = \int_{-1}^{1} f(x)dx \text{ or } I = \int_{0}^{1} f(x)dx \tag{9-92}$$

The general Gaussian approximation formula has the following form:

$$I \approx Ia = \sum_{i=1}^{n} W_i.f(x_i) \tag{9-93}$$

where W_i are weighting factors and n depends on the order of the polynomial I.

Calculating the Element Applied Load Energy Functions

The Nodal-Force Energy Potential Function

Nodal forces representing point loads are commonly applied to 3D systems defined in e.g. Cartesian coordinates, and are defined in terms of the global coordinate directions (Figure 9.6). The nodal-force energy potential of the element can be found directly from equation (9-39):

$$V_N = -\left(\sum_{i=1}^{n} F_{iX}.u_i + \sum_{i=1}^{n} F_{iY}.v_i + \sum_{i=1}^{n} F_{iZ}.w_i\right) \tag{9-94}$$

where F_{iX}, F_{iY} and F_{iZ} are the X,Y and Z components of the forces acting on node i, and V_N is the element force energy potential due to nodal loading.

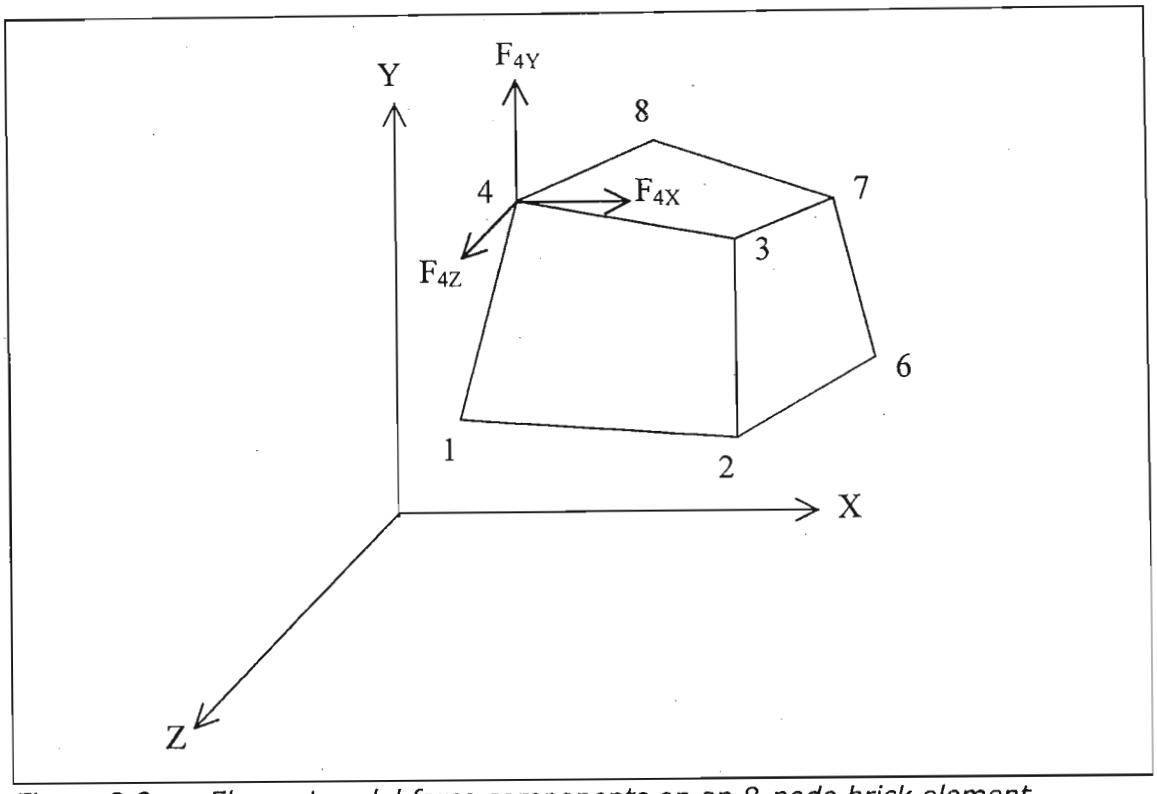


Figure 9.9: Element nodal force components on an 8-node brick element.

Equation (9-94) can be written in symbolic notation:

$$V_N = -\{Q\}_N^T\{q\} \tag{9-95}$$

$\{Q\}_N^T$ is the vector of element nodal forces. Some or all of the entries will equal zero.

$$\{Q\}_N = \begin{Bmatrix} F_{1X} \\ F_{1Y} \\ F_{1Z} \\ \vdots \\ F_{nX} \\ F_{nY} \\ F_{nZ} \end{Bmatrix} \tag{9-96}$$

The only point loads that can be applied to axisymmetric elements are at those points along the z-axis in the z-direction. Forces applied to nodes not lying on the z-axis are in fact equivalent to edge loads (the unique case of axisymmetric forces is discussed later in this section).

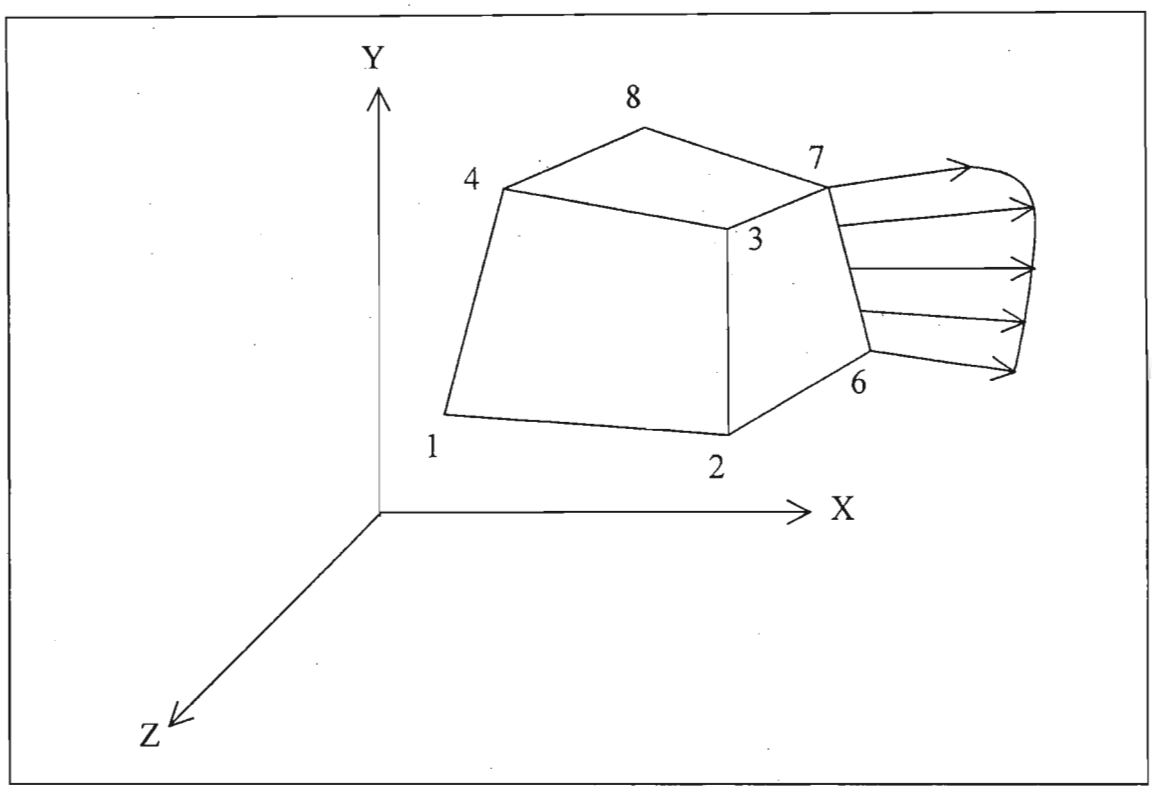


Figure 9.10: Distributed element edge load applied to an 8-node brick element.

Distributed element edge loads are specified as force per unit length. The force per unit length at a particular point along the edge is called the stress traction vector, $\bar{T}_E$, and is presented in terms of the global co-ordinates. Edge loads applied to 2D axisymmetric elements are in fact equivalent to surface loads (the unique case of axisymmetric forces is discussed later in this section).

The following entities are defined:

$\bar{F}_E$ force at a particular point along the edge of the element undergoing stress traction $\bar{T}_E$.

L length of the edge undergoing stress traction $\bar{T}_E$.

V_E force energy potential due to edge loading.

At a particular point along the edge:

$$d\bar{F}_E = (T_E dL)\hat{n} \tag{9-97}$$

where $\hat{n}$ is the unit vector in the direction of the stress traction at that point. Hence:

$$d\bar{F}_E = (T_{EX}dL)\hat{i} + (T_{EY}dL)\hat{j} + (T_{EZ}dL)\hat{k} \tag{9-98}$$

$$dV_E = -[(T_{EX}dL)u + (T_{EY}dL)v + (T_{EZ}dL)w] \tag{9-99}$$

Vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$ are unit vectors in the global co-ordinate directions X, Y and Z respectively. Integrating:

$$V_E = -\int_L [u \quad v \quad w]\begin{Bmatrix} T_{EX} \\ T_{EY} \\ T_{EZ} \end{Bmatrix} dL \tag{9-100}$$

Now from equation (9-81):

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = [N]\{q\} \tag{9-101}$$

The displacement of a point on the edge of an element experiencing edge loading is therefore approximated using the assumed element shape function [N]. For the 8-node solid element shown in Figure 9.7, the edge and the shape function are linear. Naturally, this is not generally the case in reality. The 20-node brick element for example (midside as well as corner nodes) allows displacement according to a quadratic shape function.

From equations (9-100) and (9-101):

$$\begin{aligned} V_E &= -\int_L \{q\}^T [N]^T \begin{Bmatrix} T_{EX} \\ T_{EY} \\ T_{EZ} \end{Bmatrix} dL \\ &= -\{q\}^T \int_L [N]^T \begin{Bmatrix} T_{EX} \\ T_{EY} \\ T_{EZ} \end{Bmatrix} dL \end{aligned} \tag{9-102}$$

When a distributed load acts over an element edge, the equivalent nodal forces must be calculated. From equation (9-95):

$$V_N = -\{Q\}_N^T \{q\}$$

Equation (9-102) has the same form, i.e.:

$$V_E = -\{Q\}_E^T\{q\} \tag{9-103}$$

where $\{Q\}_E$ is the vector of equivalent nodal forces for edge loading:

$$\{Q\}_E = \int_L [N]^T \begin{Bmatrix} T_{EX} \\ T_{EY} \\ T_{EZ} \end{Bmatrix} dL = \begin{Bmatrix} F_{1X} \\ F_{1Y} \\ F_{1Z} \\ \vdots \\ F_{nX} \\ F_{nY} \\ F_{nZ} \end{Bmatrix} \tag{9-104}$$

The Surface-Force Energy Potential Function

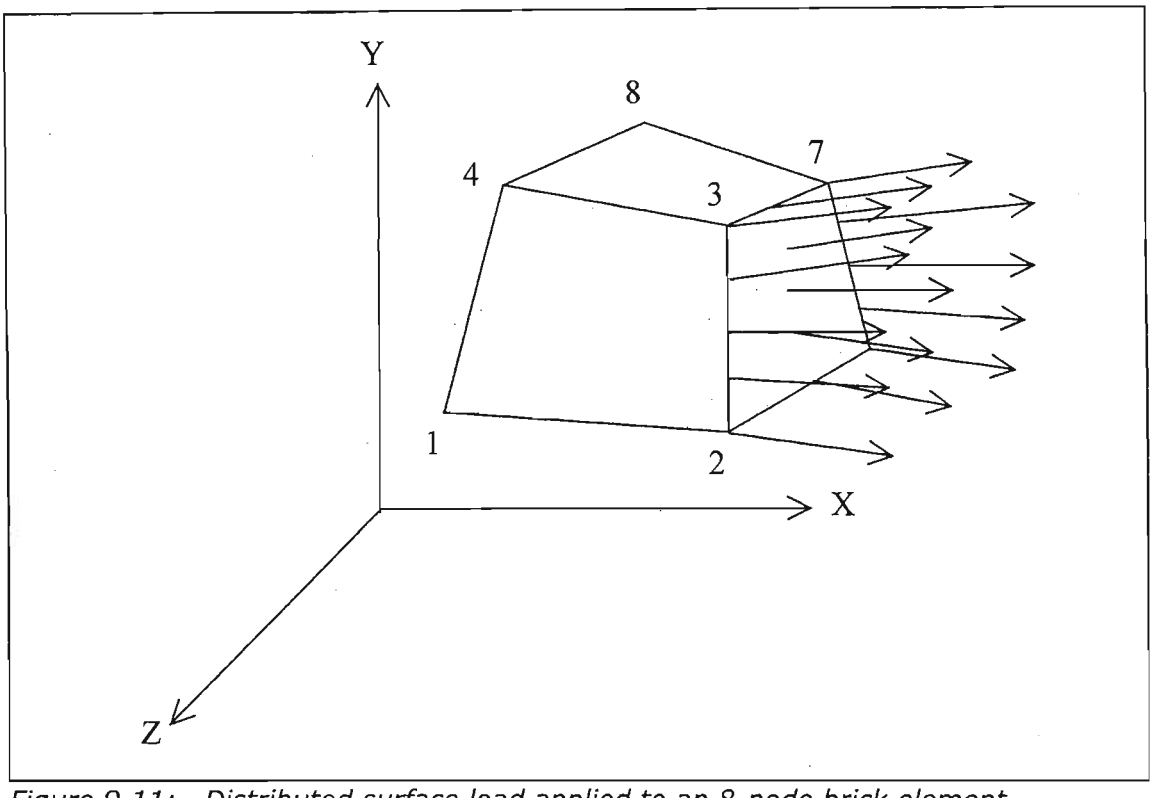


Figure 9.11: Distributed surface load applied to an 8-node brick element.

Distributed surface loads are specified as force per unit area. The force per unit area at a particular point on the surface is also called the stress traction vector, denoted $\overline{T}_A$. Surface loads applied to 2D axisymmetric elements are in fact equivalent to body loads (the unique case of axisymmetric forces is discussed later in this section).

The following entities are defined:

$\bar{T}_A$ stress traction vector at a particular point on a surface of the element.

$\bar{F}_A$ force at a particular point on a surface of the element undergoing stress traction $\bar{T}_A$.

A area of the surface undergoing stress traction $\bar{T}_A$.

V_A force energy potential due to the surface loading.

At a particular point on the surface:

$$d\bar{F}_A = (\bar{T}_A . dA)\hat{n} \tag{9-105}$$

where $\hat{n}$ is the unit vector in the direction of the stress traction at that point. Hence:

$$d\bar{F}_A = (T_{AX} . dA)\hat{i} + (T_{AY} . dA)\hat{j} + (T_{AZ} . dA)\hat{k} \tag{9-106}$$

$$dV_A = -[(T_{AX} . dA)u + (T_{AY} . dA)v + (T_{AZ} . dA)w] \tag{9-107}$$

Vectors $\hat{i}$, $\hat{j}$ and $\hat{k}$ are unit vectors in the global co-ordinate directions X, Y and Z respectively. Integrating:

$$\begin{aligned} V_A &= -\int_A [u \quad v \quad w] \begin{Bmatrix} T_{AX} \\ T_{AY} \\ T_{AZ} \end{Bmatrix} dA \\ &= -\int_A \{q\}^T [N]^T \begin{Bmatrix} T_{AX} \\ T_{AY} \\ T_{AZ} \end{Bmatrix} dA \\ &= -\{q\}^T \int_A [N]^T \begin{Bmatrix} T_{AX} \\ T_{AY} \\ T_{AZ} \end{Bmatrix} dA \end{aligned} \tag{9-108}$$

As with edge loading, the displacement of a point on the surface of an element experiencing surface loading is approximated using the assumed element shape function [N]. For the 8-node solid element shown in Figure 9.8, the shape function is linear. The 20-node brick element is more accurate albeit computationally more intensive, and allows displacements according to a quadratic shape function.

The equivalent nodal forces for element surface loading are computed as follows:

$$\{Q\}_A = \int_A [N]^T \begin{Bmatrix} T_{AX} \\ T_{AY} \\ T_{AZ} \end{Bmatrix} dA = \begin{Bmatrix} F_{1X} \\ F_{1Y} \\ F_{1Z} \\ \vdots \\ F_{nX} \\ F_{nY} \\ F_{nZ} \end{Bmatrix} \tag{9-109}$$

Therefore:

$$V_A = -\{Q\}_A^T \{q\} \tag{9-110}$$

The Body-Force Energy Potential Function

Body forces $\bar{B}$ are specified as force per unit volume (e.g. weight can be exerted on each element having mass throughout the continuum). See Figure 9.9.

At a particular point in the element continuum:

$$\begin{aligned} d\bar{F} &= B.d\Omega.\hat{n} \\ &= (B_X.d\Omega).\hat{i} + (B_Y.d\Omega).\hat{j} + (B_Z.d\Omega).\hat{k} \end{aligned} \tag{9-111}$$

$$\begin{aligned} dV_B &= -[(B_X.d\Omega).u + (B_Y.d\Omega).v + (B_Z.d\Omega).w] \\ &= -[u \quad v \quad w] \begin{Bmatrix} B_X \\ B_Y \\ B_Z \end{Bmatrix} d\Omega \end{aligned} \tag{9-112}$$

$$\begin{aligned} V_A &= -\int_\Omega [u \quad v \quad w] \begin{Bmatrix} B_X \\ B_Y \\ B_Z \end{Bmatrix} d\Omega \\ &= -\{q\}^T \int_\Omega [N]^T \begin{Bmatrix} B_X \\ B_Y \\ B_Z \end{Bmatrix} d\Omega \\ &= -\{Q\}_B^T \{q\} \end{aligned} \tag{9-113}$$

Where:

$$\{Q\}_B = \int_\Omega [N]^T \begin{Bmatrix} B_X \\ B_Y \\ B_Z \end{Bmatrix} d\Omega = \begin{Bmatrix} F_{1X} \\ F_{1Y} \\ F_{1Z} \\ \vdots \\ F_{nX} \\ F_{nY} \\ F_{nZ} \end{Bmatrix} \tag{9-114}$$

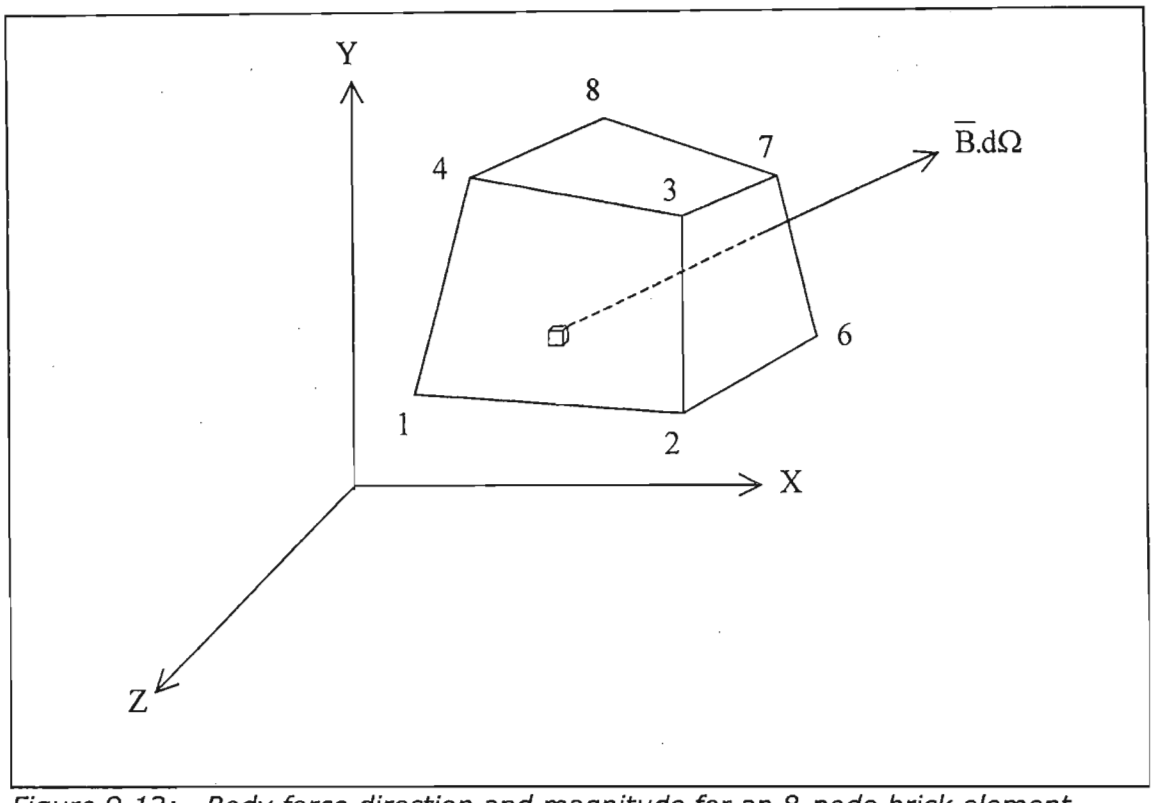


Figure 9.12: Body force direction and magnitude for an 8-node brick element.

Energy Potential Functions for Axisymmetric Elements

Because axisymmetric elements are 2D representations of cylindrical volume elements, applying a force to a particular element node is equivalent to applying a circumferential load to the element rotated 360°. This circumferential load is symmetric about the z-axis.

Similarly, applying a force to a particular axisymmetric element edge is equivalent to applying a surface load to the element volume rotated 360°.

The Total Applied-Force Energy Potential Function

From the principle of superposition:

$$V = V_N + V_E + V_A + V_B$$
$$= -\{q\}^T(\{Q\}_N + \{Q\}_E + \{Q\}_A + \{Q\}_B) \quad (9\text{-}115)$$

In general:

$$V = -\{q\}^T\{Q\} \quad (9\text{-}116)$$

where $\{Q\}$ represents the sum of the applied and the equivalent element nodal forces.

Assembling the Elements and Solving Using Variational Techniques

From equation (9-89), the element strain energy is defined as:

$$U = \frac{1}{2}\{q\}^T[k]\{q\}$$

The element applied load energy potential is defined in equation (9-116) as:

$$V = -\{q\}^T\{Q\}$$

Recall from Section 9.4.3, that the total potential energy Π of a system under physical loading equals the sum of the strain energy, U, and the applied load energy potential, V:

$$\Pi = U + V$$

Expanding for a single element:

$$\Pi_e = \frac{1}{2}\{q\}^T[k]\{q\} - \{q\}^T\{Q\} \quad (9\text{-}117)$$

To analyse the entire system of elements, the global nodal numbers are used. The local element displacement vectors $\{q\}$, which have dimensions equal to the number of degrees of freedom of the element, are now replaced by a single system displacement vector $\{\Delta\}$ with dimension dependent on the number of nodes n in the entire system. $\Delta_{i,X}$, $\Delta_{i,Y}$ and $\Delta_{i,Z}$ are the displacements of node i in the global coordinate system directions. For systems defined in cylindrical coordinates, displacements $\Delta_{i,r}$, $\Delta_{i,\theta}$ and $\Delta_{i,Z}$ apply. For problems consisting of 2D axisymmetric elements, $\Delta_{i,r}$ and $\Delta_{i,Z}$ are the global displacement variables.

Expanding for axisymmetric elements:

$$\{\Delta\} = \begin{Bmatrix} \Delta_{1,r} \\ \Delta_{1,z} \\ \Delta_{2,r} \\ \Delta_{2,z} \\ \vdots \\ \Delta_{n,r} \\ \Delta_{n,z} \end{Bmatrix} = \begin{Bmatrix} \Delta_{1,1} \\ \Delta_{1,2} \\ \Delta_{2,1} \\ \Delta_{2,2} \\ \vdots \\ \Delta_{n,1} \\ \Delta_{n,2} \end{Bmatrix} = \begin{Bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \\ \vdots \\ \Delta_{2n} \end{Bmatrix} \tag{9-118}$$

where n is the total number of nodes in the system.

The components of the element load vectors {Q} are now applied to the corresponding system nodal number (the same node) to form a system nodal force vector {P}. Forces from different element local nodes that are applied to the same system node and that share a common direction are added according to the principle of superposition.

Summing the total potential energy of all the elements in the system:

$$\Pi = \frac{1}{2}\{\Delta\}^T[K]\{\Delta\} - \{\Delta\}^T\{P\} \tag{9-119}$$

where [K] is the system stiffness matrix.

Section 9.3 introduced the concept of variational techniques to minimise functionals such as that of equation (9-119) to find the discrete variables of the field variable, which in this case is displacement. From equation (9-24):

$$\frac{\partial \Pi}{\partial \Delta_i} = K.\Delta_i - P = 0 \tag{9-120}$$

Physical Nonlinearity

The physical macroscopic behaviour of the TRIP rockbolt problem is nonlinear on several counts.

1. The large deformations are a *geometric nonlinearity*. As rockmass slip increases, so the distributed loads transmitted to the bolt and grout begin to change in magnitude and direction. This is addressed by applying an incrementally-increasing displacement load to the applicable rockmass model boundaries. The FEA analysis automatically re-calculates the distributed loads as the stiffness matrix changes.

2. The rockbolt transition from elastic to plastic in critical locations is a *material nonlinearity*. The stiffness of the system is affected by the large displacements. At each equilibrium state along the equilibrium path (minimised functional), the set of simultaneous equations $[K]\{\Delta\} = \{P\}$ is nonlinear because the stiffness matrix [K] is no longer constant. Equilibrium must be re-established as slip progresses. Therefore, a direct solution is not possible and an iterative method is required.

   Nonlinear behaviour is approximated as piecewise linear, and the linear laws are used for each piece. The applied load is broken down into several "time" steps. In the Newton-Raphson procedure, the value of the stiffness matrix is established at time $t + \Delta t$ by initially using the value at time t. The matrix [K] is re-formed for each iteration. At the end of each iteration, a check should be made to test whether the iteration converged within realistic tolerances or if it is diverging.

   An iterative FEA process also automatically incorporates a nonlinear-shaped stress-strain curve into the analysis. A yield point must still be specified to identify when the phase transformation to ferromagnetic martensite begins. A suitable yield criterion based on the principal stresses in the mesh should be selected to accurately predict the onset of this phase transformation at the correct location. No unloading of the deformed support unit takes place, therefore no "spring-back" effects need be considered.

3. Because of the potentially large displacements realised by the support unit under operating conditions, the interface between the grout and bolt is subject to separation. The interfacial bonds break in critical locations and frictional effects arise. Because the bolt is typically far stronger than the grout, the FEA model should cater for possible separation, but frictional effects can be neglected. This can be modelled as *contact nonlinearity* with initial high values of friction at the interface until a critical load value is exceeded and the friction essentially becomes zero. Also to be considered is that the grout and rockmass are subject to fracture due to the potentially large values of slip involved, notably where the bolt bridges the slip plane.

9.5. FEA FOR MAGNETOSTATIC MODELLING

9.5.1. Magnetic Field Equations

A magnetic field is established by [55]:

1. A permanent magnet (e.g. martensitic steel).
2. An electric current in a conductor.
3. Moving charges.

Magnetic field strength $\overline{H}$ is the magnetic equivalent of the electric field strength $\overline{E}$. *Magnetic flux density* $\overline{B}$ is represented in schematics (e.g. Figure 9.13) by a series of curves, called the *magnetic field lines*. Each curve is drawn so that (i) at any point the curve is tangent to $\overline{B}$ at that point, and (ii) so that the number of lines per unit area is proportional to the magnitude of $\overline{B}$ at that point. In magnetostatic response modelling, the magnetic field is considered to be time-invariant. The relationship between $\overline{H}$ and $\overline{B}$ is defined as follows:

$$\{B\} = \mu\{H\} \tag{9-121}$$

μ is the permeability of the medium, and is constant for most materials. The *permeability of a vacuum* $\mu_0 = 4\pi*10^{-7}$ H/m. Notable exceptions are ferromagnetic materials, where $\overline{B}$ is a more complex function of $\overline{H}$ that must be described graphically [56]. Introducing $\overline{B}$ therefore removes the dependence of the magnetic field calculations on the medium. The SI unit of $\overline{B}$ is the tesla, T. The SI unit for $\overline{H}$ is the ampere-turn per metre. 1 ampere-turn per metre = $4\pi*10^{-3}$ oersted (Oe).

In ferromagnetic materials, saturation begins to occur as $\overline{H}$ rises and $\overline{B}$ fails to continue to do so. This effect places a limit on the values of $\overline{B}$ that can be attained. For values well below saturation, $\overline{B}$ is approximately proportional to $\overline{H}$ with a μ value based on the initial slope of the B-H curve [56].

The magnetic flux $d\Phi$ through a small surface having area dA (illustrated in Figure 9.13) is defined as:

$$d\Phi = B_{\perp}.dA = B.\cos\theta.dA = \overline{B} \bullet d\overline{A} \tag{9-122}$$

$\overline{B}$ can therefore be described as the magnetic flux per unit area (hence the term *magnetic flux density*) passing through an area $dA \rightarrow 0$. The total magnetic flux through a surface of finite dimensions is therefore [55]:

$$\Phi = \iint B_{\perp}.dA = \iint \overline{B} \bullet d\overline{A} \qquad (9\text{-}123)$$

where $d\overline{A}$ has direction normal to the surface and pointing outwards in the same way as $\overline{B}$.

