Dynamic Recrystallization during Strip Rolling of HSLA Steels and Prediction of Roll Forces using Artificial Neural Networks

by

Jalal Biglou

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Mechanical Engineering

Waterloo, Ontario, Canada, 1997

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below and give the address and data.

Acknowledgements

The author would like to express his sincere thanks to all the members of the Department of Mechanical Engineering, University of Waterloo, who helped me throughout this research and made it possible and enjoyable.

I am deeply grateful to have had a knowledgeable and encouraging advisor who was able to guide this research to be in-line with the current research in the material processing field and to be useful to the steel industry. For this, I am greatly indebted to Dr. John Lenard for all his ideas, guidance, support, and confidence in me.

Sincere thanks go to our knowledgeable and always helpful lab supervisor, Richard Gordon, fellow graduate students, and staff in the Mechanical Engineering Department. In particular, I would like to thank George Hwu, Myung-Sik Chun, Per Munther, Andy Bugg, John Fitzpatrick, Jim Baleshta, Steve Hitchman, Fred Birckman, Ray White, and Colin Campbell. Special thanks also go to Mahta and Shahyar who helped me in the most critical time of this thesis, its completion.

During last four months I have been very fortunate to be a part of Quad Engineering team. I am grateful to have had supporting and encouraging colleagues at Quad. In particular, I would like to thank Mike Levick and Joseph Pataki.

The excellent financial support of this research and my personal expenses were provided by the Manufacturing Research Corporation of Ontario which is highly appreciated. A big part of this research was carried out in the Research and Development Department of Dofasco Inc. which proved to be invaluable to me. In this regard, I would like to extend my special thanks to Brian Nelson, Dave Hall, and Ian O'Reilley.

Finally, I am extremely fortunate to have had all the support from my family, particularly my parents and my wife, who helped me attain this degree with their emotional support, understanding, and counsel.

To my wife,

with love

Abstract

The softening mechanisms taking place during hot rolling of steels influence both the mechanical properties of the final product and the steel flow stress during deformation. The knowledge of the material's constitutive behavior is an essential requirement for the design and control of rolling processes. Steel manufacturers are looking for more accurate models being able to predict the material deformation resistance, micro-structural evolution of steel, and roll forces in order to produce strips with a more consistent output gauge and mechanical properties. In this regard; the occurrence of dynamic recrystallization during strip rolling of HSLA steels and its effects on the flow stress, roll forces, and final properties are of importance. The occurrence of dynamic recrystallization during hot strip rolling still remains controversial.

In this research , the experimental techniques were used to simulate the whole rolling process to study the occurrence of dynamic recrystallization. Axisymmetric compression tests were used to study the kinetics of static recrystallization. Torsion simulations were performed to verify the occurrence of dynamic recrystallization. An industrial mill log was analyzed which further confirmed the occurrence of the dynamic recrystallization and torsion test results.

In spite of drawbacks in terms of ease of development, adaptability, accuracy and speed, empirical stress-strain relationships and traditional roll force equations, along with look-up tables, are being commonly used. In this research, a Neural Network simulator code, based on the gradient descent learning rule, was developed. This code was used to predict the steel and aluminum flow stresses at high temperatures and strain rates, experimental rolling forces during cold and hot rolling of aluminum strips, and rolling forces during industrial strip rolling of a high Nb HSLA steel. The model predictions were compared to those of the statistical models and existing on-line industrial models. The approach based on Neural Networks is shown to be superior in terms of accuracy, speed, and ease of development.

Principal Component Analysis was used as a data pre-processor to remedy data deficiencies when there is an excessive linear correlation between input variables of a database.

This analysis was integrated to the Neural Network simulator code in order to decouple linearly correlated input data. The code was applied to an industrial hot rolling database to develop a model to predict the occurrence and the effects of dynamic recrystallization on the rolling forces. This model not only predicts the occurrence of dynamic recrystallization, but also predicts the extent of the resulting softening. The results also proved the beneficial effects of the integration of Principal Component Analysis with Neural Network modelling.

Table of Contents

List of Tables

List of Figures

1. Introduction

Development and thermo-mechanical processing of high-strength low-alloy (HSLA) steels have been the biggest milestones of the steel industry in the last two decades. With the addition of a small amount of carbide-former alloying elements and control of temperature and deformation during processing, the yield stress of a plain carbon steel can be increased from 350 MPa to 550 MPa with an improvement in toughness. This increased strength per unit weight, has maintained the status of steel as the material of choice for cars, home appliances, off-road vehicles, bridges, pipelines, etc. There are two significant features of thermo-mechanical processing which have contributed the most to the success of HSLA steels. First, the beneficial increase in strength does not harm toughness or weldability. Second, the technology of producing these steels is inexpensive because the alloying elements are only needed in very small concentrations and associated improvements in mechanical properties precludes the need for further heat treatment.

The HSLA steels gain their strength and toughness from refined ferrite grains and fine carbo-nitride precipitates. Grain refinement is achieved through different mechanisms of repeated static recrystallization, possible dynamic or metadynamic recrystallizations, or transformation to smaller ferrite grains from pancaked austenite grains. Another source of the strength of HSLA steels emanates from fine precipitates. A large number of scientific research has been conducted to investigate the mechanisms and kinetics of both recrystallization and precipitation and their interactions. This knowledge is vitally important for two reasons. First, the optimal design of draft schedule in order to benefit the most from both precipitation and

recrystallization requires a clear and quantitative understanding of these processes and their complex interactions. Second, the occurrence and kinetics of precipitation and recrystallization during the processing of HSLA steels affects the steel's flow strength. An accurate prediction of steel hot flow stress during rolling is crucially important for roll gap set-up and finish product gauge uniformity. Gauge consistency is one of the main characteristics of the high quality steel strips. Inconsistency in gauge or mechanical properties is detrimental to downstream processing and stamping of steel strips. Automobile manufacturers and stamping shops are continually demanding tighter tolerances.

In spite of the vital importance of the possible softening during HSLA strip rolling, the very nature of the occurrence or non-occurrence of dynamic recrystallization in this process is under question. Experimental verification of this question is notoriously difficult, if not impossible, with existing equipment. A valid experiment would require multi-stage deformations followed by quenching at a rate of 2000 °C per second. However, there are indirect techniques for the study of dynamic recrystallization. Flow stress-strain curves are commonly used to study work-hardening or softening. A major problem with this kind of mechanical testing result is again experimental limitations. Given these facts, no quantitative model has been developed yet which is able to predict the conditions under which dynamic recrystallization is taking place. More importantly, a model is required to predict the amount of this softening, if any, under different processing conditions.

This research work was partly devoted to the physical simulation of the entire rolling process to shed more light on the dynamic recrystallization issue. Axisymmetric compression tests were used to study the kinetics of static recrystallization. Torsion simulations were performed to verify the occurrence of dynamic recrystallization.

Recent advancements in instrumentation, data acquisition, and data management systems have enabled steel mills to collect a large amount of process data. This has created a whole new perspective for rolling process analysis and for the study of the metallurgical phenomena-taking place during hot rolling. Mill log data of the Dofasco Inc. hot strip mill was

analysed in detail to verify the occurrence of dynamic recrystallization during high niobium HSLA strip rolling. The results of this analysis have been compared with laboratory simulations.

Traditionally, linear and non-linear regression analyses and other statistical techniques have been the only available tools for process modelling applications. However, the applicability and performance of these techniques are very much dependent on the availability of the appropriate mathematical equations which can describe sufficiently the process input-output relationship. The general forms of these equations are usually derived from a physical understanding of the process, then optimal coefficients and exponents are found through regression. Therefore, for the more complex multi-variate processes, where there is no form of mathematical equations capable of describing the whole process, regression analysis faces major shortcomings. Unless a capable modelling tool is available, the collected process data cannot be utilised to any significant level.

During the last decade, a new modelling technique has been evolved. This heuristic technique is called Artificial Neural Networks (ANN). It was developed with inspirations from the human nervous system and cognition. A Neural Network learns from experiential knowledge and once it has learned, it can predict the output of the process without assuming or making any rules. This technique extracts the possible trends in the data, linear or non-linear, without resorting to any preconceived form of mathematical equations. This is exactly how a human brain learns and functions. The brain does enormous amount of data processing, most of the time without even one simple arithmetic operation.

There is no universally accepted definition of Artificial Neural Networks. However, common to all Neural Networks is a network of simple processors, the neurons, each of them having multiple inputs (from outside or from other neurons) and a single output, that is read out or can be connected to other neurons[1]. There is a weight (a numerical value) associated with each of the connections. Each neuron does quite simple arithmetic operations. It adds all the products of the output signals of the pervious neurons and the corresponding connection

weights. Then it compares this value with a threshold value and applies a non-linear activation function to the result and sends out one output to the next neuron or output node. This simple neuron, similar in functionality to the biological ones, is not able to do much by itself. However, a network of these neurons, organised in layers, is able to perform quite demanding classification and modelling tasks.

All Neural Networks have some sort of training rule whereby the weights of connections are adjusted on the basis of data. In other words, Neural Networks first learn their knowledge from examples (in the same way that a child acquires knowledge) then they generalise beyond the training data. Once sufficient numbers of data are presented, Neural Networks learn the linear and non-linear trends in the data. Hence, a Neural Network can be looked on as a non-linear procedure that maps inputs of the process to outputs. In other words, Neural Networks are universal function approximators.

Various types of Neural Networks differ by the way these neurons are connected (configuration) and by the learning procedure by which the weights are adapted. The most popular and by far the most capable Neural Network is the feed-forward network with back-propagation learning. The mathematical basis of this kind of network and its applications to the flow stress modelling and roll force predictions are presented in this thesis.

Neural Network paradigm has been applied successfully to many demanding problems in the fields of engineering and science. However, it is somewhat new to material science and processing. One of the main focuses of this research work was to explore the ways that Neural Networks can be utilised as a modelling tool for material scientists and steel mill engineers. A Neural Network simulator code was developed and used to model the rolling forces during hot strip rolling of an HSLA steel, and was found to outperform the existing on-line model.

A problem associated with most industrial process databases is the co-linearity between the process inputs. These colinearities, if are not taken into account, can cause model instability and vagueness. For statistical modelling of data with colinearities biased regression

techniques have been developed. However, for the Neural Network modelling of the data with colinearities, the author proposes pre-processing of the data through transformation into the principal component space. Principal Component Analysis is a multivariate analysis technique that detects the linear correlations between input variables of a process and decouples them through an orthogonal transformation. Upon this transformation, the original variables are transformed into principal components which are inherently linearly independent of each other. Principal Component Analysis was integrated into the Neural Network simulator code as a data pre-processor. The case studies demonstrated the increased accuracy of the models developed based not on the original variables, but based on the principal components.

2. Literature Review

The final properties and rolling quality of HSLA steels are markedly influenced by the nature of metallurgical processes taking place during and after deformation. In this chapter, a review of metallurgical concepts relevant to hot deformation of steels will be presented. Then, applications of these metallurgical events in thermo-mechanical processing of steels will be discussed. Finally, modelling of hot rolling processes and important models being used are given.

2.1 Hot Deformation Metallurgy of Steels

In this section the principles of events occurring in the process of hot deformation of steels will be reviewed, in chronological order,. A schematic of these microstructural changes, i.e., dynamic recovery and recrystallization (during deformation) and static recovery and recrystallization (after deformation) has been presented in **Figure 2-1** and will be reviewed in the following sections.

2.1.1 Metallurgical Phenomena During Hot Deformation

2.1.1.1 Work Hardening

The increased dislocation density associated with the bulk deformation of a material is responsible for work hardening; that is, dislocations themselves are obstacles to dislocation motion. Depending on the degree of interaction experienced between moving dislocations, dislocations can be either soft or hard obstacles.

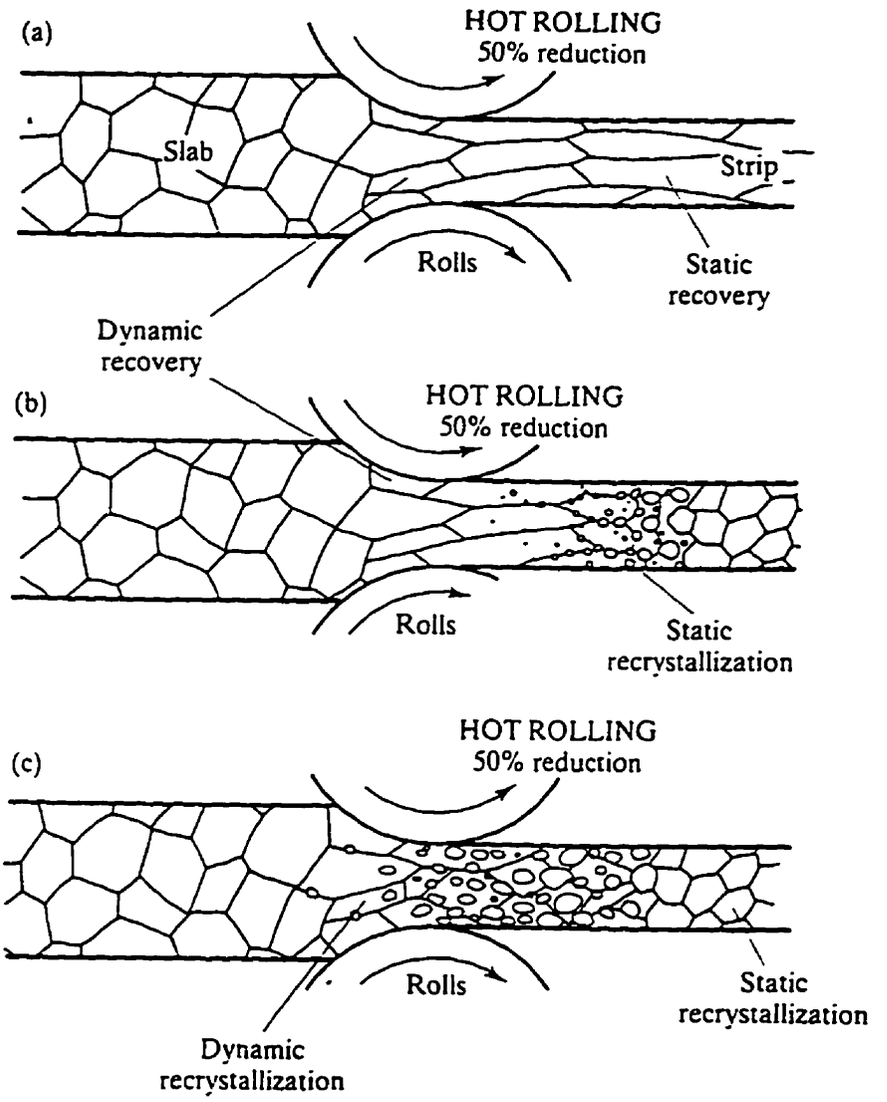

Figure 2-1 Schematic of microstructural changes possible during hot deformation[2]

Studies of FCC single crystals show that at low temperatures ($T < 0.5\ T_m$) the rate of work hardening is divided into stages I, II and III, as depicted in **Figure 2-2**. In the first stage of work hardening, the density of dislocations is still not high enough to create a considerable number of dislocation intersections. In this region, termed easy glide region, dislocations interact only through their long range stress fields and they usually move on parallel slip planes. The work hardening rate ($\delta\tau/\delta\gamma$) is relatively low in this region.

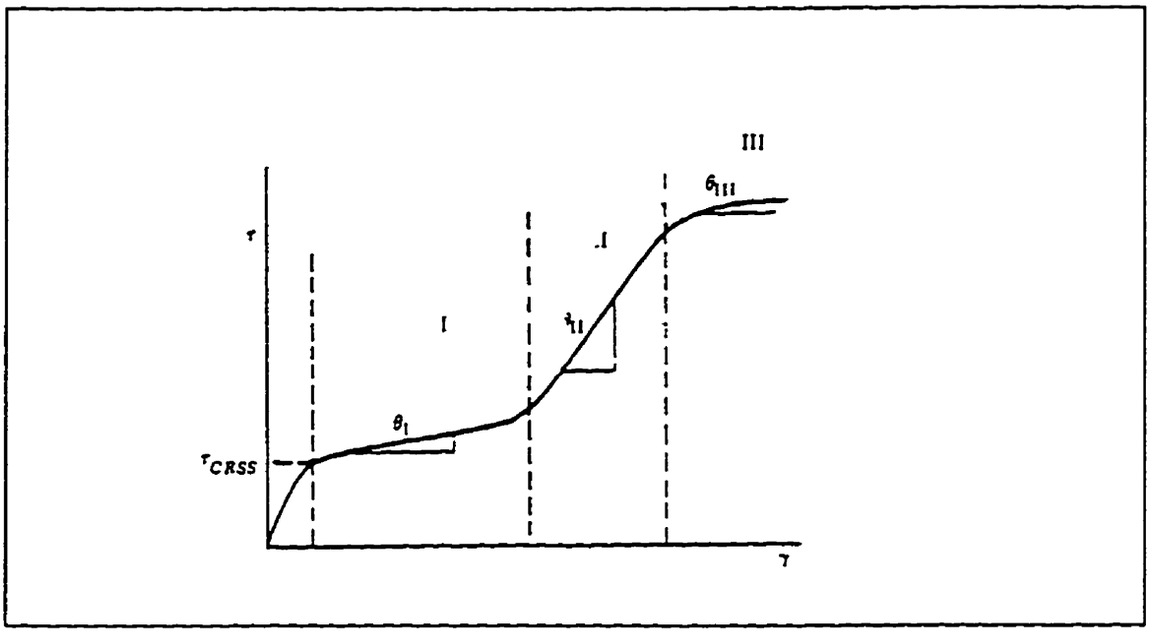

Figure 2-2 Shear stress-strain curve for a single crystal demonstrating different work hardening stages. [3]

In stage II, dislocation intersections produce hard obstacles, manifested by the high work hardening rate. With increasing strain, the flow of material involves multiple slip and the intersection of dislocations moving in intersecting slip planes creates immobile jogs.

Excessive accumulation of dislocations leads to the stage III of work hardening, where the rate of work hardening decreases in a parabolic manner due the superposition of an accumulation and rearrangement (annihilation) of dislocations, referred to as dynamic recovery[4]. Stage III is observed to be associated with the occurrence of cross-slips. According to Seegers and his co-workers[5] screw dislocations held up during stage II cross slip onto other slip systems. In this case, dislocations can by-pass dislocation motion obstacles

(such as inclusions or grain boundaries). Hence slip distance increases, with a consequent decrease in the rate of work hardening. In some cases, screw dislocations may be attracted by opposite sign dislocations and be annihilated. These facts explain the lower parabolic work hardening rate at stage III.

At high temperatures ($T>0.5\ T_m$), work hardening in pure single crystals or polycrystals is usually said to start in stage III. Stage I and II become less and less pronounced as the temperature is increased and can be entirely absent[6].

2.1.1.2 Dynamic Recovery

In general, recovery implies those changes which do not involve the sweeping of the deformed material by migrating high angle grain boundaries. Thus during recovery, each crystal retains its basic identity and boundaries, although the density and distribution of dislocations within it are changed[7].

Dynamic recovery involves the rearrangement of dislocations and consists of two processes. Dislocations of opposite signs annihilate each other or rearrange to form cells of relatively low dislocation density enclosed by boundaries of relatively high dislocation density. At high temperatures (applicable to hot rolling processes), the mechanisms responsible for dynamic recovery are the cross slip and climb of dislocations[8]. As the transition from stage II to III occurs, dislocations become entangled and a cellular structure is transformed into sub-grains.

In metals of high stacking fault energy, such as aluminium, where dislocations cannot easily dissociate, cross slip is facilitated. In these materials as well as BCC transition metals such as iron, and most HCP metals like zirconium, dynamic recovery takes place rapidly and a steady state of stress is reached, which results from the balance of work hardening and recovery. This steady state is characterised by a sub-grain size which depends only on the strain rate and temperature[9,10]. A schematic of the true stress-strain curve of a material undergoing dynamic recovery is illustrated in **Figure 2-3**. This steady state behaviour is

produced by the continuous rearrangement of sub-boundaries, the annihilation of old boundaries and generation of new boundaries (referred to as repolygonization).

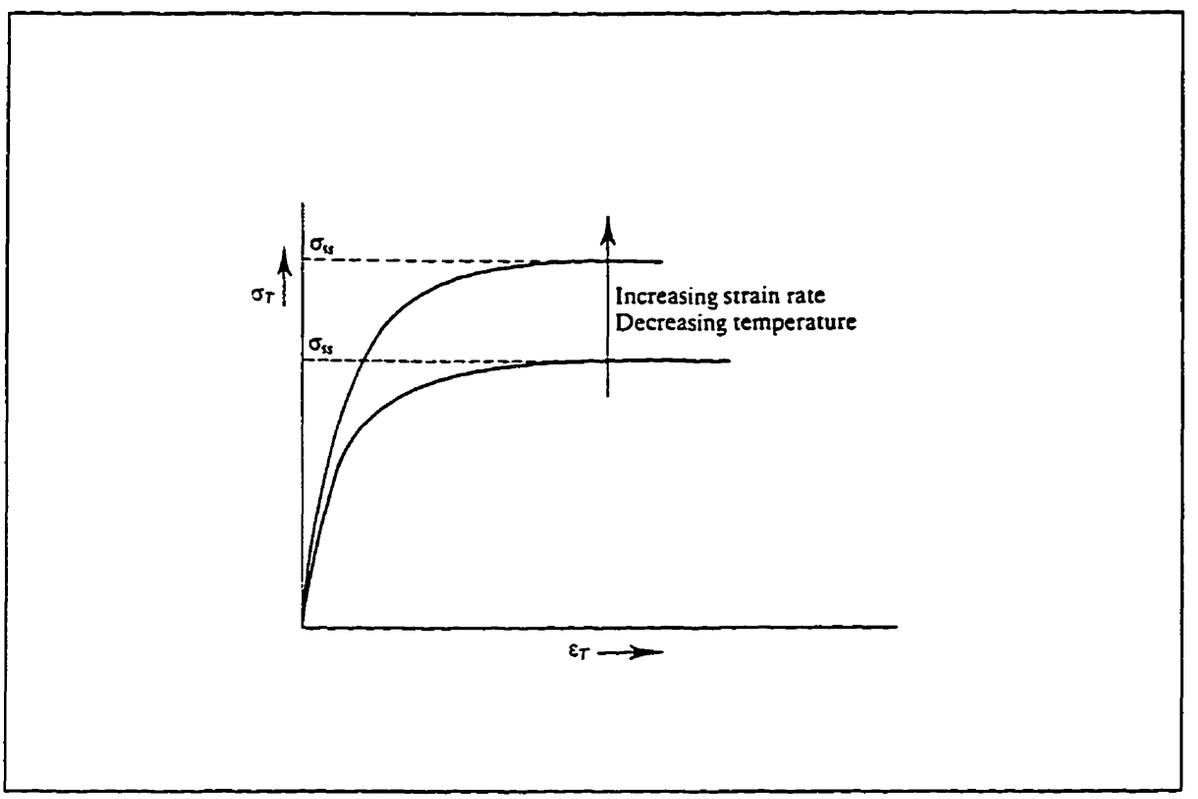

Figure 2-3 Schematic of true stress-true strain behaviour of a material undergoing dynamic recovery[2].

The steady state flow of a given material undergoing dynamic recovery is a function of strain rate and temperature with the general form of

$$\sigma_{ss} = (\dot{\varepsilon})^m \exp\left(\frac{Q}{RT}\right) \qquad (2\text{-}1)$$

The form of this equation is similar to the equation for the steady state creep flow rate. Also, the steady state microstructure closely resembles that of a material deformed during creep; i.e., grains elongate in the primary strain direction and contract in the other directions. As such, recovery has been considered to be analogous to creep.

2.1.1.3 Dynamic Recrystallization

Recrystallization is the collective term applied to those stages in which the crystal orientation is altered through the passage of high angle grain boundaries through the material. During recrystallization, new grains are nucleated and then grow at the expense of other deformed grains.

In FCC metals of low and medium stacking fault energy (e.g. Cu, Ni and γ-iron), slow dynamic recovery permits the dislocation density to increase to an appreciable level. When a critical density is exceeded, new grains are nucleated. The recognition of dynamic recrystallization is based on the existence of peaks on the stress-strain curve, before a steady state is reached, as depicted in **Figure 2-4**. A transition from cyclic to single peak is observed as the strain rate is increased or the temperature is decreased. Multiple peaks are observed during sequential grain refinement and coarsening, and single peaks during grain refinement with no coarsening, as is common in high strain rate hot deformation processes.

The nucleation of dynamic recrystallization is commonly said to begin at a critical strain, ε_c ($\varepsilon_c=\alpha\varepsilon_p$, where α is a constant between 0.8 to 0.85 and ε_p is the strain at peak stress), which corresponds to a critical dislocation density[11]. Once the critical density (which depends on strain rate, temperature and composition) is reached, dynamic recrystallization is initiated by the bulging of pre-existing grain boundaries at low strain rates. At higher strain rates, dynamic recrystallization is initiated by the growth of the high angle cell boundaries formed by dislocation accumulation[12]. The driving force for the growth of the nuclei is the difference in dislocation density in front of and behind the boundary. However, the mechanism of nucleation differs for single peak and multiple peak behaviour. In the single peak case (grain refinement), nucleation occurs essentially along existing grain boundaries and referred to as necklace structure[13]. The growth of each grain is stopped by the concurrent deformation. When all the grain boundary sites are exhausted, further new grains are nucleated within the original grain at the interface of the recrystallized and unrecrystallized grains. Progress of dynamic recrystallization is schematically illustrated in **Figure 2-5**. In the industrial hot rolling processes

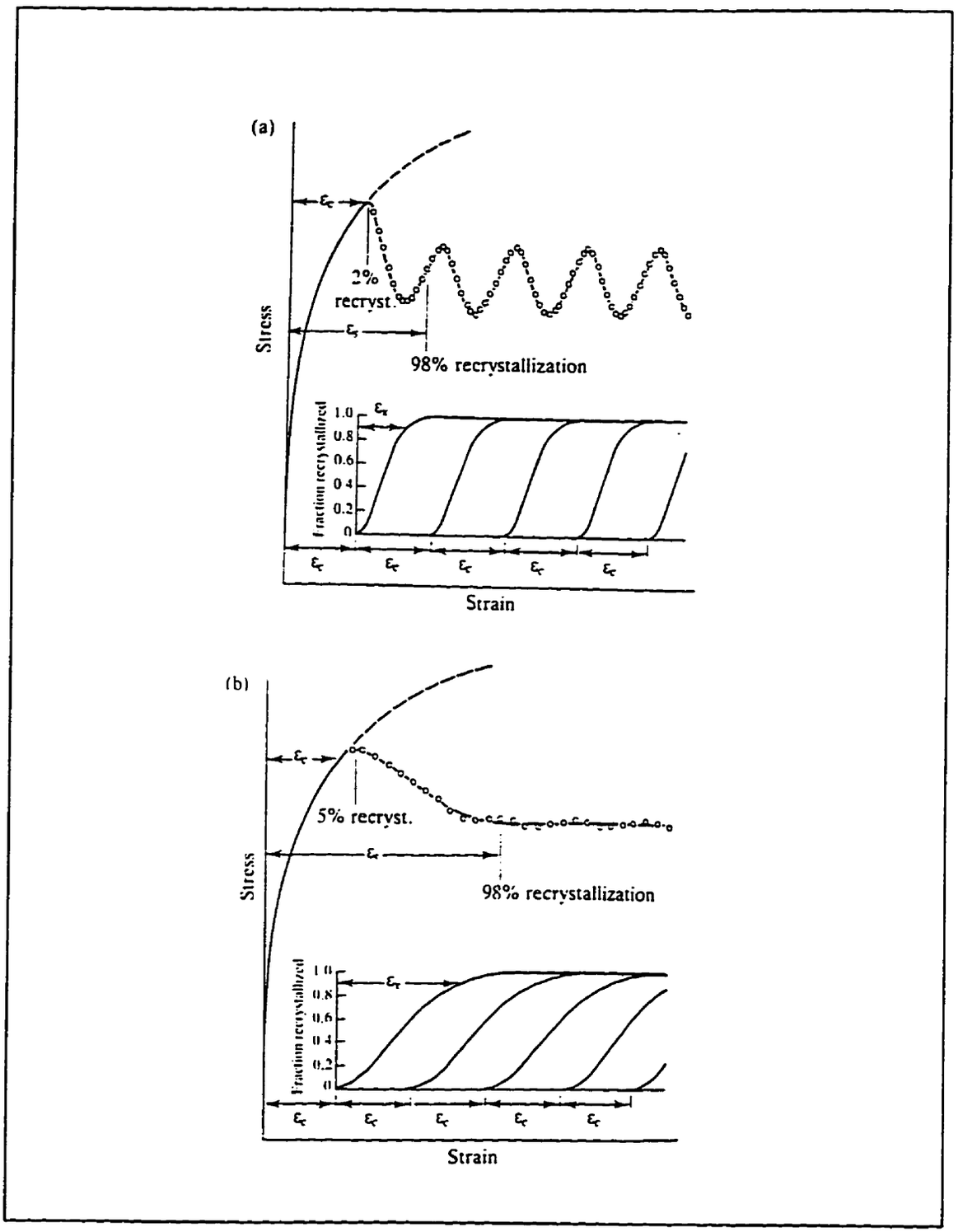

Figure 2-4 Characteristic shapes of true stress-strain curves demonstrating the occurrence of dynamic recrystallization, a) Multiple peak; b) Single peak[14].