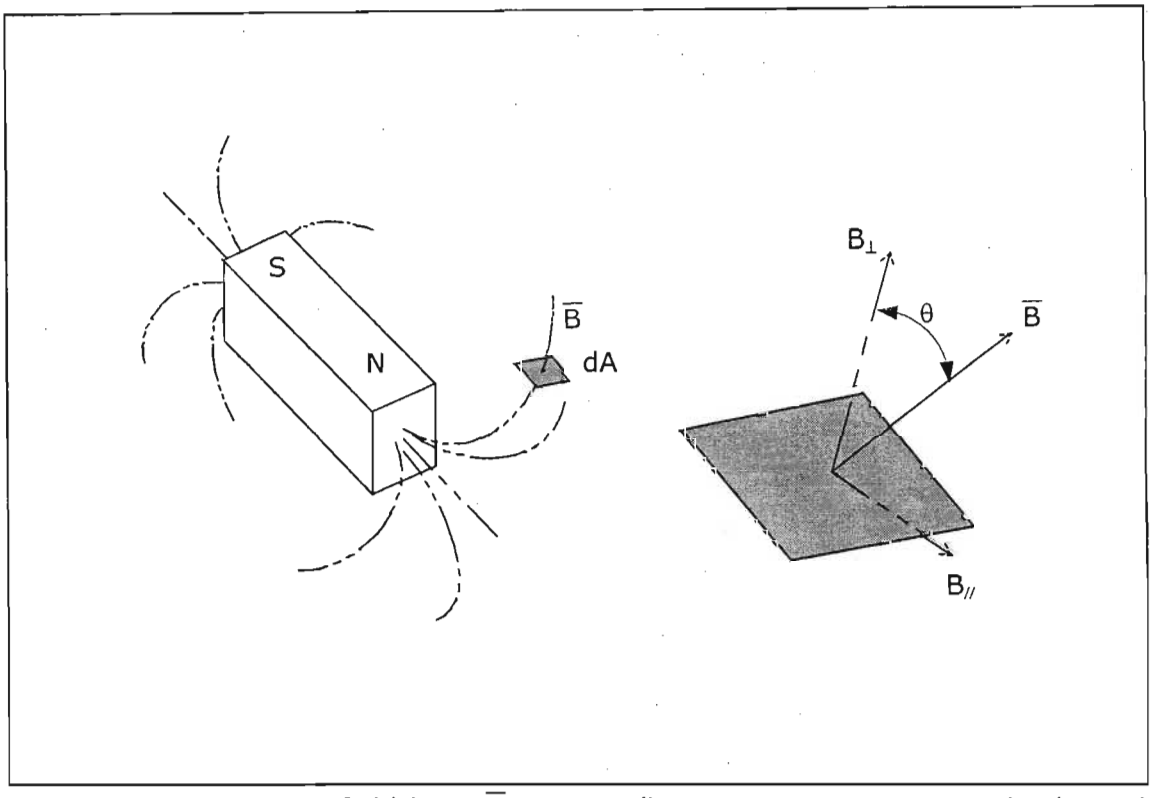


Figure 9.13: Magnetic field lines $\overline{B}$ surrounding a permanent magnet in the N-S direction

Coulomb's Force Law

Experimentally, it can be shown that a force exists between 2 charges. Mathematically, the force relationship between two point charges Q_1 and Q_2 can be stated as [58]:

$$F = k\frac{Q_1.Q_2}{R^2} \qquad (9\text{-}124)$$

where R is the distance between the point charges, and k is a constant of proportionality depending on the *permittivity* ε of the medium. In SI units:

$$k = \frac{1}{4\pi\varepsilon} \tag{9-125}$$

In a vacuum, the *permittivity of free* space ε_0 = 8.854e-12F/m. This value is not determined experimentally, but has been defined in a way that makes the SI system of units self-consistent. Relative permittivity $\varepsilon_r = \frac{\varepsilon}{\varepsilon_0}$. For quartz, $\varepsilon_r \approx 5$. Crystalline structures have a permittivity that is more complex than a scalar constant. Since force is a vector, Coulomb's law for 2 point charges immersed in a uniform material medium of infinite extent can be written as:

$$\bar{F} = \frac{Q_1 . Q_2}{4\pi\varepsilon R^2}\hat{R} \tag{9-126}$$

where $\hat{R} = \frac{\bar{R}}{R}$ is a unit vector along the line connecting the 2 charges.

The Electric Field

If the charge Q_2 described above is removed, an *electric field* remains in space around Q_1. The magnitude of the electric field at a point near Q_1 (a field point) is the force per unit charge on a positive test charge (e.g. ΔQ_2) at that field point, provided the test charge is small enough so as not to disturb the electric field around Q_1 being tested. Therefore, at the field point where ΔQ_2 is located:

$$\bar{E} = \frac{\bar{F}}{\Delta Q_2} = \frac{Q_1}{4\pi\varepsilon R^2}\hat{R} \tag{9-127}$$

It is now necessary to define vector quantity $\bar{E}(\bar{r})$ at a field point located at $\bar{r}$ with respect to a global coordinate system. Refer to Figure 9.14.

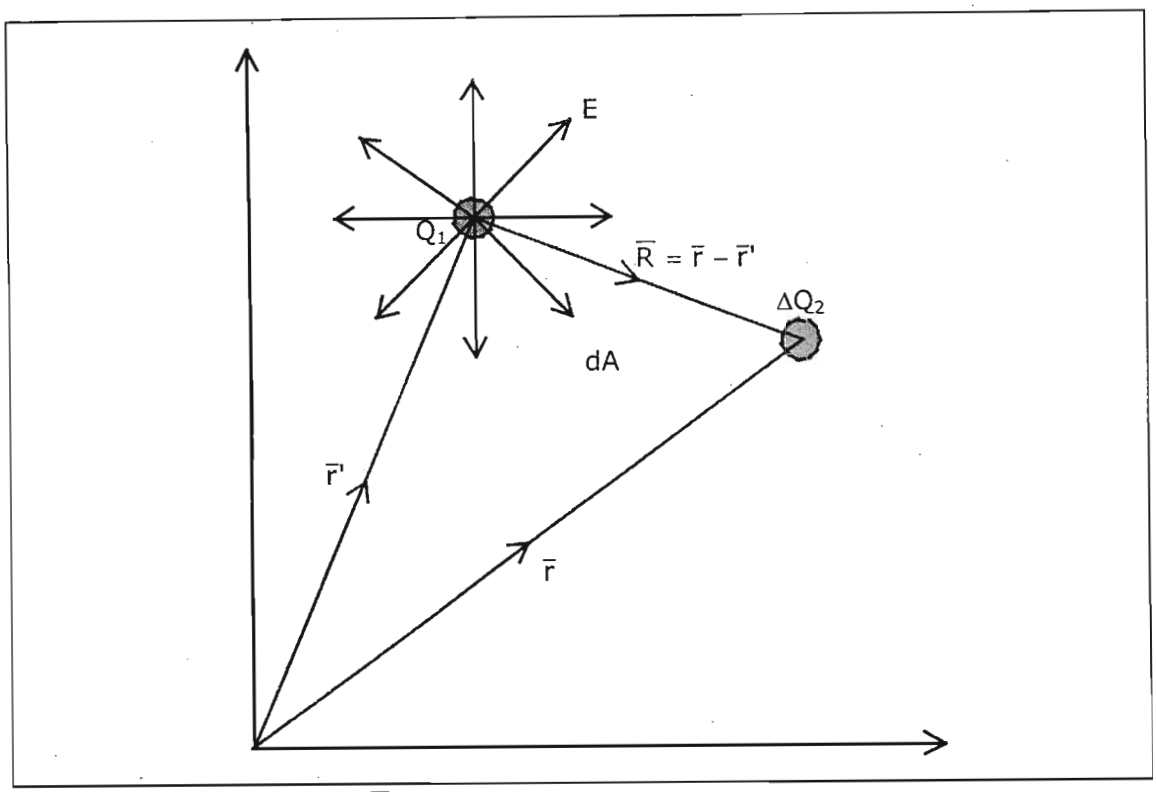


Figure 9.14: Calculating $\bar{E}$ due to Q_1 at location $\bar{r}$ in a global coordinate system

From equation (9-127):

$$\bar{E}(\bar{r}) = \frac{Q_1(\bar{r}')}{4\pi\varepsilon|\bar{r} - \bar{r}'|^2}\hat{R} = \frac{Q_1(\bar{r}')}{4\pi\varepsilon|\bar{r} - \bar{r}'|^2}\frac{\bar{r} - \bar{r}'}{|\bar{r} - \bar{r}'|} \tag{9-128}$$

Using the principle of superposition for an electric field surrounding n point charges, the value of $\bar{E}(\bar{r})$ at a point at location $\bar{r}$ is:

$$\bar{E}(\bar{r}) = \frac{1}{4\pi\varepsilon}\sum_{i=1}^{n}\frac{Q_i(\bar{r}_i')}{|\bar{r} - \bar{r}_i'|^2}\frac{\bar{r} - \bar{r}_i'}{|\bar{r} - \bar{r}_i'|} \tag{9-129}$$

Electric Flux and Gauss' Law for Electric Fields

The electric field $\bar{E}$ is dependent on the permittivity ε of the medium. The vector quantity $\bar{D}$, referred to as the *electric flux density*, is defined to remove the dependence on the medium:

$$\bar{D} = \varepsilon\bar{E} \tag{9-130}$$

For an isotropic medium (i.e. *not* for crystalline structures), ε is scalar. Since $\bar{D}$ is a flux density, the electric flux $d\Psi$ streaming through an elemental area dA (calculated in a similar fashion to $d\Phi$ in equation (9-122)) is:

$$d\Psi = \overline{D} \bullet d\overline{A} \tag{9-131}$$

Therefore the electric flux streaming through a large surface is:

$$\Psi = \iint \overline{D} \bullet d\overline{A} \tag{9-132}$$

The total electric flux streaming out of a closed surface is:

$$\Psi = \oiint \overline{D} \bullet d\overline{A} \tag{9-133}$$

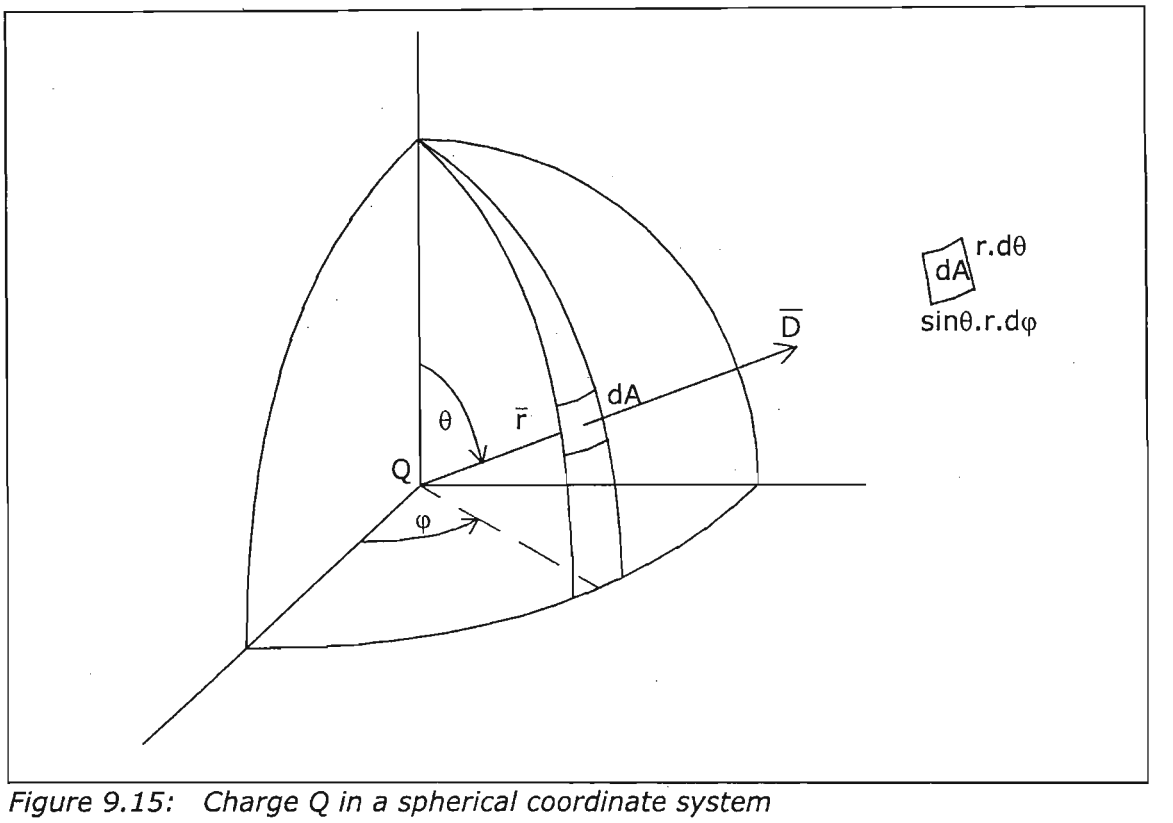


Figure 9.15: Charge Q in a spherical coordinate system

If ψ is finite in equation (9-133), then a net electric flux is leaving the surface. Consider a charge Q located at the origin as in Figure 9.15. The flux streaming out of a spherical surface having radius r about the origin can be calculated from equations (9-128) and (9-133):

$$\Psi = \oiint \frac{Q}{4\pi r^2} \hat{r} \bullet d\overline{A} \tag{9-134}$$

Now, since the normal to this surface is always in the radial direction (i.e. $d\overline{A} = \hat{r}.dA$):

$$\begin{aligned}\Psi &= \oiint \frac{Q}{4\pi r^2} dA \\ &= \frac{Q}{4\pi r^2} \oiint dA\end{aligned} \tag{9-135}$$

Now:

$$\begin{aligned}\oiint dA &= \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} (r.d\theta)(\sin\theta.r.d\varphi) \\ &= 2\pi r^2 \int_0^{\pi} \sin\theta.d\theta \\ &= 4\pi r^2\end{aligned} \tag{9-136}$$

From equations (9-135) and (9-136), the important observation can be made that the total electric flux ψ emanating from a closed surface is equal to the enclosed charge Q; that is:

$$\psi = Q \tag{9-137}$$

This is Gauss' law for electric fields. The relationship in equation (9-137) can be proved to be true for arbitrary shaped surfaces:

$$\begin{aligned}\Psi &= \oiint \overline{D} \bullet d\overline{A} \\ &= \oiint D.\bar{r} \bullet \hat{n}.dA \\ &= \oiint D.\cos\theta.dA \\ &= \oiint \frac{Q}{4\pi r^2} \cos\theta.dA\end{aligned} \tag{9-138}$$

$\hat{r} \bullet \hat{n}.dA$ is the projection of surface element $\hat{n}.dA$ onto a spherical surface that cuts the arbitrarily shaped surface at $\bar{r}$. Therefore, $\cos\theta.dA$ is simply an elemental area on a spherical surface and is given by:

$$\cos\theta.dA = r.d\theta.\sin\theta.r.d\varphi \tag{9-139}$$

Finally:

$$\begin{aligned}\Psi &= \oiint \frac{Q}{4\pi r^2} \cos\theta.dA \\ &= \frac{Q}{4\pi} \oiint \frac{\cos\theta.dA}{r^2} \\ &= \frac{Q}{4\pi} \oiint \frac{r^2 \sin\theta.d\theta d\varphi}{r^2} \\ &= Q\end{aligned} \tag{9-140}$$

It should be noted that the r^2 term in the denominator that belongs to the flux density D always cancels the r^2 term in the numerator due to the area. In other words, the proof of Gauss' law for electric fields is based upon the assumption that the electric force from a charge varies as $\frac{1}{r^2}$.

Gauss' law can now be stated as:

$$\oiint \overline{D} \bullet d\overline{A} = Q \tag{9-141}$$

Gauss' Law for Magnetic Fields

A similar law to Gauss' law for electric flux ψ through a closed surface:

$$\Psi = \oiint \overline{D} \bullet d\overline{A} = Q \tag{9-142}$$

can be written for magnetic flux Φ through a closed surface:

$$\Phi = \oiint \overline{B} \bullet d\overline{A} = Q_m \tag{9-143}$$

where Q_m is the total magnetic charge inside the closed surface. The difference between electric and magnetic fields is that no magnetic charge exists. Therefore, from equation (9-123) for magnetic fields:

$$\oiint \overline{B} \bullet d\overline{A} = 0 \tag{9-144}$$

The implication is that each magnetic field line terminates on itself (i.e. is a closed loop).

Divergence and the Differential Form of Gauss' Law for Electric Fields

For a distributed charge Q in a given region, ρ is defined as the charge density at a given point in that region. Consider for example the small volume $\Delta v = \Delta x.\Delta y.\Delta z$ of Figure 9.16 permeated by a flux density $\overline{D}$. Then, by definition:

$$\Delta Q = \rho.\Delta v \tag{9-145}$$

Hence, from equation (9-141):

$$\begin{aligned} \oiint \overline{D} \bullet d\overline{A} &= \Delta Q \\ &= \rho.\Delta v \end{aligned} \tag{9-146}$$

And:

$$\frac{\oiint \overline{D} \bullet d\overline{A}}{\Delta v} = \rho \tag{9-147}$$

The limit $\lim_{\Delta v \to 0}\left(\frac{\oiint \overline{D} \bullet d\overline{A}}{\Delta v}\right)$ is called the *divergence of* $\overline{D}$:

$$\mathrm{div}\overline{D} = \lim_{\Delta v \to 0}\left(\frac{\oiint \overline{D} \bullet d\overline{A}}{\Delta v}\right) \tag{9-148}$$

The divergence of a vector flux density is therefore a measure of the outward flow of the flux per unit volume. From (9-147) and (9-148):

$$\mathrm{div}\overline{D} = \rho \tag{9-149}$$

For a field at a point to have a finite divergence, an equivalent to a charge source must exist there. If the divergence of a field is zero, for a small volume, it implies that all the flux that enters the volume also leaves it; this region of space is then source-free.

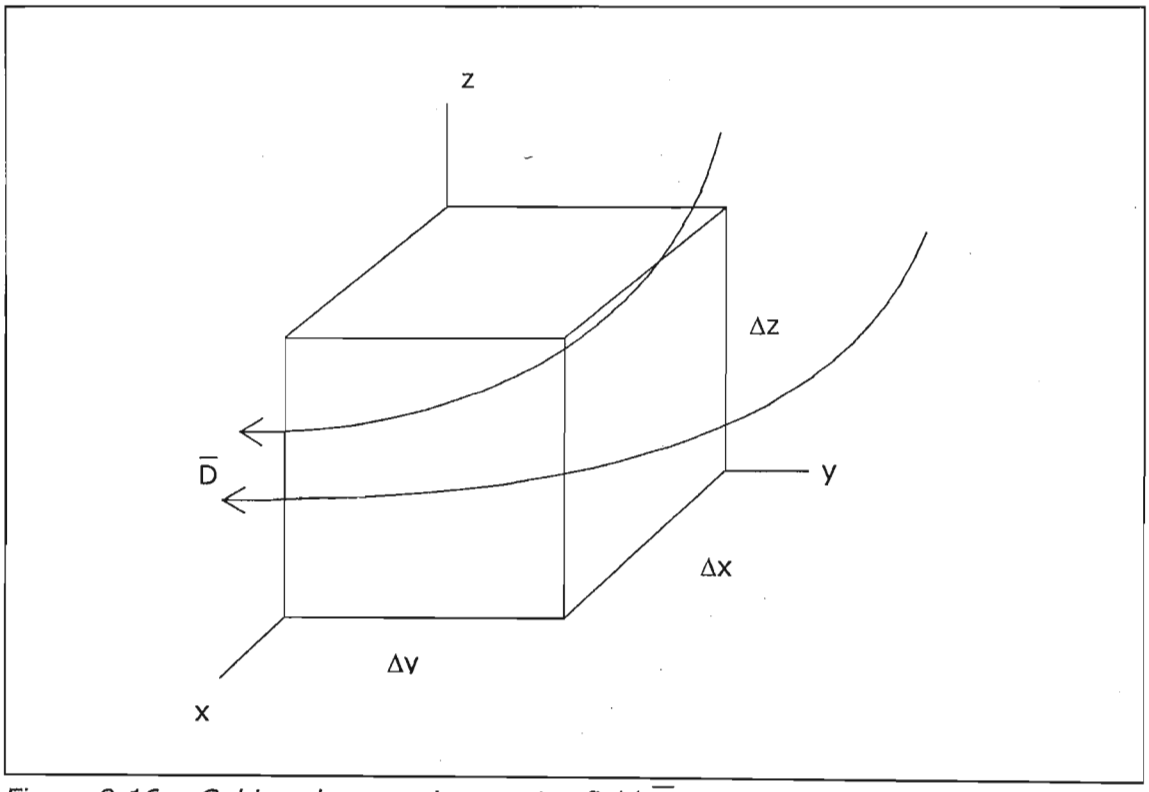


Figure 9.16: Cubic volume Δv in a vector field $\overline{D}$.

Consider now Figure 9.16. The small volume Δv in a region through which $\overline{D}$ permeates has 3 sets of opposing faces. For flux entering and leaving the opposing pair of faces $\Delta z \Delta y$:

$$\text{Flux entering} = \overline{D} \bullet \hat{n}.\Delta A = -D_x.\Delta y.\Delta z \tag{9-150}$$

The minus sign is due to D_x and $\hat{n}$ pointing in opposite directions.

$$\text{Flux leaving} = \overline{D} \bullet \hat{n}.\Delta A = \left(D_x + \frac{\partial Dx}{\partial x}\Delta x\right)\Delta y.\Delta z \tag{9-151}$$

$$\text{Flux entering + flux leaving} = \left(\frac{\partial Dx}{\partial x}\Delta x\right)\Delta y.\Delta z \tag{9-152}$$

Considering also the remaining 2 sets of faces, then:

$$\oiint \overline{D} \bullet d\overline{A} = \left(\frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z}\right)\Delta x.\Delta y.\Delta z \tag{9-153}$$

$$\frac{\oiint \overline{D} \bullet d\overline{A}}{\Delta v} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \tag{9-154}$$

Hence:
$$\text{div}\overline{D} = \frac{\partial D_x}{\partial x} + \frac{\partial D_y}{\partial y} + \frac{\partial D_z}{\partial z} \tag{9-155}$$

Therefore:
$$\text{div}\overline{D} = \left(\frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial y}\hat{j} + \frac{\partial}{\partial z}\hat{k}\right) \bullet \left(D_x\hat{i} + D_y\hat{j} + D_z\hat{k}\right) \tag{9-156}$$

i.e.
$$\text{div}\overline{D} = \overline{\nabla} \bullet \overline{D} \tag{9-157}$$

Finally, the differential form of Gauss' law for electric fields is:

$$\overline{\nabla} \bullet \overline{D} = \rho \tag{9-158}$$

In summary, if a charge density exists at a point in space, then the electric flux density must have a divergence at this point. In other words, if the point is surrounded by a small volume, a net flux must emerge from it. Conversely, if the electric flux density $\overline{D}$ shows a divergence at a point, a charge density must exist at such a point.

Gauss Divergence Theorem

The methods listed in the section above can be extended to derive a theorem that will be useful in manipulating vector equations in order to reduce them to their most convenient forms. This is *Gauss' divergence theorem*. From equations (9-148) and (9-157), it can be stated that the outward flux from the surface of an infinitesimal cube is equal to the divergence of the flux-density vector multiplied by the volume of the cube, i.e.:

$$\frac{\oiint \bar{\Gamma} \bullet d\bar{A}}{\Delta v} = \bar{\nabla} \bullet \bar{\Gamma}$$

or:
$$\oiint \bar{\Gamma} \bullet d\bar{A} = \left(\bar{\nabla} \bullet \bar{\Gamma}\right)\Delta v \qquad (9\text{-}159)$$

where $\bar{\Gamma}$ is an arbitrary vector field. For a finite volume, the divergence can be integrated over the entire volume [61]:

$$\oiint_{A(v)} \bar{\Gamma} \bullet d\bar{A} = \iiint_{v(A)} \bar{\nabla} \bullet \bar{\Gamma}.dv \qquad (9\text{-}160)$$

where A(v) is the surface area bounding volume v, and v(A) is the volume v enclosed by area A.

Gauss' Divergence Theorem and Magnetic Fields

By applying the Gauss divergence theorem described in equation (9-160) to equation (9-144):

$$\oiint \bar{B} \bullet d\bar{A} = \iiint \bar{\nabla} \bullet \bar{B}.dv = 0 \qquad (9\text{-}161)$$

Hence:
$$\bar{\nabla} \bullet \bar{B} = 0 \qquad (9\text{-}162)$$

Current Density and Ohm's Law

Whenever a force field exists, a work field W can be associated with it. The work done by a force $\bar{F}$ in moving a test charge ΔQ_2 a distance $\overline{dl}$ in the electric field $\bar{E}$ generated by a charge Q_1 can be stated as:

$$\Delta W = \bar{F} \bullet \overline{dl}$$
$$\frac{\Delta W}{\Delta Q_2} = \frac{\bar{F}}{\Delta Q_2} \bullet \overline{dl} \qquad (9\text{-}163)$$

Then from equation (9-127):

$$\frac{\Delta W}{\Delta Q_2} = \bar{E} \bullet \overline{dl} \tag{9-164}$$

The work done per charge is defined as a scalar potential V. V is the difference in electric potential between 2 points. Hence:

$$V = -\bar{E} \bullet \overline{dl} \tag{9-165}$$

The negative sign arises because when a positive test charge is moved against (in the opposite direction to) the force field of a positive charge, the work done is positive. When moving a test charge from l_1 to l_2:

$$V = V_2 - V_1 = -\int_{l_1}^{l_2} \bar{E} \bullet \overline{dl} \tag{9-166}$$

If a charge is moved around any closed path, the final and initial positions coincide and no net work is done:

$$\oint \bar{E} \bullet \overline{dl} = 0 \tag{9-167}$$

Electric current I is defined as the amount of charge that flows through a point in a conductor in a given period of time. Ohm's law relates the current I flowing between 2 points in a conductor to the electric potential difference V:

$$V = I.R \tag{9-168}$$

where R is the electric resistance of the conductor between the 2 points. In order to express Ohm's law for a field point, the concept of *resistivity* ρ of the material is introduced. For a cross-sectional area A of a given conductor having length l:

$$\rho = \frac{R.A}{l} \tag{9-169}$$

Hence equation (9-168) can be rewritten as:

$$V = I.R = I\frac{\rho.l}{A}$$
$$\frac{V}{l} = \rho\frac{I}{A} \tag{9-170}$$

The charge is flowing in the same direction as the force field. Therefore, for small dimensions of the material:

$$E = \rho.J \tag{9-171}$$

where J is the current density $\frac{I}{A}$. Usually it is convenient to use the electrical conductivity σ, which is the reciprocal of ρ, i.e. $\sigma = \frac{1}{\rho}$. For homogeneous and isotropic materials, Ohm's law in vector form can then be expressed as:

$$\bar{J} = \sigma\bar{E} \tag{9-172}$$

Now the current passing dI through an area dA has density J. The following relationship represents the general condition where the flow of current is not necessarily normal to dA:

$$\begin{aligned} dI &= \bar{J} \bullet d\bar{A} \\ I &= \iint \bar{J} \bullet d\bar{A} \end{aligned} \tag{9-173}$$

Magnetic Field Generated by a Moving Charge

In the study of electric and magnetic fields, experiments have shown that the electric field $\bar{E}$ of a point charge q at a field point located at a distance R from the charge is proportional to q and to $\frac{1}{R^2}$. For positive q the direction of E is along the line from source point to field point. The magnetic field $\bar{B}$ of a point charge q moving with velocity $\bar{v}$ is also proportional to q and to $\frac{1}{R^2}$, as well as to $\bar{v}$. However, the direction is perpendicular to the plane containing the line from source point to field point and the velocity vector $\bar{v}$:

$$\bar{B} = k' \frac{q.\bar{v} \times \hat{R}}{R^2} \tag{9-174}$$

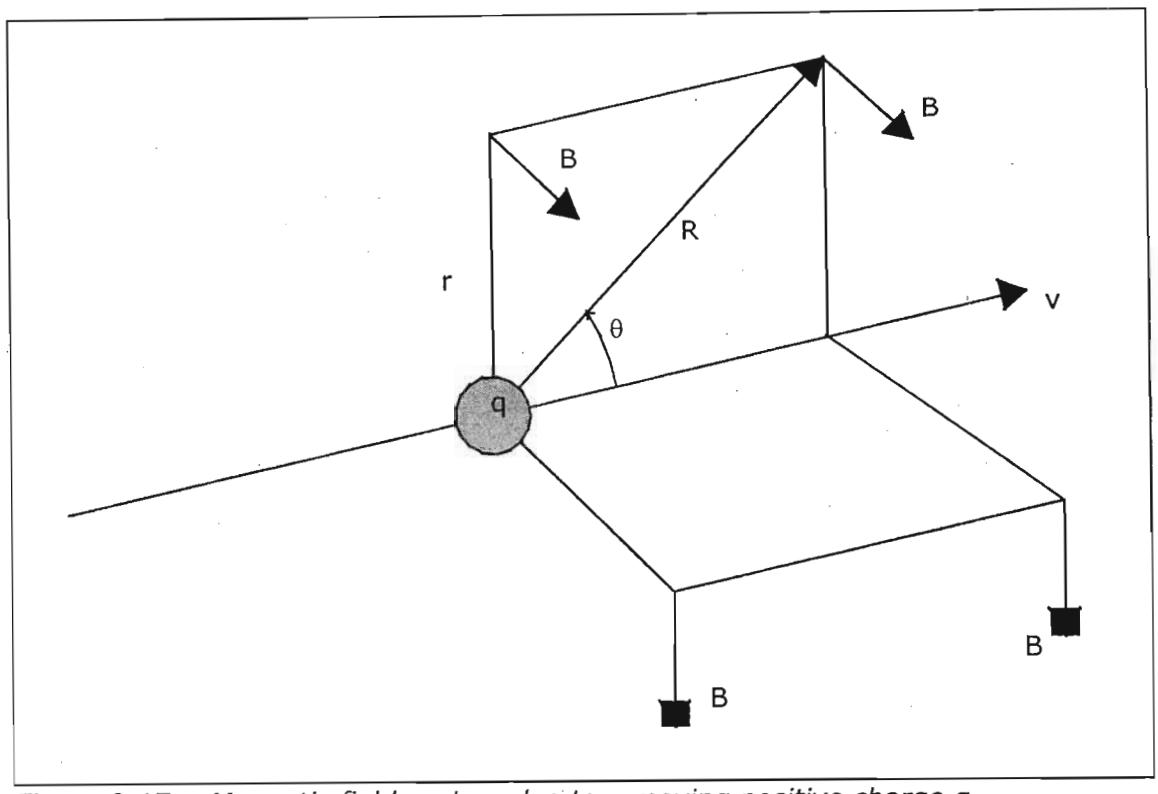


Figure 9.17: Magnetic field vectors due to a moving positive charge q.