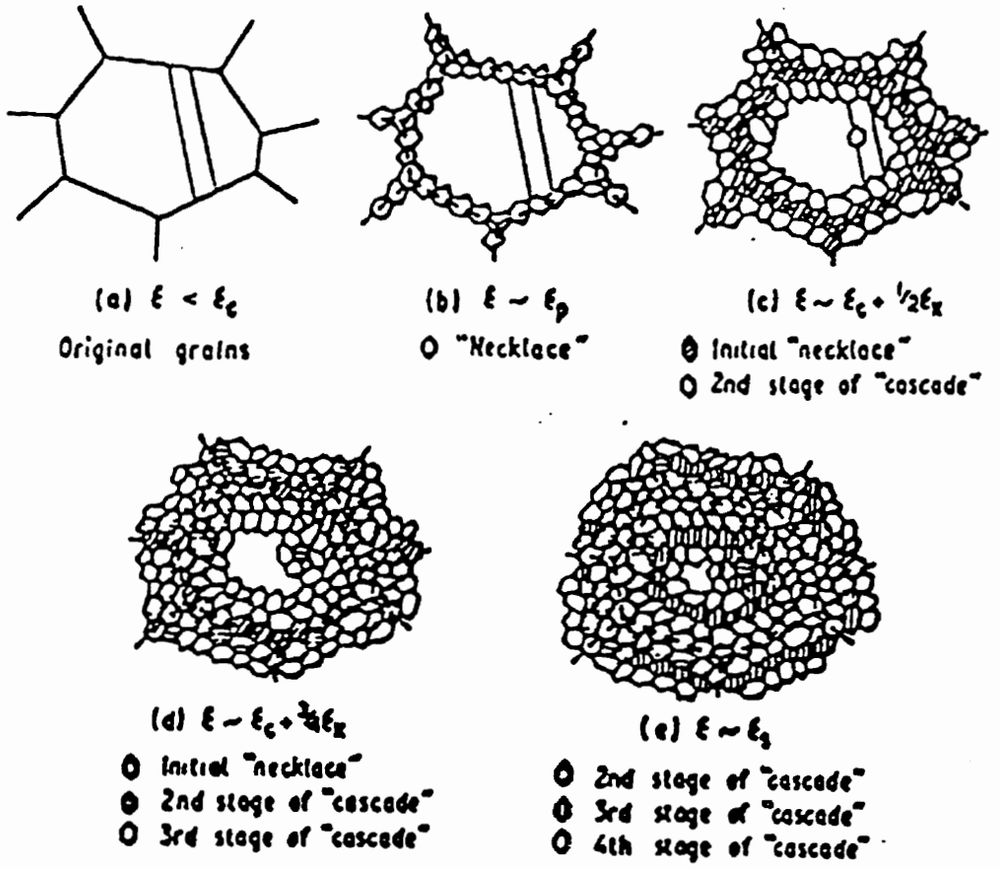

Figure 2-5. Schematic illustration of the progress of the dynamic recrystallization when the recrystallized grain size is much finer than the original grain size[13].

the strain rates are relatively high, such that only single peak dynamic recrystallization is likely to occur, if any.

In a given material, the characteristics of dynamic recrystallization depend on three parameters: initial grain size, D_o, temperature, T, and strain rate, $\dot{\varepsilon}$. The initial grain size affects the critical strain, ε_c, the peak strain, ε_p, and the kinetics of dynamic recrystallization. The finer the initial grain size, the lower are the critical and peak strains. This is because dislocations accumulate more rapidly and the higher specific grain boundary area (per unit volume) leads to faster recrystallization kinetics[14]. Peak stress is also found to be dependent on the initial grain size, however, the steady state stress and final grain size are independent of the initial grain size[14].

The effects of temperature and strain rate are commonly described using the Zener-Hollomon parameter:

$$Z = \dot{\varepsilon} \exp\left(\frac{Q_{def}}{RT}\right) \tag{2-2}$$

where Q_{def} is the activation energy of deformation and R is the universal gas constant and T is absolute temperature. The critical and peak strains, ε_c and ε_p, are found to be dependent on Z. Results obtained for copper, stainless steel and γ-iron have shown that the peak strain is related to the initial grain size and Z as follows[14]:

$$\varepsilon_p = A\, D_o^{0.5}\, Z^p \tag{2-3}$$

where A and p are constants. Softening due to dynamic recrystallization, X_{dyn}, generally follows an Avrami type dependence on strain and strain rate as follows[15]:

$$X_{dyn} = 1 - \exp\left[-\beta_{dyn}(\varepsilon - a\,\varepsilon_p)^n\right] \tag{2-4}$$

where β_{dyn}, a and n are constants, ε is the strain, and ε_p is a function of strain rate, temperature and grain size, as given by equations (2.2) and (2.3). As a consequence, the dynamic recrystallization kinetics is also a function of Zener-Hollomon parameter. Finally, the steady state grain size, D_{ss}, and steady state stress, σ_{ss}, are found to be dependent only on Z, as given below[16]:

$$D_{ss} = A' Z^{-m} \tag{2-5}$$

$$\sigma_{ss} = A'' Z^{-q} \tag{2-6}$$

where A', A", m and q are constants.

2.1.2 Metallurgical Phenomena After Hot Deformation

In industrial hot deformation processes, usually the strain rates are so high that there is not enough time for complete dynamic softening of the work hardened material. Hence, concurrent static recovery, accompanied by either static or possibly metadynamic recrystallization, is taking place after deformation.

2.1.2.1 Static Recovery

Static recovery is defined as a softening process in which the decrease in density and change in the distribution of the dislocations after hot deformation or during annealing are the operating mechanisms. These changes do not involve the sweeping of the deformed material by migrating high angle boundaries[17]. The mechanism operating in the low temperature range involves vacancy motion; those operating in the intermediate temperature range involve dislocation motion without climb; and those in the high temperature range (>$0.5T_m$) involve dislocation motion with climb and cross slip. It is found that if the dislocations have a low mobility at the temperature of deformation, they appear as a fairly random array in the deformed metal. However, if the dislocations are able to cross slip (high stacking fault energy), they immediately begin to condense into tangles so that the metal contains regions of high and

low dislocation density, as shown in **Figure 2-6**[18]. The regions of low dislocation density are called cells or sub-grains. It has been shown that in high stacking fault energy materials, the dislocation network can lower its energy by forming tangles rather than by random distribution. After deformation, the dislocation tangles isolate cellular regions of relatively low dislocation density as shown in **Figure 2-6**. These cells are slightly mis-oriented with respect to each other (a few degrees) and are in a size range in the order of 0.1-1 μm. Upon annealing after cold working or after hot deformation, dislocation tangles condense into sharp boundaries and the dislocation density within the cells decreases.

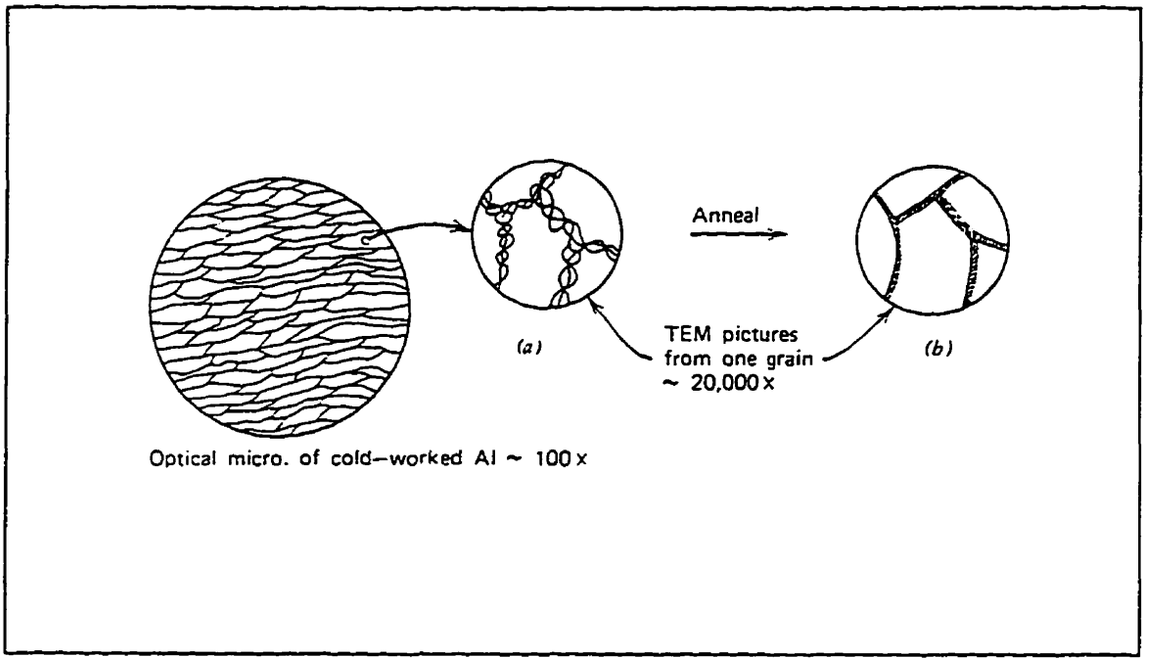

Figure 2-6 Schematic representation of grain structure and sub-grain structure in aluminum[18].

2.1.2.2 Static Recrystallization

Austenite does not undergo dynamic recovery to the same degree as other metals. This ability to have extensive work hardening, without softening by recovery may lead to dynamic recrystallization at higher strains and lower temperatures. However, it is much more common in hot rolling for the material to be deformed only in the work hardening regime[19], i.e. the strain per pass is not large enough (usually less than 0.5) to initiate dynamic

recrystallization. As such, there is a high driving force for static softening to take place between rolling passes and during cooling after the final pass prior to transformation. Both static recovery and recrystallization have been observed in austenite, although the extent of the former is rather limited. There is a general consensus that the maximum amount of softening during holding time attributable to recovery is approximately 20%[20].

Static recrystallization is usually described as taking place in two stages: nucleation of new grains, and the growth of these grains at the expense of deformed ones. Some features of static recrystallization are:

a) A minimum amount of deformation (critical strain) is necessary before static recrystallization can take place.
b) The lower the degree of deformation, the higher the temperature required to initiate static recrystallization.
c) The final grain size depends upon the degree of deformation and to a lesser extent upon the annealing temperature.
d) The larger the original grain size, the slower the rate of recrystallization.

2.1.2.3 Nucleation

The nucleation of new grains takes place preferentially where the local deformation is the highest, i.e. on grain boundaries, deformation bands and inclusions. The process of nucleation is thermally activated and requires an incubation time before nuclei become detectable. Three different mechanisms of nucleation have been proposed for static recrystallization[21]:

a) *subgrain growth or polygonization:* The nuclei are formed by a process of cell or subgrain growth. In an area of high deformation, the dislocations rearrange themselves by polygonization into cells. As the cells grow, more dislocations accumulate into the walls and eventually high angle boundaries are formed.
b) *subgrain coalescence:* A nucleus is formed by the rotation of a subgrain, so that the misfit with its neighbour is decreased, following by disappearance of their common boundary. A subgrain structure prior to nucleation is demonstrated in **Figure 2-7-a**. By some rotation,

subgrain A with B and subgrain C with D coalesce (**Figure 2-7-b**), followed by coalescence of subgrain B with C (**Figure 2-7-c**). Once a few of these subgrains coalesce together, they form a nucleus with high angle boundaries, as depicted in **Figure 2-7-d**.

c) *strain induced grain boundary migration*: The boundary separating two grains marked by different sub-grain sizes bows out and away from the grain with the coarser substructure into that with the finer sub-grains. This leads to the creation of a strain free area. This mechanism has been found to occur without the need for a finite incubation period.

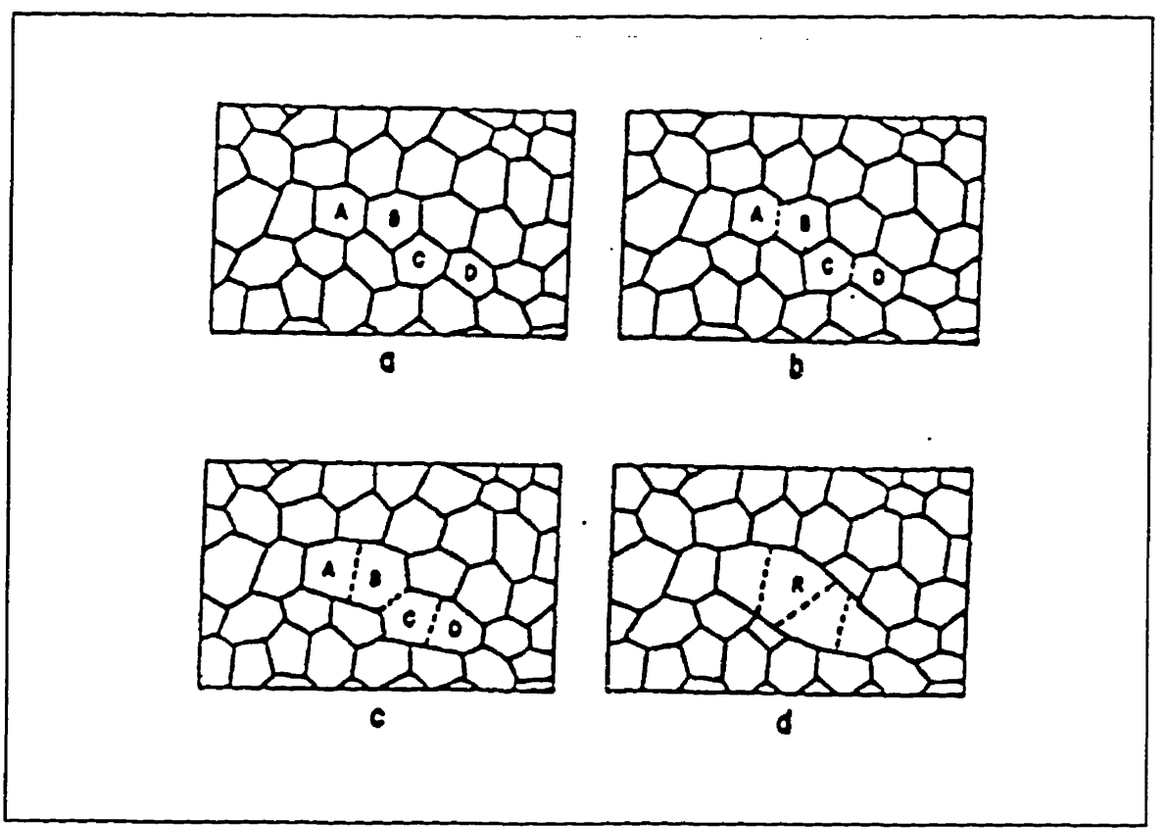

Figure 2-7 Nucleation of new grains by the coalescence of subgrains[17].

2.1.2.4 Growth of New Grains

Once a high angle boundary is formed, it is capable of moving into the deformed material. The migration rate of such boundaries is quite sensitive to the presence of impurities, the structure of the grains into which they are migrating, and the orientation relationship between the growing grains and deformed matrices. Attempts at understanding the effects of impurities are based on the postulate that dissolved impurities retard a moving grain boundary by the elastic attraction of impurity atoms towards the grain boundary. The moving grain

boundary must drag the impurities along, or break away if the concentration of impurities is small enough or the driving force or temperature high enough[17].

2.1.2.5 Metadynamic Recrystallization

Metadynamic recrystallization can be defined as static growth of grain nuclei formed during deformation. It is very important to differentiate between static and metadynamic recrystallization, as the rate of the latter one is much faster and the type of recrystallization has important consequences on the final properties of steel. Once dynamic recrystallization is initiated during deformation, the dynamically recrystallized nuclei continue to grow, even after straining is interrupted. This mechanism was first identified by Petkovic et al.[22] and was termed metadynamic recrystallization. Three distinct softening processes take place after dynamic recrystallization; static recovery, metadynamic recrystallization and static recrystallization. The difference between static and metadynamic recrystallization is in the nucleation mechanism of the new grains. Unlike static recrystallization, metadynamic recrystallization does not require an incubation time for grain nucleation, as dynamically recrystallized nuclei already exist.

2.1.3 Role of Alloying Elements in Recrystallization

Alloying elements in solid-solution form or as fine carbonitride precipitates retard all softening mechanisms, particularly recrystallization. The extent of softening retardation depends on the temperature and the element type. Grain boundary mobility has been shown to be quite sensitive to the presence of small concentrations of impurities. The speed of migration is drastically reduced by the impurities which segregate to the grain boundaries. This mechanism has been used effectively in controlled rolling of steel to retard recrystallization and hinder grain growth after recrystallization by alloying elements like titanium, molybdenum, vanadium and niobium. If recrystallization is retarded and the strain is accumulated and maintained after hot rolling and during the cooling through the transformation, it would produce finer ferrite grain sizes, resulting in a higher room temperature strength and toughness. This is the basis for conventional controlled rolling. In C-Mn steels, due to very fast recrystallization kinetics and a narrow range between the no-recrystallization temperature and A_{r3}, only limited controlled rolling is possible.

2.1.3.1 Alloying Elements in Solution

A quantitative treatment of the interaction between grain boundaries and solute atoms was given by Lucke and Deret[23]. At high concentrations or low temperatures, the impurities are dragged along by the grain boundaries and the boundary speed is controlled by the diffusion of the impurities behind the boundaries. At low concentrations or high temperatures, the boundary breaks away from its atmosphere and moves faster.

2.1.3.2 Alloying Elements in Precipitate Form

The existence of the solute-drag effect of microalloying elements on recrystallization retardation has been generally acknowledged. However, it has been shown that the most important effect of microalloying elements is the formation of fine carbonitrides, which in turn, retard recrystallization much more than solutes. As such, the carbonitride precipitation, its kinetics and interaction with recrystallization are of extreme importance for plate and strip rolling schedule design and a great deal of research has been focused on this issue.

Recrystallization can be retarded by precipitates via the pinning of grain boundaries or pinning of individual dislocations. For grain boundary pinning to occur, the distribution of the precipitates must satisfy two criteria:

a) The particle size and spacing must be below appropriate critical values.
b) A sufficient volume fraction of precipitates is necessary to maintain the spacing below the critical value[23].

In practice, this means that a fine dispersion of precipitates delays the onset of recrystallization by pinning the boundaries and dislocations and by restraining grain growth. The precipitates encountered in austenite can be separated into three types:

a) Precipitates which are not dissolved during reheating.
b) Precipitates formed dynamically during deformation (dynamic precipitation).
c) Strain-induced precipitates formed after deformation.

Undissolved precipitates have little effect on recrystallization, as they are too coarse. However, both dynamic and strain induced precipitates can be responsible for the recrystallization retardation. The rate of dynamic precipitation in Nb steels is generally one order of magnitude faster than strain induced precipitation[24]. The interaction between precipitation and recrystallization can be followed by comparing the recrystallization start time with precipitation start time using a recrystallization-precipitation-time-temperature (RPTT) diagram, as depicted in **Figure 2-8**. When precipitation takes place before the start of recrystallization, the nucleation and growth of recrystallized grains are severely impeded. However, once precipitation starts after the end of recrystallization, recrystallization is unaffected. The temperature at which precipitation retards static recrystallization is referred to as the "no-recrystallization temperature", or T_{nr}. The no-recrystallization temperature depends on the nature and amount of alloying elements as well as prior deformation and heat treatment history. Amongst microalloying elements, Nb has the most profound effect in retarding recrystallization and this accounts for its selection as the most common microalloying element in hot rolling. It has been reported that the addition of 0.03 Nb leads to approximately an order of magnitude retardation in time for 50% softening at temperatures above 900 °C[19].

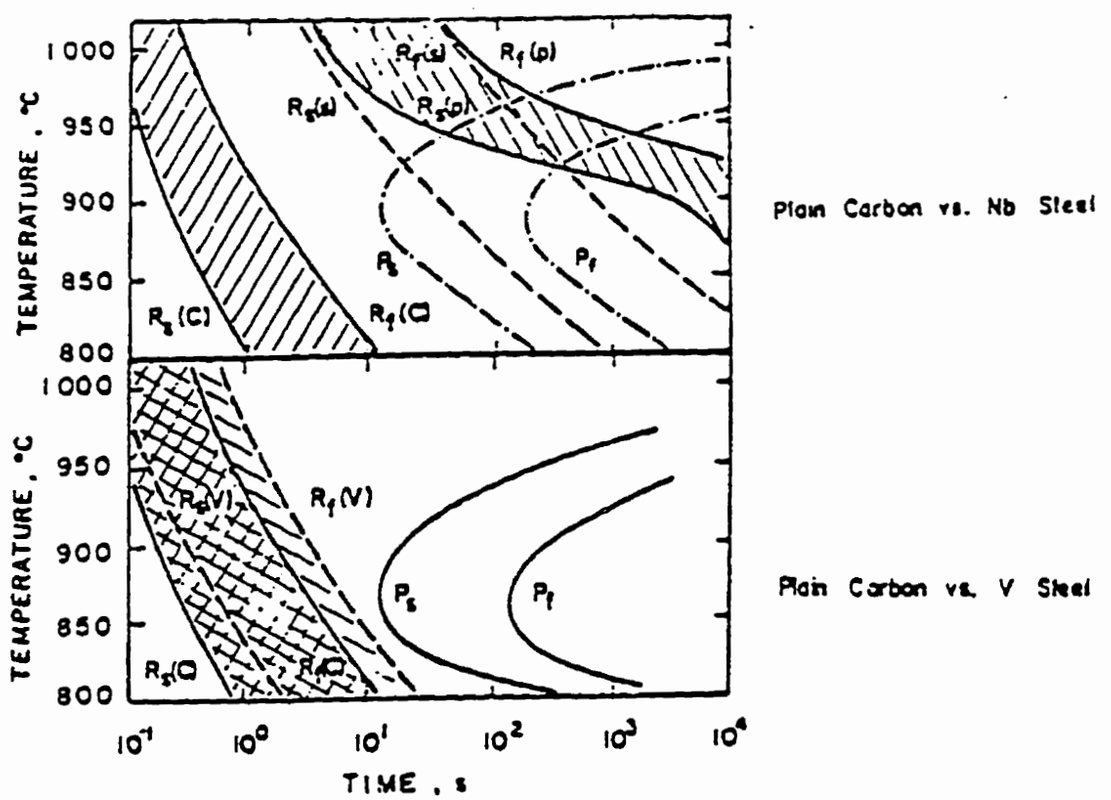

Figure 2-8 Comparison of a PTT diagram for a Nb and a V steel with a RTT diagram for plain carbon and Nb and V modified steels[24].

2.2 Thermo-Mechanical Processing of Steels

For most commercial products in the steel industry, their external shapes are the result of hot-deformation, such as hot-rolling, while the necessary mechanical properties are obtained by alloying elements and heat treatment after deformation. However, metallurgical changes caused by hot deformation may result in additional beneficial effects on the mechanical properties of steels, and sometimes can eliminate heat treatment after deformation. Thermo-mechanical processing is a technique to combine shaping and heat treating of steel. Controlled rolling is a typical example of thermo-mechanical processing in which austenite is conditioned to produce a fine ferrite grain size.

2.2.1 Controlled-rolling of C-Mn steels

After 1945, it became clear that notch ductility and yield strength can both be improved by α grain refinement.[25] Among other techniques for grain refinement, European mills utilised controlled low temperature hot-rolling in order to refine α grains and increase toughness. The following features were generally applied in this controlled rolling process:

a) Interrupting the hot-rolling operation when the slab had been reduced to the prescribed thickness, e.g., 1.65 times the final thickness.
b) Recommencing hot-working when the slab has reached a prescribed temperature and finishing at temperatures in the austenite (γ) range, above the A_{r3} but lower than the conventional finishing temperatures, e.g. down to 800°C.

The low temperature finish rolling practise refines the γ grains, hence, the transformed α grains. A considerable additional grain refinement can be achieved by rolling in the non-recrystallized γ region, where deformation bands increase the nucleation sites for α grains. However, the temperature range for non-recrystallized austenite in C-Mn steels is relatively narrow, and this mechanism for grain refinement cannot be effectively utilised, due to the risk of getting into the two-phase region deformation.

2.2.2 Controlled-Rolling of Nb-Treated Steels

2.2.2.1 Conventional Controlled Rolling

In 1958, the first Nb-treated steel plates were produced. By a small addition of Nb, (up to 0.06%) the strength was increased by 10kgf/mm^2 (100 MPa) but the no-recrystallization temperature was raised by 40°C. However, by normalising, the tensile strength was decreased to the original level but the yield strength and no-recrystallization temperature were improved by 50-80 MPa and 25-70 °C, respectively, compared with normalised niobium free steels. In order to minimise deterioration in notch toughness, the following measures were reported:

a) An increase in Mn content
b) Lower reheating temperatures of the slabs down to the 1200°C or less
c) Lower finish rolling temperature and an increase in total percentage reduction at low temperatures. It was also observed that low-temperature finish rolling was more effective in Nb-bearing steels than in C-Mn steels.

The lower the hot-deformation temperature, the heavier the draught and the finer the starting grain size during hot-working in the recrystallization temperature range, the finer the recrystallized γ grains. The smaller γ grains will lead to the smaller α grains after transformation. These are the reasons for the improvement in toughness by low temperature finish-rolling of C-Mn steels, low temperature slab reheating and low temperature roughing in niobium treated steels.

Further research on precipitation hardening of niobium and vanadium steels conducted in the1960's led to the conclusion that the remarkable strengthening by niobium or vanadium additions was caused by precipitation of fine planar Nb(C,N) or VN, coherent with the α matrix.[26] These carbides and nitrides are in solution in γ grains at the reheating temperatures of the slab. The solubility of niobium in austenite is small, so excessive amount of niobium, e.g. above 0.05% is ineffective in strengthening of the α matrix, since it cannot be in solution in the austenite. The solubility of vanadium nitride and vanadium carbide in α and γ is higher than the solubility of niobium carbo-nitride. So the strength of vanadium containing steels can increase

with larger amounts of vanadium, i.e. it does not show the pronounced saturation effect as niobium does.

Another fundamental result was the effect of precipitation on recrystallization retardation[27]. The addition of niobium retards recrystallization of the γ grains during hot-rolling. The addition of niobium raises the critical temperature for γ recrystallization by about 100°C. This is mainly due to fine niobium carbo-nitrides which have been in solution at reheating temperatures and re-precipitated during hot deformation[28]. Precipitation of carbo-nitrides is accelerated by hot deformation. Recrystallization of austenite in steels containing 0.02-0.05% of niobium can start after one or a few reductions at temperatures between 1050 to 1100 °C, but it takes more than 100 seconds to complete the recrystallization at these temperatures. A similar but weaker effect was also observed for vanadium treated steels. This will be discussed in more detail later, as it is one of the fundamental metallurgical phenomena which can be exploited for remarkable improvement of mechanical properties of niobium treated steels.

Kubota *et* al.[28] reported that improvements in notch ductility and α grain refinement were progressive with increasing total reduction below 950-900°C, where γ recrystallization could not start during intervals between reductions.

Addition of manganese to niobium treated steels intensifies controlled-rolling at non-recrystallization temperatures in the γ region, as it decreases the A_{r3} temperature. In this way the temperature range for straining the γ grains without recrystallization is increased. This is similar to the effect of increasing the no-recrystallization temperature by the addition of niobium.

Deformation of γ grains in the non-recrystallization temperature range elongates these grains in the rolling direction and creates deformation bands and annealing twins. The α grains nucleate in these deformation bands as well as at prior austenite grain boundaries, giving an

increased surface area for the austenite-ferrite transformation[29]. This is the reason for the α grain refinement in niobium treated steels finish-rolled with heavy reductions below 900°C.

In summary, the role of conventional controlled-rolling using niobium treated steels is to retard recrystallization by the help of carbo-nitride precipitates, then impose higher reductions in the non-recrystallization austenite region to produce pancaked grains prior to transformation.

2.2.2.2 Dynamic Recrystallization Controlled Rolling

In rod and bar rolling, using high strain rates (100-1000 s^{-1}), short interpass times (between few tens of milliseconds to few hundreds of milliseconds) and large strains per pass (0.4-0.6) dynamic recrystallization has been found to occur[30]. It has been proposed that under appropriate conditions dynamic recrystallization also occurs during strip rolling of niobium HSLA steels[31]. The occurrence of dynamic recrystallization during simulated strip rolling of HSLA steels has been cited by several other authors.[32,33] The results of an analysis of the events during strip rolling also indicated that dynamic recrystallization is happening during rolling of Nb steels[34]. Dynamic recrystallization affects rolling loads and is reported to produce considerably finer ferrite grains (~3μm) than those ferrite grains transformed from pancaked austenite (~7μm)[35,36,37]. However, there are concerns regarding the validity and applicability of the results obtained in all of the above studies to the real mill practice, primarily due to the low strain rates employed in the experiments.