In SI units, the numerical value of k' is arbitrarily defined as 10^{-7}. In equation (9-124), electric field relations are expressed in terms of constant k, where $k = \frac{1}{4\pi\varepsilon}$.

Similarly, in magnetic field relations:

$$k' = \frac{\mu}{4\pi} \tag{9-175}$$

Therefore:

$$\bar{B} = \frac{\mu}{4\pi}\frac{q\bar{v} \times \hat{R}}{R^2} \tag{9-176}$$

The Biot-Savart Law

Consider a conductor with uniform cross-sectional area A divided into short segments of length dl. The volume of the segment is $d\Omega = A.dL$. Suppose there are n charges q per unit volume; then the total moving charge dQ in the segment is:

$$dQ = n.q.A.dL \tag{9-177}$$

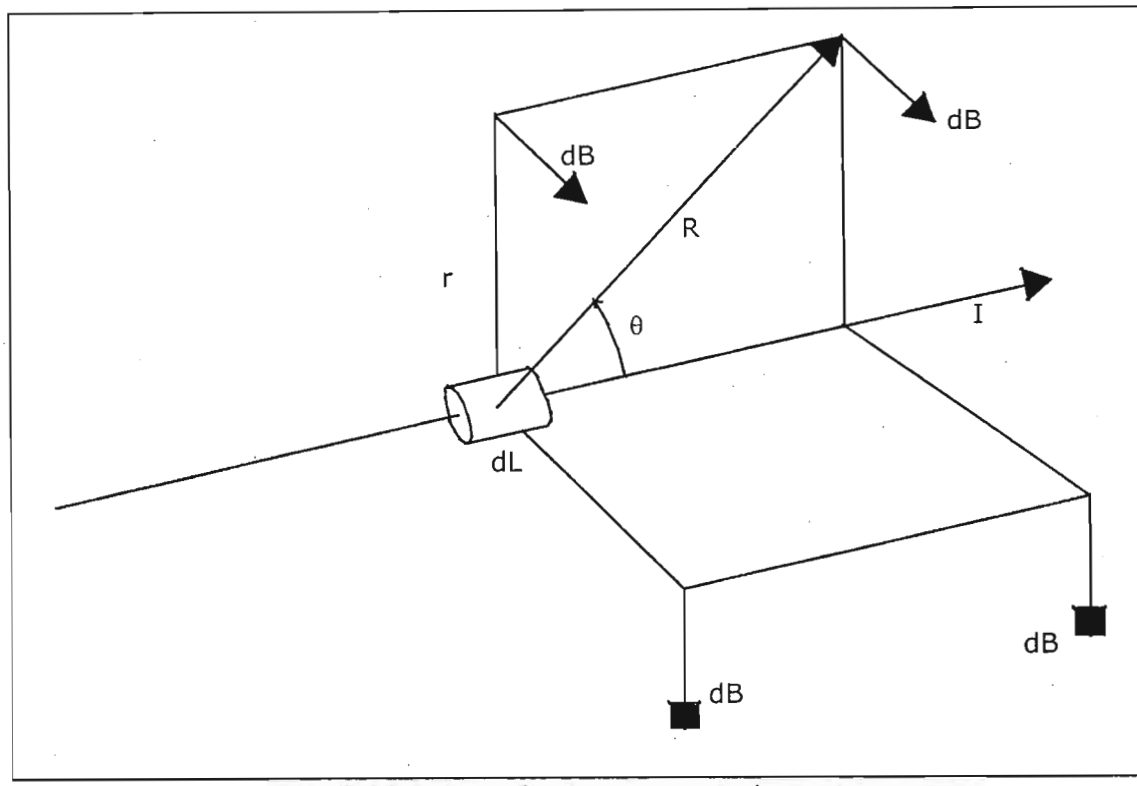


Figure 9.18: Magnetic field vectors due to a current element.

The moving charges q are therefore equivalent to a single charge dQ travelling at *drift* velocity v. From equation (9-176), the magnitude of the resulting field $d\bar{B}$ at any point is:

$$\begin{aligned} d\bar{B} &= \frac{\mu}{4\pi}\frac{dQ.\bar{v}\times\hat{R}}{R^2} \\ &= \frac{\mu}{4\pi}\frac{n.q.A.dL.\bar{v}\times\hat{R}}{R^2} \\ &= \frac{\mu}{4\pi}\frac{\bar{I}\times\hat{R}}{R^2}dL \end{aligned} \qquad (9\text{-}178)$$

This is the Biot-Savart law. The total magnetic field $\bar{B}$ at any point in space due to the current I in a complete circuit

$$\bar{B} = \frac{\mu}{4\pi}\int\frac{\bar{I}\times\hat{R}}{R^2}dL \qquad (9\text{-}179)$$

The magnitude of $d\bar{B}$ is:

$$dB = \frac{\mu}{4\pi}\frac{I.dL}{R^2}\sin\theta \qquad (9\text{-}180)$$

where θ is the angle between $\bar{I}$ and $\bar{R}$.

Ampere's Law

Ampere's law is a means of describing the relationship between a magnetic field $\overline{B}$ and its current-carrying sources. Consider for example that a current-carrying wire induces the magnetic field in Figure 9-19. Ampere's Law relates the *tangential* component of *magnetic* field at points on a *closed* curve to the current passing through the *area* bounded by that curve [55]. In general, $\overline{B}$ varies from point to point along the curve.

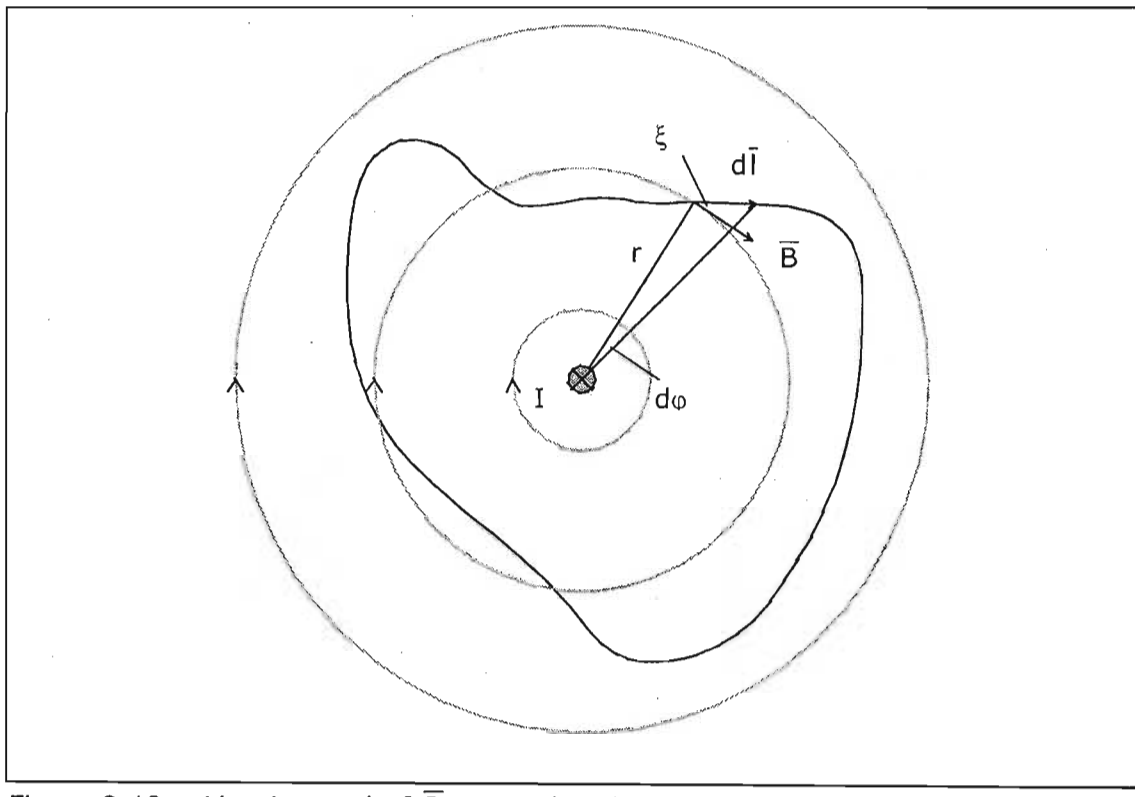


Figure 9.19: Line integral of $\overline{B}$ around a closed path surrounding a long, straight conductor carrying current I

Ampere's Law is formulated in terms of the line integral of B around the closed path, denoted by:

$$\oint B_{||}.dl = \oint \overline{B} \bullet \overline{dl} \qquad (9\text{-}181)$$

From equation (9-180); and since from Figure 9.18 $R.\cos\theta = r$; and since from Figure 9.19 $dl.\cos\xi = r.d\varphi$:

$$\begin{aligned}
\bar{B} \bullet \bar{dl} &= \frac{\mu}{4\pi}\int \frac{\bar{I} \times \hat{R}}{R^2} dL \bullet \bar{dl} \\
&= \frac{\mu}{4\pi}\int_L \frac{I.\sin\theta.dL.dl.\cos\xi}{R^2} \\
&= \frac{\mu.I.dl.\cos\xi}{4\pi}\int_L \frac{\sin\theta.dL}{R^2} \\
&= \frac{\mu.I.r.d\varphi}{4\pi}\int_L \frac{\sin\theta.dL}{R^2} \\
&= \frac{\mu.I.r.d\varphi}{4\pi}\frac{2}{r^2} \\
&\frac{\mu.I}{2\pi r} d\varphi
\end{aligned} \tag{9-182}$$

Hence:

$$\begin{aligned}
\oint \bar{B} \bullet \bar{dl} &= \oint \frac{\mu.I}{2\pi r} d\varphi \\
&= \mu.I
\end{aligned} \tag{9-183}$$

This is Ampere's law and is valid for any magnetic field and current. From equation (9-174), equation (9-183) can be written as:

$$\oint \bar{B} \bullet \bar{dl} = \mu.I = \mu \iint \bar{J} \bullet d\bar{A} \tag{9-184}$$

From equation (9-121):

$$\oint \bar{H} \bullet \bar{dl} = \iint \bar{J} \bullet d\bar{A} \tag{9-185}$$

For the case where the closed path l does not enclose the current source in the plane, equation (9-183) will be a superposition of integrals:

$$\begin{aligned}
\oint \bar{B} \bullet \bar{dl} &= \frac{\mu.I}{2\pi r}\oint d\varphi \\
&= \frac{\mu.I}{2\pi r}\int_a^b d\varphi + \frac{\mu.I}{2\pi r}\int_b^a d\varphi \\
&= 0
\end{aligned} \tag{9-186}$$

Stokes' Theorem

Stokes' Theorem is analogous to Gauss' divergence theorem. Recall Gauss' Theorem from equation (9-135):

$$\iint_{A(v)} \bar{\Gamma} \bullet d\bar{A} = \iiint_{v(A)} \bar{\nabla} \bullet \bar{\Gamma}.dv$$

Stokes' theorem is:

$$\oint_{l(A)} \bar{\Gamma} \bullet \bar{dl} = \iint_{A(l)} \bar{\nabla} \times \bar{\Gamma} \bullet \bar{dA} = 0 \qquad (9\text{-}187)$$

For magnetic fields:

$$\oint_{l(A)} \bar{B} \bullet \bar{dl} = \iint_{A(l)} \bar{\nabla} \times \bar{B} \bullet \bar{dA} = 0 \qquad (9\text{-}188)$$

From equation (9-184):

$$\oint \bar{B} \bullet \bar{dl} = \mu.I = \mu \iint \bar{J} \bullet \bar{dA}$$

By applying Stokes' theorem described in equation (9-162) to the left hand side of equation (9-184):

$$\iint_{A(l)} \bar{\nabla} \times \bar{B} \bullet \bar{dA} = \mu \iint \bar{J} \bullet \bar{dA} \qquad (9\text{-}189)$$

Hence:

$$\bar{\nabla} \times \bar{B} = \mu.\bar{J} \qquad (9\text{-}190)$$

Or:

$$\bar{\nabla} \times \bar{H} = \bar{J} \qquad (9\text{-}191)$$

For the case where the closed path l does not enclose the current source in the plane:

$$\bar{\nabla} \times \bar{B} = \bar{\nabla} \times \bar{H} = 0 \qquad (9\text{-}192)$$

Maxwell's Equations for Magnetostatic Fields

The calculations performed in this Section have been used to derive the linear equations (principle of superposition applies) that relate the electromagnetic fields to their sources ρ (charge density) and $\bar{J}$ (current density). They form part of a system of equations known as Maxwell's equations. Maxwell's equations in differential and integral form can be summarised as shown in Table 9.1. Because this study is concerned only with time-invariant magnetic fields, the applicable

Maxwell's equations are reduced to those in Table 9.2. The derivation of the equations in Table 9.2 has been illustrated earlier in this Section.

Table 9.1: Maxwell's equations.

Differential Form	Integral Form
$\bar{\nabla} \times \bar{E} = -\frac{\partial \bar{B}}{\partial t}$	$\oint_{l(A)} \bar{E} \bullet \bar{dl} = -\iint_{A(l)} \frac{\partial \bar{B}}{\partial t} \bullet \bar{dA}$
$\bar{\nabla} \bullet \bar{D} = \rho$	$\iint_{A(v)} \bar{D} \bullet \bar{dA} = \iiint_{v(A)} \rho . dv$
$\bar{\nabla} \times \bar{H} = \bar{J} + \frac{\partial \bar{D}}{\partial t}$	$\oint_{l(A)} \bar{H} \bullet \bar{dl} = \iint_{A(l)} \bar{J} \bullet \bar{dA} + \iint_{A(l)} \frac{\partial \bar{D}}{\partial t} \bullet \bar{dA}$
$\bar{\nabla} \bullet \bar{B} = 0$	$\iint \bar{B} \bullet \bar{dA} = 0$

Table 9.2: Maxwell's equations for magnetostatic fields.

Differential Form	Integral Form
$\bar{\nabla} \times \bar{H} = \bar{J}$ (9-165)	$\oint_{l(A)} \bar{H} \bullet \bar{dl} = \iint_{A(l)} \bar{J} \bullet \bar{dA}$ (9-158)
$\bar{\nabla} \bullet \bar{B} = 0$ (9-137)	$\iint \bar{B} \bullet \bar{dA} = 0$ (9-119)

Material Electromagnetic Constitutive Equations

Maxwell's equations are indefinite unless the constitutive equations are known. For simple media that is linear, isotropic and homogeneous, the constitutive equations are:

$$\bar{D} = \varepsilon \bar{E} \qquad (9\text{-}130)$$

$$\bar{B} = \mu \bar{H} \qquad (9\text{-}121)$$

$$\bar{J} = \sigma \bar{E} \qquad (9\text{-}172)$$

For this study, the equation that relates to magnetostatic fields caused by the magnetisation of TRIP steel is (9-121). Because the permeability of a ferromagnetic material is nonlinear:

$$\mu = \frac{\bar{B}(H)}{\bar{H}} \qquad (9\text{-}193)$$

Poisson and Laplace Equations

The potential $\bar{P}$ for the magnetostatic flux density $\bar{B}$ is related to $\bar{B}$ as follows:

$$\bar{B} = \bar{\nabla} \times \bar{P} \tag{9-194}$$

Substituting equation (9-194) into equation (9-190):

$$\bar{\nabla} \times \bar{\nabla} \times \bar{P} = \mu.\bar{J} \tag{9-195}$$

Applying the vector identity:

$$\bar{\nabla} \times \bar{\nabla} \times \bar{P} = \bar{\nabla}\left(\bar{\nabla} \bullet \bar{P}\right) = \bar{\nabla}^2 \bar{P} \tag{9-196}$$

to equation (9-195):

$$\bar{\nabla}^2 \bar{P} = -\mu \bar{J} \tag{9-197}$$

This is Poisson's equation for magnetostatic fields. For the condition where no current exists (i.e. equation (9-192)), $\bar{J} = 0$ and equation (9-197) becomes the Laplace equation:

$$\bar{\nabla}^2 \bar{P} = \bar{0} \tag{9-198}$$

Therefore, in a Cartesian coordinate system:

$$\frac{\partial^2 \bar{P}}{\partial x^2} + \frac{\partial^2 \bar{P}}{\partial y^2} + \frac{\partial^2 \bar{P}}{\partial z^2} = 0 \tag{9-199}$$

9.5.2. FEA of Magnetostatic Fields

Energy Stored in a Magnetic Field

The energy stored in a magnetic field in a medium exhibiting linear permeability is:

$$\begin{aligned} W &= \int_{\Omega} \frac{\bar{B} \bullet \bar{H}}{2} d\Omega \\ &= \frac{1}{2} \int_{\Omega} \bar{\nabla} \times \bar{P} \bullet \mu.\bar{\nabla} \times \bar{P}.d\Omega \end{aligned} \tag{9-200}$$

Generalised Workflow to Set Up and Solve the FEA Equations for Magnetostatic Fields

A similar 4-step process to the 5-step process described in Section 9.4.4 applies. This Section is covered more briefly since the derivation of the systems of equations is similar to that of the deformation analysis.

1. Select a suitable element type and define a local natural (dimensionless) coordinate system.

   The same mesh is used as that for the deformation calculations (see Figure 9.7).

2. Assume an interpolation function in terms of the local element coordinate system to represent the governing equations.

   An approximation is required for the potential $\bar{P} \approx \{P^e\}$ within each discrete element. Again utilising Lagrange's interpolation formula, the magnetic field potential at any point within a particular element in a specified direction (depending on the global coordinate system selected) is calculated using a shape function [N]. The shape function relates the field values to those at that element's nodal values. The same shape function used for the physical analysis is suitable. For 3-node triangular axisymmetric problems, [N] = [N(a,b)] has been developed.

$$P_r^e = \sum_{i=1}^{m} N_i(a,b).P_{ri}^e$$

$$P_\theta^e = \sum_{i=1}^{m} N_i(a,b).P_{\theta i}^e \qquad (9\text{-}201)$$

$$P_z^e = \sum_{i=1}^{m} N_i(a,b).P_{zi}^e$$

   Combining the above equations:

$$\{P^e\} = [N(a,b)]\{P_i^e\} \qquad (9\text{-}202)$$

   For isoparametric problems, the derivatives of the approximated field values in terms of the local coordinate system can be calculated and used to find the derivatives of the field values in terms of the global coordinate system.

$$r = \sum_{i=1}^{m} N_i(a,b).r_i$$

$$\theta = \sum_{i=1}^{m} N_i(a,b).\theta_i \qquad (9\text{-}203)$$

$$z = \sum_{i=1}^{m} N_i(a,b).z_i$$

3. The stored field energy density for each element can be found using the cross product $\overline{\nabla} \times \overline{P} \approx \overline{\nabla} \times [N]\{P^e\}$. The following equations apply to an axisymmetric problem in a cylindrical coordinate system.

   In the radial direction:

$$\begin{Bmatrix} \frac{\partial P_r^e}{\partial a} \\ \frac{\partial P_r^e}{\partial b} \end{Bmatrix} = \begin{bmatrix} \frac{\partial r}{\partial a} & \frac{\partial z}{\partial a} \\ \frac{\partial r}{\partial b} & \frac{\partial z}{\partial b} \end{bmatrix} \begin{Bmatrix} \frac{\partial P_r^e}{\partial r} \\ \frac{\partial P_r^e}{\partial z} \end{Bmatrix} \qquad (9\text{-}204)$$

$$\begin{Bmatrix} \frac{\partial P_r^e}{\partial r} \\ \frac{\partial P_r^e}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 & z_1 - z_2 \\ r_3 - r_2 & r_1 - r_3 & r_2 - r_1 \end{bmatrix} \begin{Bmatrix} P_{r1}^e \\ P_{r2}^e \\ P_{r3}^e \end{Bmatrix} \qquad (9\text{-}205)$$

   The derivatives of the potential in the hoop direction of an axisymmetric problem equal zero. In the axial direction:

$$\begin{Bmatrix} \frac{\partial w}{\partial r} \\ \frac{\partial w}{\partial z} \end{Bmatrix} = \frac{1}{|J|} \begin{bmatrix} z_2 - z_3 & z_3 - z_1 & z_1 - z_2 \\ r_3 - r_2 & r_1 - r_3 & r_2 - r_1 \end{bmatrix} \begin{Bmatrix} P_{z1}^e \\ P_{z2}^e \\ P_{z3}^e \end{Bmatrix} \qquad (9\text{-}206)$$

   Equations (9-204), (9-205) and (9-206) are used in the derivation of the cross product $\overline{\nabla} \times \overline{P} \approx \overline{\nabla} \times [N]\{P^e\}$:

$$\overline{\nabla} \times [N]\{P^e\} = \begin{vmatrix} \hat{r} & \hat{\theta} & \hat{z} \\ \frac{\partial}{\partial r} & \frac{\partial}{\partial \theta} & \frac{\partial}{\partial z} \\ P_r^e & P_\theta^e & P_z^e \end{vmatrix} \qquad (9\text{-}207)$$

   From equation (9-200), stored field energy W in each element:

$$W^e = \frac{1}{2}\int_\Omega \bar{\nabla} \times [N]\{P^e\} \bullet \mu.\bar{\nabla} \times [N]\{P^e\} d\Omega \tag{9-208}$$

4. Assemble the element equations for the total potential energy of the system in terms of the global coordinate system and solve using variational techniques.

$$\{P_i\} = \begin{Bmatrix} P_{1,r} \\ P_{1,\theta} \\ P_{1,z} \\ P_{2,r} \\ P_{2,\theta} \\ P_{2,z} \\ \vdots \\ P_{n,r} \\ P_{n,\theta} \\ P_{n,z} \end{Bmatrix} = \begin{Bmatrix} P_{1,1} \\ P_{1,2} \\ P_{1,3} \\ P_{2,1} \\ P_{2,2} \\ P_{2,3} \\ \vdots \\ P_{n,1} \\ P_{n,2} \\ P_{n,3} \end{Bmatrix} = \begin{Bmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \\ \vdots \\ P_{3n} \end{Bmatrix} \tag{9-209}$$

Total approximated $\bar{P}$-field potential energy Π is the vector sum of the stored field energy W and the driving energy for initiating the magnetism in the system $S \approx -\{f_i\}\{P_i\}$. In a linear system (constant μ):

$$\Pi = W + S = \frac{1}{2}\{P_i\}^T[K]\{P_i\} - \{f_i\}\{P_i\} \tag{9-210}$$

Using variational calculus to calculate the nodal values of {p} that minimise Π:

$$\frac{\partial \Pi}{\partial P_i} = K.P_i - f_i = 0 \tag{9-211}$$

For TRIP steels the magnetic field potential is "triggered" by the transformation of paramagnetic austenite to ferromagnetic martensite as a result of applied strain. The Olson and Cohen model describing the nucleation of the martensite phase may be used to characterise the kinetics of the strain-induced transformation reaction in terms of the temperature and plastic strain (refer to Section 6.2.4).

For ferromagnetic materials, the permeability μ in equation (9-208) is a nonlinear function of $\bar{H}$ that must be experimentally measured and iteratively updated in conjunction with the physical state of each element (paramagnetic or ferromagnetic, as determined by an iterative nonlinear static analysis of the quasi-static deformation) in the solution of equations (9-24) and (9-211).

Boundary Conditions

There are 3 categories of boundary conditions.

1. Dirichlet. Here, the value of the vector potential $\bar{P}$ is explicitly defined on the boundary. The most common use is to apply a zero-value to keep flux from crossing the boundary. This is done for example to simulate symmetry conditions.

2. Neuman. The normal derivative of $\bar{P}$ is specified along the boundary. It is usually applied as a zero-value to force flux to cross the boundary in the normal direction (e.g. $\frac{\partial P}{\partial \theta} = 0$). This facilitates the modelling of interfaces between materials with widely differing permeabilities.

3. Mixed. A combination of Dirichlet and Neuman boundary conditions is used to prescribe a relationship between the value of the vector potential $\bar{P}$ and its normal derivative at the boundary (i.e. $c_0 \frac{\partial P}{\partial n} + c_1 \bar{P} = 0$).

Open Boundary Conditions

A difficulty encountered when applying the finite element method to the magnetostatic problem arises when attempting to realistically represent the boundary condition where $\bar{P} = 0$ for $r = \infty$. The simplest is to define a boundary sufficiently far enough away from the area of interest. Otherwise, the mixed boundary condition can be utilised.

10. DISCUSSION

The generalised aim of this study is to identify a potential solution to the problem of accurately assessing the health state of a virtually inaccessible structural component where and when required.

In Chapter 2 a description of the problem is presented. In the mining industry, excavated tunnels interfere with the local environment stability and are in effect stress raisers of the local rockmass. Excavations typically experience some combination of quasi-static and seismic loading that threatens the structural integrity of the tunnel. Rockfalls and rockbursts result when the support system fails to adequately contain the rockmass. Designers of tunnel support systems currently rely on (and are limited to) simplified mechanistic evaluation, numerical modelling, and generalised empirical or semi-empirical design methodologies that are based on studies of a wide range of environmental conditions. Due to the diverse nature of load and rockmass conditions/behaviour, it is extremely difficult to predict or accurately assess the condition of the support system over time. The cost and safety implications of this uncertainty are enormous. Associated problems are that the support units cannot be removed for testing without damaging the rockwall, nor can they be tested in-place using currently-available techniques.

The requirement is therefore to identify a suitable structural health monitoring system capable of quantifying the health state (or damaged state) of the support units at an arbitrary location within the network of tunnels at any given time. This would supplement the existing design methodologies in addition to providing an insight into a given region's rockmass behaviour and the response of the installed support units. The damage assessment should be made without affecting the operational performance of the support system.

Chapter 3 provides a broad overview of available structural health monitoring (SHM) systems. With *smart* SHM, damage assessment is usually based on the change in data measured at the same location at different times. Smart systems therefore provide in-service, real-time monitoring with minimal human involvement (reduces human error). Automated data acquisition, reporting and analysis is available. Smart SHM systems comprise of hardware and software and address 5 criteria: they contain a sensor (or sensor array), appropriate signal generation instrumentation, signal processing equipment, and signal identification and interpretation capabilities; finally, they are integrated into the structure to be monitored as well as into the business procedural practices. By utilising a suitable

health monitoring system in a given application, the support unit need not be removed from service until it is deemed to have reached the end of its working life. Cost-effective and safe decisions can be made based on an analysis of the quantitative data relating to the severity and propagation rate of the damage.

Smart materials react to changing environmental conditions in a consistent and measurable fashion. Two classes of smart material behaviour can be defined; the first is suited to *passive* SHM systems, the second to *active* SHM systems. Passive smart sensors exhibit an inherent and permanent property change that records the peak value of the environmental variable without the need for a power source. Active systems measure the environmental variable as a function of time, but do not inherently record any information; removing the power supply resets the active sensor to its reference value.

A passive system is more suited to the rockbolt problem, primarily because deformation due to rockmass slip is always at the peak value in the load history. Passive smart materials can also be used simultaneously as structural and sensor element. In fact, TRIP steel strain memory alloys exhibit far superior strength, ductility and corrosion resistance compared to existing units, so structural performance would in fact be enhanced by their implementation. Appendix A contains a material survey of smart materials that may find use in alternative applications.

The sensing property of TRIP steel is an increase in measurable *ferromagnetism* as a function of the applied *strain* environmental variable. The increase in magnetic field strength is the result of a microstructural phase transformation from metastable paramagnetic austenite to ferromagnetic martensite.

Chapters 4 and 5 explain the dependence of material properties on crystal microstructure, which is manipulated by suitable heat treatment through specific temperature ranges at controlled rates, and by adjusting the alloy composition. In carbon steels, heat treatments are used to increase the hardness of the material, typically by controlling the austenite-ferrite and austenite-martensite transformations. In TRIP steels, however, the requirement for a metastable austenite parent phase at room temperature implies that normal heat treating procedures do not apply. The description of common heat treatments is therefore included as Appendix 2.

Typically, phase transformations, including the transition from metastable austenite to martensite, are *temperature*-dependent. Chapter 6 describes the mechanism

whereby this reaction becomes *strain*-dependent. In TRIP steels, a so-called M_D temperature exists (above room temperature) below which cold work (plastic deformation) increases the temperature required to initiate the austenite-martensite reaction M_s (initially below room temperature). At some incubation strain magnitude, the martensitic start temperature reaches a temperature just above ambient (superheated) and the transformation proceeds as a function of strain at constant temperature. The transformation is irreversible, so that transformed locations are plastic. In general, the higher the alloying element content, the lower the M_s and M_D temperatures. In the absence of cold working, the M_s temperature can be approximated using empirical equations:

$$M_s(^\circ C) = 539 - 423(\%C) - 30.4(\%Mn) - 17.7(\%Ni) - 12.1(\%Cr) - 7.5(\%Mo) \quad (6\text{-}1)$$

As deformation of the metastable austenitic progresses, the strained areas transform from austenite to the stronger martensite. The transformation is inhomogeneous because local weaker areas transform first. Necking in a test specimen is therefore delayed until the entire gauge length and beyond has transformed to the martensitic phase. Once the transformation progresses to a significant extent along the entire gauge length, final fracture occurs when the UTS of the martensite is exceeded. Failure typically occurs with very little or no necking.

The Olson and Cohen model describing the nucleation of the martensite phase may be used to characterise the kinetics of the strain-induced transformation reaction in terms of the temperature and plastic strain.

Chapter 7 describes the influence of alloy content on manipulating the TRIP steel microstructure at ambient temperatures in order to facilitate the strain memory effect. Other material properties are also controlled in this manner. TRIP steels typically exhibit open- γ-fields on their phase diagrams, which are promoted by the addition of elements such as Ni and Mn that are referred to as *austenitic stabilisers*.

The Fe-Cr-Mn group of austenitic stainless steel have an advantage over the more expensive Fe-Cr-Ni group in that they reduce the sensitivity of the transformation to changes in the ambient temperature (common in strain memory alloys). This temperature dependence is minimised by the initial formation of a hexagonally close-packed ε-martensite phase (which like austenite is paramagnetic). The

subsequent nucleation of ferromagnetic BCT α'-martensite from within HCP ε-martensite is a function of the applied strain only.

An important characteristic of TRIP steels is their excellent response to thermo-mechanical processing. Elements such as Mn that lower the stacking fault energy promote the work-hardenability. However, it is the transformation to martensite that undoubtedly has the largest effect on work hardening rates.

It is possible to create a metastable austenitic bay in the material's TTT curve if certain alloying elements, notably chromium, are added. Work-hardening processes in this bay (referred to as *prior deformation of austenite*) dramatically increase the strength and ductility of TRIP steels. By allowing the formation of some carbides on cooling after work-hardening, the strength is higher than that obtained through PDA and ductility is high because of the partial transformation to martensite during deformation. However, this *ausformed* material is not suitable for sensitive strain memory requirements. By cycling repetitively between a martensitic and reverted austenite structure in a process referred to as the Koppenaal alternative, strengthening occurs without affecting the strain memory.

Chapter 8 investigates the suitability of TRIP steels to smart SHM applications in general, and specifically to smart mining bolts. As mentioned above, TRIP steel exhibit far superior strength, ductility and corrosion resistance compared to current rockbolt units. Replacing current units with self-sensing TRIP steel units would in fact enhance the structural performance of the support unit.

A TRIP steel SHM system would comprise a TRIP steel sensor that produces a magnetic field as the diagnostic data. A *ferromagnetic sensor* (e.g. a Hall sensor) produces a localised measurement of the magnetic field at the location where the sensor is positioned. This location is virtually impossible to predict in mining bolts. When taking measurements from TRIP steels, the Hall sensor will register stepped measurements for small strains of generally below 2%. An alternative *induction coil* measurement arrangement produces a reading that reflects the overall behaviour of the deforming gauge length (i.e. an "averaged" response). The magnetic field strength is then plotted on a curve corresponding to the TRIP material that measures magnetic field strength as a function of applied strain. If the magnetic field is affected by e.g. temperature variations, these variations must be taken into account to improve the accuracy of the results. The instrumentation as well as the bolts themselves would need to be integrated into the support

structure; the information obtained would require integration into the existing design, inspection and maintenance practices.

In Chapter 9 it was shown that numerical FEA techniques are available to analyse both the deformation of the rockbolt under specified rockmass loading conditions as well as the magnetic field that results from the induced strain. Numerical models assist in investigating the response of the support system to a variety of environmental conditions.

11. CONCLUSION

The aim of this study is to identify a suitable smart structural health monitoring system to assess the cumulative damage experienced by rockbolt units in mining tunnels.

TRIP steels exhibit the high strength, ductility and corrosion-resistance required to replace the existing units and thereby have the capability to act simultaneously as sensor and structural component. Thermo-mechanical processes have been developed to enhance the TRIP steel material properties and to reduce the sensitivity to changes in temperature and other environmental conditions.

Mature numerical techniques are available to supplement the data obtained from in-situ components by extrapolating data to a wide range of potentially hazardous environmental conditions.

As with other controlled transformation steels, TRIP steels require extremely good metallurgical control and are very expensive to make. Costs would be reduced should significant amounts of e.g. TRIP rockbolts be manufactured for commercial use. These costs would be significantly offset by the improvement of the inspection data available. The benefit to the safety of mine personnel would be immeasurable.

12. APPENDIX A: SMART MATERIAL SURVEY

In this investigation, attention is focused on a passive smart SHM system using transformation-induced plasticity (TRIP) steel. The decision to select TRIP steel as the preferred solution could only be made once a material survey had been conducted into a wide range of potential candidates. This Chapter discusses the merits and disadvantages of some of these alternative smart materials.

12.1. PIEZOELECTRICITY

12.1.1. Piezoelectric Behaviour

Refer to [62], [68] - [70]. When a piezoelectric element is stressed electrically by a voltage, its dimensions change. When stressed mechanically by a force, it generates an electric charge; if the electrodes are not short-circuited, a voltage associated with the charge appears. A piezoelectric element is therefore capable of acting as a sensing or transmitting element, or both.

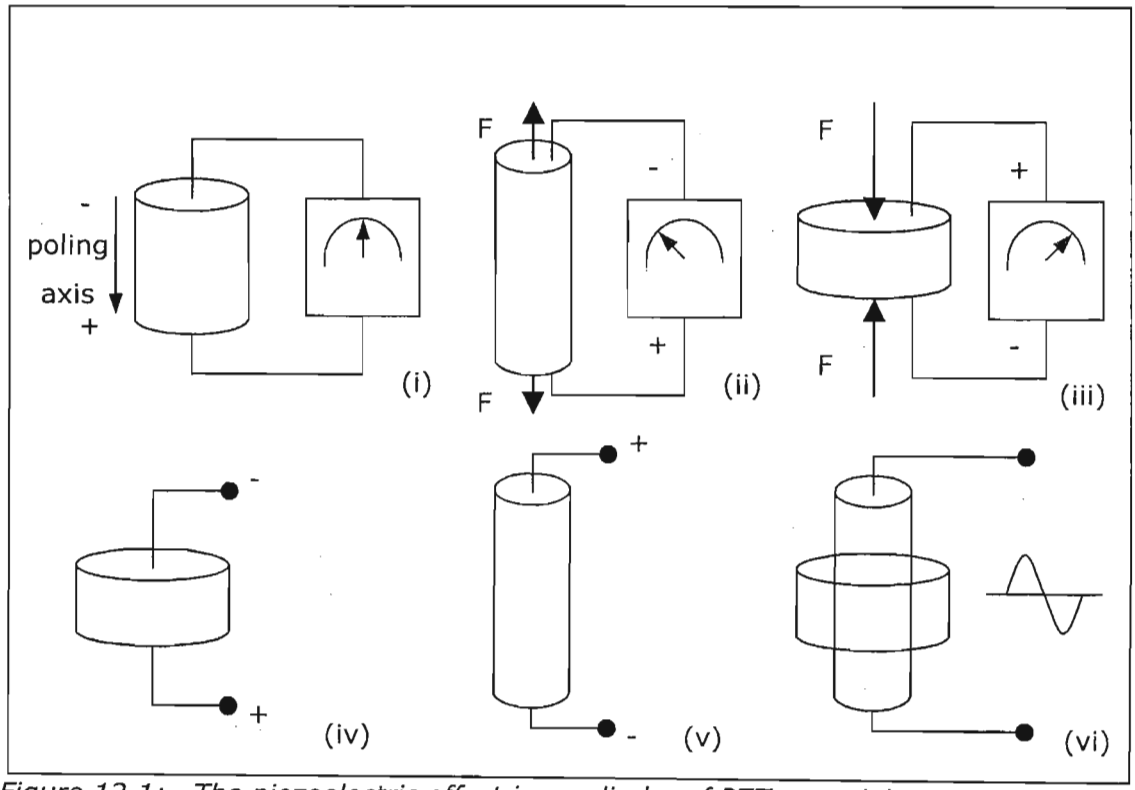


Figure 12.1: The piezoelectric effect in a cylinder of PZT material

If an external force produces compressive or tensile strain, a voltage appears between the electrodes. In tension, the voltage across the electrodes will have opposite polarity to the poling voltage (Figure 13.1(ii)). In compression, the voltage will have the same polarity as the poling voltage (Figure 13.1(iii)). These

are examples of generator action, the conversion of mechanical energy into electrical energy. Generator action can be found in cigarette and gas lighters, gramophone pick-ups, accelerometers, hydrophones and microphones.

If a voltage of opposite polarity to the poling voltage is applied to the electrodes (Figure 13.1(iv)), the cylinder will shorten. If the applied voltage has the same polarity as the poling voltage, the cylinder will lengthen (Figure 13.1(v)). If an alternating voltage is applied (Figure 13.1(vi)), the cylinder will grow and shrink at the same frequency as that of the applied voltage. These are examples of motor or actuator action, the conversion of electrical energy into mechanical energy. Motor action is found in transducers for ultrasonic cleaning equipment, ultrasonic atomisers, fuel injection systems and piezoelectric motors.

12.1.2. History

In 1880, Jacques and Pierre Curie discovered a surface electric charge on specially prepared crystals (notably quartz, tourmaline, topaz, cane sugar and Rochelle salt) that were subjected to mechanical stress. This effect was quickly dubbed as "piezoelectricity" in order to distinguish it from other areas of scientific phenomenon such as "contact electricity" (friction-generated static electricity) and "pyroelectricity" (electricity generated from crystals by heating).

In 1881, using mathematical deductions from fundamental thermodynamic principles, Lippmann predicted that these crystals would also exhibit the converse piezoelectric effect (stress in response to an applied electric field). The Curies immediately confirmed the existence of the "converse effect" and continued on to obtain quantitative proof of the complete reversibility of electro-elasto-mechanical deformations in piezoelectric crystals.

In the years leading up to 1910, the 20 natural crystal classes in which piezoelectric effects occur were defined, and all 18 possible macroscopic piezoelectric coefficients were determined using appropriate tensorial analysis.

In 1917, P. Langevin and French co-workers invented the first piezoelectric transducer, a steel-quartz sandwich with a resonant frequency of ~50kHz, for use as an ultrasonic submarine detector. Over the next 25 years, the success of sonar stimulated intense development activity, both resonating and non-resonating. A new class of materials testing methods was developed based on the propagation of ultrasonic waves. For the first time, elastic and viscous properties of liquids and gases could be measured with comparative ease, and previously invisible flaws in

solid metal structural members could be detected. Also, new ranges of transient pressure measurement were opened up, permitting the study of explosives and internal combustion engines, along with a host of other previously unmeasurable vibrations, accelerations and impacts. In fact, during this period, most of the classic applications were conceived and reduced to practice (microphones, accelerometers, ultrasonic transducers, bender element actuators, phonograph pick-ups, signal filters, etc.). However, available materials often limited device performance and commercial exploitation.

12.1.3. Piezoceramics

To obtain the desired results, natural crystals must be cut precisely. Synthetic piezoelectric materials called piezoceramics were developed during World War 2, when it was discovered that certain ceramic materials (prepared by sintering metallic oxide powders) exhibited dielectric constants up to 100 times higher than common cut crystals.

Since piezoceramic elements are capable of generating very high voltages, they are compatible with today's generation of solid-state devices – rugged, compact, reliable and efficient.

Piezoceramics are very versatile and can be manufactured in a variety of shapes and sizes. Physical, chemical and piezoelectric characteristics can be tailored to achieve desired results. They are chemically inert and immune to moisture and other atmospheric conditions.

12.1.4. Manufacture For Specific Response

Relationships between applied forces and the resultant responses depend upon the piezoelectric properties of the ceramic; the size and shape of the specimen; and the direction of the electrical and mechanical excitation.

To identify directions in an element, an orthogonal set of axes (1, 2, 3) is used (Figure 12.2). The polar axis (3^{rd} axis) is defined as parallel to the direction of polarisation within the ceramic. This direction is established during manufacturing by a high DC voltage that is applied above the Curie temperature (described later in this section) between a pair of electroded faces to activate the material. In shear operations, these poling electrodes are later removed and replaced by electrodes deposited on a second pair of faces. In this event, the polar axis is not altered (Figure 13.3).

Piezoelectric coefficients have double subscripts that link electrical and mechanical properties. The first subscript gives the direction of the electrical field associated with the applied voltage or the charge produced. The second subscript gives the direction of the mechanical stress or strain. Several piezoceramic material constants, discussed below, may be written with one of the following superscripts that specifies either a mechanical or electrical boundary condition:

T constant stress = mechanically free

E constant field = short circuit

D constant electrical displacement = open circuit

S constant strain = mechanically clamped

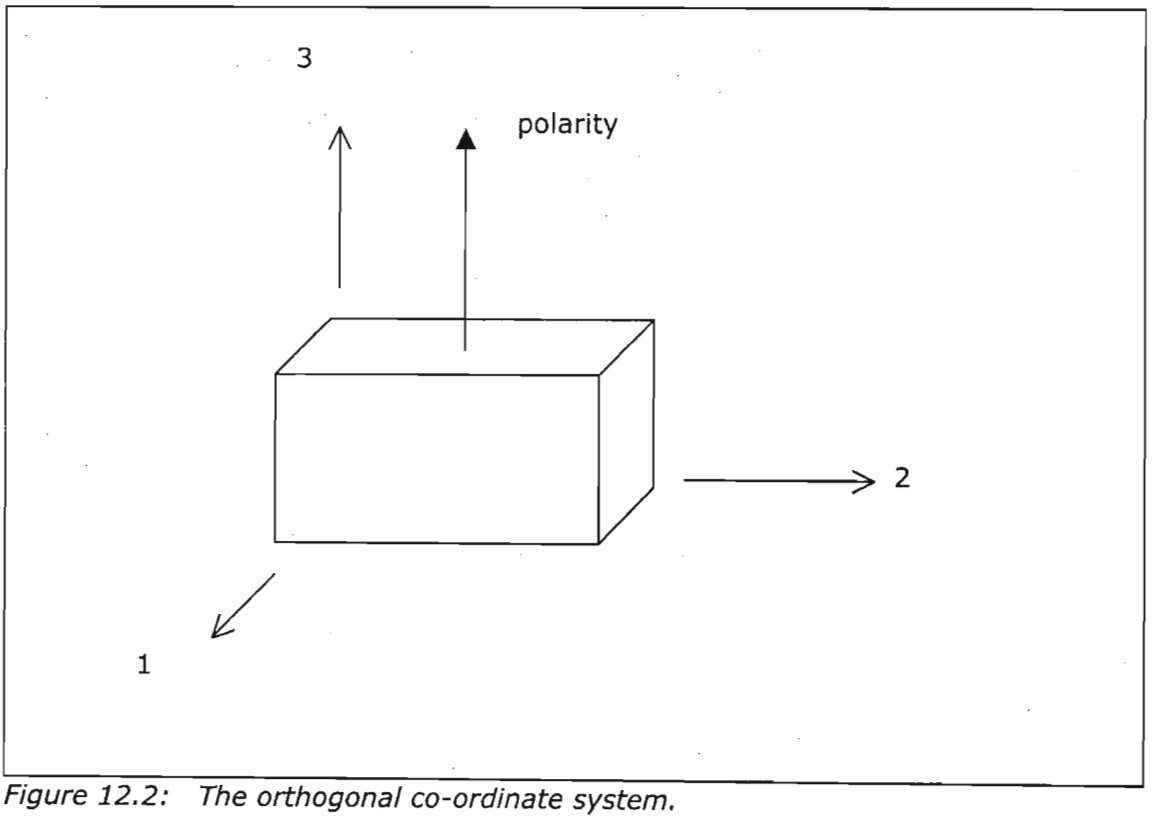


Figure 12.2: The orthogonal co-ordinate system.

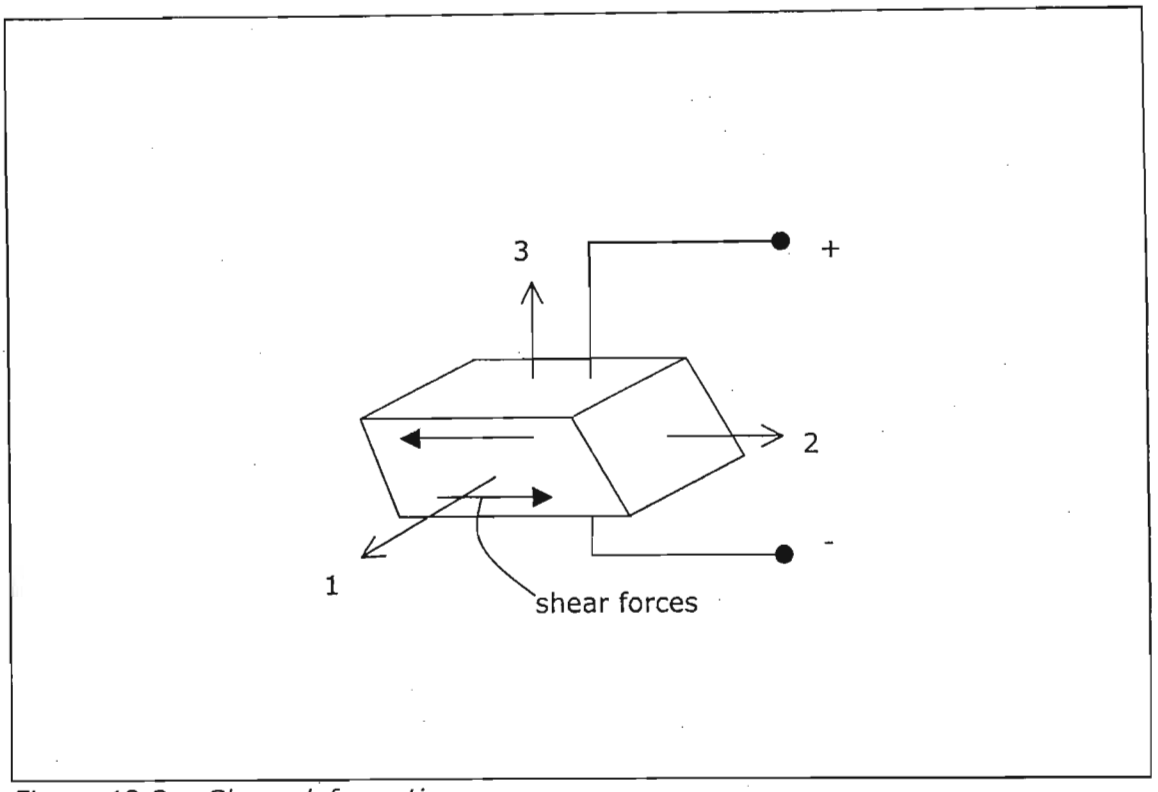


Figure 12.3: Shear deformation

The "d" Constant

The piezoelectric constants relating the mechanical strain to an electric field are termed the strain constants, or the "d" coefficients:

$$d = \frac{\text{strain developed (m/m)}}{\text{applied electric field (V/m)}} \qquad (13\text{-}1)$$

Large d_{ij} constants relate to large mechanical displacements, which are usually sought in motional transducer devices. d_{33} refers to a force in the 3 direction (along the polarisation axis) on the same surface on which the charge is collected, which in this case is also the polarisation axis. For d_{31} the charge is collected on the same surface, but the force is at right angles to the polarisation axis. d_{15} indicates that the charge is collected on electrodes that are at right angles to the original poling electrodes, and that the applied mechanical stress is shear.

The units for the d_{ij} coefficients are commonly expressed as coulombs/square metre per newtons/square metre:

$$d = \frac{\text{short circuit charge density (C/m}^2\text{)}}{\text{applied mechanical stress (N/m}^2\text{)}} \qquad (13\text{-}2)$$

This is useful when charge generators are contemplated, e.g. accelerometers.

The "g" Constant

The piezoelectric constants relating the electric field produced to a mechanical stress are termed the voltage constants, or the "g" coefficients:

$$g = \frac{\text{open circuit electric field (V/m)}}{\text{applied mechanical stress (N/m}^2\text{)}} \qquad (13\text{-}3)$$

Output voltage is obtained by multiplying the calculated electric field by the thickness of ceramic between electrodes.

g_{33} indicates that the electric field and the mechanical stress are both along the polarisation axis. g_{31} indicates that that the pressure is applied at right angles to the polarisation axis, but the voltage appears on the same electrodes as in the "33" case. g_{15} implies that the applied stress is shear and that the resulting electric field is perpendicular to the polarisation axis. High g_{ij} values favour large voltage output, popular for sensors.

Note also that :

$$g = \frac{\text{strain developed (m/m)}}{\text{applied charge density (C/m}^2\text{)}} \qquad (13\text{-}4)$$

Relative Dielectric Constant

The relative dielectric constant is the ratio of the permittivity of the material, ε, to the permittivity of free space, $\varepsilon_0 = 8.9*10^{-12}$ farads/metre, in the unconstrained condition (i.e. well below the mechanical resonance of the part):

$$K = \frac{\varepsilon}{\varepsilon_0} \qquad (13\text{-}5)$$

Capacitance

Whereas the relative dielectric constant is strictly a material property, the capacitance is dependent on the type of material and its dimensions.

$$C = \frac{K.\varepsilon_0.A}{t} \qquad (13\text{-}6)$$

A electrode surface area

t thickness separating the electrodes

At frequencies far below resonance, piezoelectric ceramic transducers are fundamentally capacitors. Consequently, the voltage coefficients g_{ij} are related to the charge coefficients d_{ij} by the dielectric constant K_{ii} as, in a capacitor, the voltage V is related to the charge Q by the capacitance C:

$$Q = C.V \tag{13-7}$$

K_{33} is related to the capacitance between the original poling electrodes. K_{11} is related to the capacitance between the second pair of electrodes applied after removal of the poling electrodes for the purposes of shear excitation.

$$\begin{aligned} d_{33} &= K^T_{33}.\varepsilon_0.g_{33} \\ d_{31} &= K^T_{33}.\varepsilon_0.g_{31} \\ d_{15} &= K^T_{11}.\varepsilon_0.g_{15} \end{aligned} \tag{13-8}$$

At resonance, the dielectric constant is reduced by the factor $(1-k^2)$ where k is the coupling coefficient of the mode in question.

Coupling Coefficients

Electromechanical coupling k_{33}, k_{31}, k_p and k_{15} describe the conversion of energy by the ceramic element from electrical to mechanical form or vice versa:

$$k = \sqrt{\frac{\text{mechanical energy stored}}{\text{electrical energy applied}}} \tag{13-9}$$

or

$$k = \sqrt{\frac{\text{electrical energy stored}}{\text{mechanical energy applied}}} \tag{13-10}$$

Subscripts denote the relative directions of electrical and mechanical properties and the type of motion involved. They can be associated with vibratory modes of certain simple transducer shapes. k_{33} is appropriate for a long, thin bar, electroded at both ends, polarised along the length, and vibrating in a simple length expansion and contraction. k_{13} relates to a long, thin bar, electroded on a pair of long faces, polarised in thickness, and vibrating in simple length expansion and contraction. k_p represents the coupling of electrical and mechanical energy in a thin round disc, polarised in thickness and vibrating in radial expansion and contraction. k_{15} describes the energy conversion in a thickness shear vibration.

Young's Modulus

Because mechanical stressing of the ceramic produces an electrical response that opposes the resultant strain, the effective Young's modulus with electrodes short-circuited is lower than with the electrodes open circuited. In addition, the stiffness is different in the 3 direction from that in the 1 or 2 direction. Therefore, in expressing such quantities, both direction and electrical conditions must be specified.

Y^E_{33} is the ratio of stress to strain in the 3 direction with electrodes shorted.

Y^D_{33} is the equivalent with electrodes open circuited.

Y^E_{11} and $Y^D_{11}$are the moduli in the 1 or 2 direction.

Y^E_{55} and Y^D_{55} are the ratios of shear stress to shear strain.

Dissipation Factor

A measure of the dielectric losses in the material, defined as the ratio of parallel resistance to parallel reactance, and expressed in percent.

Mechanical Q_m

The ratio of reactance to resistance in the equivalent series circuit representing the mechanical vibrating resonant system. The shape of the part affects the value.

Curie Temperature

The temperature at which the crystal structure changes from a non-symmetrical (non-piezoelectric) to a symmetrical (piezoelectric) form, expressed in degrees Celsius.

Ageing Rate

Ageing is the attempt of the ceramic to change back to its original state (i.e. prior to polarisation). Ageing is a logarithmic function with time. The ageing rate defines change in the material parameters per decade of time, i.e. 1-10 days, 5-50 days, etc.

Performance (Deflection and Force) and Mathematical relationships

Piezoelectric actuators are usually specified in terms of their free deflection and blocked force. Free deflection, X(f), refers to displacement at a given voltage level without the actuator working against any external load. Blocked force, F(b), refers to the force exerted at a given voltage level when the actuator is not allowed to move. Since the force at maximum deflection is zero, all other values of simultaneous displacement and force (for a given voltage level) are determined by a line drawn between these points on a force versus deflection graph.

In practice, a bending motor must move a specified amount and exert a specified force, which determines its operating point on the force versus deflection diagram. Work is maximised when the deflection performed permits one half the blocked force to be developed. This occurs when the deflection equals one half the free deflection.

For cantilevered bending motors, X(f) and F(b) are approximated by observing the tip deflection after energising, and by holding a force gauge against the tip during energisation.

The expressions listed in diagrams 13.4 and 13.5 represent the linear relationships measured at low voltages.

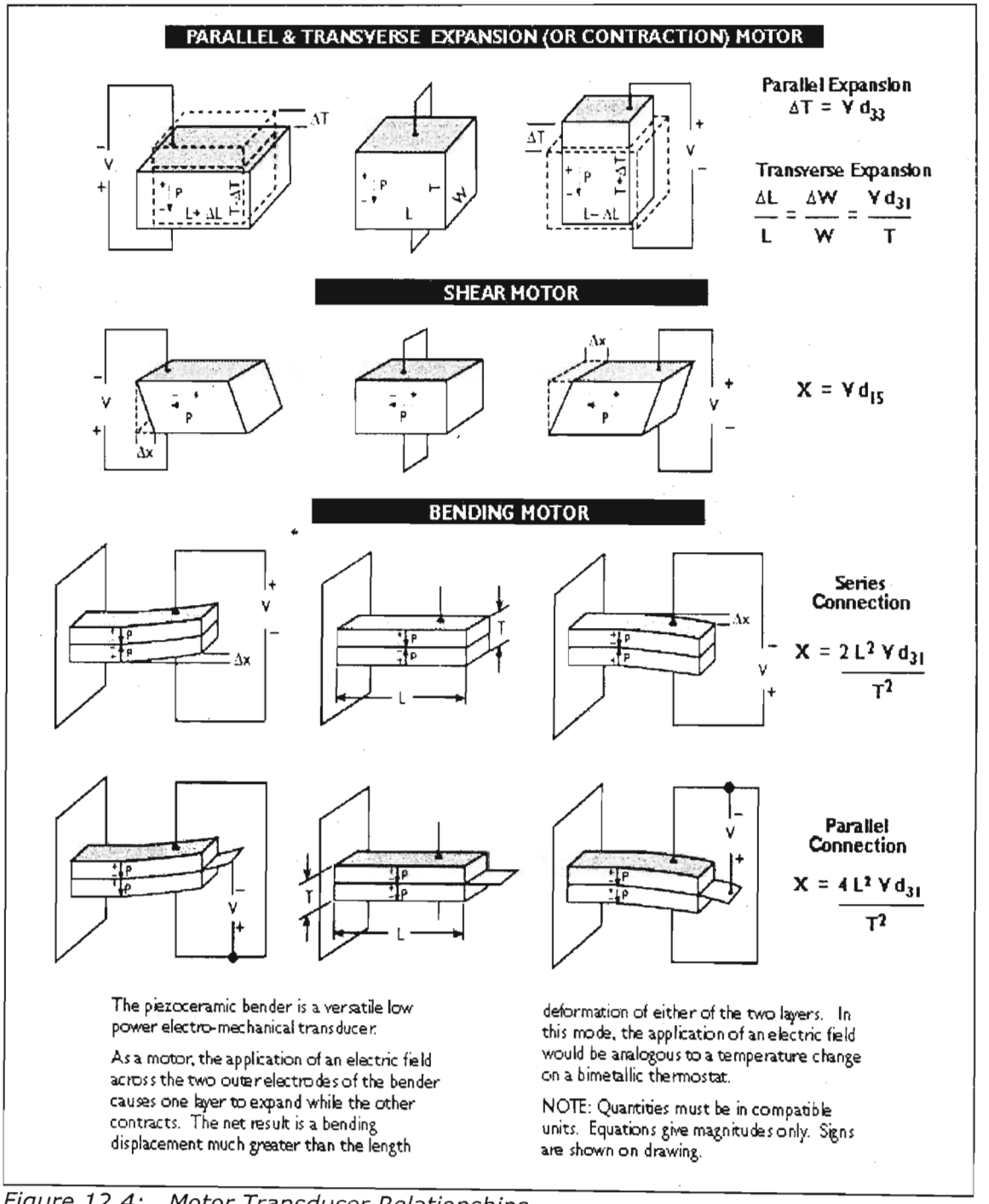


Figure 12.4: Motor Transducer Relationships.