Conventional controlled rolling relies on static recrystallization in the early stages of finish rolling to refine austenite and pancaking of the austenite in the last stages to enhance ferrite grain nucleation during transformation. In contrast, dynamic recrystallization favours higher reductions in the first few stands to exceed the critical strain for the onset of dynamic recrystallization. Dynamic recrystallization controlled rolling leads to greater ferrite grain refinement through austenite grain refinement.[37] Another advantage of the initiation of dynamic recrystallization during rolling is a marked reduction in roll forces and torques, which

in turn translates to savings in energy consumption and less roll wear. Also, the gauge accuracy will be enhanced due to the lower reductions required in the last stands. The only justification that the author found in the literature regarding the better grain refinement through dynamic recrystallization is the higher nucleation rate and formation of "necklace structure" during deformation[37]. High density of grain nucleation already incorporated into the matrix, expedites the post dynamic recrystallization compared to static recrystallization, as no incubation time is required.

2.2.2.3 Effects of Recrystallization Type on Grain Size

Many different authors have attempted to develop models predicting grain sizes produced by static, dynamic and metadynamic recrystallization for different materials. An excellent compilation of these models has been done by Hodgson[30]. The general observation, common in all these models, is that statically recrystallized grain size is a function of the initial grain size, temperature, and amount of strain, while dynamically and metadynamically recrystallized grain sizes are only a function of Zener-Hollomon parameter, i.e. temperature and strain rate, in an inverse power law form. This indicates that increasing strain rate and decreasing rolling temperature lead to more grain refinement provided dynamic and metadynamic recrystallization are in place. Another common understanding is that rolling schedules with dynamic and metadynamic recrystallization produce finer final grain sizes compared to schedules with only static recrystallization. This idea is appealing to the steel manufacturers to achieve further grain refinement.

2.2.2.4 Controversies Regarding the Type of Recrystallization in Strip Rolling

The occurrence of dynamic recrystallization by strain accumulation during industrial hot strip rolling schedules has been questioned[38]. It has been argued that the kinetics of static recrystallization approaches those of dynamic recrystallization as the strain increases. In addition, interpass times are generally much greater than deformation times. Hence, softening of the material during strip rolling may be due to enhanced static recrystallization.

This controversy, in spite of its practical importance in terms of final mechanical properties and mill set-up, still remains. The physical proof of the possibility of dynamic

recrystallization during strip rolling is notoriously difficult, since it requires extremely fast quenching of the steel after deformation to freeze the structure and look for dynamically created grain nuclei. Most of the mill engineers do not believe in the possibility of dynamic recrystallization in any kind of steel during strip rolling schedules. This belief has been reinforced by the fact that the possibility of dynamic recrystallization has not been taken into account in the conventional strip mill set-up and in control modules developed by General Electric and Westinghouse. In these control modules, which are in use in North America, it is assumed that the steel repeatedly goes through only work hardening during deformation and static softening during interpass times. This assumption may lead to erroneous roll force prediction if the steel actually softens in one or more stands instead of hardening.

2.3 Modelling of Hot Rolling Processes

The past two decades have seen a marked increase in the ability to control and optimise the final properties and dimensional accuracy of rolled products. These improvements have been brought about partly through the application of improved predictive models of rolling forces and microstructural events that take place during and after rolling. These models enable mill engineers to reduce expensive mill trials, study the effects and interactions of different variables, improve rolling schedules, and control the process. Amongst the more important models are: temperature, hot flow stress, roll force, and microstructural evolution models. This thesis is concerned with the material flow stress and roll force prediction models, and to lesser extent with microstructural evolution models.

2.3.1 Modelling Methodologies in Hot Rolling

Similar to other industrial process modelling methodologies, there are four types of models[30]:

a) **Phenomenological**: These models describe actual physical processes in the form of a mathematical equations derived from the basic principles of the process.

b) **Empirical:** These are characterised by the statistical analysis of process data to provide relationships between the process variables and the parameters of interest. Regression analysis has been the most commonly applied technique, however, there

are increasing number of sophisticated methods available for constructing models of this type.

c) **Semi-empirical:** This combines certain features of the two methods above, i.e., the general form of the model equation is representative of the physics of the process with some coefficients and exponents found through regression.

d) **Heuristic:** These include the rule-based models, expert systems, and diagnostic models which are now gaining widespread applications within the steel industry[30]. These models do not contain mathematical representation of a physical process, but can still predict process parameters of interest and the process outcome, based on previous experience.

There are few examples of phenomenological models being used in hot mills, due to the complexity of the process and the lack of true phenomenological models for even some of the simpler microstructural changes occurring during rolling[30]. Historically, there has been widespread application of empirical and semi-empirical models, based on simple statistical regressions of large data sets of plant measurements. The accuracy of these models depends on the availability of a general form of mathematical equation which can fit the data well. Data is usually nonlinear and inter-related which limits the applicability and accuracy of regression models. It is expected that the next decade will see extensive use of both the semi-empirical and heuristic models within the steel industry[30].

2.3.2 Hot Flow Stress Models

A number of studies has been conducted over the past two decades to develop constitutive relations to describe the hot flow stress of steels as a function of significant deformation process variables: strain, strain rate and temperature.[39,40,41,42,43] From the results of isothermal creep tests, it is found that the relationship between stress and strain rate varies according to the stress levels. At low stress levels the power relation $\dot{\varepsilon} = A_1 \sigma^{n'}$, and at high stress levels the exponential relation $\dot{\varepsilon} = A_2 \exp(n''\sigma)$ appears to fit the data well. However, it has also been shown that the hyperbolic sine relation $\dot{\varepsilon} = A(\sinh \alpha\sigma)^n$ is able to

represent the whole range of stresses satisfactorily[44]. The parameters A, A_1, A_2, n, n', n" and α are material constants. Sellars and Tegart[40] expanded on these equations to include the effects of temperature, resulting in an Arrhenius type equation

$$\dot{\varepsilon} = A(\sinh\alpha\sigma)^n \exp\left(\frac{-Q}{RT}\right) \tag{2-7}$$

Incorporating the Zener-Hollomon parameter, $Z = \dot{\varepsilon} \exp\left(\frac{Q}{RT}\right)$, into the Eq. (2.7) results in

$$Z = A(\sinh\alpha\sigma)^n \tag{2-8}$$

where Q is the activation energy. The values of A, Q, and n are shown to be functions of the strain level.[45] The value of α has been studied by Sakai et al.[46] for plain carbon steels and shown to be a function of the carbon content. For low carbon contents and HSLA steels, the α value of 0.012 has been used by others[42,45] as well as in the present work. Wang and Lenard[41] employed a power relationship and normalized strains to model the hot strength of a Nb-V steel. The activation energy was calculated using peak stresses and the constants of the power law at different strain levels were stored in a data bank.

The Arrhenius type equations are normally used to characterize steady state flow stress data. However, they do not explicitly incorporate the softening effects of dynamic recrystallization. When dynamic recrystallization occurs, the drop in flow stress may be estimated by an Avrami equation[42,47]:

$$\Delta\sigma = (\sigma_{ss} - \sigma'_{ss})\left\{1 - \exp\left[-k\left(\frac{\varepsilon - a\varepsilon_p}{\varepsilon_p}\right)^m\right]\right\} \tag{2-9}$$

where σ_{ss} is the saturation stress, σ'_{ss} is the steady-state flow stress after dynamic recrystallization, a, k, m are constants and ε_p is the strain at the peak stress. Incorporation of a softening term in the case of highly alloyed HSLA steels is not always necessary when the flow stress does not exhibit steady state behavior[48].

Shida[49] conducted a study of the hot strength of a broad range of carbon steels and provided equations to predict the flow stresses as a function of strain, strain rate, temperature and the carbon content. Murthy and Lenard[50] analyzed industrial data, obtained during the finish rolling of 8000 slabs in a seven-stand hot strip mill as well as that of 150 laboratory rolling tests. They compared the measured and predicted roll forces obtained by using Shida's equations in conjunction with Orowan's model[51] and Ford & Alexander's model.[52] In the best case, i.e. Shida's flow stress equations with Orowan's roll force model, the mean of the difference of the measured and predicted data was 15.9% with a standard deviation of 9.68. They also concluded that Shida's equations are not directly applicable for HSLA steels but may be used with an appropriate expression for the equivalent carbon content.

The effect of constitutive modeling on the predictive capabilities of a one-dimensional model of hot strip rolling was examined by Nadkarni et al.[53] Three different approaches were compared. In the first, the flow stress at different strains, strain rates and temperatures were stored in a multi-dimensional data bank and retrieved as needed; in another an empirical relation for the flow stress was used and in the third, average values in the pass were employed. The multi-dimensional data bank was thought to be the most natural technique, as it eliminated the need for modeling the behavior of the metal.

Artificial Neural Networks have also been used to model the flow stress of steels at high temperatures. Tsoi[54] used the MARS (Multiple Adaptive Regression Spline) algorithm to predict flow stresses in a plate rolling mill, with a 5.3% accuracy, surpassing those of conventional models. Hwu et al.[55] used a backpropagation neural net to predict satisfactorily the flow stresses of an extra-low carbon steel in the austenite, ferrite and dual-phase regions, within a ±2% accuracy. Rao and Prasad[56] used a commercial Neural Network package, NWorks[57], to model the flow stress of a low carbon steel. A mean absolute percentage error of 3.01% was reported.

2.4 Scope of the Thesis

The knowledge of the material's constitutive behavior is an essential requirement for the design and analysis of deformation processes. In spite of suffering some drawbacks in terms of accuracy, ease of development, adaptability, and speed empirical stress-strain relationships

and constitutive equations describing material behaviour during deformation are being widely used. In the other hand, complex metallurgical events during hot rolling of HSLA steels affect the material flow stress, which in turn affects the rolling forces. Hence, control algorithms of hot strip mills are required to integrate the material deformation resistance model, microstructural evolution model, and roll force prediction model into a module. The main requirements for on-line control models include: high accuracy, short computational time and adaptability to incorporate new data. The current procedure for the study and modeling of the material behavior and microstructural evolution is mainly based on conventional statistical techniques. The applicability and accuracy of these techniques are limited and highly dependent on the availability of a mathematical relationship which can describe and fit the data well.

Manufacturing of higher value added products such as HSLA strip coils is gaining more importance for the North American steel industry. The standards for the uniformity of the mechanical properties and consistency of output gauge are much tighter for these products. However, the metallurgical events taking place in these grades involve complex and quantitatively not-well-understood events of strain accumulation, dynamic and metadynamic recrystallization, precipitation of carbonitride formers, and complex interaction of precipitates with softening processes. The existing knowledge of these complex metallurgical events during hot rolling of HSLA steels are limited and there is no predictive model directly applicable to the industrial rolling processes to take into account the effects of these metallurgical events on the rolling forces and final mechanical properties of the rolled products. The industrial and academic research in this field is quite active, however, far from completion and fulfilling the needs of the industry. Hence, the steel industries are looking for immediate alternative modelling solutions to address their current needs.

In reply to the above need for applicable and accurate microstructural and roll force models, this thesis is concerned with the applications of a class of heuristic modelling techniques, called Artificial Neural Networks, to hot strip mill processes. The major rolling mill process automation and control development firms are actively working on utilising this technique, however, the resulting knowledge remains mainly proprietary. Recently, Siemens of

Germany, with more than half of the global market share in minimill process automation development, announced that all of the next generation mill control algorithms developed by Siemens will be based on Neural Networks[58].

Neural Network modelling technique and its mathematical basis, along with its applications in different areas of material science and processing are described in detail in Chapter 4. Applications of this technique to: flow stress prediction, roll force prediction, and detection of dynamic recrystallization during hot strip rolling of HSLA steels are given in Chapters 6 and 7. Principal Component Analysis is used as a data pre-processor for Neural Network simulator code, in order to decouple the linearly correlated input variables. This technique is explained in Chapter 3 and the results of its application is presented in Chapter 7.

3. Principal Component Analysis

Principal Component Analysis is one of the most commonly used multivariate statistical analysis techniques. When the variation of a dependent variable is a function of more than one independent variable, then the problem is multivariate. Most real engineering and scientific problems and analyses are multivariate. The complexity of this kind of problem arises from both the multi-dimensionality, and the possible interactions of the independent variables which affect the dependent variable. Principal Component Analysis can be used to both reduce the dimensionality of the problem and remedy problems associated with the linear correlations between original variables.

The objective of the analysis is to take p variables X_1, X_2, X_3,... X_p, and find combinations of these variables to produce new variables Z_1, Z_2, Z_3,... Z_p that are uncorrelated. The lack of correlation is a useful property because it means that these new variables are measuring different 'dimensions' in the data. These new variables are called Principal Components and are ordered so that Z_1 represents the largest amount of variation , Z_2 represents the second largest and so on. That is $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \text{var}(Z_3) \geq \ldots..\geq \text{var}(Z_p)$, where $\text{var}(Z_i)$ denotes the variance of Z_i in the data set.

When performing a Principal Component Analysis there is always the hope that the variances of most of the new components will be so low as to be negligible. This will happen if there are significant linear correlations in the original variables. In that case the variation in the data set can be adequately described by the first few principal components.

To avoid confusion with the regression analysis, it should be stressed that, in regression analysis we are looking for a line or surface of closest fit to a system of points in space which indicates the best prediction of dependent variables. In Principal Component Analysis, the primary interest is in detecting any possible linear association between the independent variables. In short, the regression line indicates the best prediction of a dependent variable and the principal component line indicates the best linear association between independent variables

3.1 An Introduction to the Theory of Principal Component Analysis

Principal Component Analysis is a variable-directed multivariate technique which can be used for identifying possible linear relationships between several quantitative variables[59]. At one extreme, these variables may be completely independent of each other, while at the other extreme there may be a linear relationship between two or more apparently independent input variables (conditions). In the range between these two extremes, we can have different degrees of correlation between variables. In two-dimensional space with just two variables, we simply look at the correlation between them by plotting the data on a two-dimensional graph. However, once we want to analyse data that is dependent upon more than two input variables, looking at their interdependency becomes more complicated. In this case, the "Principal Component Analysis" technique could be utilised, provided there is a considerable correlation between the variables. This technique aims to transform n-number of observed variables with possibly some degree of interdependency, to a new set of n-uncorrelated variables, arranged in decreasing order of variability. These new uncorrelated variables are called "Principal Components" and are linear combinations of the original variables, with coefficients equal to the eigenvectors of the correlation or standardised covariance matrix. The principal components are orthogonal to each other and are derived in decreasing order of variability, such that the first component accounts for as much as possible of the variation in the original data and the second component represents the second largest variance. In geometric terms, the m-dimensional sub-space of an n-dimensional space spanned by the first m principal

components gives the best possible fit to the data points, as measured by the sum of the squared perpendicular distances from each data point to the sub-space.

The usual objective of the Principal Component Analysis is to determine whether the first few principal components account for most of the variation in the original data. If they do, the problem can be studied in a lower dimensional space without losing much of the variation of the data. This is done by ignoring the variation in the direction of the last principal components. In other words, if some of the original variables are highly correlated, then they are effectively saying the same thing and there may be nearly linear constraints on the variables.

In practice, it is not always easy to give meaningful labels to the principal components, but they can still be used effectively to reduce the dimensionality of later analysis. Indeed, Principal Component Analysis (PCA) makes use of the interdependency between the original variables and transforms them to a smaller set of uncorrelated variables. It is therefore worth mentioning that, if the original variables are nearly uncorrelated, then there is almost no point in carrying out a PCA. In this case the PCA will simply find new components which are close to the original variables, but arranged in decreasing order of variance. In short, the two main objectives of PCA are:

1) Reduction of the dimensionality of the problem in order to facilitate further analysis and modelling of the data.
2) Orthogonal transformation of linearly correlated variables to uncorrelated principal components.

An important feature of Principal Component Analysis is exploited in conjunction with regression analysis. In multiple regression, one of the major difficulties with the usual least square estimators is the problem of multi-collinearity, which occurs when there are near-constant linear functions of two or more of the variables[60]. If multi-collinearities exist, the variances of some of the estimated regression coefficients can become very large, leading to unstable and often misleading estimates of the regression equation. Various biased regression techniques such as "ridge regression" have been developed to overcome this problem, but an

alternative approach is to regress, not on the original variables, but on the first few important principal components[61]. This technique is known as "Principal Component Regression".

Existence of strong multi-collinearities poses similar problems to the development and training of Artificial Neural Network models. Once a dependent variable is to be modelled as some function of two or more independent variables, and these independent variables are some constant linear function of each other, then, the network learning algorithm cannot properly learn the input-output relationship. This is because the network is not being presented with the data providing some information about the variation of the output with each individual and independent variable.

Part of this research is to investigate the beneficial effects of performing Principal Component Analysis prior to neural network training in order to avoid network confusion due to the collinearities in input variables. The results are presented in Section 7.4.4. A detailed discussion, geometric interpretation, and important characteristics of Principal Component Analysis have been presented in the author's Master of Applied Science thesis[62]. The calculation of Principal Components is briefly described in the following section.

3.2 Calculation of Principal Components

Given a data set with p numeric variables and n observations, p principal components can be computed. Each principal component is a linear combination of the original variables, with coefficients equal to the eigenvectors of the correlation or covariance matrix. The first principal component is the linear combination of the variables X_1, X_2, X_3,... X_p, i.e.,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \qquad (3\text{-}1)$$

such that Z_1 varies as much as possible for the individuals, subject to the condition that

$$a_{11}^2 + a_{12}^2 + \cdots\cdots + a_{1p}^2 = 1. \qquad (3\text{-}2)$$

The constraint is actually a standardisation scheme and is introduced to avoid an increase in the variance of Z_1 by simply increasing any one of the a_{1j} values. The second principal component,

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \cdots\cdots + a_{2p}X_p, \qquad (3\text{-}3)$$

is such that the variance of Z_2 is as much as possible, subject to the constraint that

$$a_{21}^2 + a_{22}^2 + \cdots\cdots + a_{2p}^2 = 1, \qquad (3\text{-}4)$$

and also the condition that Z_1 and Z_2 are uncorrelated. The rest of the principal components are derived in the same manner.

The coefficients a_{11}, a_{12}, ... a_{1p} are actually the elements of the first eigenvector of the covariance matrix of original variables, a_1, standardised to the sum of the squared of one ($a_{21} + a_{22} + \ldots + a_{2p} = 1$). The coefficients a_{21}, a_{22}, ... a_{2p} are the elements of the second eigenvector, a_2, and so on. Hence, a principal component analysis basically involves calculation of the eigenvalues and eigenvectors of the covariance matrix and a simple matrix multiplication.

To form the covariance matrix, the arithmetic mean of each variable, i.e. $\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p$, is computed. Based on these, the covariance matrix is formed. The general form of the covariance matrix, **A**, is

$$\mathbf{A} = \begin{Bmatrix} \text{variance}(X_1) & \text{cov}(X_1, X_2) & . & . & . & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & \text{variance}(X_2) & . & . & . & \text{cov}(X_2, X_p) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \text{cov}(X_p, X_1) & . & . & . & . & \text{variance}(X_p) \end{Bmatrix} \qquad (3\text{-}5)$$

where variances and co-variances are:

$$\text{variance}(X_1) = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_1 - \overline{X}_1)^2 \tag{3-6}$$

$$\text{cov}(X_1, X_2) = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_1 - \overline{X}_1)(X_2 - \overline{X}_2) \tag{3-7}$$

and so on.

This matrix is always symmetric, because $\text{cov}(X_1, X_2) = \text{cov}(X_2, X_1)$ and so on. To compute eigenvalues, we form the characteristic equation for the covariance matrix A

$$\det|\mathbf{A} - \lambda \mathbf{I}| = 0 \tag{3-8}$$

Having matrix A, the solution of the above equation leads to p (number of variables) distinct solutions which are the eigenvalues. Expanding the above determinant yields a polynomial in λ. The roots of this polynomial are the solutions for the eigenvalues. A wide variety of methods are available to estimate the roots of the polynomials[63], however, for symmetric matrices, there are better approaches with no need to form the polynomial and find its roots. For example, the *Jacobi* Method[64] transforms a symmetric matrix to a diagonal matrix by eliminating off-diagonal terms in a systematic fashion. This method consist of a sequence of orthogonal similarity transformations. Each transformation is a plane rotation designed to annihilate one of the off-diagonal matrix elements. Successive transformations undo previously set zeros, but the off-diagonal elements nevertheless get smaller and smaller, until the matrix is diagonal to the desired precision. Elements of the final diagonal matrix are the eigenvalues.

Assuming that the eigenvalues are ordered as $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_p \geq 0$, then λ_i corresponds to the *i*th principal component,

$$Z_i = a_{i1}X_1 + a_{i2}X_2 + \cdots\cdots + a_{ip}X_p, \tag{3-9}$$

and demonstrates the variance of Z_i. The constants $a_{i1}, a_{i2}, \ldots, a_{ip}$ are the elements of the corresponding standardised eigenvectors. Substituting the eigenvalues into the fundamental eigenvalue equation, $Ax = \lambda x$, we find p distinct eigenvectors, corresponding to p eigenvalues.

A linear algebra theorem[62] indicates that if the non-zero eigenvectors $a_1,a_2,...,a_n$ correspond to different eigenvalues $\lambda_1,\lambda_2,....,\lambda_n$, then these eigenvectors are linearly independent. This is another reason why principal components are not linearly correlated. The combination of these eigenvectors form the p by p square matrix of eigenvectors denoted by H. Having this matrix and the mean of each variable, and recalling that principal components are a linear combination of the original variables with coefficients equal to the eigenvectors of the covariance matrix, we can compute the principal components for each individual observation:

$$\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \\ . \\ . \\ Z_p \end{bmatrix} = \mathbf{H} \begin{bmatrix} X_1 - \bar{X}_1 \\ X_2 - \bar{X}_2 \\ . \\ . \\ X_p - \bar{X}_p \end{bmatrix} \qquad \textbf{(3-10)}$$

Inversely, if we want to calculate the original variables $\mathbf{X} = X_1, X_2,..., X_p$ for a given set of principal components, $\mathbf{Z} = Z_1, Z_2, ..., Z_p$, of an observation, we invert the above equation and we get:

$$\mathbf{Z} = \mathbf{H}(\mathbf{X} - \bar{\mathbf{X}}) \qquad \textbf{(3-11)}$$

or

$$\mathbf{H}^{-1}\mathbf{Z} = \mathbf{X} - \bar{\mathbf{X}} \qquad \textbf{(3-12)}$$

Since the eigenvector matrix **H** is orthogonal (eigenvector matrix of a real symmetric matrix, i.e. covariance matrix, is orthogonal), then its inverse and transpose are equal, so that:

$$\mathbf{X} = \mathbf{H}^t \mathbf{Z} + \overline{\mathbf{X}} \qquad (3\text{-}13)$$

Here it is worthwhile to mention an important characteristic of any matrix A, called the trace of A. The trace of A is the sum of the eigenvalues of A which can be proved to be equal to the sum of the diagonal elements of matrix A. This is another way to derive the fundamental fact that the sum of the variances of the original data is equal to sum of the variances of the principal components (Eigenvalues of covariance matrix are variances of principal components). Of course this is expected due to the orthogonal nature of the transformation from the original co-ordinates to the principal components space.

4. Artificial Neural Networks

A new and robust modelling method using Artificial Neural Networks is evolving. Neural networks have been proven to be universal function approximators if a sufficient number of processing units are employed.[65] This method is applicable in many areas of behavioural science and has produced promising results in the areas of: data analysis, modelling, signal processing, weather and financial forecasting, process automation, robotics, and many others. Well-trained Neural Network models provide fast, accurate and consistent results. The technique has been well developed[66,67] and extensively used in many areas of science and engineering, however, it is somewhat new to the materials science and processing field.

4.1 Introduction

Neural Networks provide a new approach to data and process modelling. They belong to the heuristic class of modelling techniques, that is, these models do not contain a mathematical representation of the physical process, but still predict an outcome of a process based on previous experience. A neural net can be looked at as a non-linear procedure that maps an input vector to an output vector. However, knowledge of the process, in contrast to phenomenological and statistical methods, is not captured within a mathematical expression, but stored within a matrix of connection weights between the process inputs and outputs. These weights, determined randomly at first, are continuously modified so that the predictions fit the experimental data more accurately. The iterative process of modifying the weights is

called training. Upon completion of the training, there exists one weight matrix which gives the minimum error between actual output data set and the network predictions.

In the following section, the backpropagation neural network paradigm will be explained in detail and an algorithm and flow chart will be provided to develop and train a neural net model of this kind.

4.2 Feed Forward Back-Propagation Neural Networks

The multi-layered feedforward perceptron using the back-propagation learning algorithm is central to much work on modelling and classification by neural networks. This technique has been evolved from Rosenblatt's simple perceptron model[68]. The backpropagation model is the predominant supervised-training algorithm. Supervised learning implies that a good set of data or pattern associations is used for training of the network. Input-output pairs are presented to the network, and weights are adjusted to minimise the error between network output and actual value. The knowledge that a neural network possesses is stored in these weights.

4.2.1 Neuron's Function

The artificial neurons, called nodes or Processing Elements (PE), have been modelled by some inspiration from biological neurons. The neurons are the basic processors in a neural network; still, they perform quite simple arithmetic operations. Each node receives several inputs over the connections to previous layer's nodes, called synapses. There is a weight (a numeric value, functionally similar to the coefficients in a regression model) associated with each connection. Each input is a product of a previous layer node output and the weight of the connecting synapse. The activation of the neuron (output) is computed by applying an activation function (refer to section 4.2.3) to the sum of these products. McCulloch and Pitts[69] proposed the first model of a neuron as a binary threshold unit. A simple generalisation of this model has the form of

$$n_i = g(\sum_j w_{ij} n_j - \mu_i). \quad \textbf{(4-1)}$$

The value n_i is called the *state* or *activation* of unit *i*. The non-linear function *g(x)* is called the *activation function* or *transfer function*. In this model, the neuron computes a weighted sum of its inputs from all neurons of the previous layer and compares this value to a certain threshold value (μ_i), then, applies the activation function to the result to produce the neuron output. An abstract model of the neuron is shown in **Figure 4-1**.

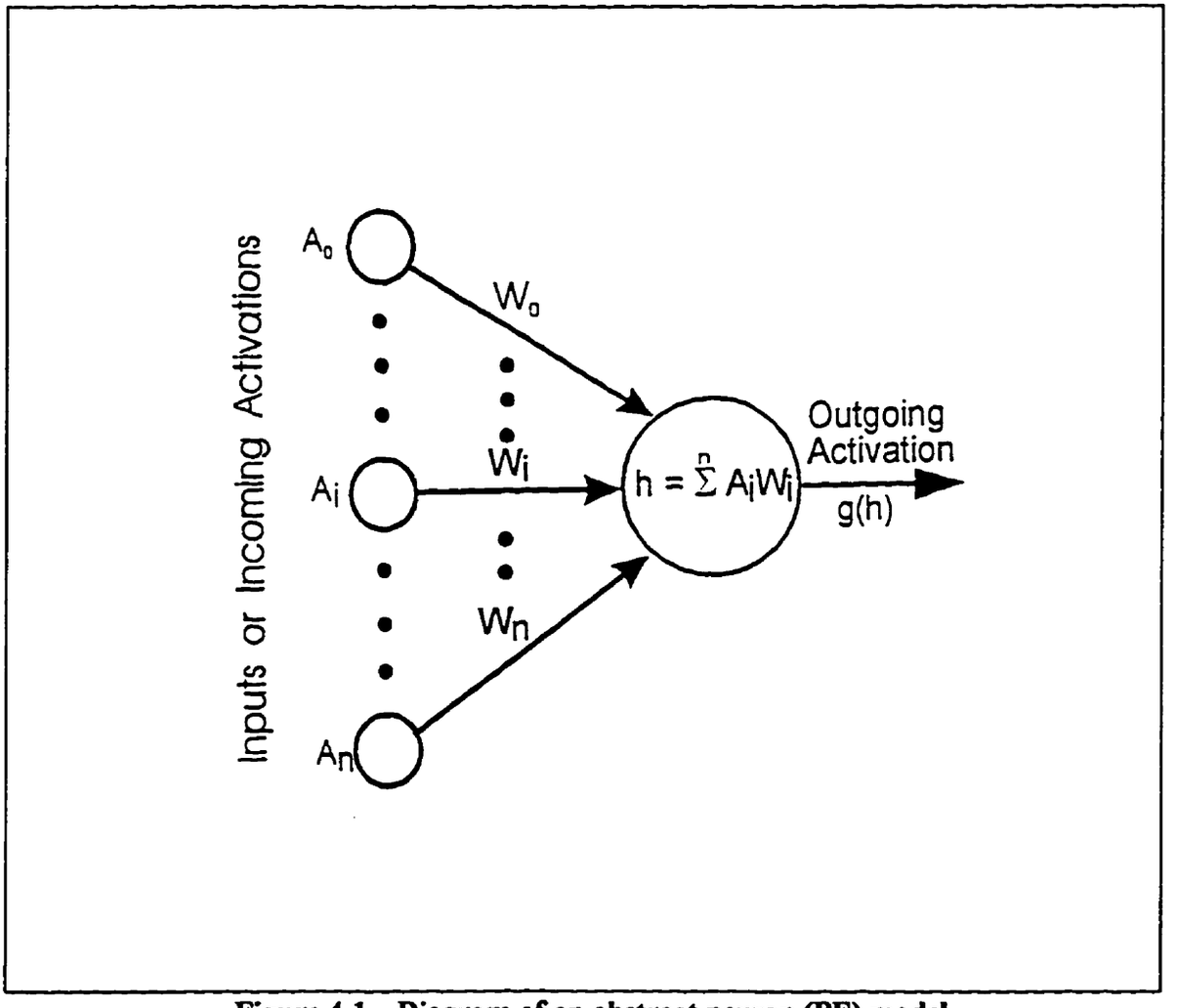

Figure 4-1. Diagram of an abstract neuron (PE) model

4.2.2 Network Architecture

Supervised neural networks generally consist of a large number of processing nodes (neurons) organised in input, hidden and output layers. The input layer receives the signals (the

input values in the given data set) and the output layer provides the final outputs (network predictions) to the environment. The hidden layer, with no outside connections, increases the functionality of the network in handling linearly inseparable data sets and enhances the function approximation and classification capability of the network. The number of the nodes in the input and output layer is dictated by the problem. However, the optimum number of hidden layers and number of nodes in each hidden layer depends on the complexity of the problem and number of input and output variables. For non-linear function approximation, generally one hidden layer of nodes suffices. There are a few methods available to aid network designers in deciding the optimum number of hidden nodes, however, there are no rigid rules for deciding the optimum number of hidden nodes for any given problem. The two more common techniques are node pruning and dynamic node allocation.[70,71] In node pruning, we start with a large number of nodes and train the network. Then, the number of nodes is progressively reduced and each time the performance of the network is compared to the performance of the network with the previous number of nodes. This procedure is continued until the accuracy of network predictions starts to deteriorate. The optimum number of nodes is then reached. The same result can be obtained through dynamic node allocation, that is, to start with a small number of nodes and progressively increase the number of nodes and monitor the performance. After arriving at optimum number of nodes, the accuracy of predictions will not increase with an increase in the number of nodes.