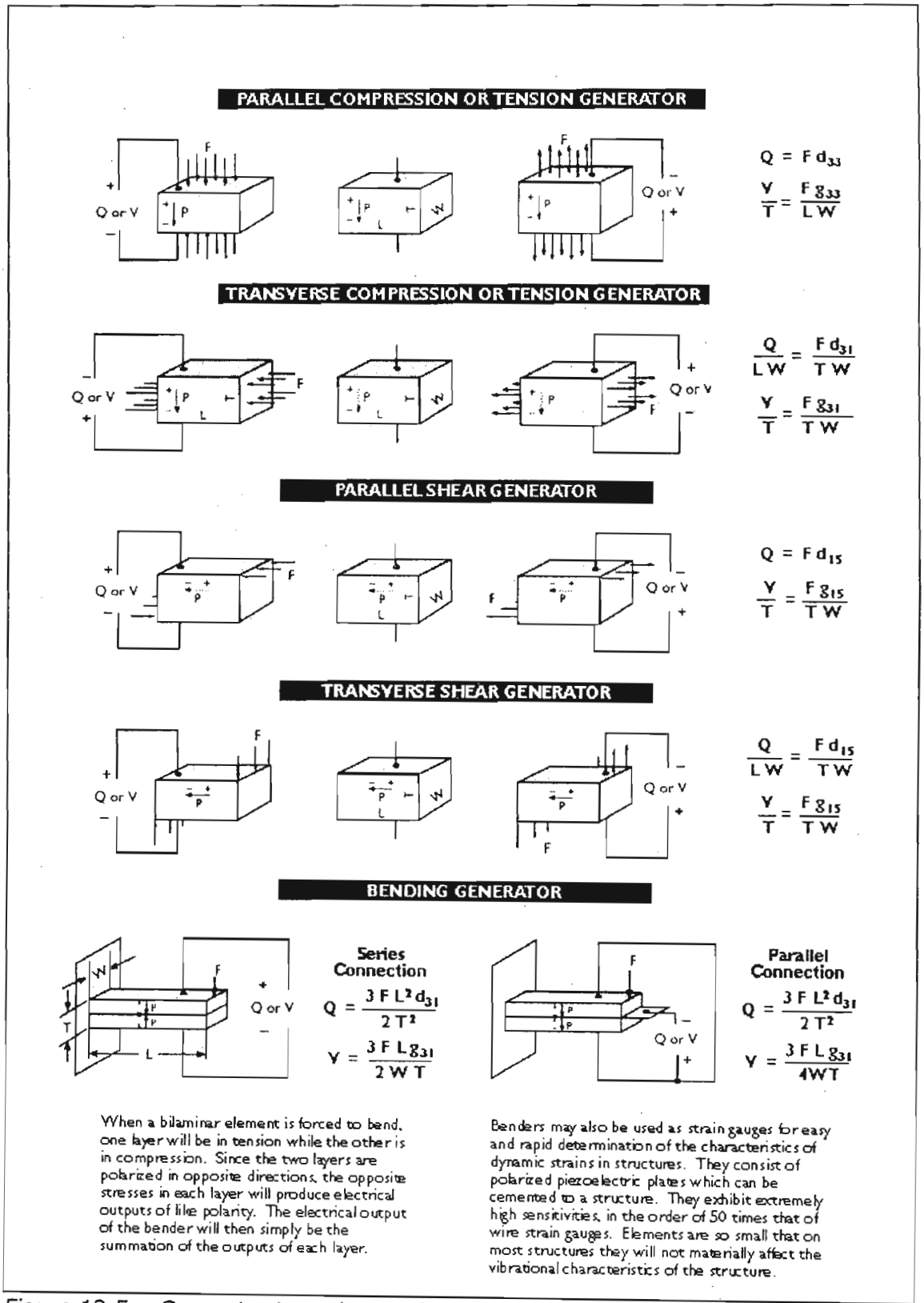


Figure 12.5: Generator transducer relationships.

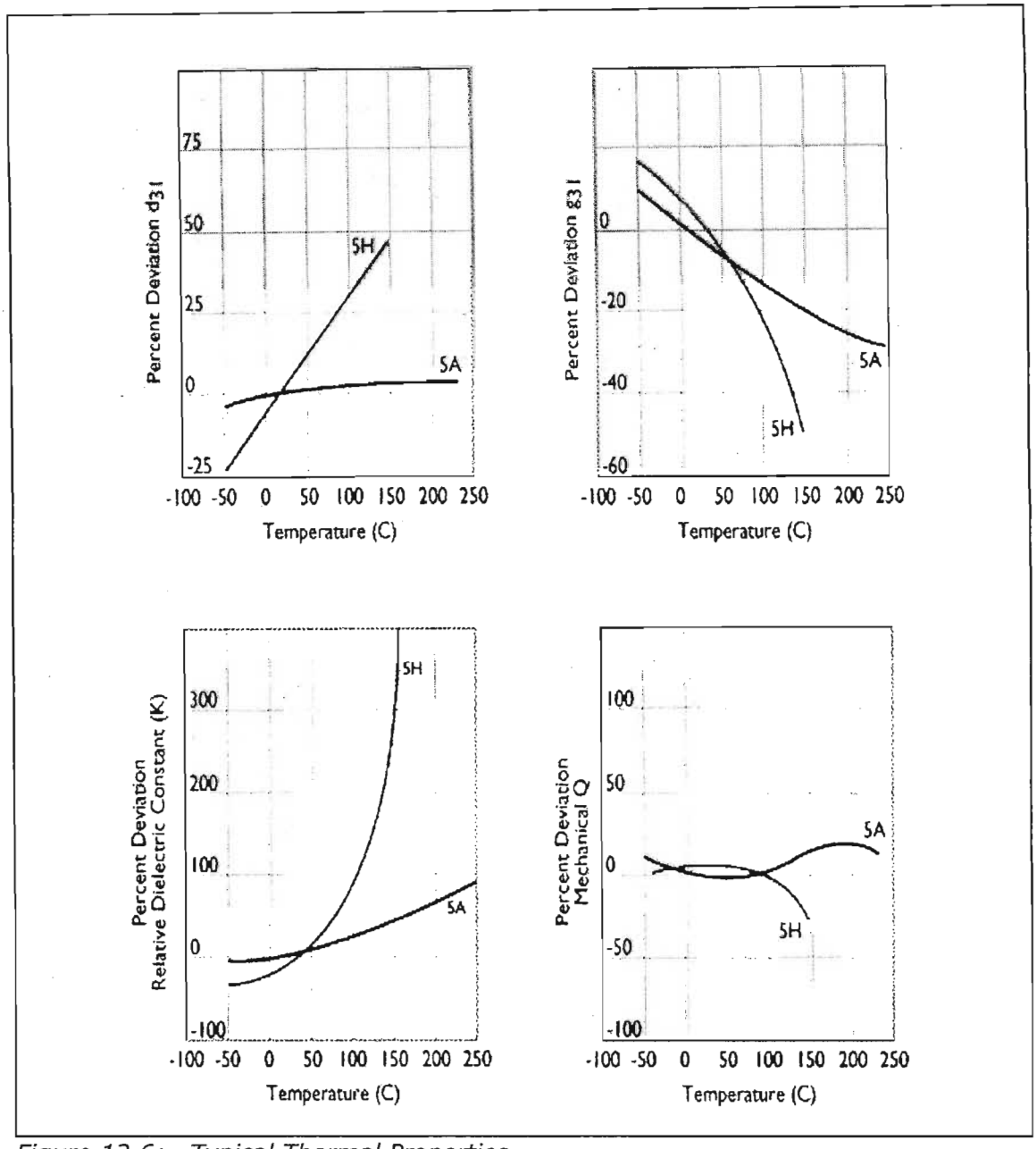


Figure 12.6: Typical Thermal Properties

Piezoelectric Materials and Their Properties

The following materials are recognised for their piezoelectric properties: polyvinylidene fluoride (PVDF), polytrifluroethylene) (PTrFE), lead metaniobate, bismuth titanate, barium titanate, and lead zirconate titanate (PZT). PZT is the material of choice because of its relative ease of use and maturity.

Summary of Material Properties

Table 12.1: Material properties

Physical Properties	
Curie temperature	The temperature above which spontaneous polarisation and piezoelectric behaviour cease to exist.
Maximum operating temperature	The safe use temperature for piezoelectric materials to ensure continued piezoelectric behaviour. Generally, the operating temperature is 50-60% of the curie temperature.
Mechanical Q	A dimensionless number that describes the merit of the material as a harmonic oscillator. Q is the reciprocal of the mechanical damping factor.
Dielectric Properties	
Dielectric constant, K	This is the ratio of the amount of charge a material can store compared to the charge stored by a comparable vacuum. (Units of capacitance*length/area).
Dissipation factor	A measure of the electrical loss in the material. It is the ratio of the effective series resistance to the effective series reactance.
Ageing	The polarisation of the material gradually decreases with time. Ageing is the time decay given as a percentage per decade of time.
Piezoelectric Properties	
Coupling coefficient	A measure of converting electrical energy to mechanical energy, and vice versa. It is the square root of the ratio of stored electrical energy to the mechanical energy input.

Applications

Shape changes can occur very quickly, hence piezoelectric components have been utilised in high frequency applications like vibration control, audio speakers and ultra-sound generators.

Piezoelectric measuring devices are used in almost every conceivable application requiring accurate measurement and recording of dynamic changes to mechanical variables such as pressure, force and acceleration.

Piezoelectric actuators are good choices for intelligent structures because their actuation strain mechanism can be easily regulated to control extension bending and twisting motion in the material. They are easily integrated with load-bearing

structures by surface bonding or embedding and do not significantly alter the passive static and dynamic stiffness characteristics of the structure.

Piezoelectric transducers are used mostly in ultrasound medical imaging, non-destructive testing and underwater acoustics (for tracking vessels, finding fish and deep-sea seismology). Low-frequency transducers for underwater acoustics are known as hydrophones.

Aerospace applications include modal testing, wind tunnel and shock tube instrumentation, landing gear hydraulics, rocketry and ejection systems.

Ballistics applications include combustion, explosion, detonation and sound pressure distribution.

Biomechanics applications include multi-component force measurement for orthopaedic gait and posturography, sports, ergonomics, neurology, cardiology and rehabilitation.

Engine testing applications include combustion, gas exchange and injection, indicator diagrams and dynamic stressing.

Engineering applications include materials evaluation, control systems, reactors, building structures, ship structures, auto chassis structural testing, shock and vibration isolation and dynamic response testing.

Industrial uses include machining systems, metal cutting, press and crimp force, automation of force-based assembly operations, and machine health monitoring.

Other applications include transportation systems, plastic moulding, rockets, machine tools, compressors, engines, flexible structures, oil/gas drilling and shock/vibration testers. Piezoelectric ceramics are the brains behind birthday cards that sing when opened.

Techniques have been developed for macro damage detection in complex structures using electro-mechanical impedance theory. Neural network methods have been used to detect impact locations and magnitudes.

Piezoelectric-actuated fixtures have shape-adaptive clamping surfaces so that the handwork required to finish the clamping surfaces can be eliminated. When the clamping surfaces do not perfectly match those of the work piece, the piezoelectric elements sitting on the clamping surfaces will sense abnormal stress distribution due to the clamping forces or the weight of the work piece. A computer analyses

the stress signals and then instructs the piezoelectric elements to exert appropriate forces to fine-tune the position of the work piece and tightly clamp the parts.

Ultra-high-fidelity stereo speakers are being developed using piezoelectric actuators, which expand and contract in thousandths of a second in response to applied voltage. Whole house walls or car interiors may be turned into speakers by embedding them with tiny actuators for built-in surround-sound.

In 1992, the sporting goods company K2 embedded a piezoelectric actuator/sensor package in skis. The equipment improves control on downhill runs by converting mechanical energy into electricity, which is used to increase vibration damping.

In health care, a smart gel made from piezoelectric polymers is being developed for artificial muscles. An electric charge will initiate their movement.

High power applications (ultrasonic or high-voltage energy and sonar devices) include ultrasonic cleaning, underwater echo-sounding, ultrasonic micro-bonding, high frequency transducers, high stress pressure sensors, gas ignitors (squeeze) transformers, vibratory motors, and high power actuators.

Low power resonant and non-resonant devices where high coupling and/or high charge sensitivity are required include flow meters, thickness gauges, underwater hydrophones, pressure sensors, accelerometers, gas igniters (impact), precise movement control, microphones, medical monitoring and bimorph structures. Other devices may have high permittivity and high piezoelectric charge coefficients.

Materials with excellent dielectric and low mechanical loss at high drive conditions are used in devices for underwater welding, ultrasonic mixing and dispersion, ultrasonic surgery and cryogenic SEMs.

Dynamic Behaviour of Transducers

Piezoelectric transducers for measuring pressure, force and acceleration may be regarded as under-damped spring-mass systems with a single degree of freedom. They are modelled by the classical second order differential equation whose solution is:

$$\frac{a_o}{a_b} \cong \frac{1}{\sqrt{\left[1-\left(\frac{\omega}{\omega_n}\right)^2\right]^2 + \left(\frac{1}{Q^2}\right)\left(\frac{\omega}{\omega_n}\right)^2}} \qquad (13\text{-}11)$$

ω_n undamped natural (resonant) frequency

ω frequency at any given point on the curve

a_0 output acceleration

a_b mounting base or reference acceleration

Q factor of amplitude increase at resonance

Quartz transducers have Q of around 10 to 40, hence:

$$\text{Phase lag} \cong \frac{60}{Q}\left(\frac{\omega}{\omega_n}\right) \text{ for } \left(\frac{\omega}{\omega_n}\right) \leq \frac{2}{5} \qquad (13\text{-}12)$$

Monitoring Techniques

Piezoelectric measuring systems are active electrical systems, i.e. the crystals produce an electrical output only when they experience a change in load. For this reason, they cannot perform true static measurements. However, it is a misconception that piezoelectric instruments are suitable for only dynamic measurements. Quartz transducers, paired with adequate signal conditioners, offer excellent quasi-static measuring capability for minutes or even hours.

Due to the transversely isotropic nature of the piezoelectrics, the strain invariant allows a non-directional measurement of strains, which is independent of the co-ordinate systems. Hence, instead of measuring all the strain components at a location, only one piezo-sensor is needed, and the measurement will not be affected by the co-ordinate system selected.

The use of piezoelectric strain sensors, instead of rosettes of strain sensors, greatly simplifies the identification process.

Finite Element Analysis

Even though initial applications of finite element methods for piezoelectric material configurations can be traced back to the late 1960s, the field is still in its infancy.

12.1.5. Common Piezoelectric Materials

Quartz Transducers

Quartz transducers exhibit remarkable properties that justify their use in research, development, production and testing.

1. They are extremely stable, rugged and compact.
2. They have a high material stress limit of ~140 MPa.
3. Temperature resistance up to 500°C.
4. They exhibit very high rigidity, high linearity and negligible hysteresis.
5. Almost constant sensitivity over a wide temperature range.
6. Ultra high insulation resistance (10^4 ohms) allowing low frequency measurements (<1 Hz).

High impedance transducers have a charge output that requires a charge amplifier or external impedance converter for charge-to-voltage conversion.

Low impedance types use the same piezoelectric sensing element as high impedance units and also incorporate a charge-to-voltage converter. An external power supply coupler is required to energise the electronics and de-couple the subsequent DC bias voltage from the output signal.

Lead Zirconate Titanate, PZT

There are many variants of PZT, both hard and soft. PZT has excellent machinability, high dielectric and mechanical strength. Optical transparency is also available.

12.1.6. Piezoelectricity and Smart SHM

Piezoelectric materials are used as sensing elements in a wide range of mechanical fields. The material itself does not retain any "load history" information; therefore, for long-term structural health monitoring, additional equipment is required to record any data. Hence piezoelectric materials would form part of an *active* health monitoring system.

Piezoceramics in particular are very versatile and can be manufactured in a variety of shapes and sizes. Physical, chemical and piezoelectric characteristics can be tailored to achieve desired results. They are also chemically inert and immune to moisture and other atmospheric conditions. PZT is an example of a piezoceramic with high mechanical strength and excellent machinability.

Piezoelectrics are already being used in health monitoring systems as well as in aerospace applications. They are capable of taking measurements in quasi-static conditions, and their dynamic response behaviour can be modelled by the classic second-order differential equation. In the case of the smart mining bolt, however, a passive solution such as that offered by TRIP steel is preferred.

12.2. ELECTROSTRICTION

Refer to [81]-[83]. Electrostrictive materials are non-poled ceramics with displacement (or stress, or strain) proportional to the *square* of the applied electric field strength (or voltage).

12.2.1. Electrostrictive Materials

The lead magnesium niobate (PMN) ceramic and its derivatives are the classic electrostrictive materials. These "relaxor ferroelectrics" possess high relative permittivity (20,000 to 35,000) and extraordinarily high electrostrictive coefficients. They are not piezoelectric and possess no spontaneous polarisation.

Other electrostrictive materials include polyurethane, polyether, amorphous polybutadiene, polydimethylsiloxane, and liquid crystal elastomers.

Electrostrictive materials are centro-symmetric at zero volts (see Figure 13.7). An electric field separates the positively and negatively charged ions, changing the dimensions of the cell and resulting in an expansion.

Conventional piezoceramics are stiff: they can develop high stresses but only at small strains. Research is being conducted into electrostrictive elastomers, which offer the possibility of highly compliant actuators with some of the properties of muscle. Elastomers have high permittivity leading to large electric stresses and low modulii leading to large strains and hence high electrostrictive coefficients.

12.2.2. Electrostrictive vs. Piezoelectric Devices

Electrostrictive materials are similar to piezoelectric materials in that both are ferro-electric crystals that expand and contract with applied voltage. While there are symmetry restrictions on which materials can exhibit piezoelectricity, there are no such restrictions on electrostriction – all materials are in principle electrostrictive, the only question is to what degree.

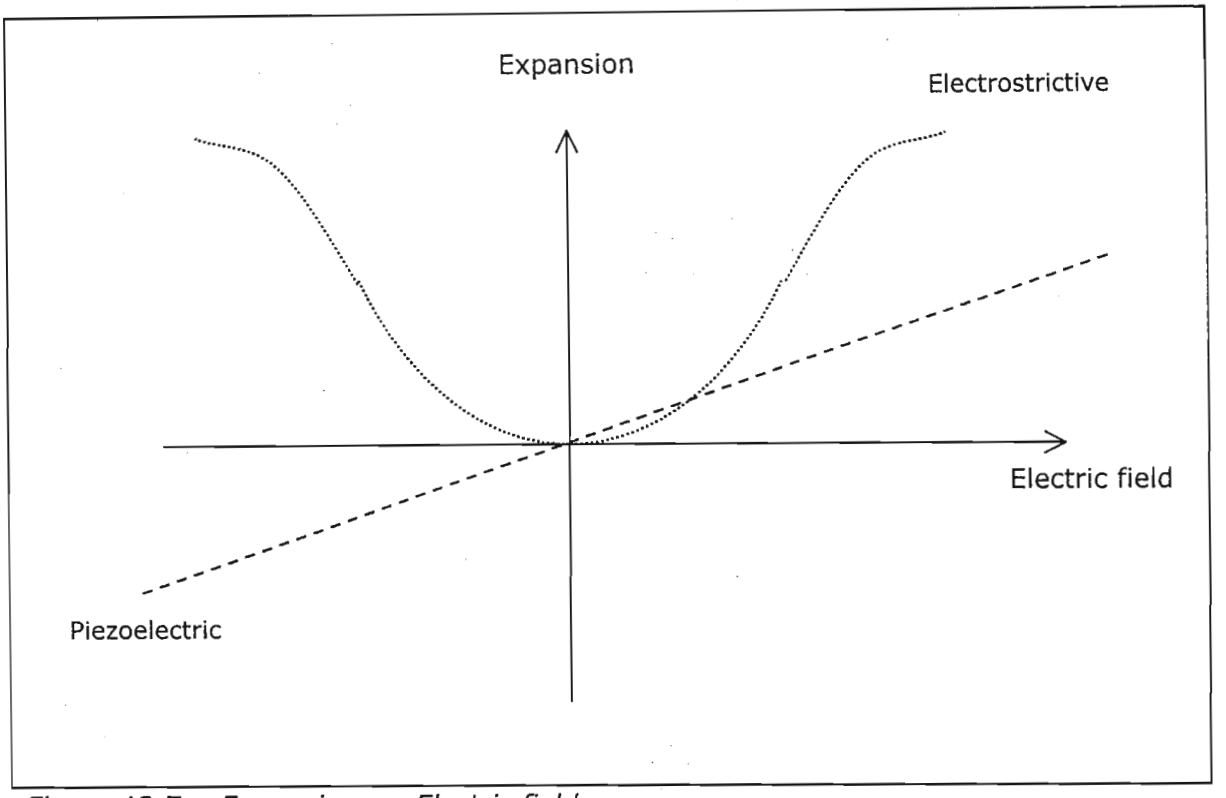


Figure 12.7: Expansion vs. Electric field

PZT ceramics are isotropic and not piezoelectric prior to poling. Upon poling, they become anisotropic and exhibit directionally dependent piezoelectric and mechanical properties. For electrostrictive PMN materials, change in length is proportional to the square of the applied voltage. Unlike piezoelectric devices, PMN ceramics are not poled. Positive or negative voltages result in an elongation in the direction of the applied field, regardless of the ceramic's polarity. Because PMN is unpoled, it is an inherently more stable device without the long-term creep associated with PZT. Creep ranges up to 15% in PZT and 3% in PMN. See Figure 13.8.

Electrostrictive actuators are operated above the Curie temperature. With respect to thermal stability, PMN has a coefficient of thermal expansion of $1*10^6$, which is twice as good as that of PZT. However, PZT's strain sensitivity to temperature is much more robust than PMN's, especially over large temperature variations (Figure

13.8). Both displacement and hysteresis of PMN materials are strongly dependent on the actuator temperature.

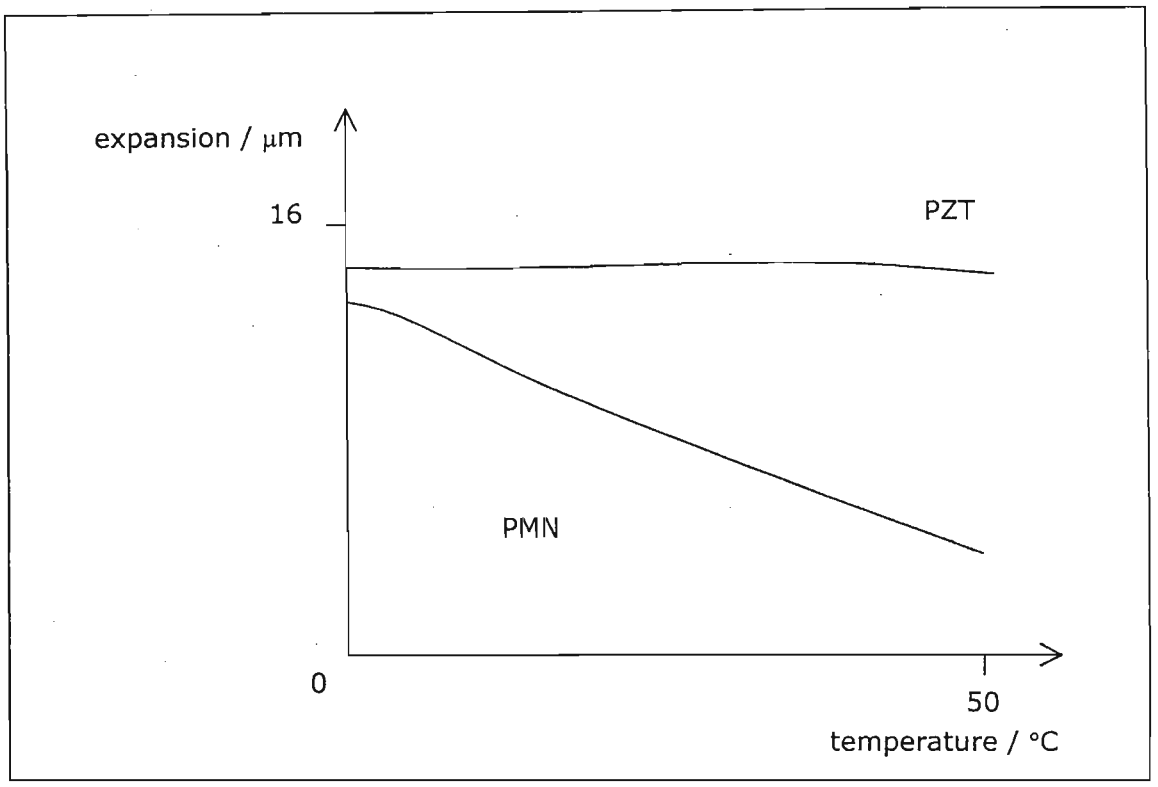


Figure 12.8: Expansion vs. temperature of PZT and PMN actuators.

PMN actuators show an electrical capacitance 4 to 5 times as high as PZT actuators, requiring significantly higher driving currents for dynamic applications.

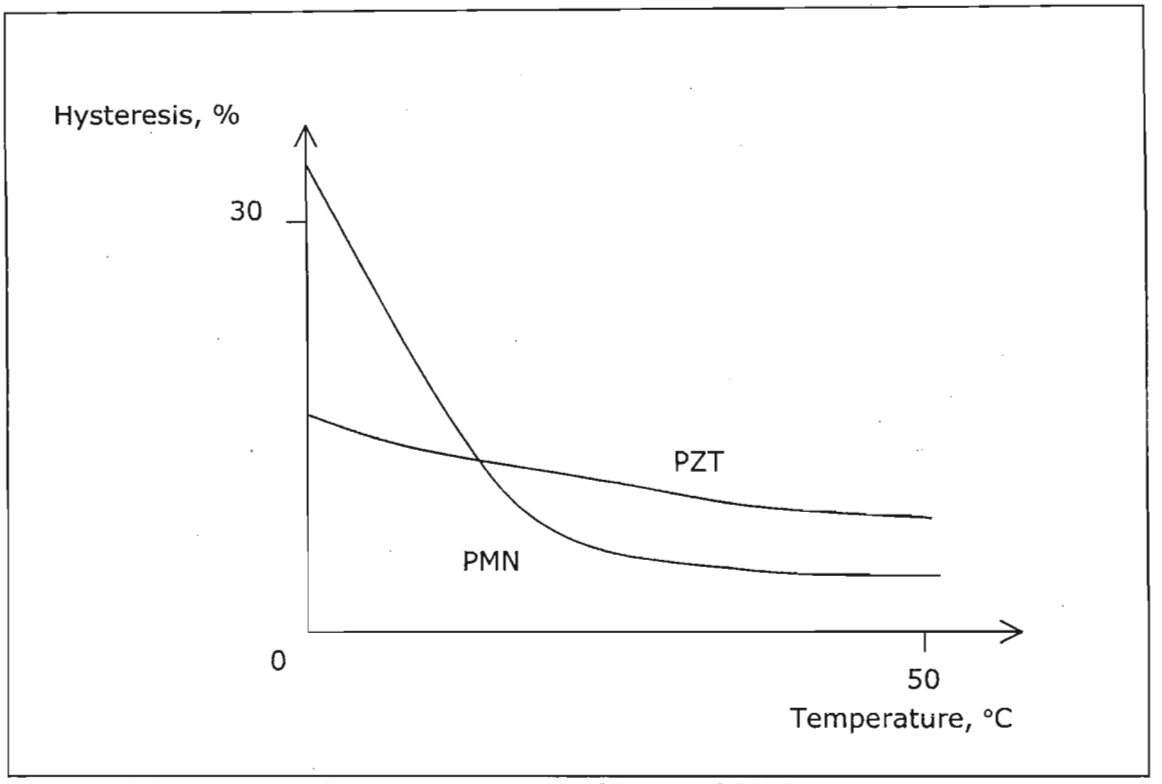


Figure 12.9: Hysteresis vs. temperature of PZT and PMN actuators.

Since both materials are generally used in open-loop, encoderless designs, any hysteresis will result in an apparent backlash or non-repeatability upon reversal of motion. Hysteresis depends on the recent motions performed by the device and is difficult to model, predict or compensate. Electrostrictive actuators experience less hysteresis (2%) than piezoceramics (12% to 15%). See Figure 13.9.

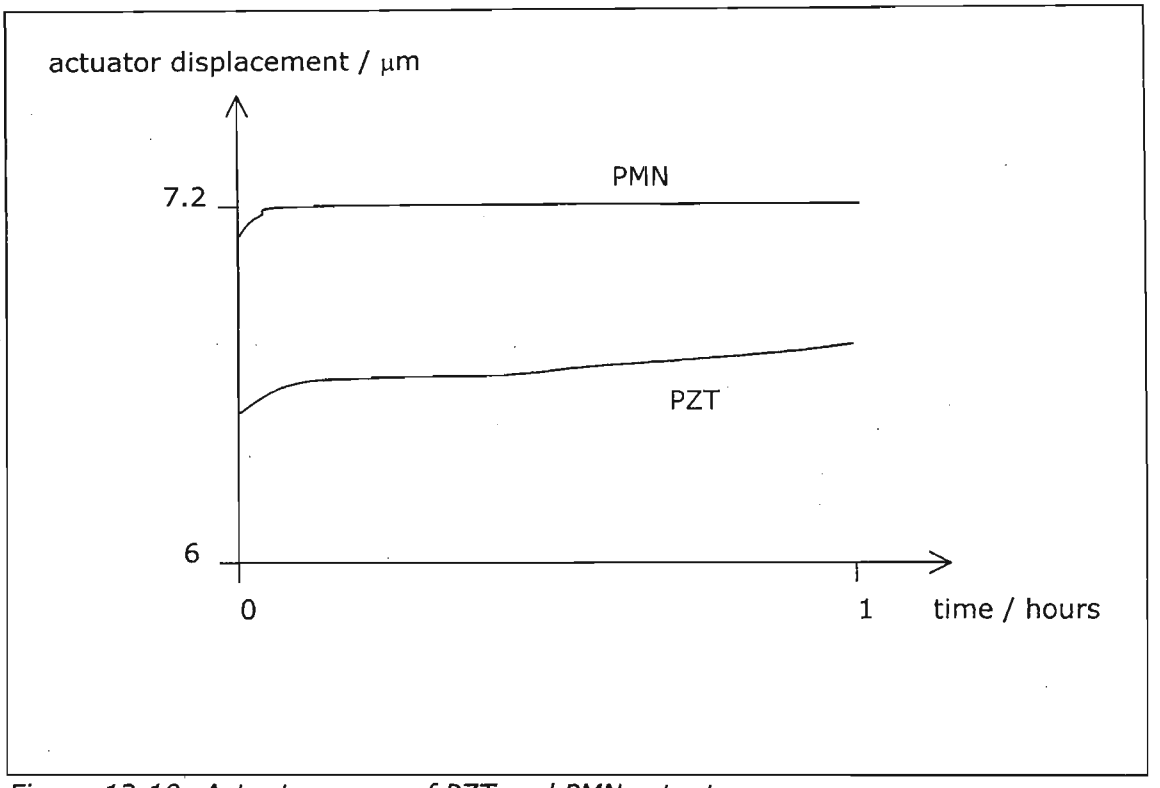


Figure 12.10: Actuator creep of PZT and PMN actuators .