The term, feedforward, identifies the direction of information propagation. In the multi-layered perceptron the only direction permitted is from input towards the output. The term, backpropagation, describes the learning algorithm in which the weights are modified by comparing the network output with the actual output and minimising the error. A fully connected three-layer network of this kind is illustrated in **Figure 4-2**. The output units are denoted by O_i, hidden units by V_j, and the input terminals by ξ_k. Notational conventions are shown in the figure and at the end of this chapter.

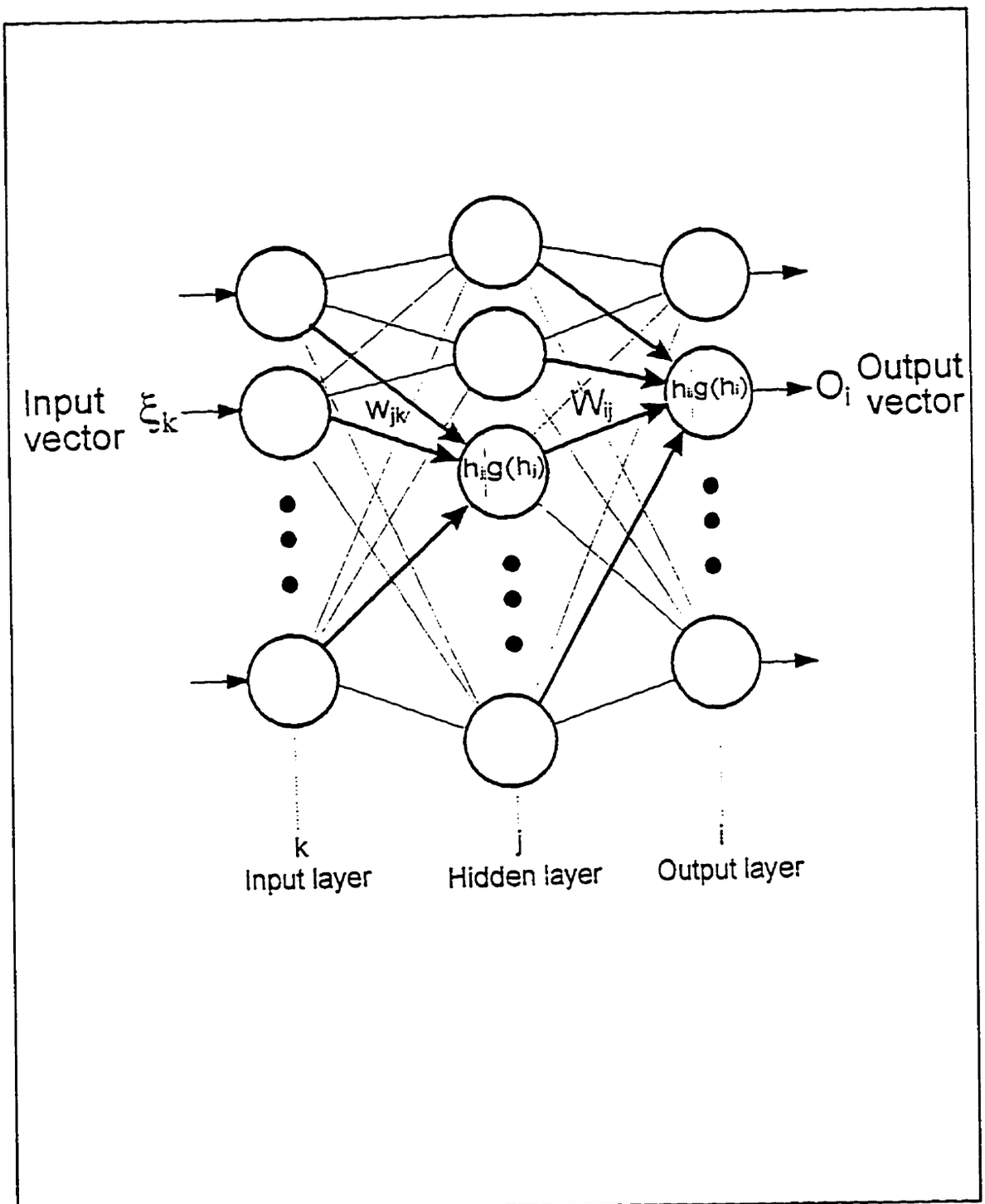

Figure 4-2 A three layer feedforward network, showing notations used for nodes and weights.

4.2.3 Activation Functions

Activation functions map the possibly infinite domain of the PE's output to a pre-specified range, usually (0,1) or (-1,1). Important requirements for these functions are that they be non-linear, differentiable, monotonic and bounded. The most popular activation functions are the *logistic sigmoidal* and *hyperbolic tangent* (*tanh*) functions, shown in **Figure 4-3**. The non-linear mapping capability of neural nets emanates from the non-linear nature of these functions. Both of them introduce a non-linearity in the network dynamics by bounding the output values within a fixed range.

4.2.4 Data

Neural networks, similar to natural brains, cannot operate unless they are provided with experiential knowledge or data. Feed-forward back-propagation networks require a sufficient number of pattern pairs (input-outputs) for training purposes. Both inputs and outputs could be multi-dimensional, binary or continuous valued. Training pairs are usually normalised to values between (-0.9 to 0.9) or (0.1 to 0.9), depending on the type of activation function being used.

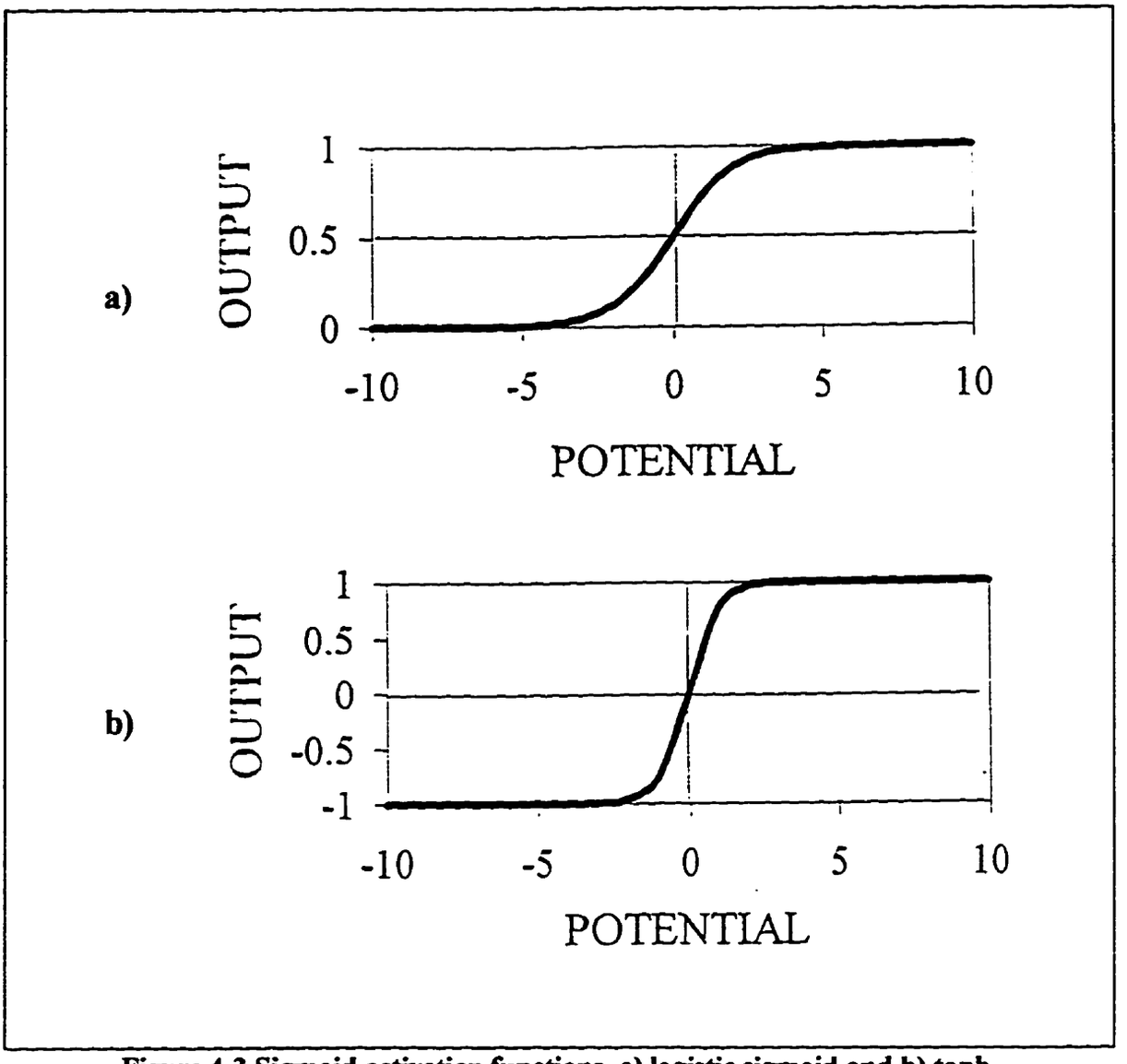

Figure 4-3 Sigmoid activation functions, a) logistic sigmoid and b) tanh.

4.2.5 The Learning Paradigm

After a tentative decision about the topology of the network (number of layers and nodes), the network should be trained to find the optimum weights. The algorithm to train networks without a hidden layer was developed long ago[68], however, these kinds of networks had limited capabilities. Although the greater power of multi-layer networks was realised then, it was not until the 1980's that the first algorithms to teach multi-layered networks were developed.[72,73,74] The algorithm consists of the forward propagation of data and an error

calculation, then an error minimisation scheme based on the gradient descent method. These steps are explained in the following sub-sections.

4.2.6 Feedforward Propagation of Data and Error Calculation

We first consider a three-layer network such as that illustrated in **Figure 4-2**. We label different input patterns by a superscript μ, so input k is set to ξ_k^μ when pattern μ is being presented. In this phase of training, an input is presented and propagated through the network, generating a set of values as network outputs, O_i^μ. The required steps are shown below.

Given pattern $\mu\{\xi_k^\mu, \zeta_i^\mu\}$, a set of input-output pairs in the training data, the hidden unit j receives a net input of

$$h_j^\mu = \sum_k w_{jk}\xi_k^\mu \tag{4-2}$$

and applies the activation function to this value, minus the threshold, to produce the node output or activation V_j

$$V_j^\mu = g\left(h_j^\mu - \theta_j\right) = g\left(\sum_{k=1} w_{jk}\xi_k^\mu - \theta_j\right) \tag{4-3}$$

Equation (4.3) can be rewritten as

$$V_j^\mu = g\left(\sum_{k=1} w_{jk}\xi_k^\mu - \theta_j\right) = g\left(\sum_{k=0} w_{jk}\xi_k^\mu - w_{j0}\xi_0^\mu - \theta_j\right) \tag{4-4}$$

To omit thresholds from the analytical descriptions, we may consider an imaginary input ξ_0, fixed to $\xi_0 = -1$, and choose connections strengths $w_{j0} = \theta_j$, then:

$$V_j^\mu = g\left(\sum_{k=0} w_{jk}\xi_k^\mu\right) \tag{4-5}$$

Now this output from the hidden layer will be propagated to the output layer nodes. Output unit i thus receives

$$h_i^\mu = \sum_j W_{ij} V_j^\mu = \sum_j W_{ij} g\left(\sum_k w_{jk} \xi_k^\mu\right) \tag{4-6}$$

and the final output from the output nodes will be

$$O_i^\mu = g(h_i^\mu) = g\left(\sum_j W_{ij} V_j^\mu\right) = g\left\{\sum_j W_{ij} g\left[\sum_k \left(w_{jk} \xi_k^\mu\right)\right]\right\} \tag{4-7}$$

Note here that the output is an explicit function of the inputs and weights of the connections. Now the training task is to find the weight vectors for the input-to-hidden layer and hidden-to-output layer to minimise the error between the network calculated output, O_i^μ, and the desired value (training data output), ζ_i^μ.

4.2.7 Cost Function

In order to systematically optimise the weights, we should utilise a cost function as a measure of total error and try to minimise it in an appropriate way. A very common cost function is the *squared error* cost function. The best solution for network weights can be attained when the sum of the errors for all the given patterns and across all outputs is minimised. Then, we may use the following form of the squared error cost function:

$$E[w] = \frac{1}{2} \sum_{\mu,i} \left(\zeta_i^\mu - O_i^\mu\right)^2 \tag{4-8}$$

This cost function is the most commonly used and is based on the squared difference between computed, O_i^μ, and desired output values, ζ_i^μ, of all output nodes across all patterns in the data set. The smaller the value of the E, the better are the weights. When E is zero, the mapping from input to output is perfect for the given pattern. Substituting Equation (4.7) into (4.8) gives:

$$E[w] = \tfrac{1}{2}\sum_{\mu,i}\left\{\zeta_i^\mu - g\left[\sum_j W_{ij} g\left(\sum_k (w_{jk}\xi_k^\mu)\right)\right]\right\}^2 \qquad \textbf{(4-9)}$$

4.2.8 Gradient Descent Learning

Given this error measure E[*w*], we can minimise the cost function by modifying the weights through 'sliding downhill' on the surface it defines in *w*-space. The Gradient Descent algorithm suggests changing each weight, w, by an amount Δw, proportional to the gradient of E at the present location. The 'downhill sliding' and gradient descent algorithm are demonstrated in **Figure 4.4**.

First, let us consider applying this algorithm to the hidden layer-to-output layer weights, W_{ij}. For these connections, the gradient descent rule gives

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}} \qquad \textbf{(4-10)}$$

The parameter η is a positive constant controlling the learning rate. Substituting for E from Equation (4.9) and using the chain rule, we obtain:

$$\begin{aligned} \Delta W_{ij} &= -\eta \frac{\partial E}{\partial W_{ij}} = -\eta \frac{\partial E}{\partial h_i} \times \frac{\partial h_i}{\partial W_{ij}} \\ &= \eta \sum_\mu [\zeta_i^\mu - O_i^\mu]\, g'(h_i^\mu)\, V_j^\mu \end{aligned} \qquad \textbf{(4-11)}$$

If we define

$$\delta_i^\mu = g'(h_i^\mu)[\zeta_i^\mu - O_i^\mu] \qquad \textbf{(4-12)}$$

then our final form of the weight correction for hidden layer to output node connection would be:

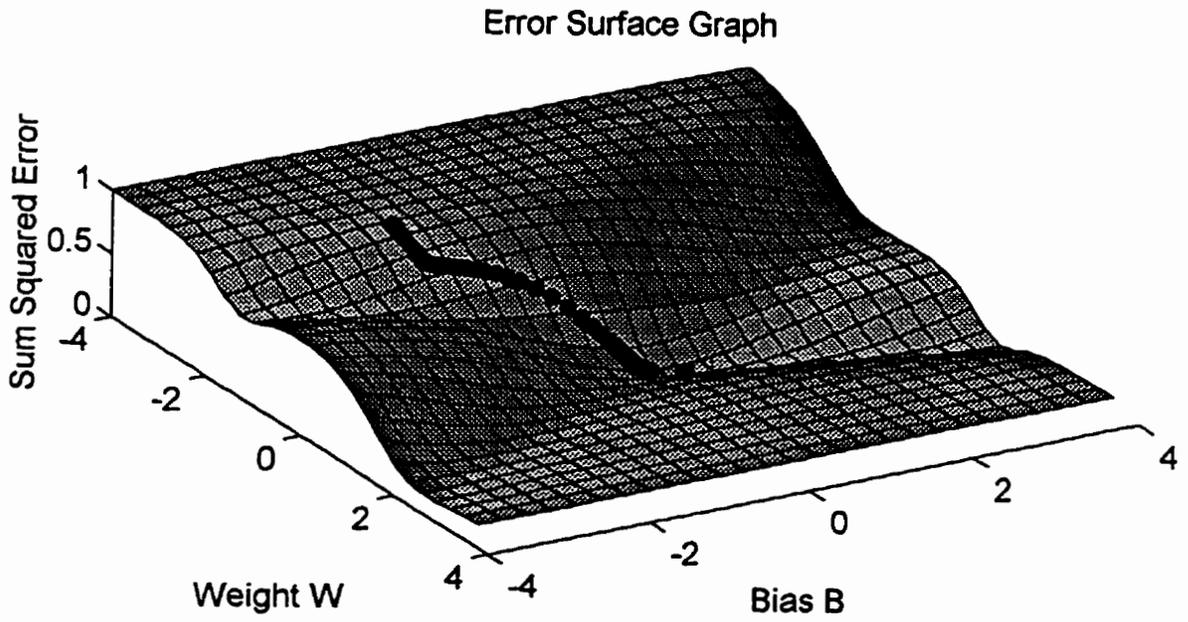

Figure 4-4 An error surface and downhill sliding through gradient descent algorithm.[75]

$$\Delta W_{ij} = \eta \sum_{\mu} \delta_i^{\mu} \, V_j^{\mu} \tag{4-13}$$

This equation is commonly referred to as the *delta rule*, the *adaline rule* or *LMS* (least mean square) rule.

We notice from Equation (4.12) that the change in weights is a function of the net error between desired and network output, and the derivative of the activation function *g(h)*. With a sigmoid form for *g(h)*, this derivative, $g'(h) = 1 - g^2$, is largest when $|h_i^{\mu}|$ is small (notice that the output of a sigmoid function is always less than 1). Thus the changes are made most strongly on connections feeding to units with small $|h_i^{\mu}|$'s, those which are in doubt about their output. This is a reason why sigmoid functions are the preferred form of activation functions.

For the input-to-hidden layer weight corrections, we must differentiate the cost function, Equation (4.9), with respect to the w_{jk}'s. Using the chain rule, we obtain

$$\begin{aligned} \Delta w_{jk} &= -\eta \frac{\partial E}{\partial w_{jk}} = -\eta \frac{\partial E}{\partial V_j^{\mu}} \times \frac{\partial V_j^{\mu}}{\partial w_{jk}} \\ &= \eta \sum_{\mu} [\zeta_i^{\mu} - O_i^{\mu}] \, g'(h_i^{\mu}) \, W_{ij} \, g'(h_j^{\mu}) \, \xi_k^{\mu} \end{aligned} \tag{4-14}$$

Defining

$$\delta_j^{\mu} = g'(h_j^{\mu}) \sum_i W_{ij} \delta_i^{\mu} \tag{4-15}$$

then the final form of the weight correction for input to hidden layer connections becomes:

$$\Delta w_{jk} = \eta \sum_{\mu} \delta_i^{\mu} \, W_{ij} \, g'(h_i^{\mu}) \, \xi_k^{\mu} = \eta \sum_{\mu} \delta_j^{\mu} \, \xi_k^{\mu} \tag{4-16}$$

Equations (4.13) and (4.16), which are the weight update rules, are written as the sum of all patterns μ, however, they are usually used incrementally: a pattern μ is presented at the input and then all weights are updated before the next pattern is considered. The incremental update

rule decreases the cost function at each step and lets successive steps adapt to the local gradient. In this case Equations (4.13) and (4.16) can be rewritten as

$$\Delta W_{ij} = \eta \delta_i^{\mu} V_j^{\mu} \tag{4-17}$$

$$\Delta w_{jk} = \eta \delta_j^{\mu} \xi_k^{\mu} \tag{4-18}$$

The above analysis demonstrates that with repeated presentation of training data points to the network and updating the weights based on the *delta rule*, we can minimise the error between network predictions and the desired outputs after a sufficient number of data presentations (iterations).

4.2.9 Local Minima

The gradient descent algorithm is guaranteed to find a local minimum. In many practical problems, this local minimum is the global minimum as well. However, there is a possibility that the network will get trapped at a local minimum and stop learning without finding the global minimum. This means that there is another set of weights which can provide a better fit to the data. If this local minimum provides results with acceptable error levels, then this local minimum is close enough to the global one.

To minimise the risk of getting trapped in a local minimum, the choice of initial random weights is important.[66] If they are too large, the sigmoids will saturate from the beginning, and the system will become stuck in a local minimum near the starting point. A sensible strategy is to choose random weights so that the magnitude of the typical net input h_i to input i is less than-but not too much less than-unity. This can be achieved by taking the weights w_{ij} to be of the order of $(1/\sqrt{k_i})$ where k_i is the number of j's which feed forward to i.

Another practical measure to increase the chance of finding the global minima is to initialise the initial weights and thresholds to different random values. In this way, the starting

point on the error hill (refer to **Figure 4-4**) will change. Hence, the probability of sliding down to the global minimum without falling into a local one will increase.

4.2.10 Generalisation and Overfitting

During learning, the outputs of a supervised neural network attempt to approximate the target values, given the input-output pairs in the training set. This ability is useful in itself, but the purpose of using a neural net or any other model is to approximate target values, given inputs that are not in the training set. This feature is called *generalization*. There are two conditions that are typically necessary, although not always sufficient, for satisfactory generalization. The first condition is that the function relating the inputs to correct outputs be smooth. In other words, a small change in the inputs should produce a relatively small change in the outputs. For continuous inputs and targets, smoothness of the function implies continuity and restrictions on the first derivative over the input space.

The second necessary condition for good generalization is that the training cases be a sufficiently large and representative subset, that is, a "sample" in statistical terminology, of the set of all cases that is needed to be generalized to, i.e. the "population". The importance of this condition is related to the fact that there are two different types of generalization: interpolation and extrapolation. Interpolation applies to data points that are surrounded by nearby training data points. Cases that are outside the range of the training data require extrapolation. Cases inside large "holes" in the training data may also effectively require extrapolation. Interpolation can often be done reliably, but extrapolation is notoriously unreliable. Hence it is important to have sufficient training data to avoid the need for extrapolation.

Neural Networks are proven to be universal function approximators[65]. They are actually able to approximate any arbitrary function to any degree of accuracy required for a given "sample". However, it should be born in mind that a good fit to a "sample" subset does not necessarily constitute a good fit to the whole "population". Indeed, the increased accuracy with the increased number of iterations or excessive number of processing nodes will harm the generalisation of the model to the unseen data in the population. Hence, in addition to careful consideration for sampling of the data, measures should be taken to avoid "*overfitting*" while

training[76,77]. An operational definition of overfitting is that the out-of-sample error starts to increase with training time after having gone through a minimum. The overfitting to a 'sample' data by neural networks is analogous to fitting an excessively large order polynomial to this sample set. This polynomial will fit the sample data well, however, it will not provide accurate and reliable predictions for unseen data in the population.

A simple way of detecting overfitting is to use the split-sample technique, i.e. splitting data into training and testing data. The mean difference of the network predictions and desired outputs of the unseen test data starts to increase once the increased number of iterations begins to cause overfitting. In the case of large databases with abundant data points, split-sample validation can be readily and reliably used. However, with limited size of available data, the cross-validation technique is advisable, since it allows all of the data to be used for training, yet providing a reasonable generalisation. In a more elaborate version of this technique, called leave-one-out cross-validation, the network is trained with only one data point left out. The training cycle is repeated as many times as the number of data points. The optimum generalisation is achieved once the sum of the errors (prediction difference for the left out point) of all training cycles is minimum.

Therefore, in assessing the performance quality of a neural network, the most important factor is to examine whether the network can successfully *generalise* what it has learned from the training data to the unseen test data.

4.3 Back-Propagation Encoding Algorithm

The simplest back-propagation algorithm minimises the squared error cost function (delta or LMS rule) in a three-layered, fully connected, feedforward topology, similar to the one shown in **Figure 4-2**. A step-by-step procedure for encoding the algorithm using incremental updates (one pattern μ at a time) is presented below.

4.3.1 Pre-Processing of the Data

All the input and output values in the dataset first have to be normalised into the range [-0.9 to 0.9], to avoid premature saturation of the sigmoid function. Similar to a linear interpolation scheme, the following formulae are used for normalising the input and output data:

$$Y = Y_{min} + \frac{Y_{max} - Y_{min}}{X_{max} - X_{min}}(X - X_{min}) \quad \textbf{(4-19)}$$

where X is the original value, Y, is the normalised value, Y_{min} and Y_{max} are 0.1 and 0.9 respectively, and X_{min} and X_{max} are the minimum and maximum values of the data. After training, the inverse operation should be performed to recover the original data:

$$X = X_{min} + \frac{X_{max} - X_{min}}{Y_{max} - Y_{min}}(Y - Y_{min}) \quad \textbf{(4-20)}$$

4.3.2 Network Generation and Initialisation

a) Define vector variables to represent the activations of input, hidden and output layer nodes.
b) Define vector variables to represent thresholds of hidden and output nodes.
c) Define a 2-dimensional matrix variable to represent connection weights, from input-to-hidden layer and from hidden-to-output layer.
d) Assign random values in the range [-0.9 - 0.9] to all weights and thresholds.

4.3.3 Forward Propagation of Data and Error Calculation

a) Randomly, take one pattern (ξ_k^μ, ζ_i^μ) from the training dataset and apply the input vector ξ_k^μ to input nodes.
b) Propagate the input signals forward through the network to calculate new activations of each hidden node, V_j^μ, using Equation (4.3)

$$V_j^\mu = g\left(h_j^\mu\right) = g\left(\sum_{k=1} w_{jk}\xi_k^\mu - \theta_j\right) \quad \textbf{(4-21)}$$

where

$$g(h) = \frac{1}{1 + \exp(-2\beta h)} \quad \textbf{(4-22)}$$

is the logistic sigmoid threshold function, with the steepness parameter β, often set to 0.5.

c) Propagate the activations of hidden nodes, V_j^μ, to calculate the new activations of output nodes, O_i^μ, using a version of Equation (4.7):

$$O_i^\mu = g\left(h_i^\mu\right) = g\left(\sum_{j=0} W_{ij} V_j^\mu\right) = g\left(\sum_{j=1} W_{ij} V_j^\mu - \Gamma_i\right) \tag{4-23}$$

where Γ_i is the current threshold value of output node i.

d) Compute the error between computed output value, O_i^μ, and desired output value, ζ_i^μ, through Equation (4.12):

$$\delta_i^\mu = g'(h_i^\mu)\,[\zeta_i^\mu - O_i^\mu] \tag{4-24}$$

The derivative of the logistic sigmoid function with steepness factor of 1/2 is:

$$g'(h) = 2\beta g(1-g) = g(1-g) \tag{4-25}$$

Thus Equation (4.12) can be used in the convenient form of:

$$\delta_i^\mu = g(1-g)[\zeta_i^\mu - O_i^\mu] = O_i^\mu\left(1 - O_i^\mu\right)[\zeta_i^\mu - O_i^\mu] \tag{4-26}$$

e) Compute the error of each hidden node output relative to each output node error, using Equation (4.15):

$$\delta_j^\mu = g'(h_j^\mu)\sum_i W_{ij}\delta_i^\mu = V_j^\mu\left(1 - V_j^\mu\right)\sum_i W_{ij}\delta_i^\mu \tag{4-27}$$

4.3.4 Weight and Threshold Adjustments

a) Adjust the hidden-to-output connections weights, according to the *delta rule* given in Equation (4.17):

$$\Delta W_{ij} = \eta \delta_i^{\mu} V_j^{\mu} \tag{4-28}$$

b) Adjust the output node thresholds according to gradient descent scheme:

$$\Delta \Gamma_i^{\mu} = \eta \delta_i^{\mu} \tag{4-29}$$

c) Modify input-to-hidden connections weights according to the Equation (4.18):

$$\Delta w_{jk} = \eta \delta_j^{\mu} \xi_k^{\mu} \tag{4-30}$$

d) Adjust the output node thresholds according to gradient descent scheme:

$$\Delta \theta_j^{\mu} = \eta \delta_j^{\mu} \tag{4-31}$$

Notice that the learning rate, η, is chosen to be the same for both hidden and output layers, however, it doesn't necessarily need to be.