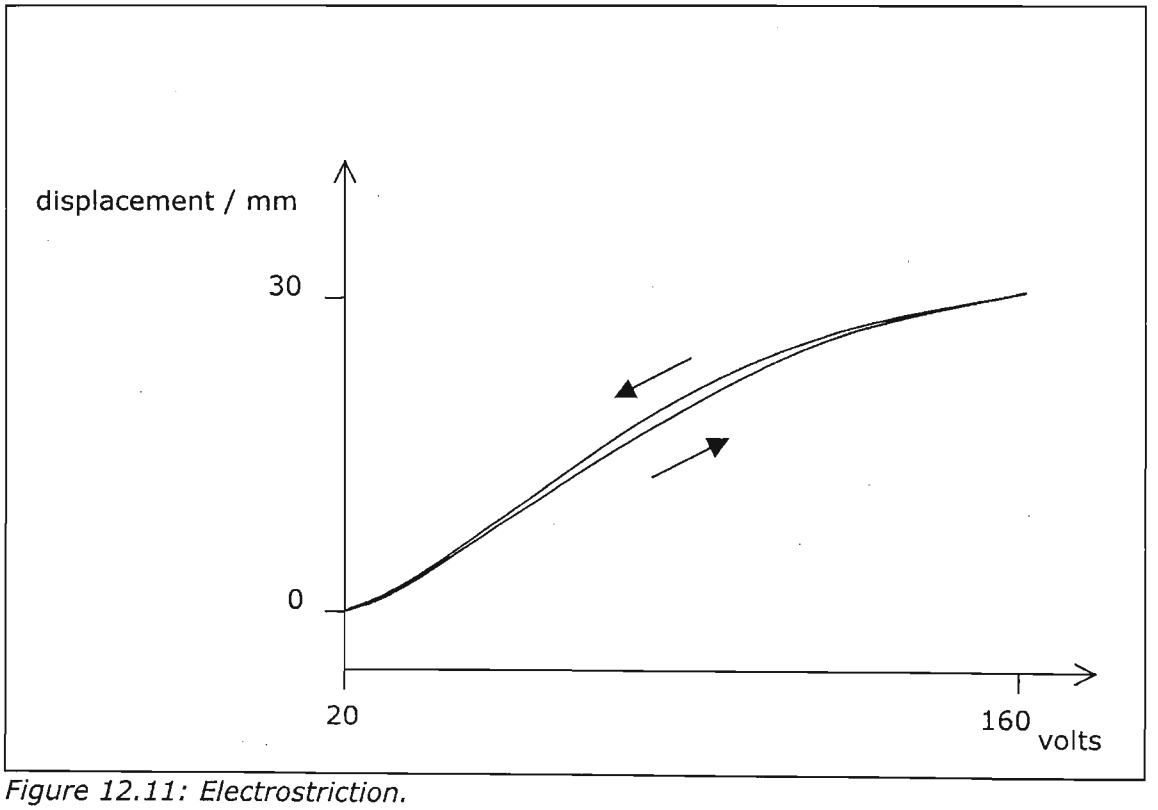


Figure 12.11: Electrostriction.

12.2.3. Measurement Techniques

The definition of the coupling constant for piezoelectric ceramics is straightforward and depends directly on the square of the piezoelectric constant and inversely on the compliance and dielectric constants appropriate for the boundary conditions.

However, these definitions are not directly applicable to electrostrictive materials because the material is nonlinear and therefore the equivalent piezoelectric and dielectric constants are not universally defined. Also, the presence of bias voltages about which the material would operate complicates the situation dramatically.

Energy methods are used to calculate k for electrostrictive materials using several different constitutive models for the material.

12.2.4. Lead Magnesium Niobate, PMN

PMN exhibits the following properties:

1. Macroscopically defect-free ceramic.
2. High relative permittivity.
3. Ideal for high voltage/high power situations with high breakdown strength.
4. Negligible hysteresis.
5. Minimal heating even at very high frequencies.
6. Many variants (high and low temperature of operation).
7. Available as discs, plates and other geometries.

12.2.5. Conclusion

Electrostrictive materials are similar to piezoelectrics except that displacement is proportional to the square of the voltage. Electrostrictive materials could therefore also form part of an active monitoring system, although piezoelectrics are more useful if compression is a factor.

Comparisons would have to be made to determine whether piezoceramics or electrostrictive materials would be more suitable in a particular active monitoring system. Comparing electrostrictive PMN properties to those of piezoelectric PZT:

1. Because PMN is unpoled, it is an inherently more stable device without the long-term creep associated with PZT (15% vs. 3%).

2. PMN actuators experience less hysteresis than PZT actuators (2% vs. 12% to 15%).

3. PMN has a coefficient of thermal expansion twice as good as that of PZT.

4. PZT's strain sensitivity to temperature is much more robust than PMN's, especially over large temperature variations. Both displacement and hysteresis of PMN materials are strongly dependent on the actuator temperature.

Recently, these materials have been used for active vibration control of thin plates. Electrostrictive PVDF-PTrFE co-polymer based materials can be used for high strain actuators. Elastomeric polymers can be used for ultra high strain response. Films made of PVDF-PTrFE co-polymers can be used for high load, high frequency applications. These materials can also be used in structural health monitoring systems. However, as in the case of piezoelectric materials, any monitoring system using electrostrictive materials is active; the benefits arising from TRIP steels as passive monitoring sensors are considered to be more attractive for the purposes this study.

12.3. MAGNETOSTRICTION

Refer to [71]-[73].

12.3.1. Definition

Magnetostriction is a property encountered in certain ferromagnetic materials whereby the sample changes shape (constricts or expands) in the presence of a magnetic field. This property can be usefully engineered for a variety of actuator applications.

12.3.2. Terfenol

Terfenol is a metallic compound containing iron and 2 rare-earth elements, terbium and dysprosium. Only available in commercial quantities since 1987, it has unique magnetostrictive properties.

With the application of minimal magnetic fields, Terfenol will generate huge magnetostrictive strains amounting to more than a 0.1% change in length when exposed in a 500 oersted (39.79 kiloampere-turn per metre) magnetic field. This is approximately 10 times greater than any previously known magnetostrictive material. Such movement can also be generated under load, converting magnetic energy to mechanical energy with efficiencies greater than 60%.

Terfenol properties include:

1. High strain
2. High force
3. Wide bandwidth
4. Microsecond response time
5. Terfenol is powerful, requires only a low voltage power, and performs more reliably than any other competing smart material in a package less than ½ the size and weight.

12.3.3. Applications

High force actuators
Industrial vibrators
Active vibration control
Seismic sources
Micro-positioning
Pressure control
Machine control
Sonar and tomography
Linear and rotary motors
Valve actuation
Aircraft control
Robotics
Speakers
Fuel injectors
Ultrasonics
Space structures
Fast acting relays
Sonic communication
High pressure pumps
Mirror positioning
Noise and vibration suppression
Sonic welding and cleaning

12.3.4. Conclusion

As with piezoceramics and electrostrictive materials, magnetostrictive materials would be used in active monitoring systems. The active sensing material of choice would depend on the particular application, the type of information required, and the environmental conditions.

12.4. ELECTRORHEOLOGICAL AND MAGNETORHEOLOGICAL FLUIDS

Refer to [74]-[77].

12.4.1. History

The phenomenon of electrorheology was first observed by W.M. Winslow in 1947, and hence is sometimes referred to as the Winslow effect.

12.4.2. Electrorheology (ER)

Rheology is the study of the deformation and flow of matter. When an external electric field is applied to an ER fluid, the viscosity of the fluid increases remarkably along with the shear yield stress. When the electric field is removed, the fluid viscosity returns to its original state. In other words, when an electric signal (electric field) is applied across an ER fluid, a mechanical signal (fluid viscosity) can be outputted. The response is fast and reversible.

Classification

There are 2 broad classes of ER fluids: the *dispersion* type and the *uniformity* type.

The dispersion type contains colloidal suspensions of dielectric particles in electrically insulating oil. The dielectric particles are so-called "*dispersed phase*" and the insulating oil is a so-called "*dispersion medium*". On the application of an electric field, the viscosity changes from Newtonian to something similar to Bingham plasticity (Figure 12.12), exhibiting a finite yield stress with the shear stress dependent on the shear rate.

Uniformity type ER fluids are liquid crystals. Upon applying an electric field the viscosity coefficient increases.

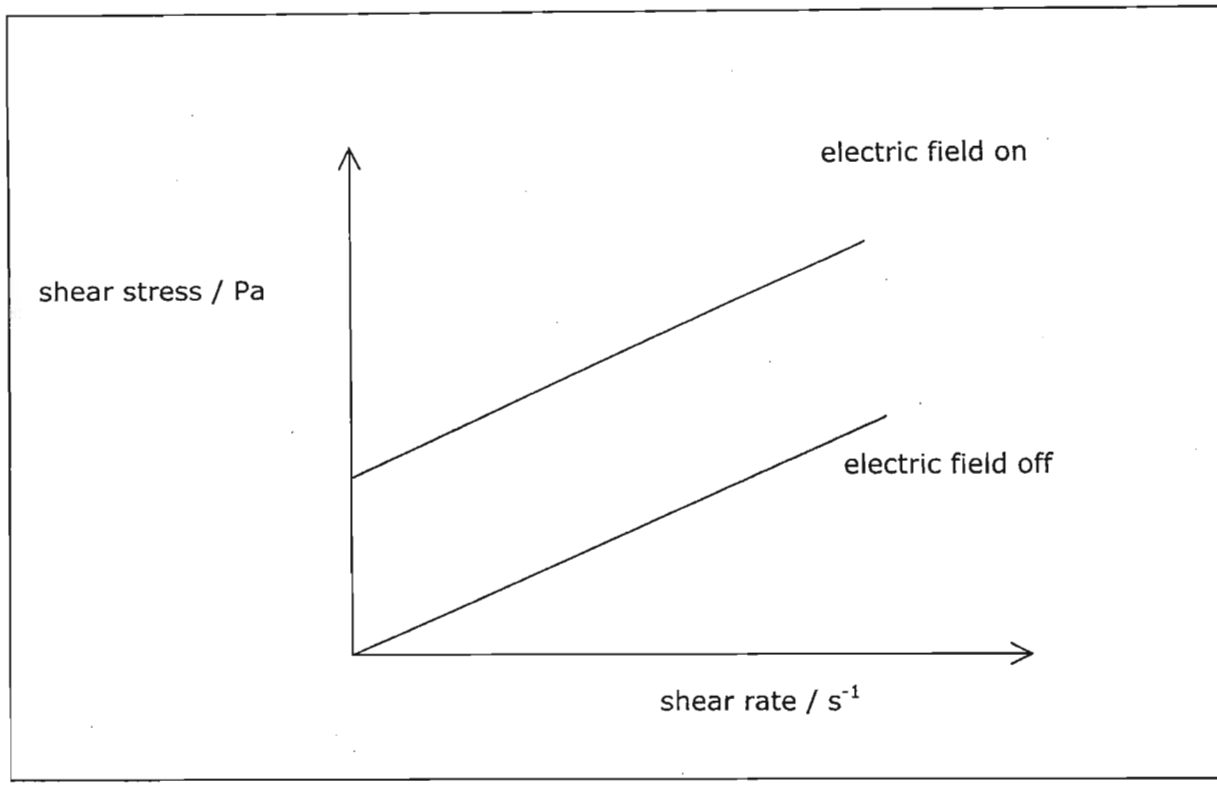


Figure 12.12: Viscosity changes

The Electrorheological Mechanism

Most ER fluids are of the dispersion type. With no applied electric field, ER dispersion behaves like a common dispersion (Figure 12.13(i)). When an electric field is applied, the dispersed particles are electrically polarised. The polarised particles attract each other due to the electrostatic interaction. Consequently, particle cluster chains form as the suspended particles fibrillate between electrodes (Figure 12.13(ii)). These chains cause the resisting power against shear flow and pressure flow as they inhibit the free flow of ER fluids.

For electric field strengths of the order of 3 kV/mm, ER fluids "solidify", with static and dynamic yield stresses as high as 10 kPa and 5 kPa respectively. These changes can be reversed in time intervals on the order of a few milliseconds, thereby offering excellent control capabilities.

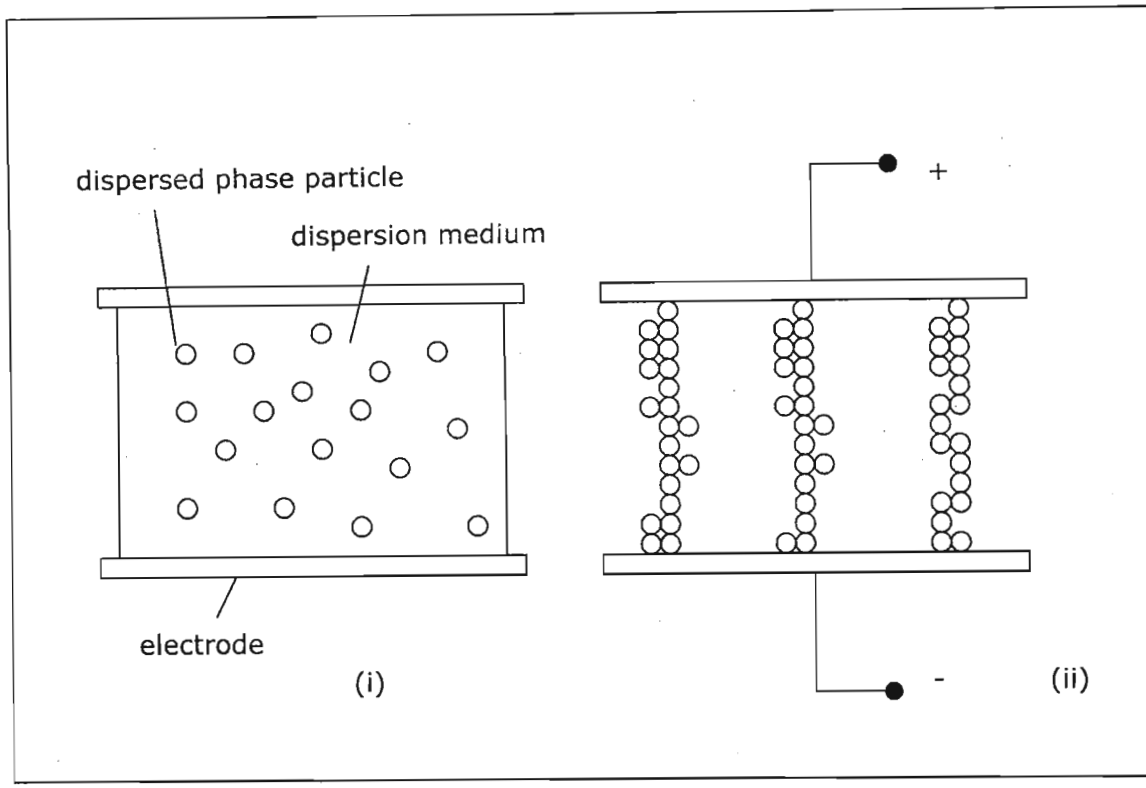


Figure 12.13: ER mechanism

Properties of Dispersion-Type ER Fluids

1. Induced shear stress is large.
2. Current density is small, hence power consumption is small.
3. Zero-field viscosity is low and fluidity is fairly good.
4. Response is very fast (in the order of milliseconds).
5. Durability is fairly good.
6. Dispersed phase particle take a long time to settle.
7. Both AC and DC voltages are available as electric fields.
8. Operation temperature range is very wide.
9. Induced shear stresses hardly depend on shear rate.
10. Large induced shear stress is obtained even under high shear rate.

12.4.3. Magnetorheology (MR)

MR fluids are very similar to ER fluids except that a magnetic field has to be applied. When a magnetic field is imposed, colloidal particles (e.g. iron) suspended in the organic medium (e.g. mineral or silicone oil) get induced a virtual magnetic dipole. Due to the dipolar interaction, chainlike structures form that change the rheological behaviour of the suspension from liquidlike to solidlike ($\sim 10^5$ to 10^6 times the viscosity).

These fluids are becoming increasingly popular due to the significantly reduced field strengths needed to activate the fluids. Moreover, the yield strengths are of a higher order than those of ER fluids.

12.4.4. Applications

Applications include dampers and shock absorbers, engine mounts, brakes, clutches, torque converters, tension controllers, valves, position and speed controllers, rehabilitation devices, games, training simulators, cutting machines, etc.

"Smart gels" are offshoots of magnetorheological fluids that have a reverse viscosity effect: when cooled they become more liquid, but heat tempers them. In orthopaedic shoe inserts, the heat from a wearer's foot will cause the insert to stiffen in a perfect mould of the sole.

12.4.5. Measurement/Monitoring Systems

Systems are available that explore the properties of any liquid by measuring both the viscosity and elasticity over a wide range of values of frequency, shear stress, shear rate and shear strain. The broadest view of liquid rheology is obtained by using oscillatory flow at a selected frequency because both viscous and elastic properties are revealed. Steady flow reveals only viscous properties.

Measurements are based on controlled oscillatory flow generated at a selected frequency in a precision-measurement cylindrical tube. Water is used as a rheological standard for instrument calibration. Sensors monitor the pressure drop across the tube and volume flow through the tube. Resolution of magnitude and phase of the pressure and volume flows allows calculation of the viscous and elastic components of the shear stress, shear rate and shear strain at the tube wall. From these values the viscosity, elasticity and storage and loss moduli are obtained.

12.4.6. Conclusion

Fluid properties are easily measured and modelled, but a rheological system does not store any "load history" information without the aid of additional recording equipment. A rheological health monitoring system would fall into the same *active* monitoring category as piezoelectrics and electrostrictive and magnetostrictive materials. Some combination of these smart materials may be used as the core of an active monitoring system.

Fluids obviously need special packaging and would require different measurement equipment than solid-state sensors. Cost is therefore a major contributing factor; maintenance of any active health monitoring system is also a consideration.

Recent developments showed that ER and MR fluids and devices based on these materials can be used as an integral part of a SHMS. These materials can be used for alternative designs of smart mining bolts.

12.5. SHAPE MEMORY ALLOYS

Refer to [64]-[66], [78]-[80].

12.5.1. Shape Memory and Crystallography

Shape memory alloys (SMAs) are a group of metallic materials that demonstrate the ability to return to some previously defined shape or size when subjected to the appropriate thermal procedure.

Generally, these materials can be plastically deformed at some relatively low temperature, and upon exposure to some higher temperature will return to their shape prior to the deformation.

The thermomechanical behaviour of SMAs is due to a martensitic-austenitic phase transition. A phase is that part of a material that is distinct from other parts of the same material in structure or composition. Austenite is the higher temperature phase and martensite the more deformable, lower temperature phase.

When in its martensitic form, a SMA is easily deformed to a new shape. When heated through its transformation temperatures, it reverts to austenite and recovers its previous shape. The temperature at which the alloy remembers its high temperature form when heated can be adjusted by slight changes in alloy composition and through heat treatment.

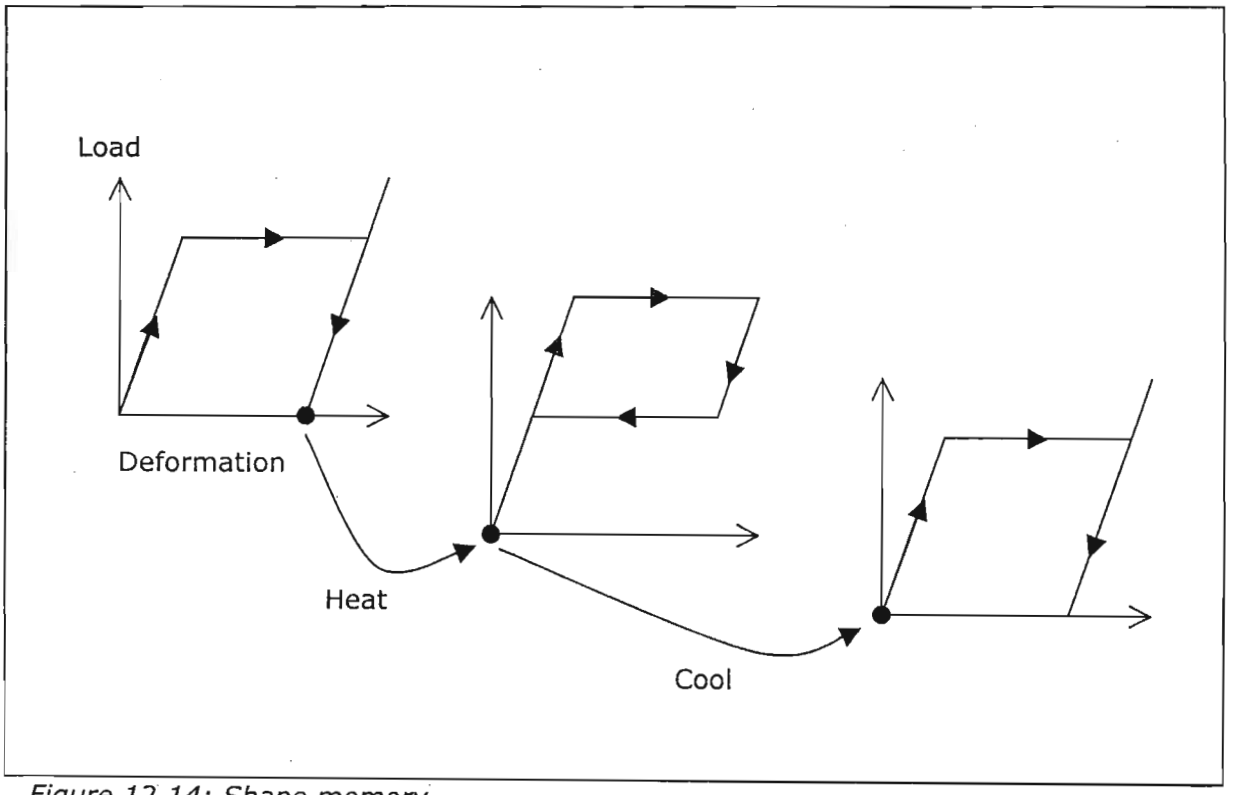


Figure 12.14: Shape memory

The transformation temperature hysteresis is generally defined as the difference between the temperatures at which the material is 50% transformed to austenite on heating and 50% transformed to martensite on cooling. This value can be approximated as the difference between A_p and M_p on a DSC curve (described later).

12.5.2. Materials That Exhibit Shape Memory

Ag-Cd	44 - 49 atomic% Cd
Au-Cd	46.5 - 50 atomic% Cd
Cu-Al-Ni	14 - 14.5 weight% Al, 3 - 4.5 weight% Ni
Cu-Sn	~15 atomic% Sn
Cu-Zn	38.5 - 41.5 weight% Zn

Cu-Zn-X	a few weight% X where X = Si, Sn, Al
In-Ti	18 - 23 atomic% Ti
Ni-Al	36 - 38 atomic% Al
Ni-Ti	49 - 51 atomic% Ni
Fe-Pt	~25 weight% Pt
Mn-Cu	5 - 35 atomic% Cu
Fe-Mn-Si	32 weight% Mn, 6 weight% Si

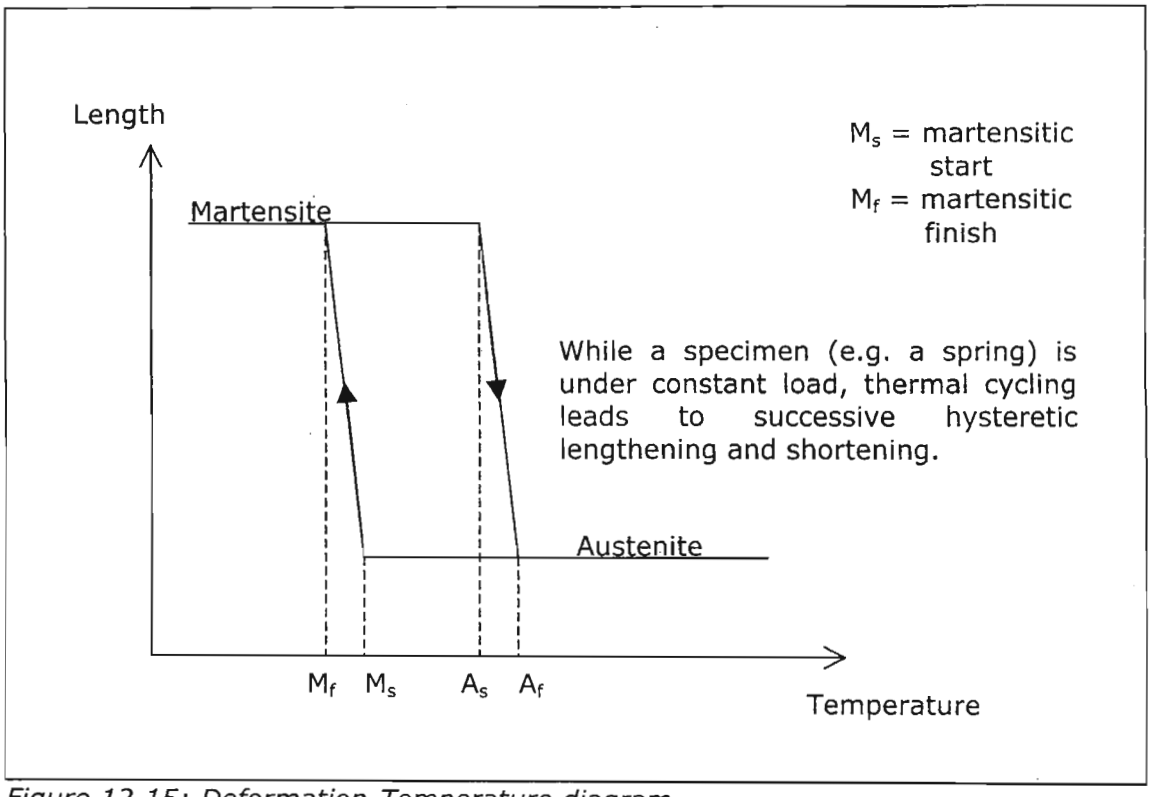
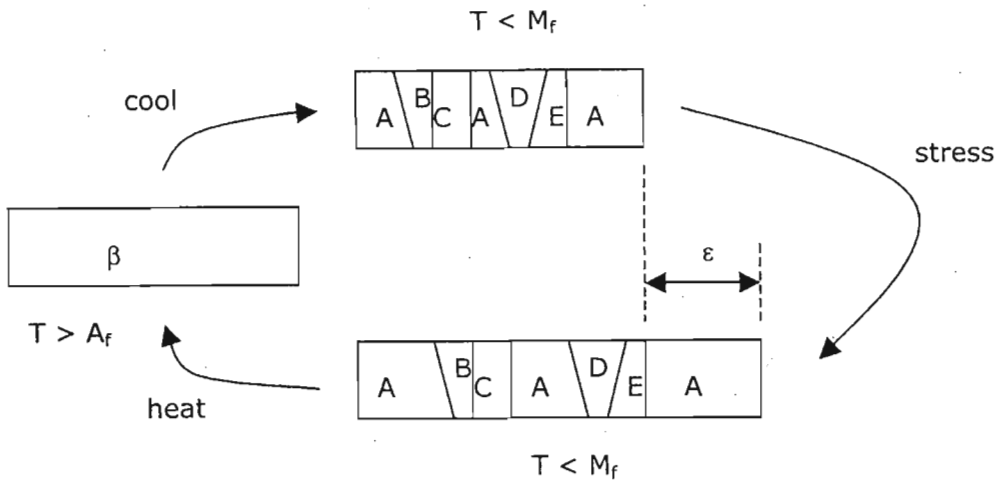


Figure 12.15: Deformation-Temperature diagram



When stress is applied to martensite, variant A becomes dominant. Upon heating to austenite, the material reverts to the β phase and recovers its original shape.

Figure 12.16: Crystallography

12.5.3. Commercial Availability

Although a wide variety of alloys are known to exhibit shape memory, only those that can recover substantial amounts of strain, or that generate significant force upon changing shape through the transformation temperatures, are of commercial interest. To date, this has mostly been the NiTi alloys and the copper-based alloys such as Cu-Zn-Al and Cu-Al-Ni. The NiTi alloys have greater shape memory strain (up to 8% vs. 4 - 5%), tend to be much more thermally stable, have superior corrosion resistance and lower susceptibility to stress-corrosion-cracking, and have much higher ductility compared to the copper-based alloys. On the other hand, the copper-based alloys are much less expensive, can be melted and extruded in air with ease, and have a wider range of potential transformation temperatures.