4.3.5 Iterative Learning

With the random selection of initial weights and thresholds, it is obvious that in the first iterations, there is a large error, δ_i^{μ}, in the network predictions. However, after repeating steps listed in the Sections 4.3.4 and 4.3.5 for new random patterns the error value will be reduced until desired accuracy is achieved without sacrificing the network generalisation.

4.3.6 Overfitting and Generalisation

With continued iterative learning, the network prediction error on training data decreases. However, excessive learning will cause the network to learn not only the real trend of the data, but also the noise which is almost always present in practical datasets. Once the network starts mapping the noise as well, the sum of the error on the test data will start increasing, an indication of excessive training or overfitting. To avoid overfitting, the network

performance after each learning cycle should be tested on the test data and learning should be terminated once the error starts to increase.

4.3.7 Recall

The recall procedure consists of a one-shot forward propagation of any given pattern. Once the final weights and thresholds are determined, the input of the data pair is plugged into the input nodes, and the output is computed through Equations (4.3) and (4.22).
The above-mentioned encoding algorithm has been summarised in the form of the flow chart of **Figure 4-**.

4.4 Comparison of Neural Net Models with Statistical Models

There is a considerable overlap between the fields of neural networks and statistics since both are concerned with data analysis. Most neural nets that can learn to generalise from noisy data are functionally similar to statistical methods.

For example, the feedforward nets with no hidden layer are basically generalised linear models and feedforward nets with one or more hidden layers are similar to non-linear regression models. However, the fundamental difference in their mechanics is that the success of statistical methods is very much dependent on availability of an underlying mathematical equation while neural nets do not require *a priori* knowledge of functional relationships between the independent and dependent parameters

There are a large number of studies comparing the neural network models and their statistical counterparts[78,79]. Most such studies involve one or two data sets and no general conclusions can be drawn. However, there are some more involved comparative studies of these two techniques, discussing the important issues of data distribution and sampling, over-training and under-training, noise effects, and generalisation[80]. It is concluded that neural networks are becoming increasingly popular as modern modelling and process automation tools because of their higher accuracy, versatility and speed.

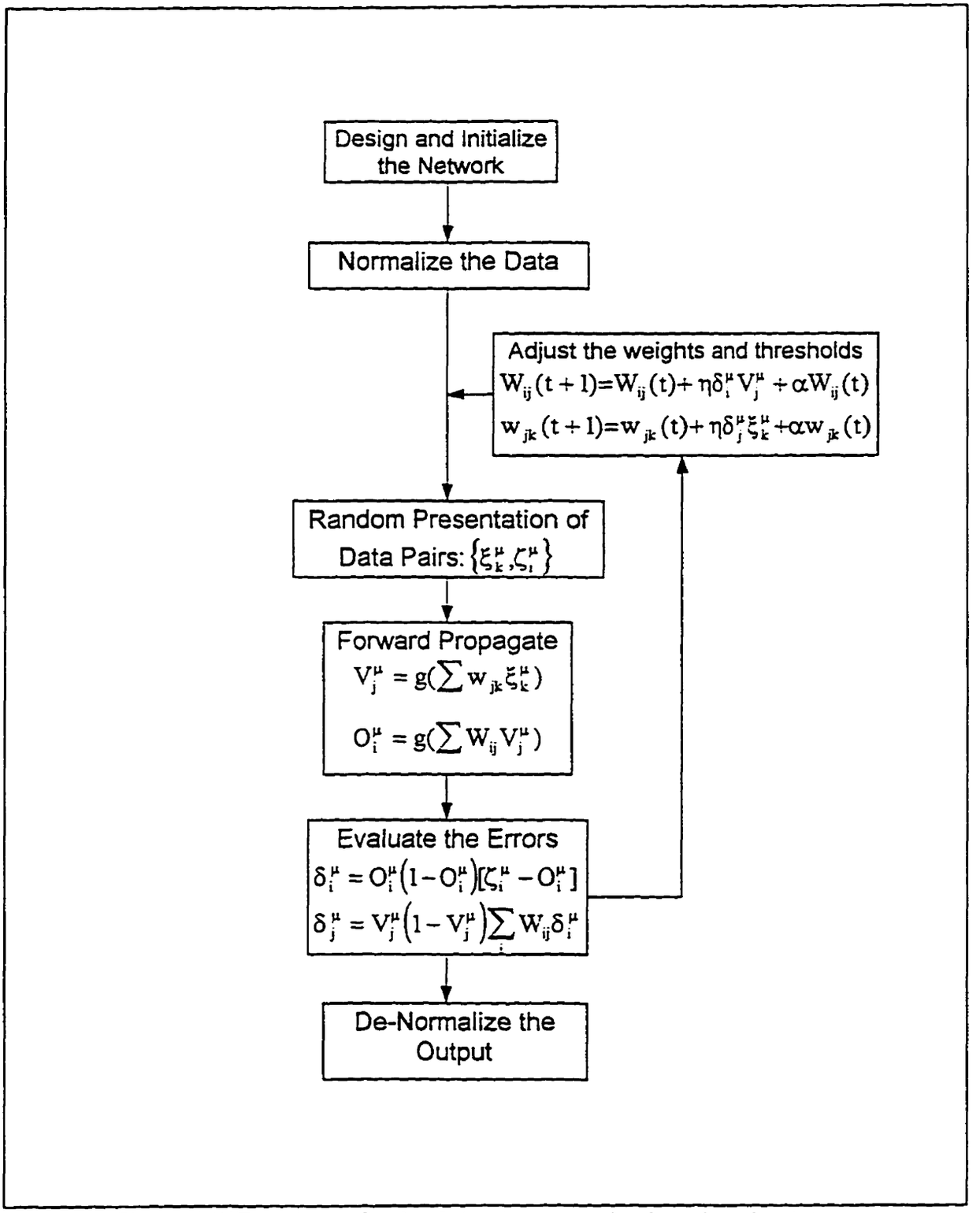

Figure 4-5 Flow chart of a simple backpropagation learning algorithm

4.5 Applications in Materials Science and Processing

Artificial Neural Networks can be looked at as a modelling and classification tool. Hence, it can replace regression based models of any kind. However, the advantages of the technique are more pronounced once the process is highly multi-dimensional and there is no form of known mathematical equation for representation of the input-output relationship. Some of these successful applications are listed below:

a) Hwu et al.[81] used a backpropagation neural net to predict satisfactorily the flow stresses of an extra-low carbon steel in the austenite, ferrite and dual-phase regions, within a ±2% accuracy.

b) Rao and Prasad[82] used a commercial Neural Network package, NWorks[83], to model the flow stress of a low carbon steel. A mean absolute percentage error of 3.01% was reported.

c) Tsoi[84] used a special class of Neural Networks, called MARS (Multiple Adaptive Regression Spline), to predict flow stresses in a plate rolling mill, with a 5.3% accuracy, surpassing those of conventional statistical models.

d) Schmitter[85] used Neural Networks to analyse computer images of polished steel in order to automate grain size determination and classification of iron carbides.

e) Yun and Chang[86] reported successful application of feedforward backpropagation networks to dynamic prediction of a BOF process in Pohang Iron and Steel Company.

f) Siemens Energy and Automation is actively involved with design and implementation of Neural Network based controllers for rolling mill gauge and width control, prediction of amount of lateral spread in flat rolling, prediction of strip thickness profile, and automatic strip classification by relation to deformation resistance.[87,88] They applied these models to the Westfallen mill in Germany and reported that large improvements in accuracy have been achieved through the application of Neural Networks in combination with algorithmic models in the control of this hot strip mill.

g) Wu et al[89] modelled the behaviour of concrete in the state of plane stress under monotonic biaxial loading and under compressive uniaxial cyclic loading with backpropagation neural networks.

h) Armand et al.[90] used neural networks to develop inverse dynamic model of a GMA welding process and successfully controlled the weld width using this model.

4.6 Summary

Artificial Neural Networks are a new class of heuristic modelling methods. One of the main advantages of this approach is that there is no need to make any *a priori* assumptions about material behaviour or the physics of the process, even though more sophisticated neural network modelling schemes may take advantage of existing knowledge of the process in the network design.

Although the multi-layered neural network models cannot guarantee a global minimum solution for any given problem, it is a reasonable assumption that the resulting model will approximate all of the laws of mechanics which the actual material or process must obey, if the neural network is trained using a "comprehensive" database with an appropriate representation scheme.

Another versatility of the technique is its ability to continuously update through a training session that includes new experimental or process data. Well-trained Neural Network models provide fast, accurate and consistent results. It is superior to all other techniques in terms of speed of mathematical manipulation. The trained network will produce an output for any given input by one-step propagation of the inputs through the processing nodes to the output nodes. This operation can be performed in milliseconds, making the technique favourable for on-line control applications.

There are some drawbacks to this technique. The network does not represent the knowledge in the familiar form of symbolic equation-based models. Thus, reasoning about material or process behaviour and desired changes would not be as intuitive as they are in regression methods. Network training relies on a "comprehensive" database of the given problem. However, the criteria to judge what constitutes the comprehensiveness of the data is not well defined and obvious.

4.7 Nomenclature

O_i^μ output of output node I for pattern μ

V_j^μ output of hidden node j for pattern μ

W_{ij} connection weights from the hidden units to the output units

$g(h)$ activation function (sigmoid)

h_j^μ sum of the inputs to the node j for pattern μ

i output layer index

j hidden layer index

k input layer index

p number of input patterns or data

w_{jk} connection weights from the inputs to the hidden units

α momentum parameter

η learning rate

μ pattern or observation number

$\{\xi_k^\mu, \zeta_i^\mu\}$ input-output pattern pair

θ_j threshold of hidden nodes

Γ_i threshold of output nodes

5. Experimental Equipment and Techniques

The experimental work reported in this thesis was performed to simulate thermo-mechanical processing of high niobium grade HSLA steels during hot strip rolling. The main objectives were to verify the possibility of the occurrence of dynamic recrystallization during strip rolling and to investigate the effects of process parameters on the softening mechanisms acting during this process. There are a number of testing techniques for the purpose of physical simulation of hot rolling processes. An understanding of the advantages and limitations of these testing techniques and selection of the right test method for any given task is crucial.

5.1 Physical Simulation Techniques of Thermo-Mechanical Processing

Physical simulation based on laboratory tests is the most common and in many cases the only reliable method to study the process of hot rolling, particularly during the development of new alloy grades. The usual testing methods are torsion testing, plane strain and axisymmetric compression. Torsion is the testing technique that is most widely used for the determination of the characteristics of flow curves and for carrying out rolling simulations. This is because torsion based testing allows large strains without tool-specimen friction problems. With regard to kinetics of recrystallization and precipitation, the axisymmetric compression test has turned out to be the most popular, even though the rolling of flat products involves plane strain deformation. Advantages and disadvantages of the different testing techniques are discussed in detail by Lenard et al.[91]

The principal limitation of physical simulation arises from the fact that laboratory testing methods are unable to reproduce the successive, rapid, and large deformations that occur in industrial rolling[92]. For example, while strain rates during hot strip rolling can exceed 100 s^{-1} and rod mill strain rates can reach 1000 s^{-1}, most laboratory torsion machines are operated at 20 s^{-1} or less. This leads to discrepancies in the deformation times of one to two orders of magnitude. Plane strain compression machines and cam plastometers can be operated at 100 s^{-1} or higher, but they are not well suited to carrying out seven or more successive deformations required for the modelling of the entire strip rolling process.

Attempts have been made to apply the principles of the "Similarity Law" to extrapolate from laboratory experimental results with lower strain rates to values that would be attained in the industrial mill practice[93]. The proposed scheme is to scale down the interpass times to compensate for the lower strain rates of experimental equipment. The similarity law is applicable only when there is no mechanism change between the model and the prototype. However, it is expected that changes in the strain rates will influence the restoration mechanisms (dynamic versus static), thus making the applicability of the similarity law debatable. The advancements in test equipment remain the most promising way of overcoming the strain rate limitation for rolling simulations.

A review of different test techniques for physical simulation of thermo-mechanical processes, including their advantages and limitations, is given in the next sub-sections. Then the principles of the Similarity Law to compensate for low strain rates of laboratory tests, used during the torsion simulations, are described.

5.1.1 Plane Strain Compression

Rolling of flat products involves plane strain deformation. Hence, plane strain compression is the first prospective test method for hot rolling simulation. Ford[94] devised a plane strain compression test (known as the Ford Test) in which only part of the specimen deforms and the rigid metal outside the plastic zone prevents the spread of deformation in one direction. The results of his experiments were quite consistent with the data from cold rolling

mills. The most comprehensive work on plane strain compression testing at elevated temperatures, however, has been carried out by Sellars and co-workers[95].

Using cam plastometers, large strains and strain rates of up to 140 s^{-1} are attainable for one or two stages of compression, however, for simulation of the whole rolling process multistage compression of seven or so hits are required. The pressure experienced by the tools in plane strain compression testing is shown to vary with the w/h ratio (tool width over specimen thickness) both theoretically and experimentally. This variation arises from the fact that different amounts of redundant work are involved in the deformation at different w/h ratios. **Figure 5-1** demonstrates this variation. Approximate solutions based on slip line field theory are available for rigid plastic materials to account for the redundant work done in metal working operations. Another drawback to be mentioned is the non-uniform deformation along slip zones (different from crystallographic slip planes) as depicted in **Figure 5-2**.

Friction between the tools and the specimen increases the pressure required to produce yielding in the specimen. This effect has been dealt with analytically and approximate solutions have been obtained which have been experimentally verified for lower coefficients of friction and low strains.[96] From a practical point of view, due to the larger specimens involved in this method, heating, deformation and quenching of them are more difficult than specimens of other techniques.

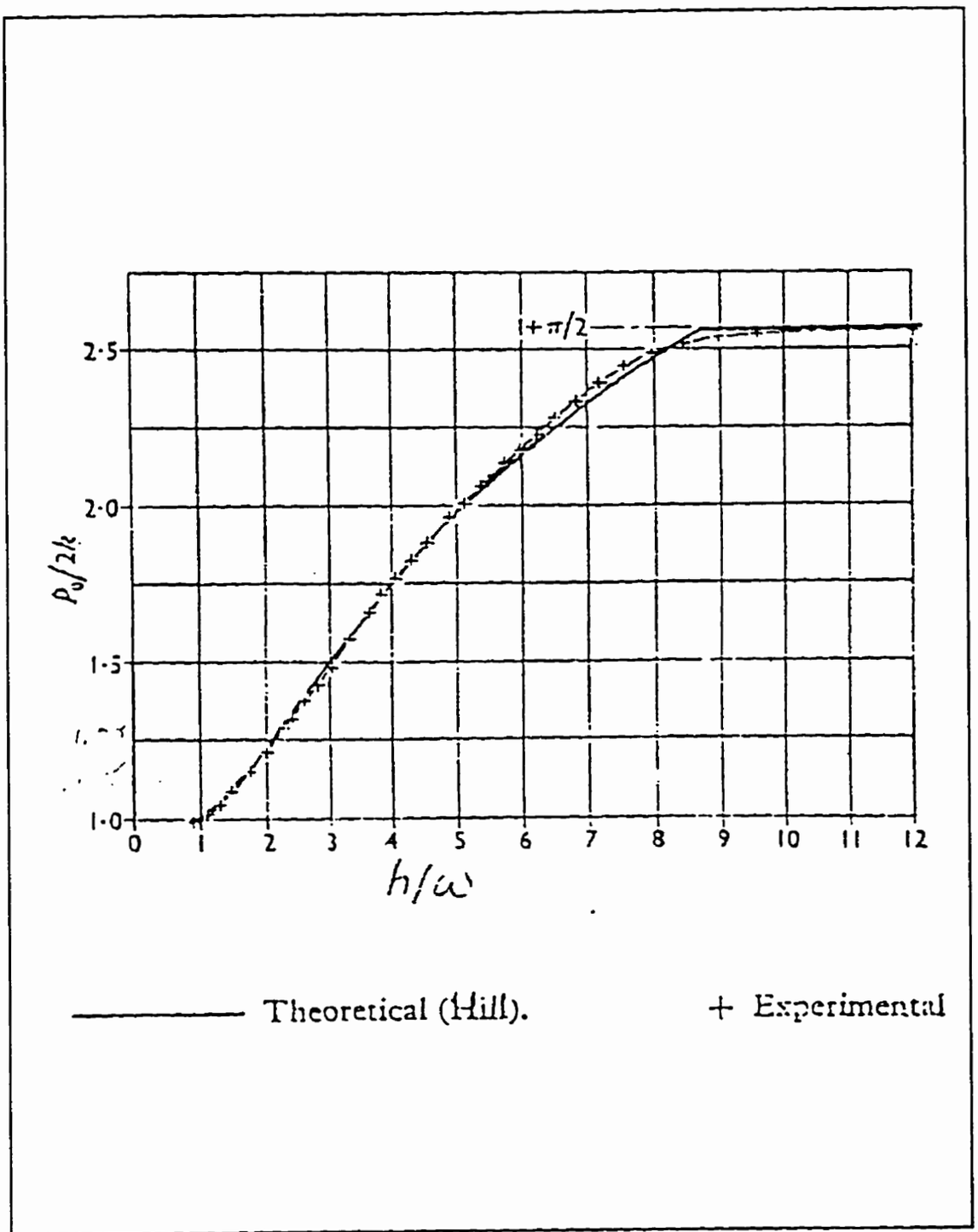

Figure 5-1 Variation of tool pressure with h/w ratio in plane strain compression testing[95].

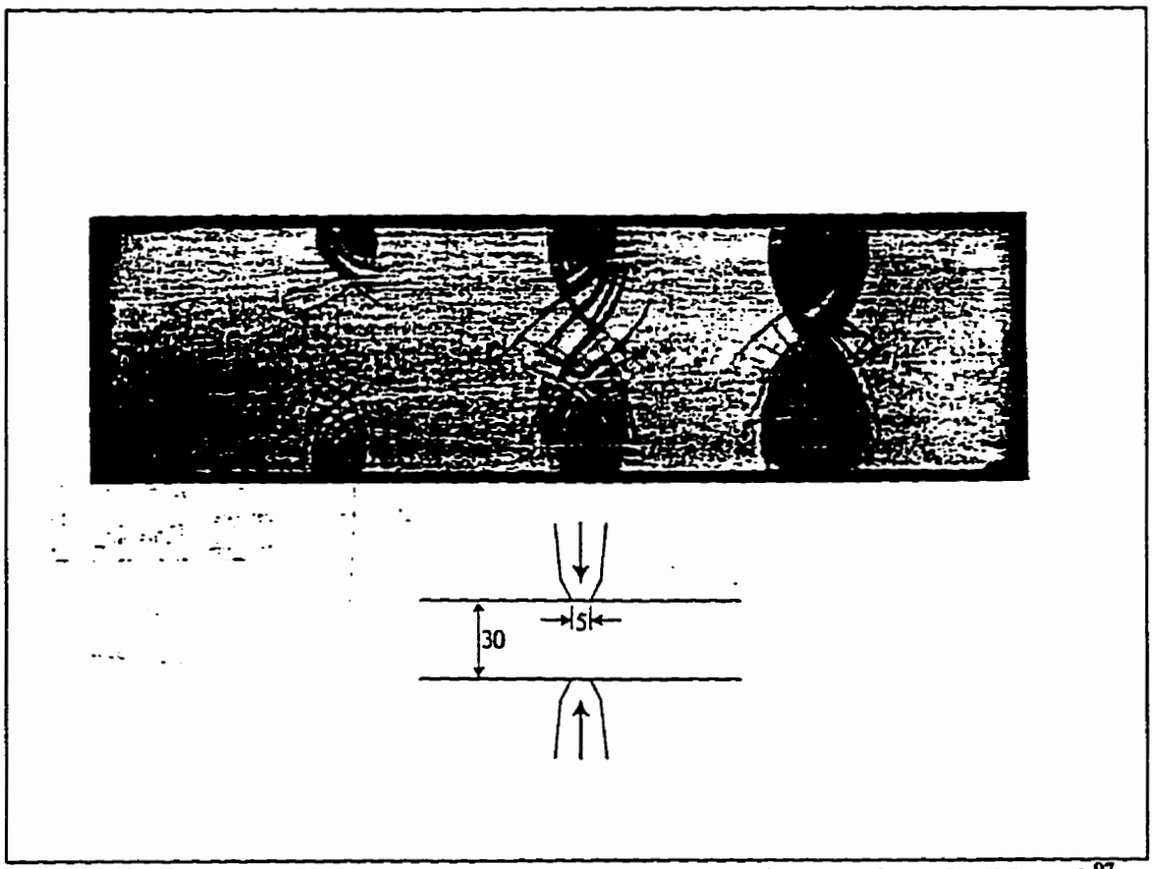

Figure 5-2 The development of the plastic zone in the indentation of mild steel.[97]

5.1.2 Axisymmetric Compression

As mentioned above, for the study of kinetics of precipitation and recrystallization, this method has turned out to be the most popular one because the experimental equipment is less expensive and the procedure is faster than plane strain testing. In work done by Sellars and coworkers[95], there was good agreement between the flow curves obtained by plane strain and axisymmetric compression, as depicted in **Figure 5-3**.

Ford[94] has also compared the results of plane strain compression testing with uniaxial compression testing. The stress-strain curves obtained by the two methods diverge progressively beyond a strain of 0.2 as shown in **Figure 5-4**.

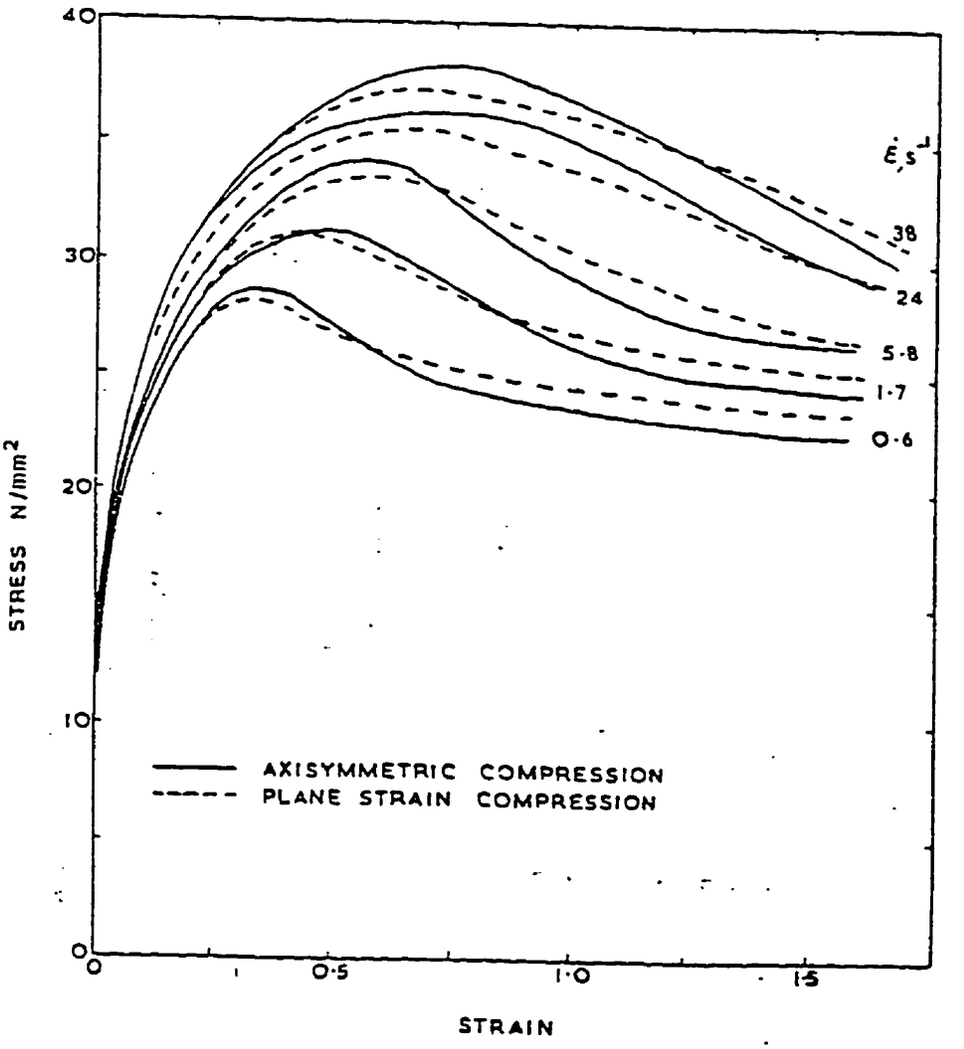

Figure 5-3 Stress-strain curves for lead obtained from axisymmetric compression tests at constant strain rate.[95]

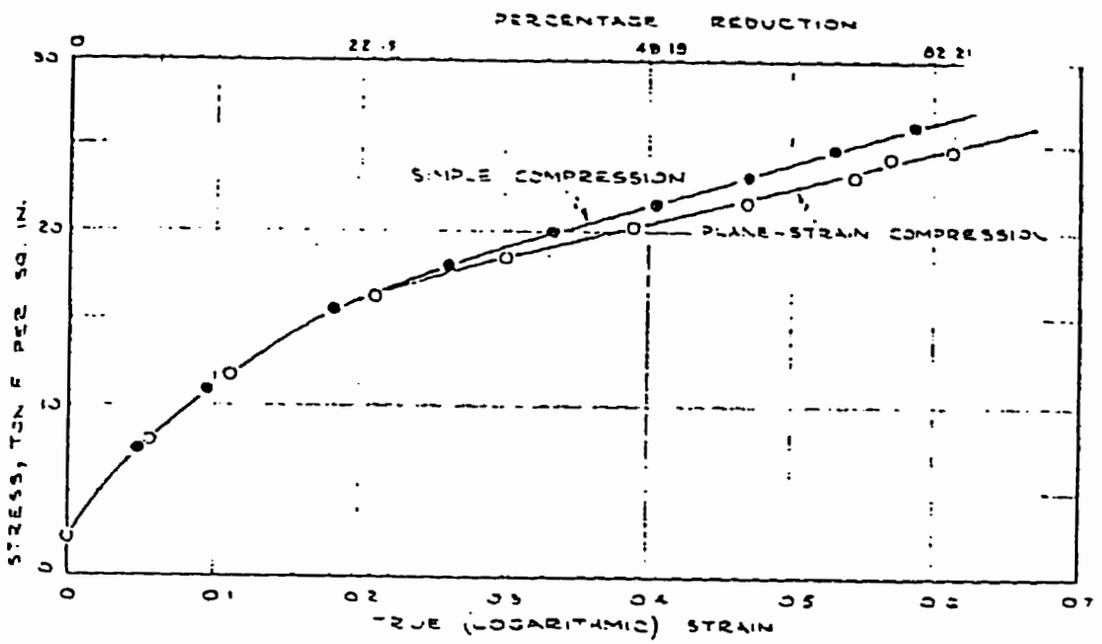

Figure 5-4 Comparison of stress-strain curves of a mild steel in axisymmetric compression and plane strain compression tests[95].

Higher stresses in axisymmetric compression are attributed to higher frictional forces and the different manner of anisotropy development in this method. This anisotropy is caused by barrelling due to high frictional forces at the specimen-tool interface. Friction increases progressively with strain and is considered as a limiting parameter when large deformations are being studied.

5.1.3 Torsion

Large successive strains and relatively high strain rates (20 s^{-1}) with the aid of simple equipment have made this test method popular for the simulation of hot rolling processes. Although it is relatively straightforward to convert test results from one mode of deformation to another using the von Mises effective stress and effective strain, specific problems arise when torsion tests are carried out on solid bars. Because of the strain gradient present in solid torsion bars, the strain at the peak stress, determined in torsion, has been shown to be about 1.3 times higher than the more accurate values determined in tension or compression[98]. Still, an even more serious problem is the overestimation of torsion-based recrystallization critical strain, ε_x, up to a factor of 3 too high, compared to those measured by other techniques[99]. To clarify this problem, one should consider the different ways of deriving stress-strain curves from the torque-twist data determined in the torsion testing of solid bars. In the most frequently used method, developed by Fields and Backofen[100], the shear strain in the surface layer is given by:

$$\gamma_R = \theta\frac{R}{L} \tag{5-1}$$

where θ is the angle of twist, and R and L are the radius and gauge length of the specimen, respectively. When the von Mises convention is applied, effective strain is obtained as:

$$\bar{\varepsilon} = \frac{\gamma_R}{\sqrt{3}} = \frac{R\theta}{\sqrt{3}L} \tag{5-2}$$

The surface shear stress τ_R is deduced from the measured torque or moment M through the relationship:

$$\tau_R = \left(\frac{M}{2\pi R^3}\right)(3+m+n) \tag{5-3}$$

where m and n are, respectively, the twist rate sensitivity of the torque and the normalised torque hardening rate of the material. Again, the effective stress, $\bar{\sigma}$, in the surface layer can be found from τ_R with the aid of the von Mises assumption, leading to:

$$\overline{\sigma_R} = \sqrt{3}\ \tau_R \qquad (5\text{-}4)$$

For the materials obeying simple work hardening laws, the m and n values are constant and the Equation (5-4) yields exact solutions. However, it is only an approximation when applied to materials being deformed at elevated temperatures, under which conditions the work hardening rate n is itself deformation rate dependent and m is dependent on strain. Some work[101] has been done to develop exact methods for the calculation of τ_R, but unfortunately they require more involved analysis, as well as the use of several samples of increasing diameter for each condition to be investigated, and so are considerably less convenient to use. In this work, formulations developed by Fields and Backofen were used to convert the torque-twist data to stress-strain relationship.

The approximations involved in the application of Equation (5-4) to high temperature flow lead to an overestimation of ε_p and an even larger overestimation in the determination of ε_x.[99] This arises because the peak strain is attained at first in the outermost layer of a torsion bar and n becomes zero at this layer, followed by softening ($n < 0$) while the inner layers are still experiencing work hardening at different rates. The moment developed in the torsion bar is of course the sum of the moments developed in different layers, however, when some of these are decreasing while some are increasing, the behaviour of a particular layer can be masked by the influence of the others. Thus, when the outer shell attains the peak stress and begins to soften, the bar as a whole continues to show hardening. As a result, the peak in the overall bar is attained later than the real instant associated with the τ_{max} in the outermost shell, leading to an overestimation of ε_p and ε_x.

This averaging error can be mostly eliminated by using tubular torsion specimens, if enough care is taken to avoid buckling at high strains. However, even with tubular specimens there remain differences between the flow curves of torsion testing and those determined by other techniques. This discrepancy is attributable to differences in textures developed at large strains by deformation along different strain paths. In plane strain and axisymmetric compression testing methods, the material directions associated with the axes of symmetry of

the sample remain fixed. In torsion, by contrast, the material direction at the surface gradually changes from longitudinal to transverse[102]. Thus, even if the current value of shear stress would be the same for different testing methods, the associated macroscopic flow stress is slightly different. Another source of discrepancy is the slight difference in work hardening rate of the material in torsion testing compared to others in which it is lower.

5.2 Strain Rate Corrections

The repetition of industrial strain rates is not possible by the existing simulation equipment. Hence, the effects of slower strain rates on metallurgical events during hot rolling should be taken into account and proper corrections should be made on test procedures. Several workers[103,104,105] have attempted to determine strain rate corrections to the pass strain and interpass times when simulating high strain rate mill processing, using relatively low strain rate laboratory torsion tests. The principles of these corrections are based on the variation of the kinetics of recrystallization with strain rate. In order for direct application of laboratory test results to mill practice, the similarity of the softening that occurs in the lab and in the mill, should be maintained.

5.2.1 Corrections for Strain Rate during Static Recrystallization

There are reports indicating the slight effect of strain rate on the kinetics of static recrystallization in C-Mn steels[106,107], however, most of the other works[108,109,110] have confirmed the results of Sellars[111] denoting the independence of static recrystallization kinetics from strain rate and a strong dependence on the strain, initial grain size, and temperature. This kinetics in terms of time for 50% softening for C-Mn steels has been formulated as:

$$t_{0.5} = 2.5 \times 110^{-19} \varepsilon^{-4} d_0^2 \exp\left(\frac{Q}{RT}\right) \qquad \textbf{(5-5)}$$

and the time t for X percent softening can be calculated from:

$$X = 1 - \exp\left[-0.693\left(\frac{t}{t_{0.5}}\right)^{n}\right] \tag{5-6}$$

where n depends on steel composition.

However, it has been shown that strain at the peak stress, ε_p, is increasing systematically with the Zener-Hollomon parameter[111]. Under isothermal test conditions, strain to the peak stress is a function of strain rate as follows:

$$\varepsilon_p = A d_o^{1/2} Z^n \tag{5-7}$$

where A is a material constant and n varies from 0.125 to 0.175 with no apparent systematic trend with the composition of the steel. The mean value of 0.15 has been used by other authors[103,105]. Zener-Hollomon parameter, Z, is given as:

$$Z = \dot{\varepsilon} \exp\left(\frac{Q_{def}}{RT}\right) \tag{5-8}$$

The peak in flow stress occurs after some low fraction of recrystallization has taken place so the strain to the peak stress, ε_p, is always greater than ε_c, the critical strain for dynamic recrystallization. It has been suggested[112] that $\varepsilon_c = a\varepsilon_p$ is a reasonable approximation for most of the steels, where a is a constant around 0.85. Since the peak occurs at the different levels of strain for different strain rates, the pass strain for the first few passes (prior to the initiation of dynamic recrystallization) needs to be adjusted so that the peak will be reached at the same point in the deformation process. This can be readily done by combining Equations (5-7) and (5-8). If ε_l and $\dot{\varepsilon}_l$ denote laboratory test strain and strain rate respectively and ε_m and $\dot{\varepsilon}_m$ represent mill strain and strain rate, then the corrected lab strain can be calculated as:

$$\varepsilon_l = \varepsilon_m \left(\frac{\dot{\varepsilon}_l}{\dot{\varepsilon}_m} \right)^{0.15} \tag{5-9}$$

Having adjusted the strain per pass, the interpass time must also be adjusted in order for the similarity of the softening between passes to be maintained. Considering the kinetics of static recrystallization characterised by the time for 50% softening (Equation 5-5), the corrected interpass time for the laboratory test can be calculated:

$$t_l = t_m \left(\frac{\varepsilon_m}{\varepsilon_l} \right)^4 \tag{5-10}$$

By combining Equations (5-9) and (5-10), the interpass time correction can be calculated directly from the strain rates:

$$t_l = t_m \left(\frac{\dot{\varepsilon}_m}{\dot{\varepsilon}_l} \right)^{0.6} \tag{5-11}$$

where t_l and t_m are laboratory and mill interpass times, respectively.

5.2.2 Corrections for Strain Rate During Post-Dynamic Recrystallization

The softening kinetics after deformations beyond peak strain, i.e. after occurrence of dynamic recrystallization, are considerably accelerated. This is due to the large number of new grain nuclei formed during the onset of dynamic recrystallization. Since the nature of the grain nucleation differs from that under static recrystallization, and its kinetics are marked by the characteristics of dynamic recrystallization, the static softening after dynamic recrystallization has been called Metadynamic Recrystallization. The rate of metadynamic recrystallization is highly sensitive to the strain rate, depends to a lesser degree on temperature and depends very little on strain[112,113]. This differs from the usual dependencies reported for static recrystallization, i.e. strong dependence on strain and temperature, and little on strain rate. The metadynamically recrystallized grain size is also highly dependent on strain rate and to a lesser degree on the temperature. The higher the strain rate, the finer the grain size. This is also in

contrast with equivalent expressions for static recrystallization, which indicate no dependence of the grain size on strain rate. Accepting the above facts, it is important to apply strain rate corrections to experimental interpass times when attempting to simulate high strain rate rolling processes using low strain rate laboratory tests. If the uncorrected mill interpass times are employed in the low strain rate tests, considerably less interpass softening will take place than those experienced in the mill.

The corrections for strain rate during the passes after the initiation of dynamic recrystallization only include interpass time corrections, i.e. pass strain no longer needs to be modified. To apply these corrections, the knowledge of the kinetics of metadynamic recrystallization is required. Several authors[104,105,107] have reported more or less similar form of dependence of these kinetics on strain rate and temperature for different steels. Roucoules[113] studied a Mo steel and assumed that metadynamic recrystallization starts at the end of the recovery stage and finishes at 90% softening. She developed the following expression for the time for 50% recrystallization:

$$t_{0.5}(s) = 6.66 \times 10^{-6}\, \dot{\varepsilon}^{-0.61\pm0.1} \exp\left(\frac{123(kJ/mole) \pm 13}{RT}\right) \qquad \textbf{(5-12)}$$

If it is verified that the same kinetics are applied for different percents of softening, then this relationship can be readily utilised to determine experimental interpass times to yield equivalent softening experienced in the mill. Under isothermal test conditions, the relationship between test and mill interpass times is:

$$t_i = t_m \left(\frac{\dot{\varepsilon}_m}{\dot{\varepsilon}_i}\right)^{0.61} \qquad \textbf{(5-13)}$$

A successful application of above corrections to the simulation of rod rolling has been reported by Maccagno and Jonas[104].

5.2.3 Implications of Strain Rate Corrections on Precipitation

Using Equation (5-13) for corrections required for interpass times when using laboratory strain rates one order of magnitude lower than mill strain rates, results in interpass times four times longer than mill interpass times. If the laboratory strain rates are two orders of magnitude lower, then interpass times will be increased 16 times. Longer deformation times due to the lower strain rates and longer interpass times lead to much longer times for nucleation and growth of carbonitride precipitates. Significant strain-induced and dynamic precipitation can effectively stop recrystallization from taking place. This implies that the interaction between precipitation and recrystallization should be carefully taken into account. This interaction is complex, however, it is interesting to note that the dependence of kinetics of recrystallization and precipitation on strain rate is similar. Dutta and Sellars[114] modelled the time for 0.05% strain-induced precipitation in a high Nb steel as:

$$t_{0.05} = A[Nb]^{-1}\varepsilon^{-1}Z^{-0.5}\exp\left(\frac{270(kJ/mole)}{RT}\right)\exp\left(\frac{B}{T^3(\ln K_s)^2}\right) \qquad (5\text{-}14)$$

where A and B are material constants, [Nb] is niobium concentration in steel, and K_s is supersaturation ratio, defined as the ratio of element concentration in solution to the equilibrium concentration. Considering the exponent of the strain rate in Equation (5-12) and exponent of Zener-Hollomon parameter in Equation (5-14) this similarity is evident.

Considering the definite interactions of precipitation and recrystallization, the possible effects of the above interpass time corrections on the occurrence of precipitation and its interactions with recrystallization has yet to be studied.

5.3 Compression Testing

5.3.1 Materials

Two Nb-microalloyed steels studied in the present work were received as 30 mm thick pieces, sectioned from transfer bars. The chemical compositions of these steels, called steel number 1 and 2, are shown in **Table 5-1**.

Steel Number 1								
C	Mn	P	S	Si	N	Cu	Ni	Cr
0.067	1.301	0.008	0.003	0.345	0.006	0.013	0.009	0.025
Mo	V	Nb	Al^{tot}	Al^{sol}	Ti	Sn	Ca	
0.004	0.005	0.077	0.037	0.0335	0.0245	0.007	0.003	
Steel Number 2								
C	Mn	P	S	Si	Cu	Ni	Cr	Mo
0.043	1.43	0.011	0.006	0.312	0.037	0.014	0.139	0.252
V	Nb	Al^{sol}	Sn	Ti	Ce	Ca	B	N
0.005	0.075	0.030	0.002	0.008	0.002	0.001	0.001	0.007

Table 5-1 Chemical composition (wt%) of steels number 1 and 2.

5.3.2 Specimen Preparation

Compression specimens with a 12 mm gauge length and 8 mm diameter were machined from the steel supplied. To enhance lubrication, a groove of 0.07 mm depth was machined in the top and bottom of the samples. The samples were annealed to remove prior texture and cooled in air. They were subsequently solution treated for 20 minutes at 1250 °C in an oxygen-free environment (to dissolve the existing precipitates into the matrix) and quenched. In order to control and monitor the temperature, k-type thermocouples were embedded in the middle of the specimens. A high temperature glass lubricant (Deltaglaze 19) was applied to the samples' machined grooves.

5.3.3 Testing System

The specimens were tested using a computer controlled servohydraulic testing system (Instron Model 1331). This machine enables the experimenter to specify the desired wave form (i.e. cross-head speed, constant or varying strain rates, interruptions and delay times, strain at each stage and final strain) off-line. The generated waveform is used to command the ram movement through an Instron 8500 control console. The tests were performed inside a split furnace, in between hot rams, made of INCONEL 718. Details of the testing system are given by Karagiozis and Lenard.[115]

5.3.4 Data Acquisition and Control System

The computer interface to the Instron machine was maintained through a GPIB (General Purpose Interface Board) made by National Instruments. Load and displacement analogue signals from compression tests were digitised by an ATD (analogue to digital converter) embedded in this board. True stress and true strains were calculated by a post-processing program, taking into account machine and rams hot stiffness and interfacial friction. The temperatures of the specimens and rams were recorded through a DAS-8 interface board. A universal analogue input expansion board (EXP-16) was used for cold junction compensation.

5.3.5 Experimental Procedure

In the first part, the effect of the pre-test thermal treatment on the metal's subsequent behaviour was examined using steel number 1. In order to determine the high temperature characteristics of the metal, compression tests to a strain of 1.2 were conducted at a broad range of strain rates and temperatures. In the second, the response of steel number 2 to multi-stage, constant strain rate, isothermal compression was examined. The effects of 2 and 20-seconds holding times at different strain levels on the static recrystallization and the retardation of dynamic recrystallization by precipitation were investigated.

5.4 Torsion Testing

5.4.1 Material and Specimen Preparation

The material used was steel number 1, sectioned from a transfer bar and machined to torsion samples of 25.4 mm (1") gauge length and 12.7 mm (0.5") diameter. The chemical composition is given in **Table 5-1**.

5.4.2 Experimental Equipment

The experiments were carried out using the MTS torsion simulation of Dofasco Inc. A detailed description of the equipment is given elsewhere[116]. The maximum constant surface strain rate is 22 s^{-1} (for this size of specimen), and continuous cooling capability of up to 40 °C per second is attainable with this machine. All tests were performed in high purity argon to minimise oxidation.

5.4.3 Experimental Procedure

Specimens were solution treated for 10 minutes at 1200°C prior to the tests and cooled to the test temperature. Three-stand roughing with pass strains of 0.5 and interpass times of 25 seconds was performed on all of the specimens prior to the finishing twists. It is expected that this roughing schedule would create the same grain size as the slab grain size entering the finishing mill, with the same potential for precipitation. Stress-strain curves were determined from the torque-twist data, using the Fields and Backofen method[100]. During finishing, different cooling rates were used to meet the targeted F7 (mill exit) temperature. In all experiments, the F7 temperature was set to 910±10 °C, which is the targeted exit temperature during rolling of this grade. The finish mill entry temperature (F1) was varied between 1100°C to 950 °C, typical of the strip head and tail entry temperatures. All experiments were carried out under continuous cooling conditions, simulating the conditions in the mill. Post-deformation cooling rate was set to 10 °C/s for all tests. The experimental matrix used is shown in **Table 5-2.**

Test Number 1										
F1 Temp:1100 °C	**Ferrite Grain Size: 4.42 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Strain Rate	2	2	2	2	2	2	2	2	2	2
Afterpass time	25	25	25	2	2	2	2	2	2	2

Test Number 2										
F1 Temp: 950 °C	**Ferrite Grain Size: 4.52 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Strain Rate	2	2	2	2	2	2	2	2	2	2
Afterpass time	25	25	25	2	2	2	2	2	2	2

Test Number 3										
F1 Temp:1100 °C	**Ferrite Grain Size: 4.36 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Strain Rate	10	10	10	10	18	20	20	20	20	20
Afterpass time	25	25	25	2	2	2	2	2	2	2

Table 5-2 Experimental Matrix used for Torsion Simulations

Test Number 4										
F1 Temp: 950 °C	**Ferrite Grain Size: 4.13 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Strain Rate	10	10	10	10	18	20	20	20	20	20
Afterpass time	25	25	25	2	2	2	2	2	2	2

Test Number 5										
F1 Temp:1100 °C	**Ferrite Grain Size: 5.07 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.42	0.41	0.35	0.26	0.20	0.19	0.14
Strain Rate	10	10	10	10	18	20	20	20	20	20
Afterpass time	25	25	25	3.6	2.4	1.7	1.3	1.05	0.87	-

Test Number 6										
F1 Temp:1000 °C	**Ferrite Grain Size: 4.60 μm**									
Stand No.	**R1**	**R2**	**R3**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
Strain	0.5	0.5	0.5	0.42	0.41	0.35	0.26	0.20	0.19	0.14
Strain Rate	10	10	10	10	18	20	20	20	20	20
Afterpass time	25	25	25	3.6	2.4	1.7	1.3	1.05	0.87	-

Table 5-2 Experimental Matrix used for Torsion Simulations (Continued)

Three types of strip rolling simulations were performed:

a) Average schedule (identical strains, strain rates, interpass times and cooling rates for all finishing passes) with a low strain rate of 2 s^{-1}, as other researcher's work[117],[118] (tests 1 and 2). Using an average schedule eliminates the need for taking into account the effects of pass strain and pass to pass strain rate variations.

b) Average schedule with the strain rate of 20 s^{-1} (tests 3 and 4).

c) Torsion simulations of hot strip rolling with a schedule typical of mill practice (tests 5 and 6).

Stress-strain curves, obtained in continuously cooled tests, were corrected for a standardised temperature of 950 °C. For this purpose, the effect of temperature on flow stress was investigated for temperatures between 900 to 1100 °C. These temperature-normalised stress-strain curves were plotted and analysed to examine the true material softening during rolling, independent of the effect of the temperature on the flow stress.

Ferrite grain sizes were estimated using the linear intercept method, with samples of micro-graphs taken at 0.25 mm from the specimen surface.

6. Physical Simulation of Hot Strip Rolling

One of the main purposes of the present research was to investigate the possibility and consequences of the occurrence of dynamic recrystallization during strip rolling of niobium treated HSLA steels through physical simulations. Compression testing was used to study the effects of pre-test thermal treatment on the recrystallization and precipitation kinetics of these steels. Also, interrupted compression tests were conducted to investigate the effects of different holding times on softening mechanisms, retardation of recrystallization, and final grain size. Torsion tests were carried out to determine softening mechanisms which occur during the strip rolling of niobium steels. Mill data was analysed and compared with laboratory results. This chapter presents the results of compression tests, torsion simulations and mill log analysis. These results complement and confirm each other, and lead to the conclusion that dynamic recrystallization does actually occur in strip rolling, even though concrete physical evidence is not yet available.

6.1 Compression Testing

6.1.1 Pre-test Reheat Treatment

In the present study, two different methods of pre-test thermal treatment were investigated. Each method has its own implications on the results of hot deformation simulation and its applicability to industrial practice. In the first method, the specimens are solution treated, quenched, then reheated to the test temperature. In the second method the specimens are solution treated and cooled in the furnace to the test temperature. The methods

yield different initial austenite grain sizes prior to the test, since quenched specimens go through transformation grain refinement. The austenite to ferrite grain ratio in niobium-microalloyed steels is reported to be about two to one[119,120]. In the present work, the average prior-austenite grain size of steel number 2 (refer to Table 5.1) after solution treatment and quenching was found to be 33.4 μm and after transformation to ferrite, the grain size reduced to 13.2 μm. These grains, after reheating to the test temperature and retransformation to austenite, will not grow in size considerably until the temperature rises to the grain coarsening temperature of the material (about 1100 °C for niobium steels). This will translate to higher yield stresses but lower strain to the peak stresses, ε_p, as the strain to the peak stress is directly proportional to the square of grain size.[121] The differences in the flow curves of steel number 1 once different pre-test reheat treatments were employed are shown in **Figure 6-1**. The test temperature was 900 °C and the strain rate was 0.5 s^{-1}. The solution treated and quenched specimen shows almost twice the strength than that of the furnace cooled specimen. It also can be noticed that the high niobium steel does not dynamically recrystallize at this range of temperature and strain rate.

The larger original grains result in larger statically recrystallized grains. Sellars[121] summarised the data of several authors and developed a set of equations that relate the recrystallized grain size to the initial grain size and strain for C-Mn steels:

$$d_{rex} = D\, d_o^{0.67}\, \varepsilon^{-1} \qquad (6\text{-}1)$$

and for niobium treated steels:

$$d_{rex} = D'\, d_o^{0.67}\, \varepsilon^{-0.67} \qquad (6\text{-}2)$$

where D and D' are constants (about 0.9 for niobium steels) and d_o is the initial grain size. Mishra et al.[122] developed similar equations for C-Mn and titanium steels. According to these equations, the recrystallized grain size is a function of the initial grain size.

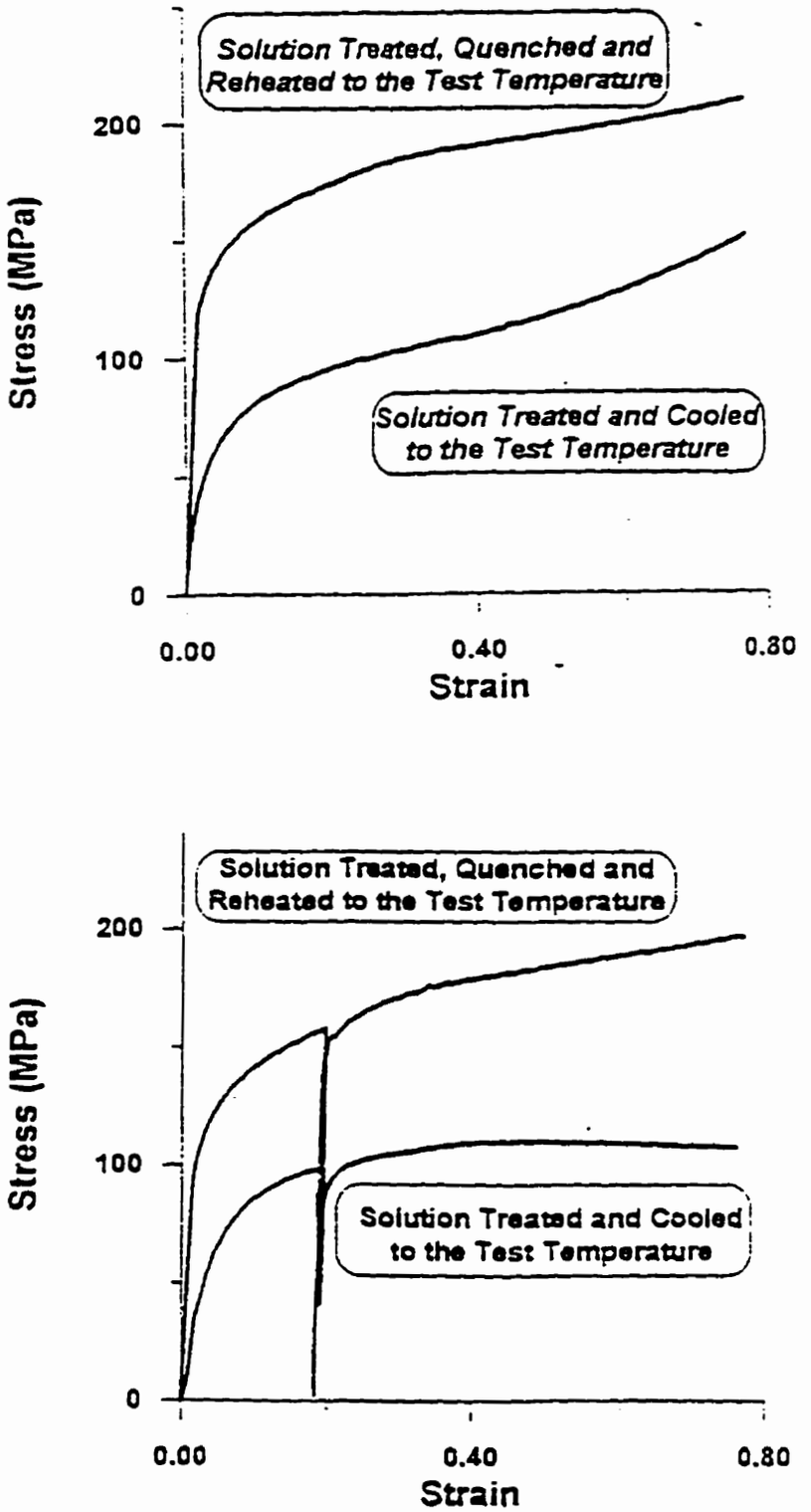

Figure 6-1 The effects of pre-test thermal treatment on the niobium steel flow characteristics at 900 °C and the strain rate of 0.5 s^{-1}.

However, when large multipass reductions are involved, the effect of the initial grain size of the austenite on the final grain size becomes negligible. For example, a niobium treated austenite with initial grain sizes of 300 and 150 μm, upon recrystallization after a strain of 0.3, yields grain sizes of 92 and 58 μm, (ratio of 1.58) respectively. The same initial grain sizes after seven passes of strains of 0.3, produce final grain sizes of 10.12 μm and 9.71 μm respectively (a ratio of 1.04). The work of Cuddy et al.[123] also reveals that, as a result of repeated recrystallization, the initially dissimilar grain sizes rapidly converge to nearly the same final grain size.

The effect of initial grain size on the no-recrystallization temperature has also been studied by several authors. It has been reported that for a given concentration of microalloying elements in solution, the initial grain size prior to multipass deformation has no effect on the no-recrystallization temperature[124]. Yue and Jonas[125] also reported the independence of A_{r3} and T_{nr} from the initial grain size of austenite when these temperatures are determined during multipass deformation.

Another implication of pre-test reheat treatment on the experimental results is the different potentials for precipitation of microalloying elements. This is due to the differences in the time spent at high temperatures and the different pattern of distribution of available elements for precipitation.

6.1.2 Single Stage Compression Tests

A series of single stage compression tests at different test temperatures (800 to 1030 °C) and strain rates (0.001 to 12 s^{-1}) was conducted to study the high temperature deformation behaviour of a niobium steel (steel number 2). The solution treated and quenched samples were held at the test temperature for seven minutes prior to compression. The average grain size after seven minutes at a furnace temperature of 1100 °C and 950 °C were found to be 16 μm and 9.3 μm respectively. Specimens were subsequently isothermally compressed to the strain of 1.2.

The results of these tests were used to determine the peak strain and the dynamic recrystallization or the strain hardening regions at different combinations of strain rates and temperatures. The test results at 800 °C indicated that the onset of dynamic recrystallization occurs at very low strains (around 0.05). This is due to the fact that precipitation of Nb(C,N) is very slow at this temperature and almost all of niobium is in solution. Increasing the test temperature to the range between 900 °C to 1000 °C results in an increase in the peak strain (from 0.1 to 0.5) and a significant retardation of dynamic recrystallization by Nb(C,N) precipitation. As expected, all test results indicate that an increase in strain rate results in an increase in strain to the peak stress, ε_p. Thus, it can be inferred that in industrial rolling practices with the strain rates of more than 20 s^{-1}, the required strain for the onset of dynamic recrystallization would be well above 0.5. However, the peak strain is dependent on deformation history and especially on the extent of carbo-nitride precipitation.

6.1.3 Interrupted Compression tests

Interrupted compression tests were used to study the effects of 2 and 20 seconds holding times (representative of strip and plate rolling interpass times) at different strain levels on the softening mechanisms, retardation of recrystallization and final grain sizes of the steel number 2.

After solution treatment, quenching, and reheating to the test temperature, the specimens were isothermally compressed at a constant strain rate of 0.1 s^{-1} to a total strain of 1.2. Samples of flow curves at both temperatures are presented in **Figure 6-2**. Fractional softening, according to the offset method, was estimated from the interrupted tests. The final average grain sizes after compression along with sample temperature, strain of interruption, fractional softening, and holding times are listed in **Table 6-1**. It can be noticed that the final grain sizes are much smaller than those reported by other workers.[126,127] This is due to the fact that solution treatment, quenching and reheating to the test temperature, cause specimens to go through transformation grain refinement.

All tests conducted at 1030 °C indicate that dynamic recrystallization occurs at a peak strain of around 0.4. However, at the testing temperature of 950°C, the flow stress levels off after the strain of 0.6, thus indicating dynamic recovery, but no dynamic recrystallization. Lowering the temperature to 900°C at the strain rate of 0.5 s^{-1} in a similar steel, results in continuous work-hardening in the range of the applied strain, as depicted in **Figure 6-1**.

When the interruption was made before the peak in the stress was reached, a softening ratio of 0.13 creating a grain size of 3.85 μm, was attained. A softening ratio less than 20% is only attributable to static recovery, and does not involve the nucleation of new grains.[128,129] However, once the same 20-second interruption was applied after the peak strain, i.e. after the onset of dynamic recrystallization, there was a considerable amount of static recrystallization (54%, 55%) and a slight improvement in grain refinement. This is an important observation, confirming the results of Sakai et al.[130], indicating larger softening after the onset of dynamic recrystallization which results in finer grains. In all the tests at the temperature of 950 °C or the interruption of 2 seconds, no measurable static recrystallization was observed. This is confirmed by other works[119,122] and leads to the conclusion that under industrial strip rolling schedules with the interpass times below 2 seconds, and strains of less than 0.5, there is not enough driving force for any appreciable static recrystallization and the metal will experience some strain accumulation.