12.5.4. Load - Deformation Diagrams

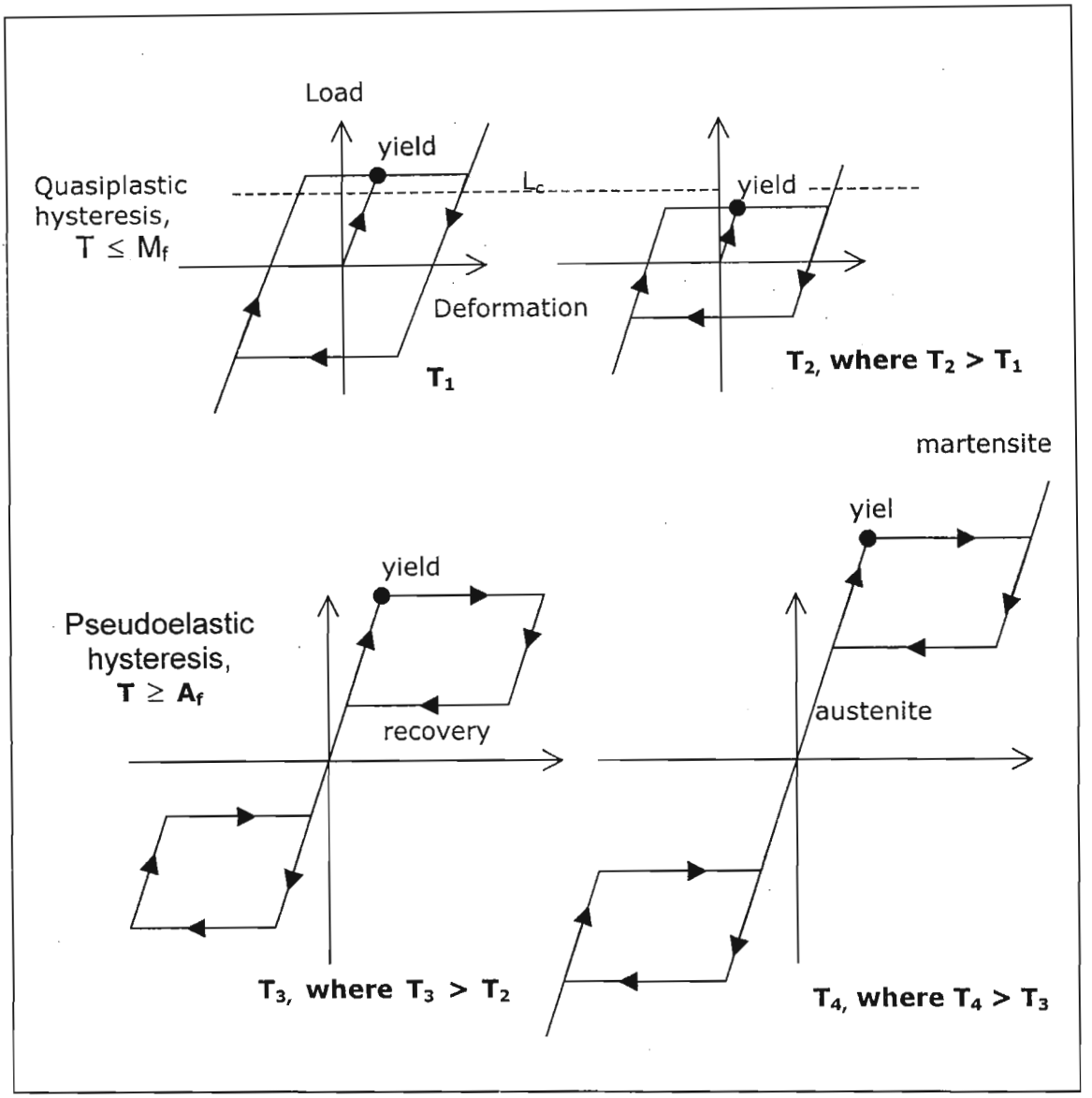


Figure 12.17: Load-Deformation diagrams

12.5.5. Superelasticity

Superelasticity is caused by the stress-induced formation of some martensite above A_f. Because it has formed above its normal temperature, the martensite reverts immediately to undeformed austenite as soon as the stress is removed. The term *superelasticity* refers to the springy, rubberlike behaviour at temperatures just above A_f.

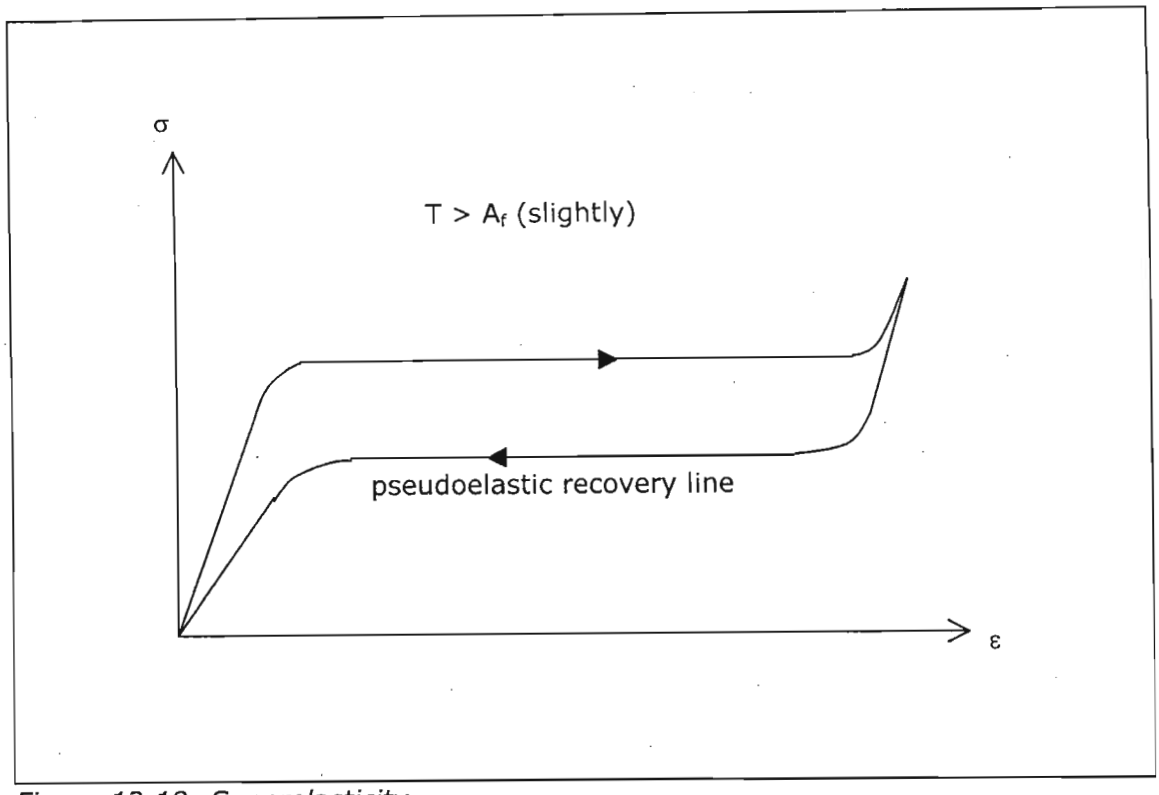


Figure 12.18: Superelasticity.

12.5.6. Applications

Applications currently making use of shape memory alloy properties:

Thermal actuators, switches, thermostats.

Heat engines

Medical applications

In heat engines, NiTi wires are dragged alternately through hot and cold oil showers. Under the hot shower they contract and their tensile force sets an oblique disk in rotation. In the cold shower the wires are soft and are pulled back to the original length.

A medical application is braces for the regulation of teeth (Figure 13.19). Important is the large contraction at constant load that occurs on the pseudoelastic recovery line.

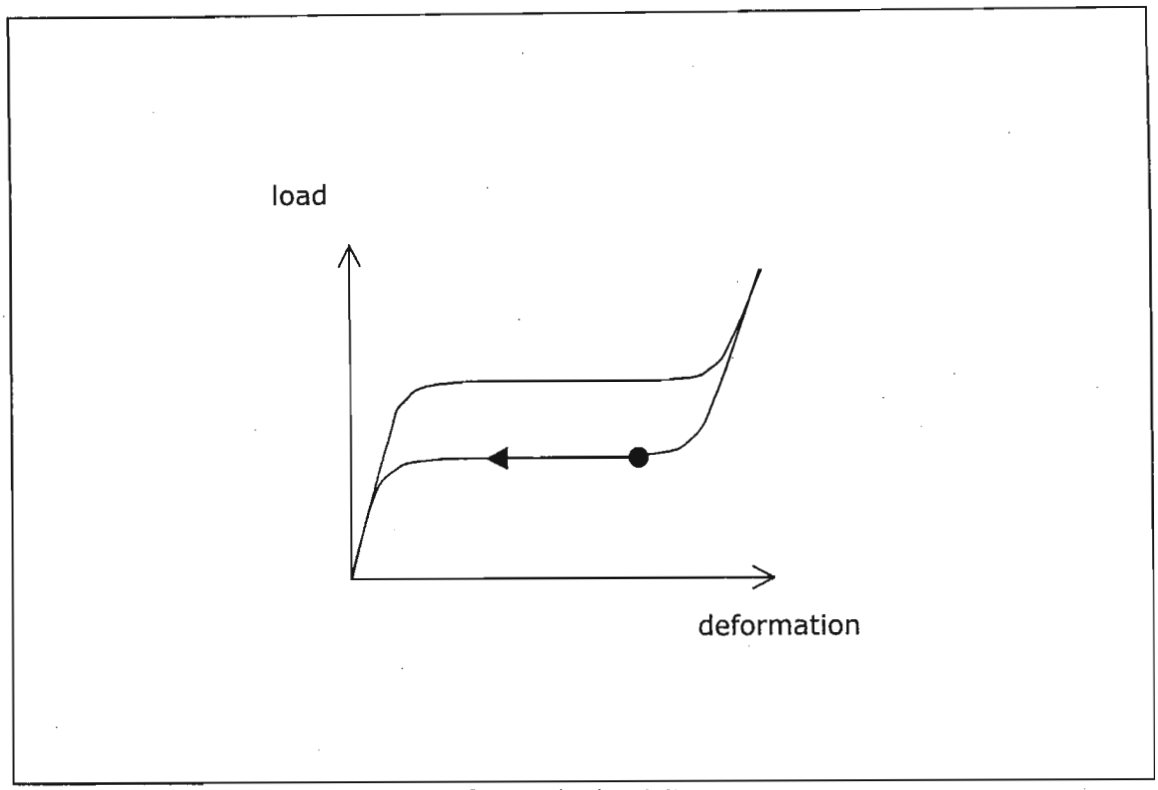


Figure 12.19: Braces make use of pseudoelasticity.

12.5.7. Measuring/Monitoring Systems

Differential Scanning Calorimetry (DSC)

DSC is used to measure the amount of heat absorbed/emitted by a tiny sample of material as it is heated through its phase transformations. It is a precise method of determining the transformation temperature range (TTR) at zero stress, but requires an expensive instrument.

The DSC method yields excellent, repeatable results on fully annealed samples. (Anneal – heating cold-worked material to produce re-crystallisation and therefore softening, i.e. stress-relief). Tests on partially cold-worked materials (e.g. those used to optimise superelasticity) can yield poor, inconclusive results.

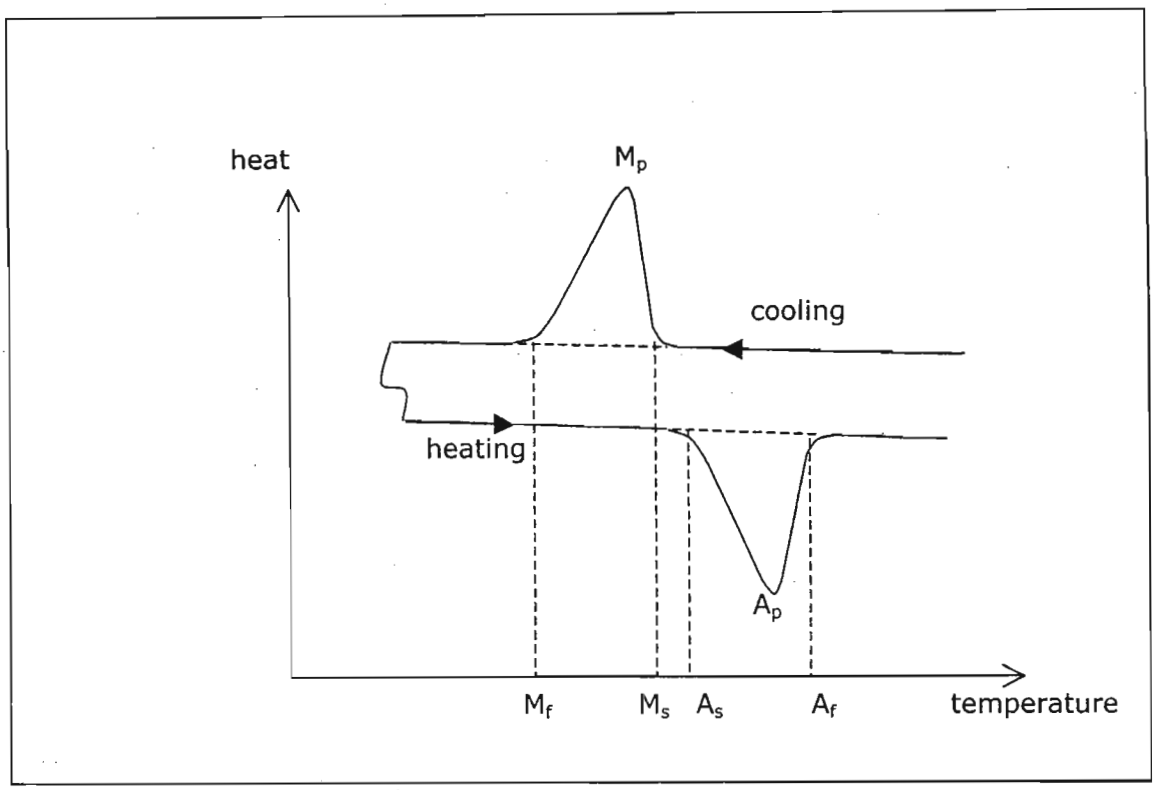


Figure 12.20: DSC test results

Resistivity

The resistivity of a sample changes as it is heated and cooled through the transformation temperature range:

$$R = \rho \frac{L}{A} \text{ for a wire in a circuit.} \qquad (13\text{-}13)$$

R resistance

L length

A cross-sectional area

ρ electrical resistivity (ω.m)

$$\sigma = \frac{J}{E} = \frac{1}{\rho} \qquad (13\text{-}14)$$

J current density

E electric field

σ electrical conductivity

Constant-load Strain Measurement

This test is generally used for applications which utilise the shape memory effect (and not superelasticity).

A constant stress is applied to a sample as it is cycled through the transformation temperature range (TTR) in both directions. The accompanying strain is measured.

Increasing the test stress leads to increasing TTRs. The DSC test yields the lowest results since there is no applied stress. To determine the TTRs at zero stress using the strain measurement technique, the curve in Figure 12.21 must be obtained at 2 or more stress levels. The particular transformation point of interest can then be extrapolated to zero stress.

A side benefit of the stress-dependent nature of the transformation is that a transformation temperature can be precisely tuned (e.g. in many actuator applications) by making small adjustments to the bias force acting against the shape memory element.

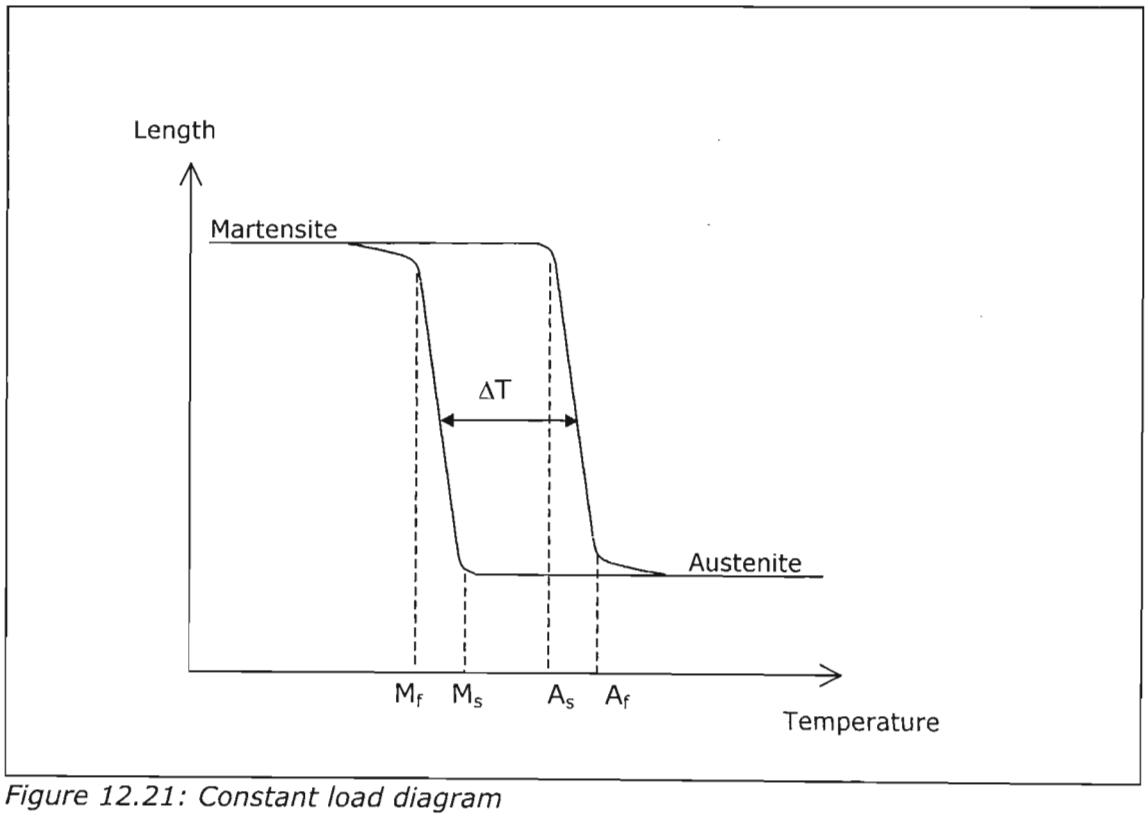


Figure 12.21: Constant load diagram

Active A_f (or Functional A_f) Test

Also known as a Water Bath or Alcohol Bath test, this test is easily conducted by merely bending an alloy sample (e.g. wire) while below M_s, and then monitoring the shape recovery while it is heated.

Surprisingly accurate, repeatable results are obtained if performed carefully. A bath of crushed dry ice and alcohol is recommended for testing at low temperatures. A water bath may be more appropriate for higher temperature alloys.

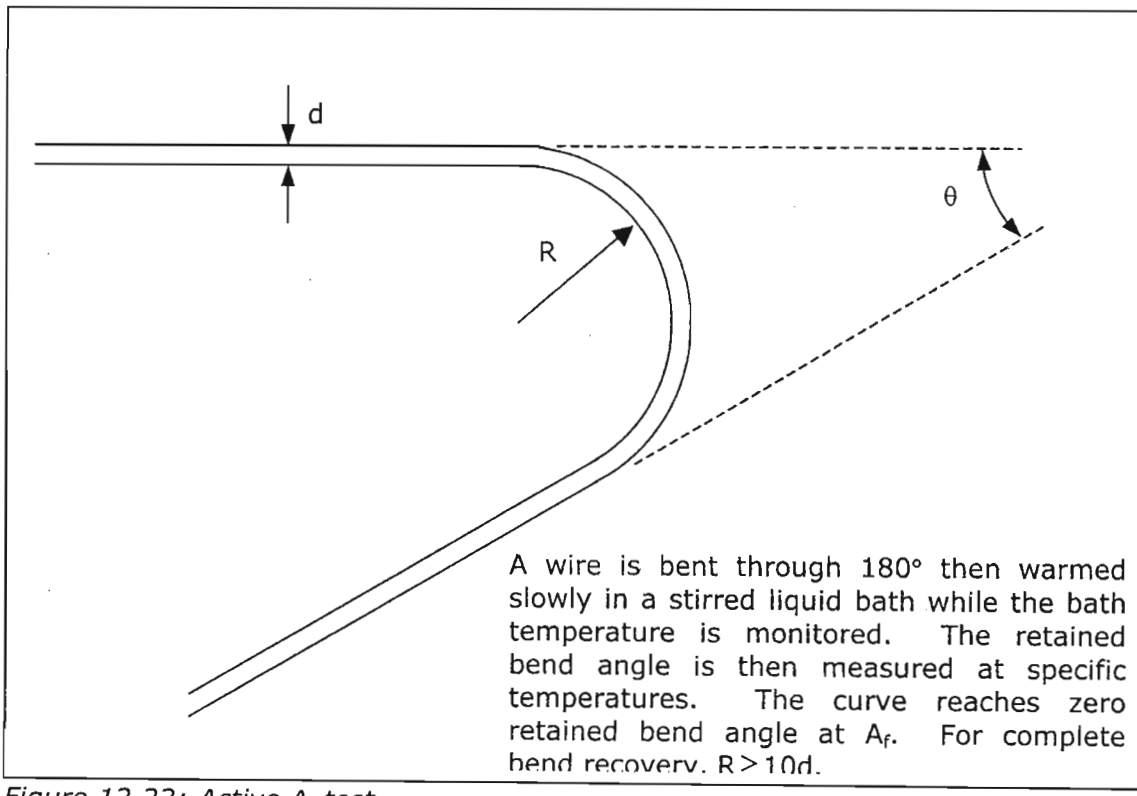


Figure 12.22: Active A_f test

Standard Tensile Test

The stress-strain properties are measured at a number of temperatures across the TTR. This is an imprecise method.

12.5.8. Nitinol

Properties

The nitinol alloy has very good electrical and mechanical properties, long fatigue life and high corrosion resistance. NiTi properties also include the shape memory effect, superelasticity, and high damping capability. The properties of NiTi can be

modified to a great extent by changes in alloy composition, mechanical working and heat treatment.

The shape recovery process occurs over a range of just a few degrees, and the start or finish of the transformation can be controlled to within a degree or two if necessary. The sensitivity of the transformation temperature to alloy composition (subtle adjustments in the ratio of Ni to Ti) is so great that the chemistry is not recommended as a way to specify the alloy of interest. Instead, the easily measurable TTR is a far more accurate means to specify the alloy of interest.

Superelastic NiTi alloys achieve incredible amounts of flexibility and kink resistance. An alloy of NiTi can behave superelastically if A_f is just below the use temperature. Superelastic properties are good from A_f to A_f + ~50°C.

Yield strength of austenite is ~250 MPa to ~ 700 MPa, while that of martensite is ~70 to ~140 MPa. Both forms of the alloy are ductile (elongation to failure > 25%) and strong. Martensite is able to absorb up to 8% recoverable strain.

Both forms are biocompatible and extremely corrosion resistant. NiTi alloys naturally form an oxide on the surface – primarily aggressive TiO_2. Even if the visible oxide is removed by pickling (etching) or mechanical polishing, there is always an invisible oxide layer that is < 100 angstroms thick.

NiTi also has two-way shape memory capability. This refers to the ability of the alloy to recover a pre-set shape upon heating above its TTRs and return to an alternate shape upon cooling. Creating two-way memory involves a somewhat complex training precess, but there are a number of limitations:

1. The amount of recoverable strain (~2%) is much lower than that achievable in one-way memory (6 to 8%).

2. The transformation forces on cooling are extremely low.

3. The memory can be erased with very slight over-heating.

4. The long-term fatigue and stability characteristics are not well known.

5. The temperature hysteresis between the heating and cooling transformations is still present.

Table 12.2: Nitinol properties

Physical Properties	
Density	6450 kg/m^3 (or 6.45 g/cm^3)
Melting temperature	1240 - 1310°C
Thermal conductivity (austenite)	0.18 W/cm.°C
Thermal conductivity (martensite)	0.086 W/cm.°C
Coefficient of thermal expansion (austenite)	11e-6 /°C
Coefficient of thermal expansion (martensite)	6.6e-6 /°C
Specific heat	0.2 cal/g.°C
Heat capacity	0.077 cal/g.°C
Corrosion resistance	Excellent. Comparable to 300 series stainless steel or titanium.
Transformation Properties	
Transformation temperature	-200 to +110°C
Latent heat of transformation	5.78 cal/g, 24.2 J/g
Transformation strain (1 cycle)	max 8%
Transformation strain (100 cycles)	6%
Transformation strain (100,000 cycles)	4%
Hysteresis (full austenite to martensite)	30 - 50°C
Electrical and Magnetic Properties	
Resistivity (austenite)	82 μohm.cm
Resistivity (martensite)	76 μohm.cm
Magnetic permeability	< 1.002
Magnetic susceptibility (austenite)	3.8 μemu/g
Magnetic susceptibility (martensite)	2.5 μemu/g
Mechanical Properties	
UTS (fully annealed)	895 MPa
UTS (work-hardened)	1900 MPa
Elongation at fracture (fully annealed)	25 – 50%
Elongation at fracture (work-hardened)	5 – 10%
Yield strength (austenite)	195 – 690 MPa
Yield strength (martensite)	70 – 140 MPa
Elastic modulus (austenite)	~75 – 83 GPa. Highly nonlinear with temperature.
Elastic modulus (martensite)	~28 – 41 GPa. Highly nonlinear with temperature.
Poisson's ratio	0.33
Hot workability	Quite good.
Cold workability	Difficult due to rapid work-hardening.
Machinability	Difficult. Abrasive techniques preferred.
Actuation	
Energy conversion efficiency	5%
Work output	1 J/g

Shape Setting

NiTi materials are generally processed by hot working (e.g. forging or hot rolling) followed by a series of cold-working processes, with a complete annealing cycle between cold-working steps. The final two processing operations for most NiTi

materials are a final cold-working step followed by a heat treatment. This establishes shape setting (or training). The heat treatment methods used to set shapes in both shape memory and superelastic forms are similar.

In general, temperatures as low as 400°C and times as short as 1 – 2 minutes can set the shape, but generally one uses a temperature between 450°C and 550°C (air furnace, inert atmosphere furnace, molten salt bath or fluidised bed), and times over 5 minutes. Rapid cooling of some form is preferred via a water quench or rapid air-cool.

Hysteresis

Both the hysteresis and the TTR (A_f – M_f) are slightly different for different NiTi alloys. Copper additions reduce hysteresis while Niobium (Columbium) additions expand hysteresis. Cold working and heat treatment have less dramatic but measurable effects on hysteresis.

Commercially Available Forms of Nitinol

Wires as small as <0.03mm and ribbon with thicknesses as thin as <0.03mm are available.

Microtubes with outer diameters as small as 0.2mm and as large as 3.2mm can be produced. The wall thickness should be >10% OD to avoid buckling.

Sheet with thicknesses down to 0.02mm and widths up to 90 mm – 100mm can be produced.

Rods, bars, stranded wire, braided wire, thin films, sputtering targets and other parts are also available.

12.6. ACTIVE DAMAGE CONTROL OF HYBRID MATERIAL SYSTEMS

Shape memory alloy, piezoelectric, electrostrictive and magnetostrictive actuators can form the basis of an induced strain actuator to control i) the propagation of cracks within hybrid composite systems and ii) fatigue damage of metallic and composite material structures [63].

Some experimental and numerical studies have been conducted to demonstrate the concept and mechanism of this active damage control technique.

Induced Strain Actuator Hybrid Composite Systems

The unique fabrication process of composite materials allows them to have embedded sensors and induced strain actuators.

An intelligent material system concept combines a composite material system with integral sensors, control processors and induced strain actuators designed to provide multifunctional capabilities including active control. The intelligent material system could be self damage sensing and be capable of self damage control or repair. Shape memory alloy hybrid composites have demonstrated in practice these unique self-sensing and damage control characteristics.

The class of composites referred to as SMA hybrid composites is a hybrid material system that contains SMA fibres (or films) in such a way that the material can be stiffened or controlled using heat (i.e. resistive heating).

The list of scientific areas that can be influenced by SMA reinforced composites includes active vibration control, active buckling control, motion and shape control, etc.

Active Damage Control Using Embedded SMA Induced Strain Actuators

The active damage control concept is to detect the occurrence of a crack with an embedded sensor system and prevent the crack from propagating or becoming dangerous by activating the embedded induced strain actuators. By changing the stress and strain field at the crack tip, or at a high stress and strain zone, the stress intensity factor could be reduced so that the propagation of the crack may be slowed or stopped.

The change in stress and strain field may also be used to reduce fatigue damage.

Active Damage Control Using Other Embedded Induced Strain Actuators

From the discussion above concerning shape memory alloy hybrid composite systems, the basic concept of active damage control using induced strain actuators is to reduce the stress and strain distribution at crack tips or around macroscopic high stress areas.

Piezoelectric, electrostrictive and magnetostrictive actuators are able to induce various amounts of strain upon activation. Piezoelectric and electrostrictive actuators generate an induced strain when an electric voltage or current is applied to them. Magnetostrictive actuators generate a deformation when subjected to a

magnetic field. These actuators can be either mounted or embedded into a composite structure similar to the SMA hybrid composites. Activation of these induced strain actuators will generate a new stress and strain field leading to changes in the stress and strain distribution of the structures.

These piezoelectric, electrostrictive and magnetostrictive actuators can also generate extension or contraction deformation depending on their activation and control schemes, such as the direction of the voltage. This may be used to increase fatigue life span of materials by reducing the cyclic stress amplitude.

Piezoelectric materials are generally not suitable for low cycle fatigue because:

1. their induced strain is not stable under a DC voltage - they have a tendency to creep with time.

2. the amount of induced strain and stress is still very limited – the maximum induced stress is about 60MPa.

13. APPENDIX B: HEAT TREATMENT PROCESSES FOR CARBON STEELS

The procedures listed in this chapter relate to the use of heat treatments to manipulate the γ-α and γ-α' transformations and thereby improve the physical properties of carbon steels [35], [67]. This chapter has been included as an appendix because the techniques described here do not directly relate to the specific thermo-mechanical procedures and alloy composition manipulations required to develop strain memory behaviour.

The more slowly a steel can be cooled from above the critical temperature in what amounts to a hardening operation, the less likelihood there is for distortion to take place. Usually a part made from cold-drawn steel will distort more in heat-treating because of cold-drawing strains than if it were made from the same steel with a hot-rolled finish. Some steels are also sensitive to the cooling medium. Most alloy steels, especially those that have been casehardened and those with over 0.3% carbon, must be cooled via oil-quenching; they are prone to excessive deformation or even cracking if quenched in water. Plain-carbon steels should be water-quenched if their maximum hardness is to be developed. The so-called *air-hardening* steels are mostly in the tool-steel category. Water-cooling rates are fast, oil-cooling rates are medium-paced, and air-cooling rates are slow.

Heat-treating specifications fit into one of the following categories:

1. *Hardness*. Resistance to wear and indentation. When ordinary carbon steel is fully hardened, the principal structure is martensite. The other products are usually the result of slower cooling leading to incomplete hardening.

   At this time it is important to note that in the case of TRIP steels it is essential that the amount of martensite be minimised and the austenite retained in order to enhance the magnetic sensor properties during deformation.

2. *Toughness*. Resistance to fatigue, tensile loads, torque and impact.