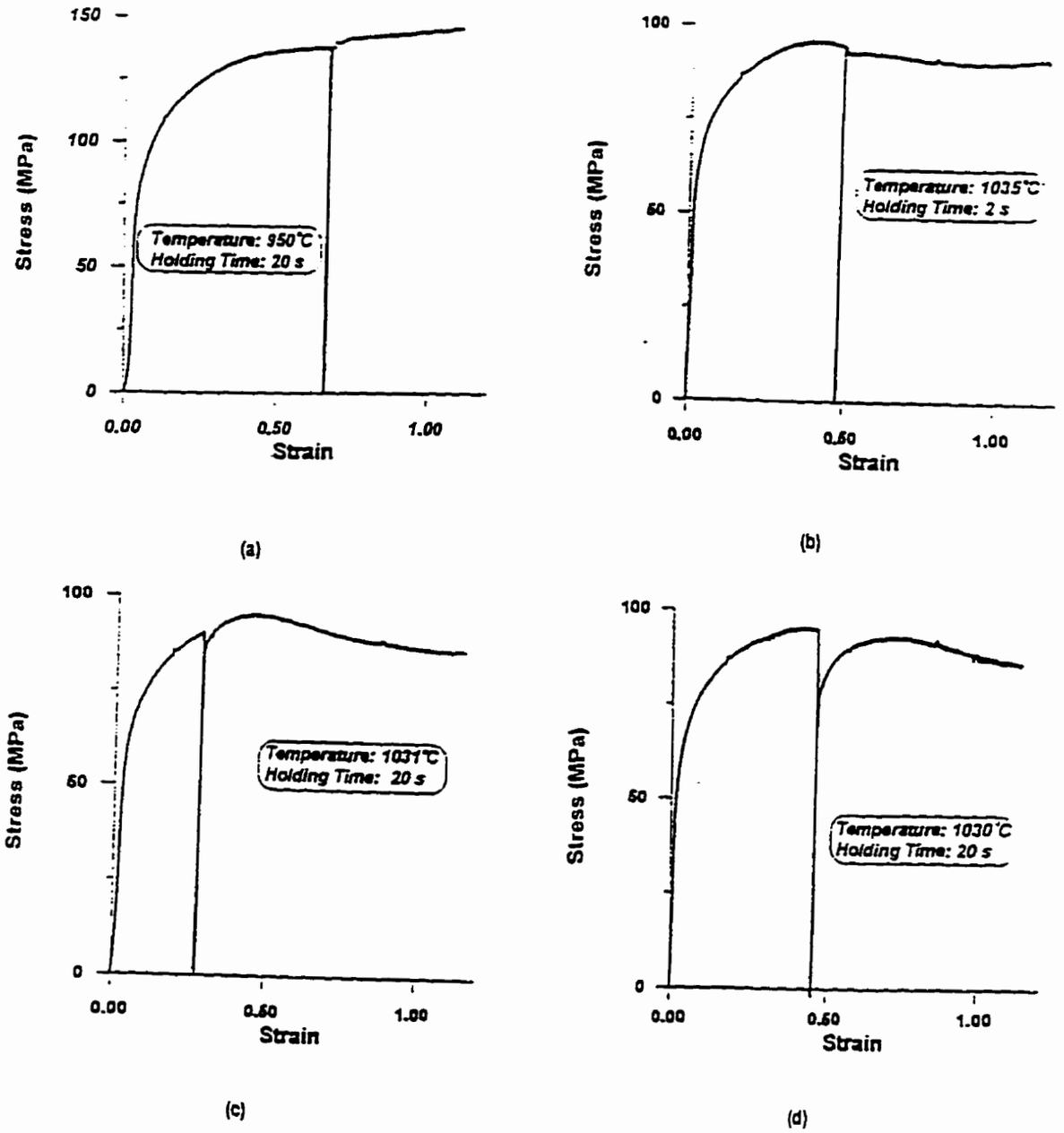

Figure 6-2 The effects of test temperature and holding time on the recrystallization of steel number 2: a, b) At the test temperature of 950 °C or holding time of two seconds there is no possibility of static recrystallization. c) At the test temperature of 1030 °C, if interruption is made before the peak strain, still there is hardly any recrystallization. d) An appreciable recrystallization is observed only at high test temperature and holding times of 20 seconds, provided interruption is made after the peak strain, i.e. onset of dynamic recrystallization.

Test Temperature (°C)	Strain of Interruption	Holding Time (s)	Fractional Softening	Grain Size (μm)
1026	—	—	—	5.31
1031	.25	20	0.13	3.85
1027	.45	20	0.54	3.78
10.29	.85	20	0.55	3.79
950	—	—	—	2.98
950	.25	20	0.06	2.68
953	.65	20	0.02	2.41
965	1.1	20	0.05	2.85
1035	—	—	—	4.6
1035	0.3	2	0.06	4.0
1035	0.5	2	0.18	4.5
1035	0.9	2	0.19	4.0
950	—	—	—	2.9
950	0.25	2	0.05	3.0
950	0.7	2	0.03	2.6
950	1.1	2	0.12	3.3

Table 6-1 Fractional softening and average prior-austenite grain sizes at different test conditions.

In the presence of solutes such as niobium, molybdenum and titanium, and in the absence of static recrystallization (relatively low temperatures and significant amount of strain-induced precipitation) the strain accumulation occurs, which may lead to dynamic recrystallization. This has been confirmed by Samuel and co-workers[131] who plotted the flow curves of a niobium steel at a constant 'standardised' temperature. This was done by taking into account the effect of temperature on the mean flow stress of steel during a multistage torsion test. The temperature corrected flow curves indicate that after a peak strain of about unity, dynamic recrystallization occurs and the stress level decreases.

The fact that there is no recrystallization at 950°C in niobium steels at small interpass times, versus considerable softening in C-Mn steels, indicates the role of niobium carbo-nitride precipitation in recrystallization retardation. Around 950°C, the kinetics of Nb(C,N) precipitation is the fastest, and once the process of precipitation starts prior to the onset of recrystallization, it retards or stops the recrystallization until all the niobium is precipitated[127]. The seven-minute temperature stabilisation time used during the tests provide enough time for the start of static precipitation at this temperature, since niobium starts to precipitate under no-load condition after 100 seconds of incubation time. During straining, the kinetics of precipitation increases by as much as two orders of magnitude, which results in higher volume fraction of precipitates.[127] Another interesting indication of precipitation during the straining and holding time is the change in the slope of flow curves, showing higher work-hardening and strength of the material following the interruption.

In comparing the flow curves at the two temperatures, it can be readily observed that lowering the temperature from 1030°C to 950°C results in a significant difference in the peak stress and strain levels. However, the most important observation is the behaviour after the peak strain has been reached. There was hardly any static or dynamic recrystallization at 950°C at holding times of 2 seconds. Only at higher temperatures and holding times of 20 seconds, some static and dynamic recrystallization was observed.

6.2 Torsion Tests

Torsion testing is the only possible physical simulation technique, if the whole rolling sequence of a hot strip mill is to be studied. The other techniques are limited to strains of near unity, due to excessive interfacial friction between the test specimen and compression platens. Torsion testing has been used extensively to characterise the flow of material and study the metallurgical events that take place during deformation. However, most of these tests have been done at strain rates of 2 s^{-1} and below, which is far from realistic when compared to the strain rates in rolling mills. Hence, the results and conclusion have been looked at with caution.

A set of torsion tests with more realistic strain rates (up to 20 s^{-1}) was designed to simulate finish rolling of a Nb-Ti steel to examine the type of recrystallization occurring. The experimental procedure, test parameters, and test setup have been described in Chapter 5. The torque-twist data was translated to true stress-strain curves by using Backofen formulations, explained in Chapter 5.

6.2.1 Average Schedule Experiments

After reheating the specimens to 1200 °C for 10 minutes, they were subjected to a three-stage roughing sequence followed by cooling to the test temperature, at a cooling rate similar to the mill cooling rates. Following this conditioning procedure, the specimens were subjected to a seven-stage twist according to the average schedule given in Table 5.2. To maximise the similarity and applicability of the test results to the real rolling practice, all tests were conducted under continuous cooling conditions, with a cooling rate similar to the cooling rate in the strip mill. This was done because isothermal tests produce quite different potentials for precipitation and recrystallization, which are not indicative of real mill conditions. Hence, the test temperature was different at each twist.

The true stress-strain curve of Test 1 is shown in **Figure 6-3**. The graph shows work-hardening throughout the test, however, part of the increase in flow stress is due to the decreasing temperature, not metallurgical changes in the material. In order to examine the

metallurgical hardening or softening in the material, the stress-strain curves should be corrected to a standardised temperature, by excluding the effect of temperature on the flow stress.

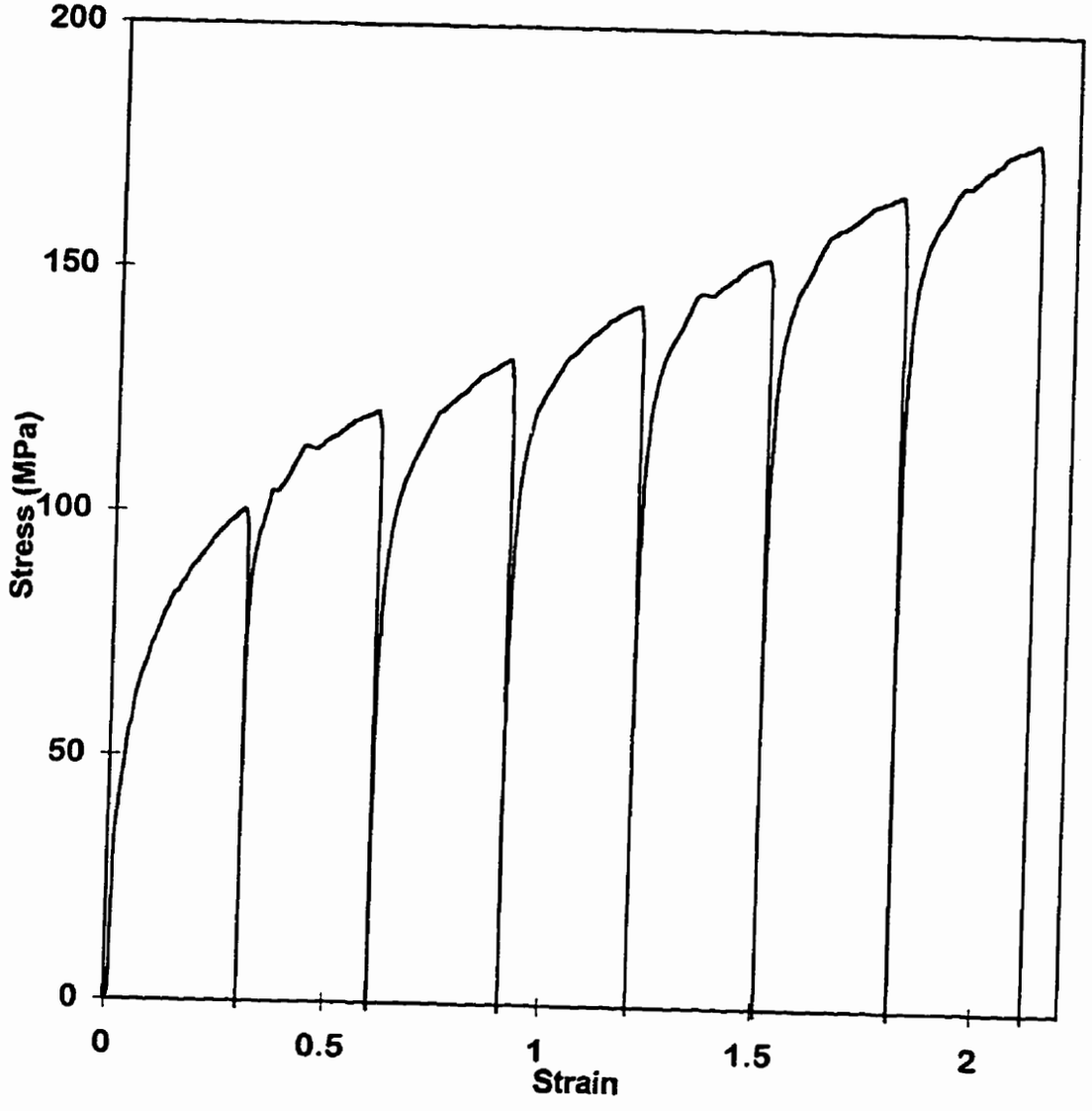

Figure 6-3 Uncorrected true stress-strain curve of Test 1.

Temperature Standardisation:

A series of isothermal continuous torsion tests was carried out between 900 and 1050°C at a strain rate of 10 s^{-1}. The mean flow stress (MFS) of the material was determined by integrating the stress-strain curves up to the strain of 0.8. A plot of the MFS versus the inverse of the temperature is given in **Figure 6-4**.

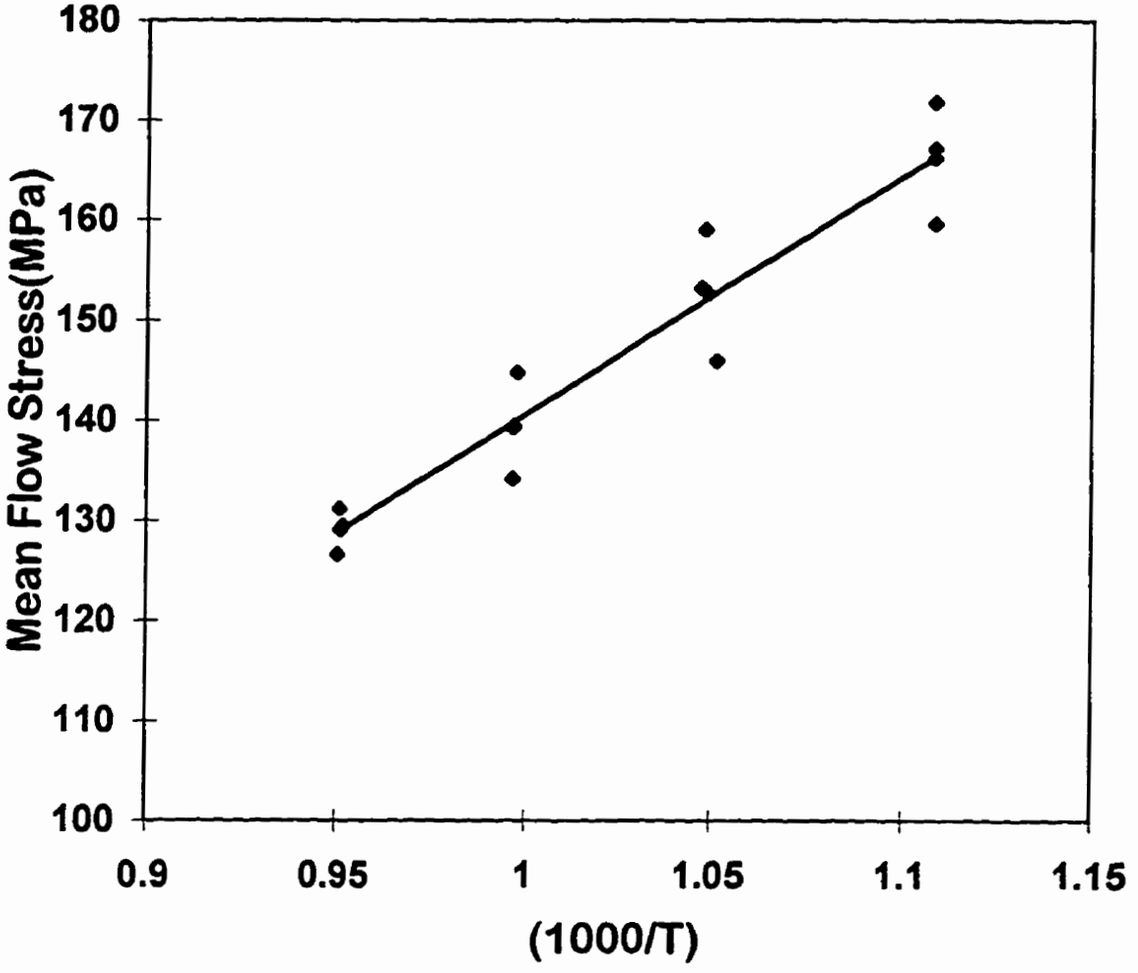

Figure 6-4 Mean flow stress variation with the inverse of the temperature.

Mean flow stress variation with the inverse of the temperature was modelled as:

$$MFS = -166.6 + 391.2\left(\frac{1000}{T}\right) \quad \textbf{(6-3)}$$

This is in good agreement with a previous report for a similar material[8]. Equation (6.3) was used to standardise the stress-strain curves of all tests to an arbitrary temperature of 950°C. A spreadsheet program was used to get the flow stress values of each test and calculate the expected flow stress if the test had been done at 950 °C.

The temperature-normalised stress-strain curve of Test 1 is plotted in **Figure 6-5**. Now, the flow stress variation from twist to twist can be attributed solely to the metallurgical softening or hardening. It can be clearly observed that there is a strain accumulation in the first two twists, then the flow stress drops in the third, fourth, and fifth twists, and rises again in the sixth and seventh twists. The maximum flow strength variation in the seven twists follows the same pattern as a flow curve which exhibits dynamic recrystallization. In each individual twist, the flow curve does not exhibit a peak or flattening of the curve. However, the level of strain reaches the critical magnitude for the creation of dynamically nucleated grains, i.e., the accumulated strain at the time of second unloading is somewhere between the critical strain, ε_c, and the strain at the peak stress, ε_p. This fact is supported by the results of double-hit compression testing, indicating that the strain to the peak stress at the temperature of 1030 °C is less than 0.5[132]. Moreover, static recrystallization kinetics is too slow to allow this level of softening to happen at these temperatures and interpass times. For example, fractional softening of a 0.027% Nb HSLA steel at 1000°C after three seconds of unloading is less than 50%[7]. At temperatures of 950°C and lower, static recrystallization is essentially arrested in Nb steels[7]. It could be argued that the nature of the softening in interstands 3 to 6 is not classic static recrystallization, but it is accelerated by nucleation during deformation.

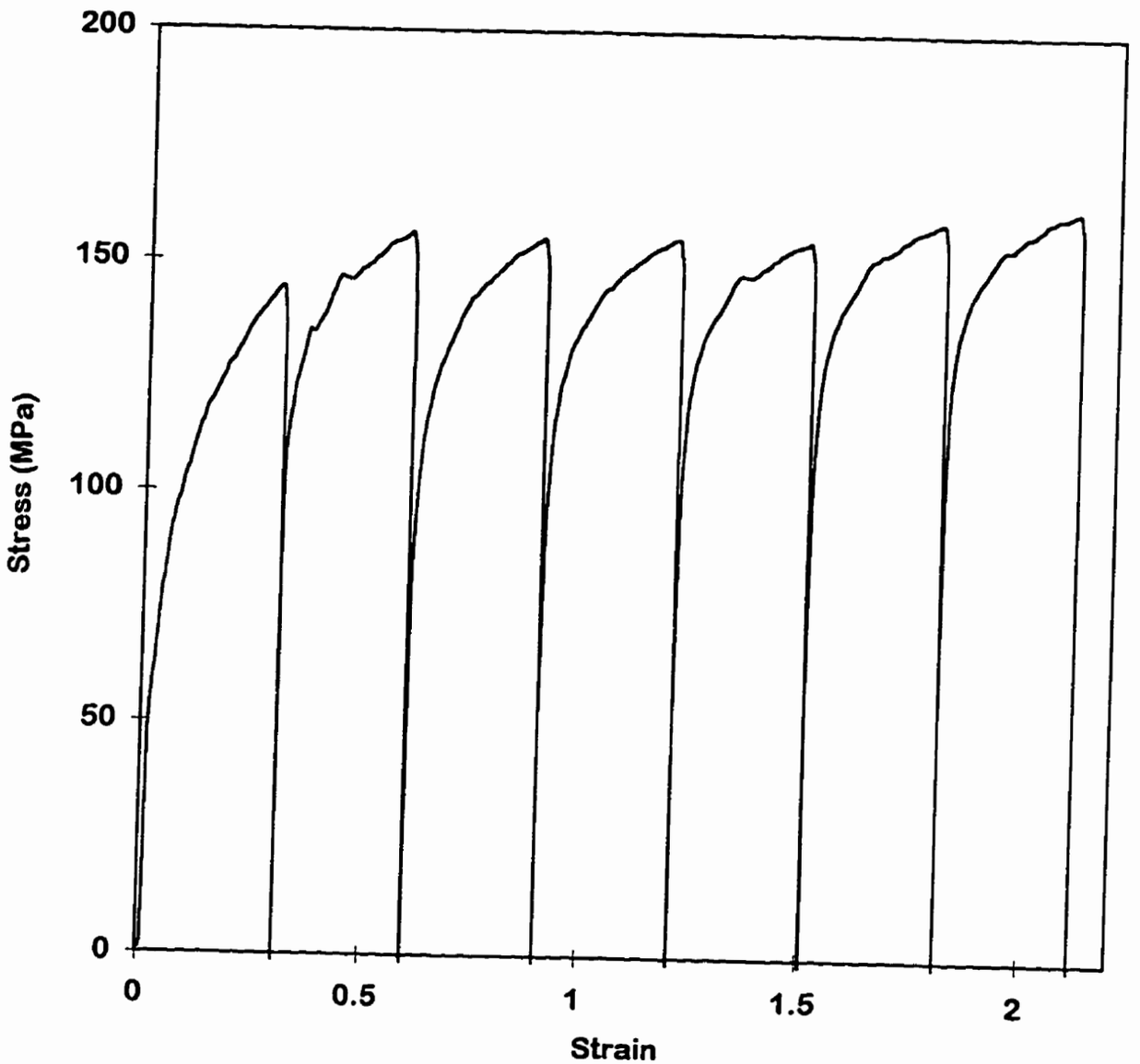

Figure 6-5. Temperature normalised true stress-strain curves of Test 1, standardised to the temperature of 950 °C.

Another fact to be noted is the over-estimation of ε_c and ε_p by the torsion testing method. This is due to the strain variation from the centre to the surface of the torsion specimens subjected to a twist. The level of strain in the outer layer reaches the strain to the peak stress while the inner layers still experience work hardening, due to lower strain levels. Hence, if the results of torsion simulations indicate the onset of dynamic recrystallization under a given strain, then the onset of dynamic recrystallization under real rolling conditions is very likely.

Comparing the results of Tests 1 and 2, it is also observed that the level of flow stress is higher at F7 in Test 2 (180 MPa) compared to test 1 (160 MPa), while the F7 temperature is the same for both tests. This difference in flow strength should be attributed to the effects of different softening and precipitation kinetics active in the previous stands. In Test 2, with F1 temperature of 950 °C (compared to F1 temperature of 1100°C in Test 1), there exist a higher retained strain and higher precipitation potential.

6.2.2 Average Schedule with Higher Strain Rates

The main concern about the results of other torsion simulation studies regarding dynamic recrystallization occurrence arises from the low strain rates employed (mainly up to 2 s^{-1})[133]. Tests 3 and 4 were designed identical to tests 1 and 2, only the strain rate was increased to 20 s^{-1}. The same temperature correction (Equation 6.3) was applied to the results of these tests which are plotted in **Figure 6-6**. As expected, the level of the flow stress is increased by about 10 to 20%. There is significant softening after the second twist. Softening after pass 2 increased with an increase in the strain rate, compared to the test 1. This observation further indicates that softening after pass 2 is not due to static recrystallization. Strain rate sensitivity of softening is characteristic of metadynamic recrystallization[1,3].

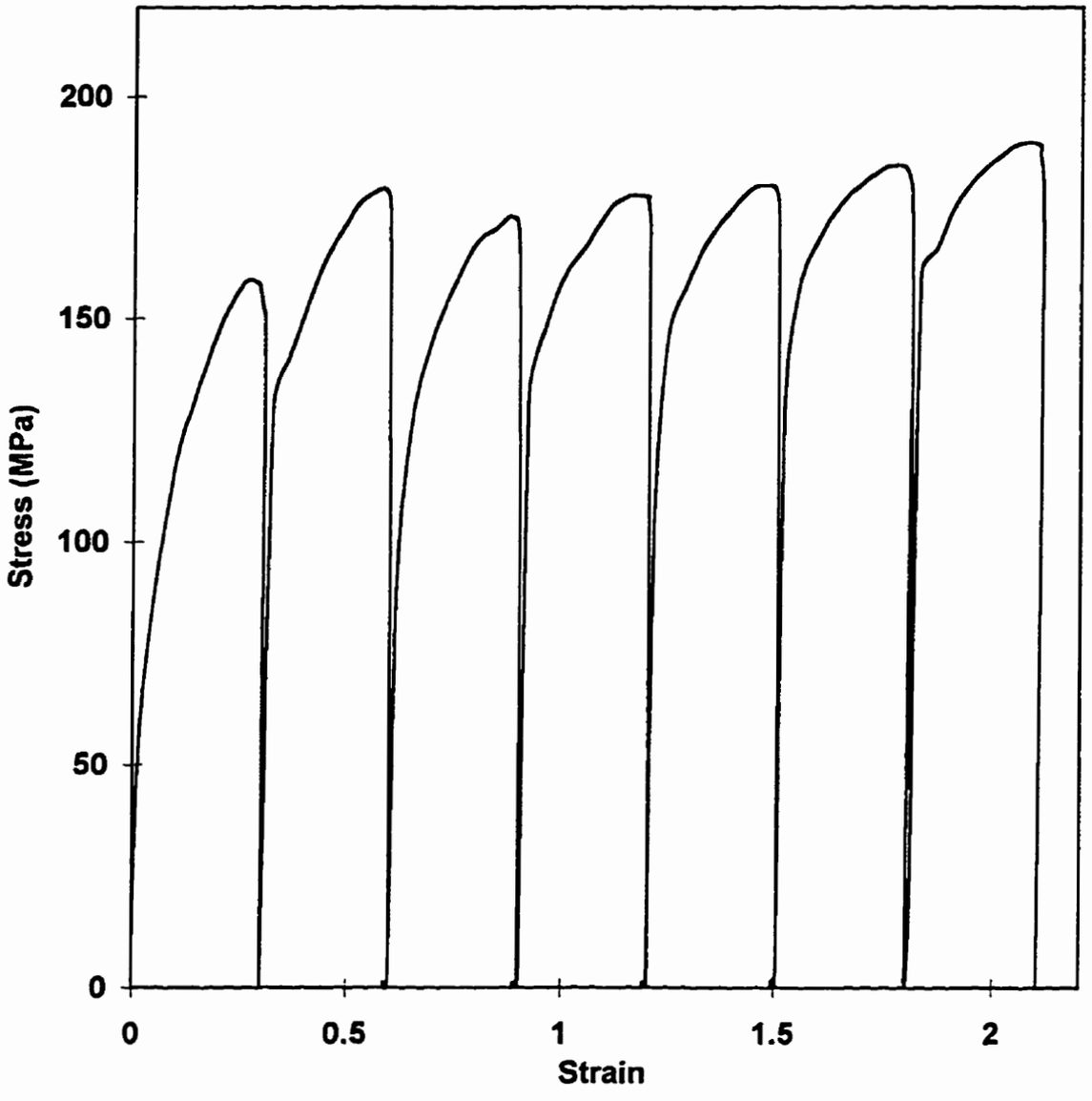

Figure 6-6 Stress-strain curves of Test 3, normalised to the test temperature of 950 °C.

Strain Rate Corrections:

A series of isothermal constant strain rate tests was conducted to evaluate the effects of strain rate on the flow stress level of austenite. Mean flow stress of each test was calculated by integrating the stress-strain curves over the strain path. Strain rate was varied between 0.1 to 20 s^{-1}. A log-log plot of the flow stress versus strain rate is given in **Figure 6-7**. The results indicate that flow stress varies exponentially with the strain rate. The slope of the best linear fit was calculated to be 0.08 and was used to correct the flow stresses to a standard strain rate of 50 s^{-1}. This value is less than Shida's model prediction for low carbon steels, which is 0.12[134].

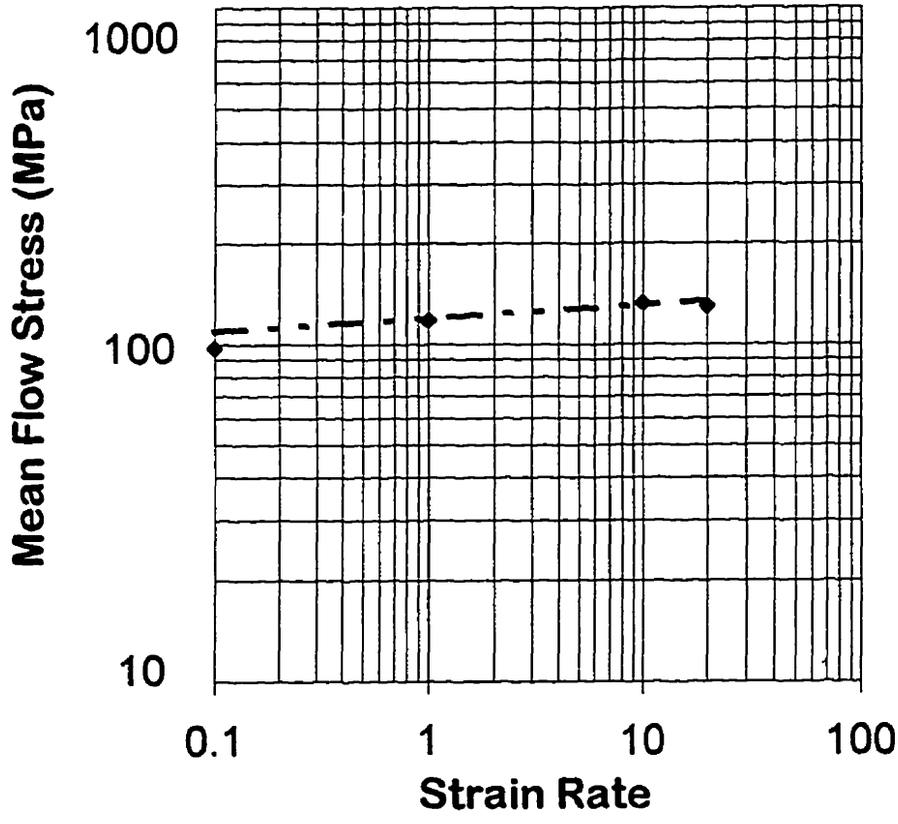

Figure 6-7 Mean flow stress variation of steel number 1 with the strain rate at the temperature of 1000 °C.