3. Combination of hardness and toughness.

The mass of a part affects its heat treatment; therefore care should be taken to avoid conflicting specifications. It is usually better to specify either the physical

properties desired or else the desired hardness rather than both, unless the ranges are sufficiently wide to take all variables into consideration.

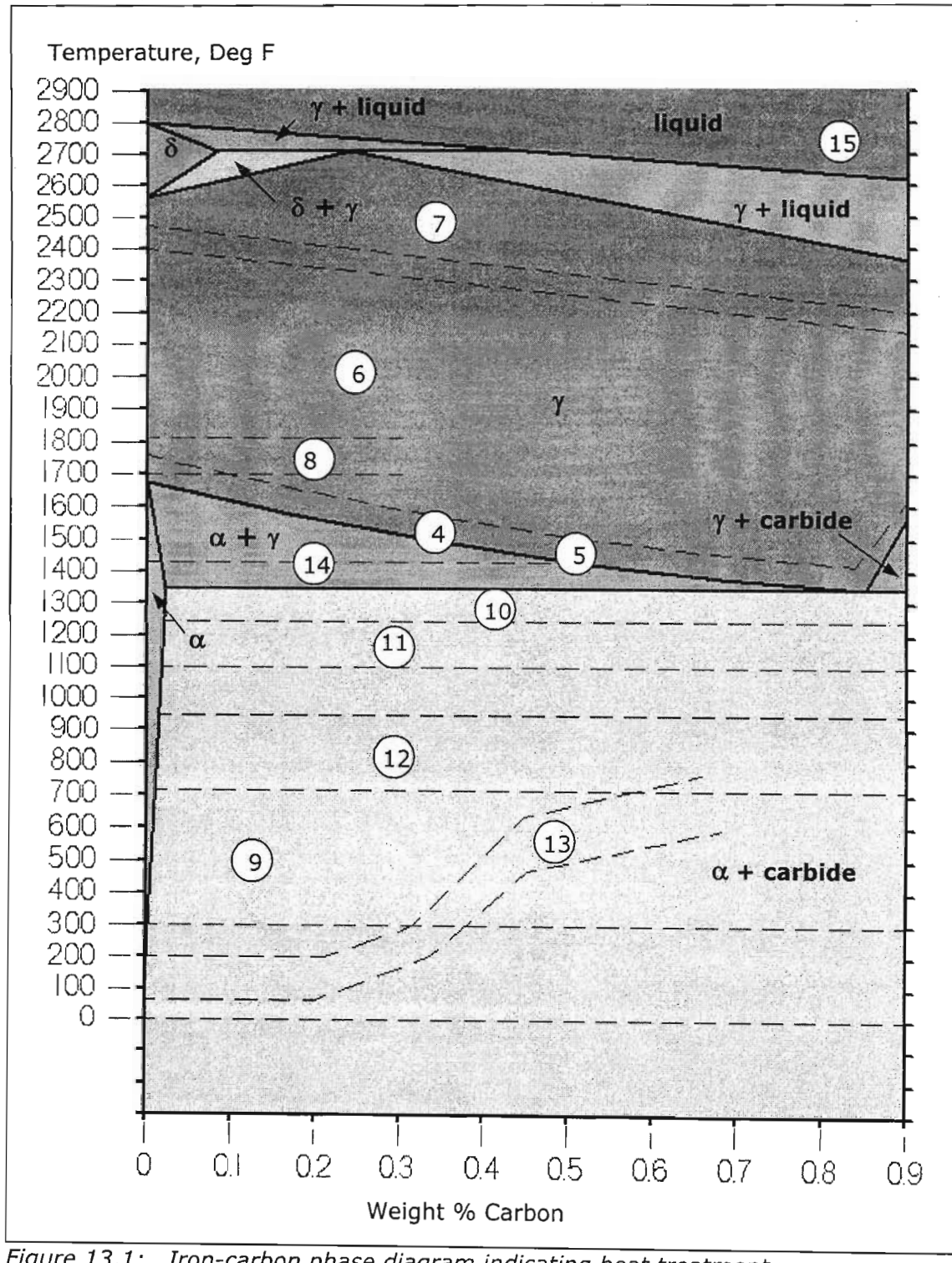


Figure 13.1: Iron-carbon phase diagram indicating heat treatment zones.

There are 6 basic heat treatments described in this section:

Non-quenching types:

1. Annealing
2. Normalising
3. Stress relieving

Quenching types:

4. Casehardening (carburising, carbonitriding, cyaniding and nitriding)
5. Surface hardening (flame and induction)
6. Through hardening

13.2. QUENCHING

Quenching describes the process of cooling from an elevated temperature (usually above A_{c3}) by contact with liquids, air or occasionally solids.

Stainless and high alloy steels may be quenched to minimize the presence of grain boundary carbides (i.e. promote intragranular rather than intergranular carbide precipitates) or to improve the ferrite distribution. Most steels including plain-carbon, low alloy and tool steels are quenched to produce controlled amounts of martensite in the structure.

13.3. ANNEALING

Item 4 in Figure 14.1. Annealing is performed to soften a material. For a *full* anneal, plain-carbon steels are held slightly above A_{c3} until austenite forms (a process called austenitising), then *slowly* cooled in a furnace to form ferrite and pearlite, or pearlite and cementite, at room temperature. The transformation results in a coarse pearlite that provides relative softness, good ductility, and good machinability.

Annealing to remove strain hardening (i.e. plastic deformation) is called *process annealing* or *isothermal annealing*. Here the steel is cooled to just below A_{r3} where it is allowed to isothermally form ferrite and coarse pearlite.

Annealing can be applied to all steels but is generally performed on those with a carbon content over 0.35% if a requirement is that they be soft enough to machine readily. This treatment reduces the tensile strength and the elastic limit, while at the same time the elongation and reduction of area are increased. Because of this it is often desirable that alloy parts be subsequently suitably heat-treated to develop their inherent physical properties; plain-carbon steels are often used in the annealed condition.

Spheroidise annealing (item 10 in Figure 13.1) consists of cooling the steel just below the critical temperature range for sufficient time for the cementite lamellae of pearlite to adopt a globular form that is about twice as tough, soft with good machinability. A sphere has the minimum surface area per unit volume; therefore a driving force exists to spheroidise the minor phase of a 2-phase microstructure.

13.4. NORMALISING

Normalising (item 5 in Figure 13.1) is an austenitising heating cycle to between 50°C and 100°C above the A_{c3} temperature followed by cooling in still or slightly agitated air. Normalising is performed to homogenise alloy steels (i.e. promote microstructural uniformity). The cooling is slightly quicker than in annealing to refine the grain size and form a relatively fine pearlite. The resulting microstructure has somewhat higher strength and hardness and slightly less ductility than obtained with annealing.

Normalising is commonly applied to ingots prior to working, and to steel castings and forgings prior to hardening. The medium carbon grades (0.35% to 0.55% carbon) are generally more responsive than the low-carbon grades (under 0.25% carbon).

13.5. STRESS RELIEVING

Residual stresses due to processes such as quenching, rolling, casting, forging, bending, grinding, welding, etc. may result in distortion and possible cracking of the steel. Stress relief is performed to allow the steel to relieve itself of these locked-up stresses. The process involves heating to a range below the lower critical temperature (lower than the recrystallisation or A_1 temperatures because it is not necessary to relocate large numbers of atoms), holding for 1 hour or more per inch of thickness, and then slowly cooling.

13.6. TEMPERING

A previously hardened or normalised steel is heated to a temperature below the critical temperature for one hour or more, and then cooled at a suitable rate. This process is performed primarily to increase ductility and toughness, but also to increase grain size.

The initial hardening process determines the specific material values; subsequent tempering relieves the quenching stresses and ensures dimensional stability. Increasing the tempering temperature increases toughness and shock resistance, but reductions in hardness and strength are tradeoffs.

A fully hardened steel is 100% martensitic and in its strongest possible condition, but freshly quenched martensite is brittle. The microstructure of such a quenched and tempered steel is referred to as *tempered martensite*. Because of the non-cubic structure, and because carbon is trapped in the lattice, slip does not occur readily in martensite. Therefore, martensite is strong, hard and brittle. This enhanced hardness is of major engineering importance, since it produces a steel that is extremely resistant to abrasion and deformation. However, martensite is too brittle to use in almost every application. It is therefore usually tempered so that α + carbide forms. The resulting microstructure is not lamellar like pearlite but contains many dispersed carbide particles because there are numerous nucleation sites within martensitic steel. Tempered martensite is much tougher than the metastable martensite, making it a more suitable for most applications, even though it may be slightly softer. The microstructure of tempered martensite becomes coarser with more tempering time or higher tempering temperatures.

13.7. THROUGH HARDENING

Refer to Figure 14.2. The steel is heated to a point above the critical temperature, quenched, and then tempered to obtain specific values of hardness, toughness or some combination thereof. 3 types of heating furnaces are used for tempering: oven, pit, and liquid bath heated with gas, oil, or electricity. Medium-carbon alloy steels are generally used for hardnesses up to about Rockwell C 50; for higher hardnesses, special oil- and water-hardening, higher-carbon steels are used.

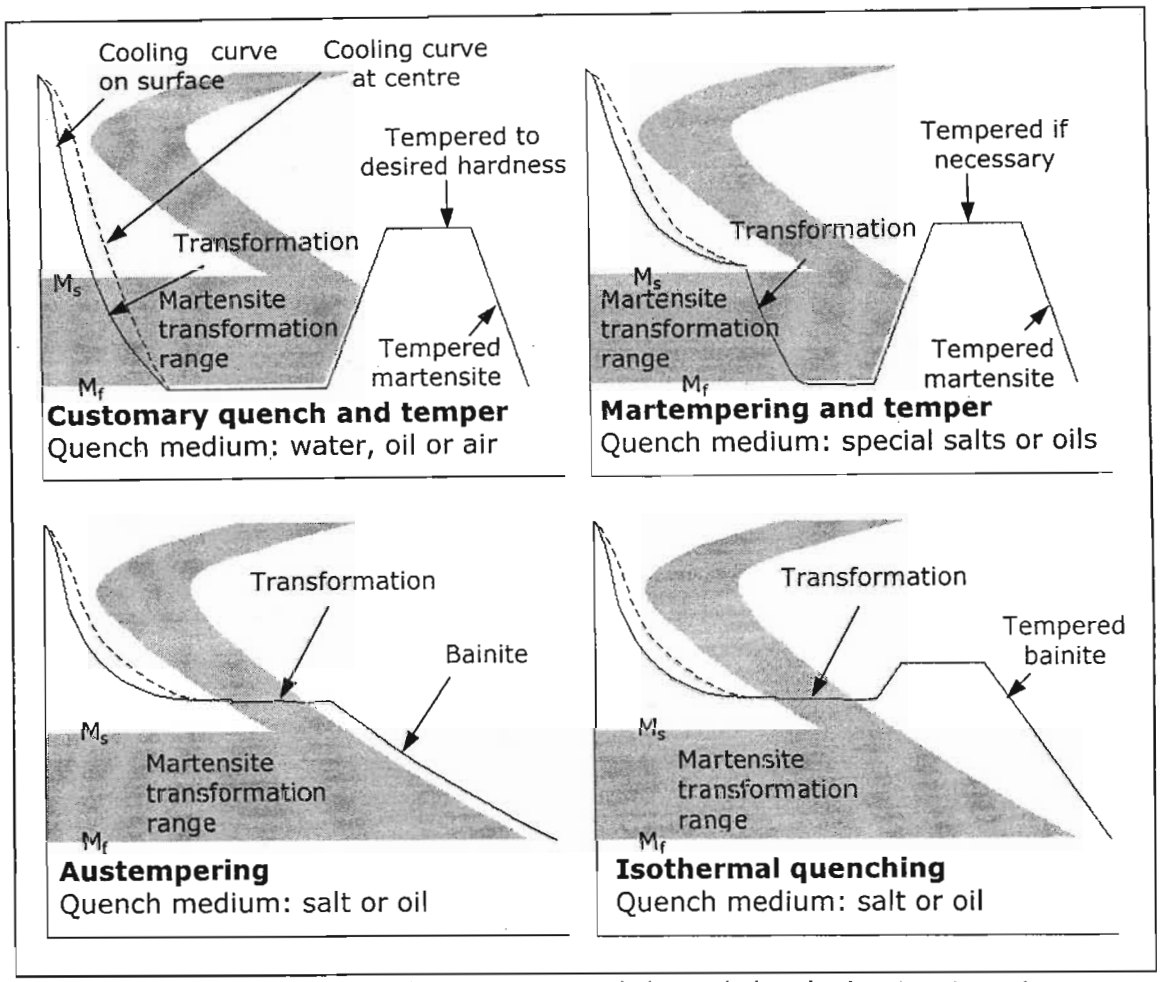


Figure 13.2: TTT diagram showing 4 typical through-hardening treatments.

Direct Quenching

This process consists of heating the steel above the critical temperature and quenching, usually in water, oil, or air so that martensite forms (first in the surface, then in the centre). The part is then tempered to the desired physical properties. Oil provides a slower and less drastic cooling operation than water and therefore usually results in less distortion. For this reason, if a part is intricate or if considerable toughness is required, the oil quench should be used, in which case an alloy steel is preferred because of better response than the plain-carbon grades. The highly alloyed air-hardening steels sold by various tool-steel companies are specially designed for minimum distortion. Resistance to wear combined with low distortion is usually the principal reason for employing an air-hardening steel on special machine parts, rolls, etc.

Interrupted Quenching

Refer to Figure 5.1. As the name implies, these treatments involve interrupting the quench after the steel has been heated to a point above the critical temperature.

Time is available for the surface and centre to transform nearly simultaneously, thus avoiding the cracks and distortion that result from direct quenching. They are best suited to relatively small parts. Three types of interrupted-quench treatments are utilised:

1. *Martempering* involves heating certain alloy, cast, tool and stainless steels above the critical temperature and quenching in a suitable salt or oil bath, which is maintained at a temperature just above martensitic formation until the temperature is equalised throughout the piece. The part is then cooled in air to form martensite and tempered to the desired hardness. Distortion, cracking and residual stress is reduced.

2. *In Austempering*, ferrous alloys are isothermally transformed at a temperature below that for pearlite formation and above that of martensite formation. The steel is initially heated to form austenite, then quenched in a suitable salt or oil bath to form bainite, and finally cooled to room temperature. The process results in increased ductility, toughness and strength at a given hardness. Distortion is reduced and overall cycle times are short.

3. *Isothermal treatment* is performed on slightly bigger parts than austempering. After being isothermally quenched the resulting bainite is subsequently tempered before the steel is cooled to room temperature.

13.8. CASEHARDENING

The very rapid diffusion of carbon and nitrogen in iron compared with that of the metallic alloying elements is readily exploited to harden the surface layer of a steel. In large sections (over ~3 inches) alloy steels tend to get a harder and more uniform surface than plain-carbon steels. The process is performed when a ductile core with a wear-resistant case is required.

Carburising

Item 8 in Figure 14.1. Carbon is added to the outer layer of a steel by a pack, gas or liquid process at a temperature where austenite has a high solubility for carbon. The steel is held long enough to obtain the desired penetration (case depth is a function of time and temperature). After carburising, the material is usually quenched to produce a hardened, high-carbon, martensitic case that is strong and

wear resistant. Anticipated distortion of the part and the amount of subsequent grinding along with service conditions will determine the depth of case required.

Carbonitriding

The steel (usually a low-carbon variety) is heated in an atmosphere that adds carbon and nitrogen simultaneously to the surface. The subsequent cooling rate determines the hardness. Distortion is usually less than experienced by carburising or cyaniding.

Cyaniding

The steel is heated in molten cyanide, which adds nitrogen and some carbon to the surface. It is then usually quenched to produce a hard and brittle, shallow case.

Nitriding

Nitriding consists of heating certain special steels (usually contain ~1% chromium, ~1% aluminium and ~0.2% vanadium or molybdenum which are nitride-forming elements) in the ferrite phase about 500-550°C for long periods in the presence of Ammonia gas and cooling without quenching. Nitrogen is absorbed into the surface to produce an extremely hard "skin" ranging up to Rockwell C 70. The case is also brittle because of the nitrides, and hence the process is more suited to parts that must principally resist wear rather than impact. Nitriding is expensive both because of the equipment and because of the extended period of time (2 to 4 days) required. An advantage to the process is that distortion is negligible because quenching is not involved. The most commonly used steels for nitriding are those marketed under the name Nitralloy, although some medium-carbon alloy steels can be used. The medium-carbon alloys produce a case somewhat less hard and deep. In most instances the steel is heat-treated to prescribed physical properties before nitriding.

13.9. SURFACE HARDENING

This is a form of casehardening that differs from the foregoing methods in that the surface of the steel is heated directly to a point above the critical temperature and quenched. It is usually performed on steels with a sufficiently high carbon content (usually over 0.30% carbon) that will respond to heating and quenching without the preliminary procedure of adding carbon or nitrogen to the surface. Occasionally this process is performed on steels that have been carburised. Normally distortion is very slight on surface-hardened parts, but there is apt to be considerable

distortion if the parts are not treated as symmetrically as possible. Surface hardening is an extremely rapid process, generally requiring only a few seconds to perform the actual treatment.

Flame Hardening

This involves rapidly heating the surface of the steel to the desired depth above the critical point by means of a flame and cooling as desired. Alloy steels parts are generally conventionally heat-treated to prescribed mechanical properties before flame hardening. The hardness of the case depends on the steel and the quenching medium.

Induction Hardening

This involves rapidly heating the surface of the steel to the desired depth above the critical point by means of electrical induction and cooling as desired. The case depth, hardness, etc., resulting from this treatment are similar to those of flame hardening.

13.10. FORGING

Item 6 in Figure 13.1. Mostly single-phase metals can be shaped by plastic deformation (i.e. above the yield strength). At temperatures above A_{c3} the material is softer and more ductile. Less energy is therefore required for deformation and there is less chance of fracturing during processing. Rolling, forging and extrusions are among the hot deformation processes. Care must be taken not to exceed the maximum forging temperature, or burning of the steel results (item 7 in Figure 13.1.).

13.11. PRE-HEATING FOR WELDING

Item 13 in Figure 13.1. Carried out to prevent crack formation. Welding of higher carbon steels leads to the easier formation of cracks within the weld zone, so it is usually necessary to limit the carbon content to not greater than 0.2% if welding is to be performed. In these circumstances, additional strength can be obtained through solid solution hardening by adding between 1% and 1.5% manganese.

13.12. MELTING

The temperature of the steel exceeds the melting temperature in the region labelled as item 15 in Figure 13.1.

14. REFERENCES

1. Haile, Andrew Thurlo, A Mechanistic Evaluation and Design of Tunnel Support Systems for Deep Level South African Mines, PhD. Dissertation, 1999.

2. Roberts, D.P., Testing of Mining Tunnel Support Elements and Systems for Hard Rock Mines, MSc. Dissertation, 1995.

3. Apsey, Jason, A Finite Element Study of Grouted Rockbolt Behaviour in a Perpendicular Slip Environment, MSc. Dissertation, 1997.

4. Hanagud, S., NageshBabu, G.L., and Zhang, J, Smart Structures – An Overview, Developments in Theoretical and Applied Mechanics, Volume 16, University of Tennessee Space Institute, 1992.

5. Thompson L., Westermo B., Tominaga M., Monitoring Based Maintenance for Civil Engineering Structures, Proceedings of the International Conference on Structural Safety and Reliability, 2001.

6. Duncan, N., Engineering Geology and Rock Mechanics, Volume 1, Leonard Hill, 1969.

7. Vutukuri, V.S., Lama, R.D., Saliya, S.S., Handbook on Material Properties of Rocks, Volume 6, Series on Rock and Soil Mechanics, Trans Tech Publications, 1974.

8. Carter, M., Geotechnical Engineering Handbook, Pentech Press, 1983.

9. Atkinson, J.H., Bransby, P.L., The Mechanics of Soils: An Introduction to Critical State Soil Mechanics, University Series in Civil Engineering, McGraw-Hill Book Company Ltd, 1978.

10. Barrow, K., Development of Rock Failure Around a Single Haulage Way, National Mechanical Engineering Research Institute, Council for Scientific and Industrial Research, Rock Mechanics Division, 1966.

11. Ortlepp, W.D., The Design of Support for Rockburst Prone Tunnels, Rock Mechanics in the Design of Tunnels, SANGORM Symposium, 1983.

12. Wagner, H., Some Considerations Concerning the Support of Tunnels in Deep Mines, Rock Mechanics in the Design of Tunnels, SANGORM Symposium, 1983.

13. Piper, P.S., The Application and Evaluation of Rock Mass Classification Methods for the Design and Support of Underground Excavations at Depth. Rock Mass Characterisation Symposium, SANGORM, MINTEK, 1985.

14. Wiseman, N., Factors Affecting the Design and Condition of Mine Tunnels, COMRO Report, Project No. G01G10, 1979.

15. Wojno, L.Z., Jager, A.J., Roberts, M.K.C., Recommended Performance Requirements for Yielding Rock Tendons, Design of Rock Reinforcing, Components and System Symposium, 1987.

16. Jager, A.J., Wojno, L.Z., Henderson, N.B., New Developments in the Design and Support of Tunnels Under High Stress Conditions, International Deep Mining Conference, Johannesburg, SAIMM, 1990.

17. Stillborg, B., Professional Users Handbook for Rockbolting, Trans Tech Publications, 1986.

18. Lang, T.A., Bischoff, J.A., Stabilisation of Rock Excavations Using Rock Reinforcement, Proceedings of the 23rd US Symposium on Rock Mechanics, Berkely, California, AIME, 1982.

19. Schach, R., Garschol, K., Heltzen, A.M., Rockbolting: A Practical Handbook, Pergamon, Oxford, 1979.

20. Barton, N., Lien, R., Lunde, J., Engineering Classification of Jointed Rock Masses For the Design of Tunnel Support, Rock Mechanics, Volume 6, 1974.

21. Barton, N., Workshop on Rock Mass Characterisation, Rustenburg, 1997.

22. Anon, An Industry Guide to Methods of Ameliorating the Hazards of Rockfalls and Rockbursts, Chamber of Mines Research Organisation, Johannesburg, 1988.

23. Beniawski, Z.T., Engineering Classification of Rock Masses, Transcripts of the South African Institute of Civil Engineering, Volume 15, 1973.

24. Beniawski, Z.T., The Geomechanics Classification in Rock Engineering Applications, Proceedings of the 4th Congress of ISRM, Montreux, 1979.

25. Stimpson, B.A., A Simplified Conceptual Model for Estimating Roof Bolting Requirements, International Journal of Mining and Geotechnical Engineering, Volume 7, 1989.

26. Laubscher, D.H., Taylor, H.W., The Importance of Geomechanics Classification of Jointed Rock Masses in Mining Operations, Proceedings of the Symposium on Exploration for Rock Engineering, Volume 1, Johannesburg, 1976.

27. Choquet, P., Charette, F., The Applicability of Rock Mass Classifications in the Design of Rock Support in Mines, Proceedings of the 15th Canadian Symposium on Rock Mechanics, Toronto, 1988.

28. Farmer, I.W., Shelton, P.D., Factors That Affect Underground Rockbolt Reinforcement Systems Design, Transcripts of the 89th Institution of Mining and Metallurgy, London, 1980.

29. Kaiser, P.K., McCreath, D.R., Tannant, D.D., Canadian Rockburst Support Handbook, Geomechanics Research Centre, Laurentian University, Ontario, 1996.

30. Vandiver, T.L., Health Monitoring of U.S. Army Missile Systems, U.S. Army Missile Command.

31. Boller, C. and Biemans, C., Structural Health Monitoring in Aircraft – State-of-the-Art, Perspectives and Benefits, Structural Health Monitoring: Current Status and Perspectives, Proceedings of the International Workshop on Structural Health Monitoring, Stanford, 1997.

32. Bartelds, G., Aircraft Structural Health Monitoring, Prospects for Smart Solutions from a European Viewpoint, National Aerospace Laboratory, The Netherlands.

33. Spillman Jr, W.B., Sirkis, J.S. and Gardiner, P.T., The Field of Smart Structures as Seen by Those Working in it: Survey Results, Journal of Smart Materials and Structures, 1996.

34. Thompson, L., Westermo, B., Tominaga, M., Monitoring Based Maintenance for Civil Engineering Structures, International Conference on Structural Safety and Reliability, Newport Beach, 2001.

35. Metals Engineering - Processes, ASME Handbook, First Edition, McGraw-Hill Book Company, 1958.

36. van Vlack, Lawrence H., Elements of Materials Science and Engineering, 6th Edition, Addison-Wesley Publishing Company, 1989.

37. Honeycombe, R.W.K., Steels: Microstructure and Properties, Edward Arnold, 1981.

38. Olson, G. B., Cohen, M., A General Mechanism of Martensitic Nucleation: Parts I, II and III, General Concepts and the FCC to HCP Transformation, Metallurgical Transcripts, Volume 7A, 1976.

39. Koppenaal, T.J., A Thermal Processing Technique for TRIP Steels, Metallurgical Transcripts, Volume 3, 1972.

40. Koppenaal, T.J., Research in Development of Improved TRIP Steels, Army Materials and Mechanical Research Centre, 1973.

41. Baik, S.C., Kim, K., Jin, Y.S., Kwon, O., Effects of Alloying Elements on Mechanical Properties and Phase Transformation of Cold-Rolled TRIP Steel Sheets, ISIJ international, Volume 41, 2001.

42. Strain Monitor Systems, Inc., Non-Destructive Bridge Evaluation Technology: Bridge Monitoring by Passive Smart-Sensors, Report Prepared for US Department of Transportation, 1999.

43. Zackay, V.E., Parker, E., Fahr, D., Busch, R., The Enhancement of Ductility in High-Strength Steels, Transcripts of ASM, Volume 60, 1967.

44. Thompson L., Westermo B., Crum D., Law W. and Trombi R., Smart Structural Fasteners For Aircraft And Construction Industries, Proceedings of the SPIE Conference on Smart Structures and Materials, Newport Beach, 1999.

45. Thompson L., Westermo B. and Waldbusser R., Smart-Materials-Based Structural Health Monitoring Systems Development for Aerospace

Applications, Proceedings of the 42nd International Instrumentation Symposium, 1996.

46. Thompson L., Westermo B., Development of Smart Structural Attachment Fixtures, Proceedings of the North American Smart Structures and Materials Conference, 1996.

47. Thompson L., Westermo B., Cost-Benefit Analysis for Utilisation of Strain-Memory Structural Materials in Civil Engineering Applications.

48. Thompson L., Westermo B., Peak Strain Sensing for Civil Structures, Technology Development Report.

49. Thomas, George B. and Finney, Ross L., Calculus and Analytic Geometry, 7th Edition, Addison-Wesley Publishing Company, 1988.

50. Spiegel, Murray R., Theory and Problems of Theoretical Mechanics, McGraw-Hill Book Company, 1982.

51. Dym, Clive L., Shames, Irving H., Solid Mechanics: A Variational Approach, McGraw-Hill, Inc., 1973.

52. Zienkiewicz, O.C., Taylor, R.L., The Finite Element Method: Basic Formulation and Linear Problems, 4th Edition, Volume 1, McGraw-Hill Book Company, 1989.

53. Grandin, Hartley R., Fundamentals of the Finite Element Method, Macmillan Publishing Company, 1986.

54. Fagan, M.J., Finite Element Analysis: Theory and Practice, Longman Scientific and Technical, 1992.

55. Sears, Francis W., Zemansky, Mark W. and Young, Hugh D., University Physics, 7th Edition, Addison-Wesley Publishing Company, 1987.

56. Schwarz, Steven E., Oldham, William G., Electrical Engineering: An Introduction, Holt, Rinehart and Winston, Inc., 1984.

57. Grossman, Stanley I., Elementary Linear Algebra, 3rd Edition, Wadsworth Publishing Company, 1987.

58. Plonus, Martin A., Applied Electromagnetics, McGraw-Hill Book Company, 1988.

59. Grant, I.S., Phillips, W.R., Electromagnetism, 2nd Edition, John Wiley and Sons, 1990.

60. Silvester, P.P., Ferrari, R.L., Finite Elements for Electrical Engineers, Cambridge University Press, 1983.

61. Sadiku, Matthew N.O., Numerical Techniques in Electromagnetics, CRC Press, 1992.

62. Peelamedu, S.M., Naganathan, N.G. and Buckley, S., Impact Analysis of Automotive Structures With Distributed Smart Material Systems, MIME Department, University of Toledo.

63. Rogers C. A., Liang C. and Li S., Active Damage Control Of Hybrid Material Systems Using Induced Strain Actuators, Proceedings of the 32nd AIAA Conference on Structures, Structural Dynamics and Materials, 1991.

64. Hodgson, D.E., Using Shape Memory Alloys, Shape Memory Alloys, Inc., 1988.

65. Birman, Victor, Theory and Comparison of the Effect of Composite and Shape Memory Alloy Structures, Applied Mechanics Reviews, Volume 39, 1997.

66. Birman, Victor, Review of Mechanics of Shape Memory Alloy Stiffeners on Stability of Composite Shells and Plates, International Journal of Mechanical Science, Volume 50, 1997.

67. Heat Treater's Guide: Practices and Procedures for Irons and Steels, 2nd Edition, ASM International, 1995.

68. www.piezo.com

69. www.kistler.com

70. www.apc.thomasregister.com

71. www.etrema-usa.com

72. www.websites.co.uk/newlands_technology/

73. www.mtssensors.com

74. www.shokubai.co.jp/english

75. www.lehigh.edu

76. www.fluiddynamics.com

77. www.vilastic.com

78. www.thermodynamik.tu-berlin.de

79. www.sma-inc.com

80. www.sma-mems.com

81. www.physikinstrumente.com

82. www.newport.com

83. www.luec.leeds.ac.uk/ElectrostrictiveMaterials

84. Naylor, D.J., Pande, G.N., Simpson, B., Tabb, R., Finite Elements in Geotechnical Engineering, Pineridge Press, 1981.

85. Herget, G., Stresses in Rock, A.G. Balkema Publishers, 1988.