6.2.3 Finish Mill Schedule

In tests 5 and 6, strains and interpass times were matched to typical mill values (refer to Table 5.2). Strain rates up to the third twist were also typical of mill values. After the third twist, the strain rate was kept constant at 20 s^{-1}, the maximum possible strain rate available with Dofasco's torsion simulator. The temperature-normalised stress-strain curve of test 5 is given in **Figure 6-8**. Compared to the previous tests, the increased strain in the first three stands increased the potential for the initiation of dynamic recrystallization. Higher interpass times of 3.6 and 2.4 seconds, however, provided more time for interstand softening and reduce the strain accumulation required for dynamic recrystallization. Temperature-corrected stress-strain curves of test 5 show a peak at twist one, followed by a slight decrease in the strength in the following twists.

Test 6 was intended to simulate rolling conditions of strip tail-end which is usually rolled at lower temperature of 950 °C. The true stress-strain curves of this simulation is given in **Figure 6-9**. Due to lower temperatures (F1 temperature of 950 °C)compared to test 5, it appears that there is a strain accumulation in stand one and there is a marked increase in the flow stress in stand two compared to stand one. In this test, the peak stress is reached at stand two and remains constant for the remaining twists.

6.2.4 Dynamic Recrystallization:

Considering the simulation results of tests one to six, the flow stress curves do not round off or decrease in any single twist in any of these tests. This indicates that classical dynamic recrystallization, i.e. flow stress flattening during deformation is not taking place. This is contrary to the hot strip mill simulations performed by some authors, in which dynamic recrystallization has been shown to occur in the finishing stands[36,37]. These authors, however, used large strains (up to 0.6) in the first two stands, very low strain rates, and longer delay times between roughing and finishing stands. This kind of rolling schedule is applicable in less common rolling mills with only five finishing stands. Due to the lower number of stands, the strain per pass has to be higher, especially for thin gauges. Higher strains and lower strain rates promote the initiation of dynamic recrystallization.

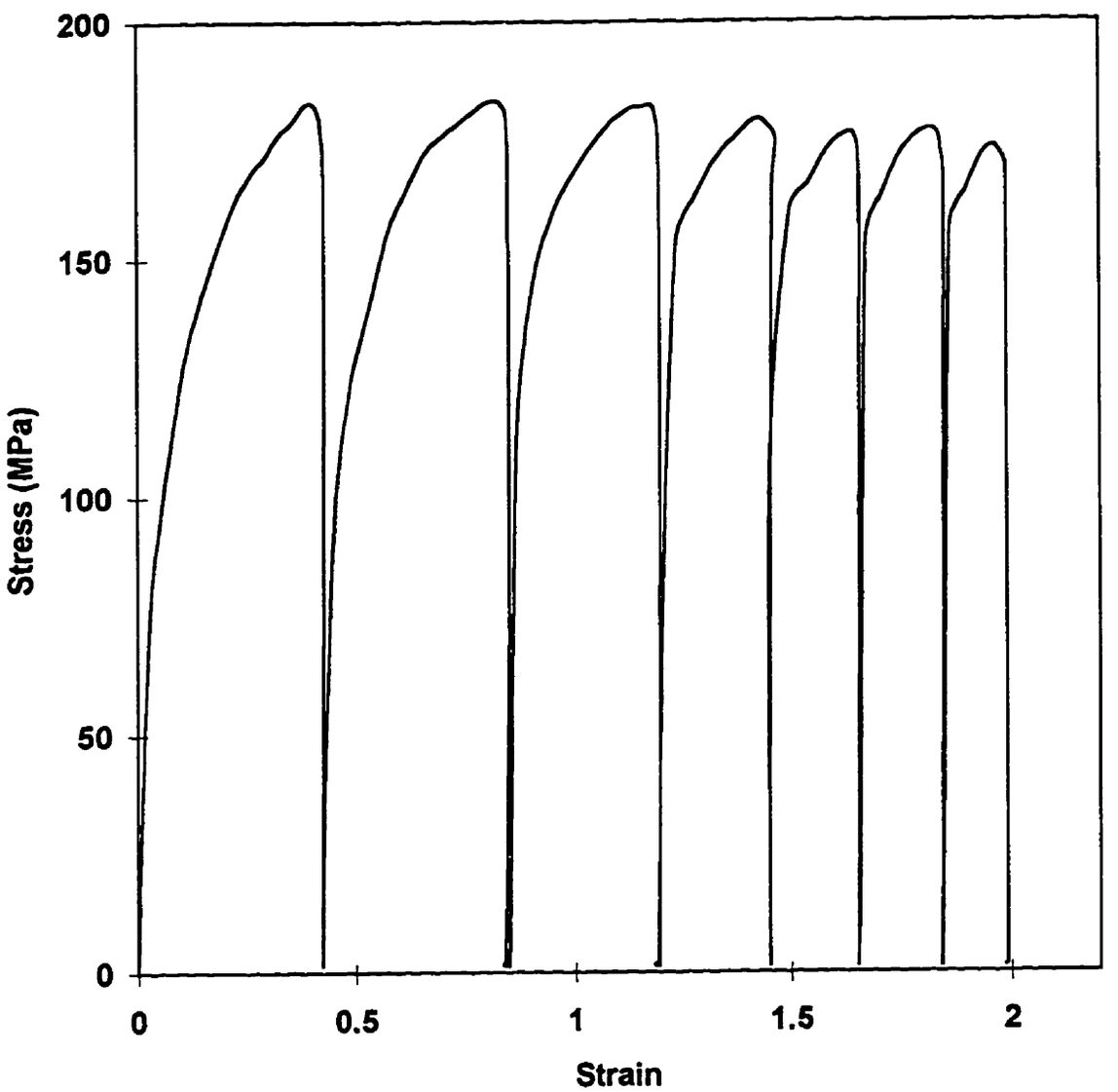

Figure 6-8 Temperature normalised true stress-strain curves of test 5 (typical of strip head-end condition).

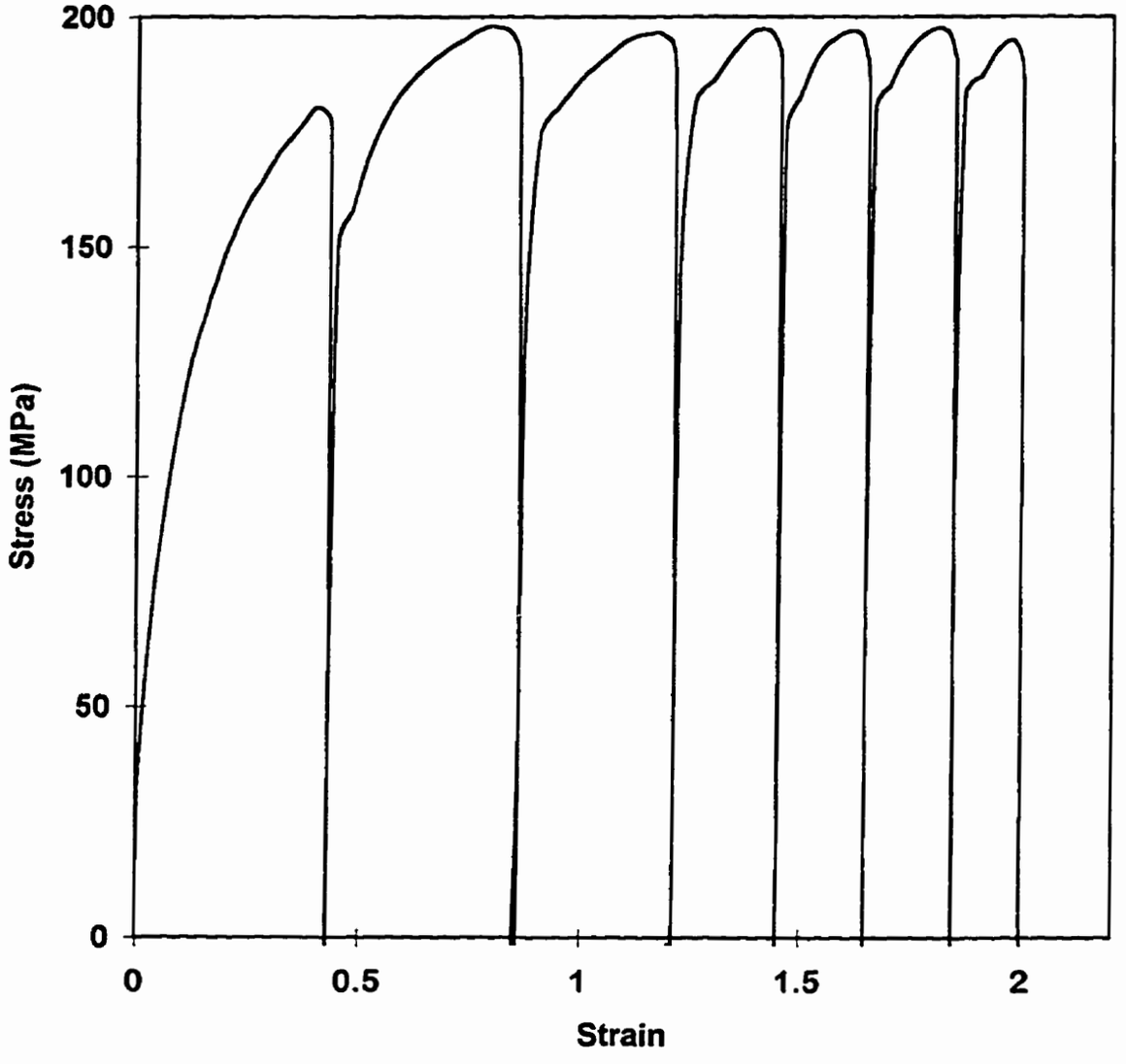

Figure 6-9 Temperature normalized true stress-strain curves of test 6 (simulating strip tail-end condition).

It has been shown that strain to the peak (commonly called peak strain), ε_p, increases systematically with the Zener-Hollomon parameter[135]:

$$\varepsilon_p = A d_o^{1/2} Z^n \tag{6-4}$$

where

$$Z = \dot{\varepsilon} \exp\left(\frac{Q_{def}}{RT}\right) \tag{6-5}$$

A is a material constant and n varies from 0.125 to 0.175 with no apparent systematic trend with the steel composition[15]. Since the peak in flow stress occurs after some low fraction of recrystallization, the peak strain is always greater than the critical strain for dynamic recrystallization (ε_c)[136]. High strain rates employed in the mill cause the critical strain for dynamic recrystallization to increase. The effect of strain rate on the peak strain and the relationship between peak strain and critical strain at higher strain rates is still unclear.

As was seen in tests 1 and 3, however, there is an unexpected softening after twist two. The time between passes is not large enough for the completion of static recrystallization. This is evident from the results of two-stage compression tests given in section 6.1.3. The softening after the second twist in tests one and three may therefore be attributed to metadynamic recrystallization. This means that softening after deformation has been accelerated by the existing grain nucleation sites formed during deformation.

6.3 Mill Data Analysis

Dofasco's Hot Strip Mill data was analysed in order to detect softening mechanisms and their effect on roll force predictions. It was of particular interest to investigate possible dynamic recrystallization during strip rolling of high niobium steels and its effect on roll force prediction. The on-line finishing mill set-up module uses a simplified Alexander and Ford[137] force model. Bar to bar feedback is used to tune-up the model. Both Alexander and Ford model (existing roll force model in Dofasco's hot strip mill), and Sims' roll force formulations are used to backcalculate the material mean flow stress variation from stand to stand using measured rolling forces at the mill. The calculated mean flow stress values were standardised to exclude the effects of varying temperature and strain rate from stand to stand.

Roll Force Model:

An accurate roll force prediction is required for the roll gap setup in order to get minimum variation in the head-end strip thickness. As discussed in Chapter 2, it is possible to simplify most rolling load models to:

$$P = K_m W L_c Q_p \tag{6-6}$$

where:

P = rolling load (N)

K_m = mean yield strength through the roll bite (N/m^2)

W = strip width (m)

L_c = arc of contact (m)

Q_p = geometric term.

The difference in the various force equations lies in the calculation of the geometric term, Q_p. Other factors that affect the rolling force, such as strip tension and friction in the roll bite, are compensated for by this term. An accurate roll force prediction is dependent on an accurate prediction of the mean yield strength.

The average yield stress in the roll bite is a function of the chemical composition of the material, strip temperature, strain rate, work hardening or softening during rolling, and interpass times. The on-line model calculates yield strength with an equation of the form:

$$K_m = f(K_n, \text{grade}, \text{temperature}, \text{strainrate}) \qquad (6\text{-}7)$$

where K_n is the average yield strength normalised for the base temperature of 950°C and a strain rate of 50 s^{-1}. This value, which is updated by a bar to bar feedback system, is stored by steel grade, finish mill gauge, stand and draught.

This yield strength model was originally developed for plain C-Mn steels. The model assumes that there is no strain accumulation between stands and full static recrystallization occurs in each interpass time. While this is the case for C-Mn steels at high temperatures and large interpass times, it may not be the case for microalloyed steels. The results of a recent study[36] indicate that fractional softening of a low niobium steel (0.027% Nb) is less than 40% during three seconds of interpass time at a temperature of 1000 °C. The results also indicate that the assumptions included in the existing yield strength prediction models disregard strain accumulation followed by possible dynamically nucleated recrystallization. The resulting error in yield strength prediction is absorbed by stand to stand feedback term, only if this softening is consistently happening for all the slabs.

In the first part of this analysis, mill data for 80 slabs of the test steel (steel number 1) was analysed. Sims' roll force model[138] was used to back-calculate the steel's flow stress at different stands. Equation (6.3) was applied to standardise the flow stresses in seven stands, which experience a continuous decrease in temperature, to an arbitrary standard temperature of 950°C. For most of the slabs (more than 80%), the average temperature-normalised flow stress of this grade of material shows some strain accumulation in the first two stands then softening in the second interpass time, followed by a decrease in flow stress in the third stand. The flow stress variation of six slabs rolled consecutively is plotted in **Figure 6-10**.

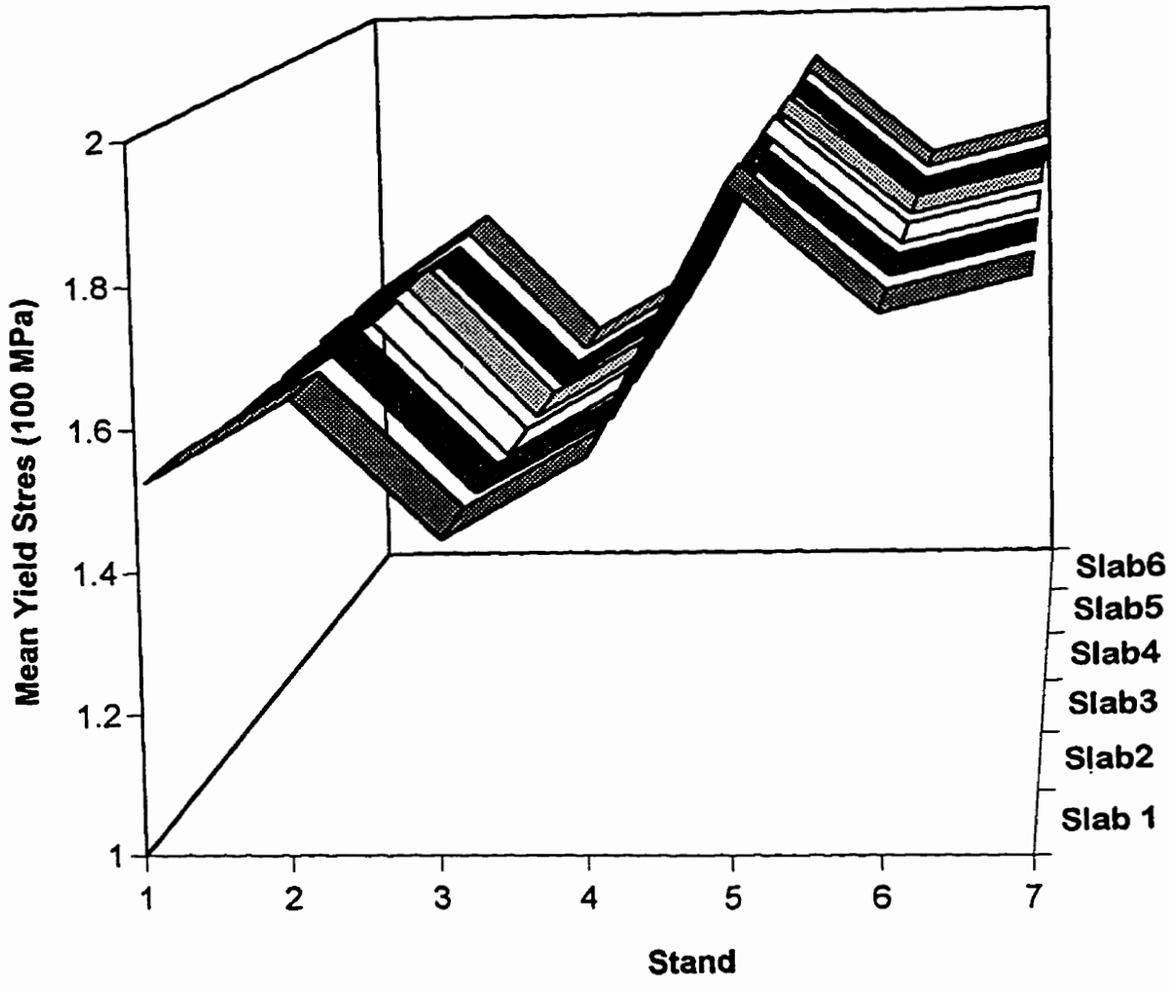

Figure 6-10 Mean flow stresses of six slabs rolled consecutively with an identical rolling schedule.

The amount of softening in the second interpass time cannot be solely attributed to static recrystallization, since the fractional softening for even lower Nb steels (<0.03) at this temperature (1000°C) and interpass time (2.5 seconds) is about 30%[36]. This drop in flow stress in the third stand implies that the softening rate has been accelerated through another mechanism, metadynamic recrystallization. After complete softening in the second interpass time, the material starts another cycle of strain accumulation until the critical strain for the onset of dynamic recrystallization is reached again. This kind of behaviour was not expected. However, this trend is not observed for all of the slabs. In some slabs, the mean flow stress increases monotonically from stand to stand. The effects of different process parameters on changing the softening regime are yet to be investigated.

To confirm the occurrence of accelerated interpass softening, the plot of the roll forces are superimposed on the plot of flow stresses in **Figure 6-11**. According to the Finishing Setup Module of Dofasco's hot strip mill control program, the force level should be decreasing monotonically from pass to pass, similar to the strain variation. However, a hump is observed in the roll force variation, that coincides precisely with the hump in the flow stress. This observation was unexpected and led to the conclusion that the flow stress in the third stand should have been affected by unpredicted metadynamic recrystallization during the second interpass time.

The knowledge that metadynamic softening is occurring in some occasions and does not occur in other occasions cannot benefit steel mills, unless this knowledge is modelled with some acceptable degree of accuracy. Linear and nonlinear regression models were used to classify the conditions under which the softening in the third stand is taking place. It is expected that for a given steel grade and roughing conditions, this softening is a function of strain, strain rate and temperature in the first three stands (interpass time for a fixed stand to stand distance is a function of strain and strain rate). All of these variables were used to develop a model to predict softening in the third stand.

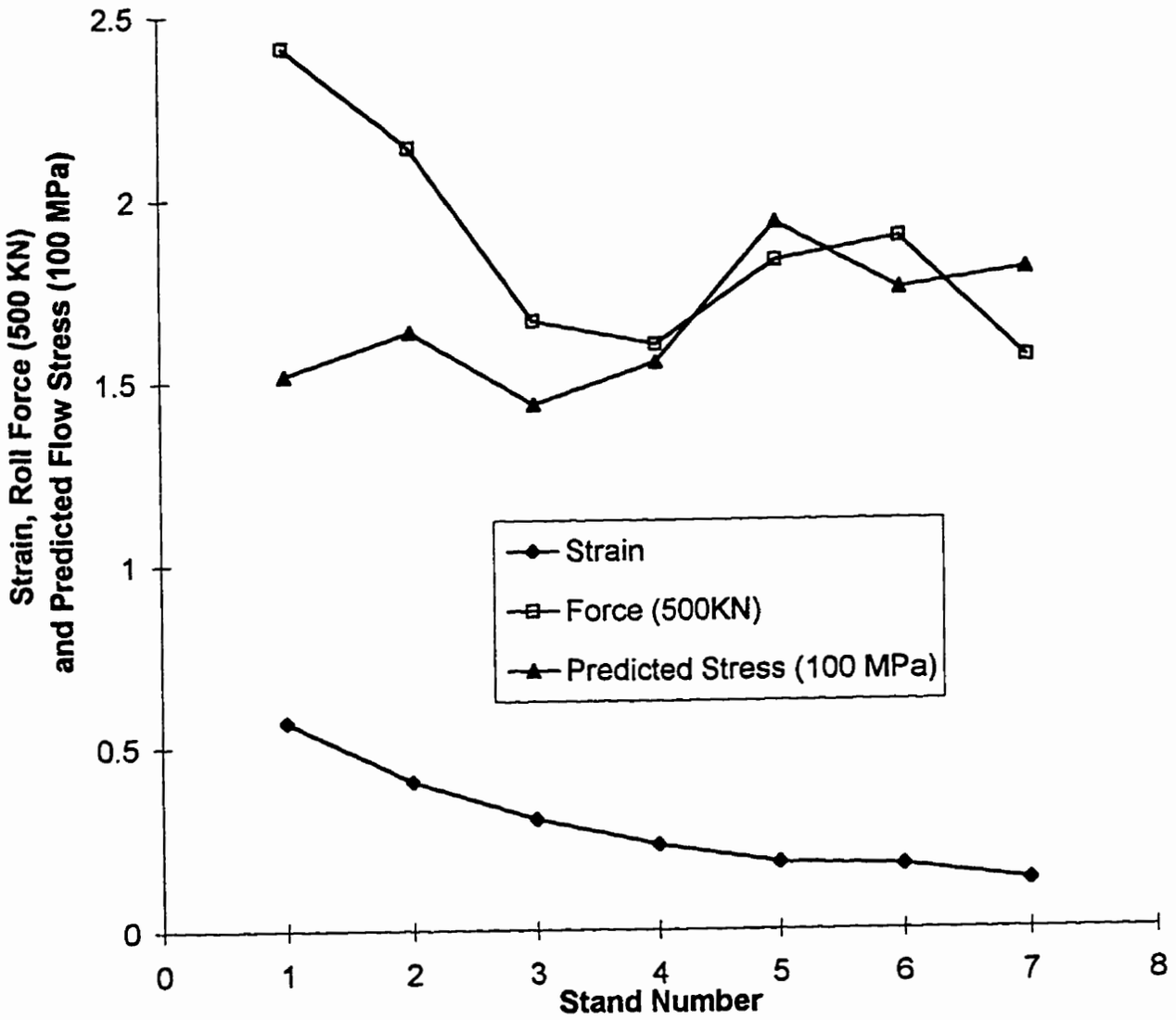

Figure 6-11 Coincidence of the hump in the rolling forces and drop in the predicted flow stress.

Also, different combinations of these variables along with Zener-Hollomon parameter were used as regressors. The author and Dofasco's engineers attempted to develop regression models to predict the occurrence and extent of this softening. However, none of the regression models provided any acceptable solution. Hence, it was decided that more advanced modelling techniques, i.e. Principal Component Analysis and Artificial Neural Networks are to be investigated for this purpose. The results of these analyses are reported in the Chapter 7. The results proved to be quite satisfactory and demonstrate the great potential of these techniques in the material science and processing applications, or generally in any data modelling and classification tasks.

7. Neural Network Modelling

A back-propagation Neural Network simulator program, based on the gradient descent learning algorithm was developed. This program was employed to develop Neural Network models for:

a) Prediction of hot flow stresses of an aluminium alloy and an HSLA steel
b) Prediction of rolling forces during aluminium cold and hot rolling
c) Prediction of rolling forces in industrial hot strip rolling of a high niobium HSLA steel
d) Prediction of the occurrence and extent of dynamic recrystallization in HSLA strip rolling during using a strip mill log.

Neural Networks, alike statistical methods and the human brain, suffer from colinearities in the training database. Principal Component Analysis was integrated into the Neural Network training procedure to decouple the linearly correlated data. This technique was used to transform linearly correlated data to non-correlated principal components. Then, Neural Networks were trained on principal components and the results were back-transformed into original variables.

This chapter is devoted to the development and results of Neural Network modelling, performance comparison of Neural Network based models with statistical models, and the results of integration of Principal Component Analysis to Neural Network learning algorithm.

7.1 Flow Stress Predictions

Knowledge of the material's constitutive behaviour is an essential requirement for the design and analysis of deformation processes. In spite of suffering some drawbacks in terms of ease of development, adaptability, accuracy, and speed, empirical stress-strain relationships and constitutive equations describing material behaviour during deformation are being widely used. In the present study, back-propagation neural networks are used to model and predict the flow stress of an aluminium alloy and a microalloyed steel in a range of temperatures, strains, and strain rates. A neural network training code, based on the gradient descent algorithm, is developed and used to train the flow stress prediction networks. The performance of this technique is compared to those of the statistical models. The approach based on neural networks is shown to be superior in its predictive capability, adaptability and speed to that using non-linear regression models.

7.1.1 Flow Stress Modelling of Hot Steel

The results of this investigation are presented in two parts: Statistical Method and Neural Network Method. The purpose of both is to develop models to predict the flow stress as a function of strain, strain rate and temperature. The results obtained are compared to the experimental values.

7.1.1.1 EXPERIMENTAL DATA

The data used in this investigation were produced by Tajima and Lenard[139]. The experiments involved the isothermal, constant true strain rate compression of cylindrical samples, made of a Nb-V steel, the chemical composition of which is given in **Table 7-1.** The test equipment consisted of a closed-loop controlled servohydraulic testing system, an attached resistance furnace and a computer-based data acquisition and control system.

Table 7-1 The chemical composition of the material (wt%)

C	Mn	Nb	V	Si	Cu	Ni	P	Ti	Ca	S	N
0.1	1.093	0.0877	0.0795	0.3645	0.0152	0.0138	0.0076	0.0042	0.0032	0.0023	0.0092

Samples, 10 mm in diameter and 15 mm in length, were machined with recesses of 0.1 mm depth at each end to retain the glass lubricant. Type-K thermocouples, embedded in each sample at midheight, were used for temperature measurements. All samples were annealed and solution treated prior to reheating to the test temperature. The test results used in this work are those of single-stage compression at temperatures of 875, 900, and 950 °C. The experimental matrix, giving the details of the process parameters, is presented in **Table 7-2.**

Table 7-2 Experimental matrix used in steel hot flow strength tests.

Strain Rate (s^{-1}) ⇨ Temperature⇩	0.01	0.1	1	2
875 °C	A1	A2	A3	A4
900 °C	B1	B2	B3	B4
950 °C	C1	C2	C3	C4

7.1.1.2 Statistical Method

As discussed in Chapter 2, the type of the constitutive relation that can adequately describe hot flow stresses of steels, depends on the stress level. The power relation is applicable below 110 MPa, while the exponential relation is valid for higher magnitudes. The stresses in hot deformations vary over a wide range, hence, it is expected that the hyperbolic sine rate equation will lead to the best results[140]. Rao et al[141] verified the better fit of the hyperbolic sine relationship when compared to the power and exponential relationships for a 0.34 % carbon steel. The procedure for determination of the values of A, n and Q in Eq. (2.7) as suggested by Rao leads to

$$\ln(\dot{\varepsilon}) = \ln(A) + n \ln\left[\sinh(\alpha\sigma)\right] - \frac{Q}{RT} \qquad \textbf{(7-1)}$$

For a given temperature and strain, the plot of $\ln(\dot{\varepsilon})$ versus $\ln\left[\sinh(\alpha\sigma)\right]$ gives the strain rate sensitivity parameter, n. Plots for the present steel at a strain level of 0.3 and three different

temperatures are given in **Figure 7-1**. The lines are nearly parallel, indicating that the strain rate sensitivity is not strongly influenced by the temperature. The activation energy at a given strain level is obtained by rearranging Eq. (7.1) and differentiating with respect to (1/T), giving

$$Q = R\,n\,\frac{d\left\{\ln\left[\sinh(\alpha\sigma)\right]\right\}}{d(1/T)} \qquad \textbf{(7-2)}$$

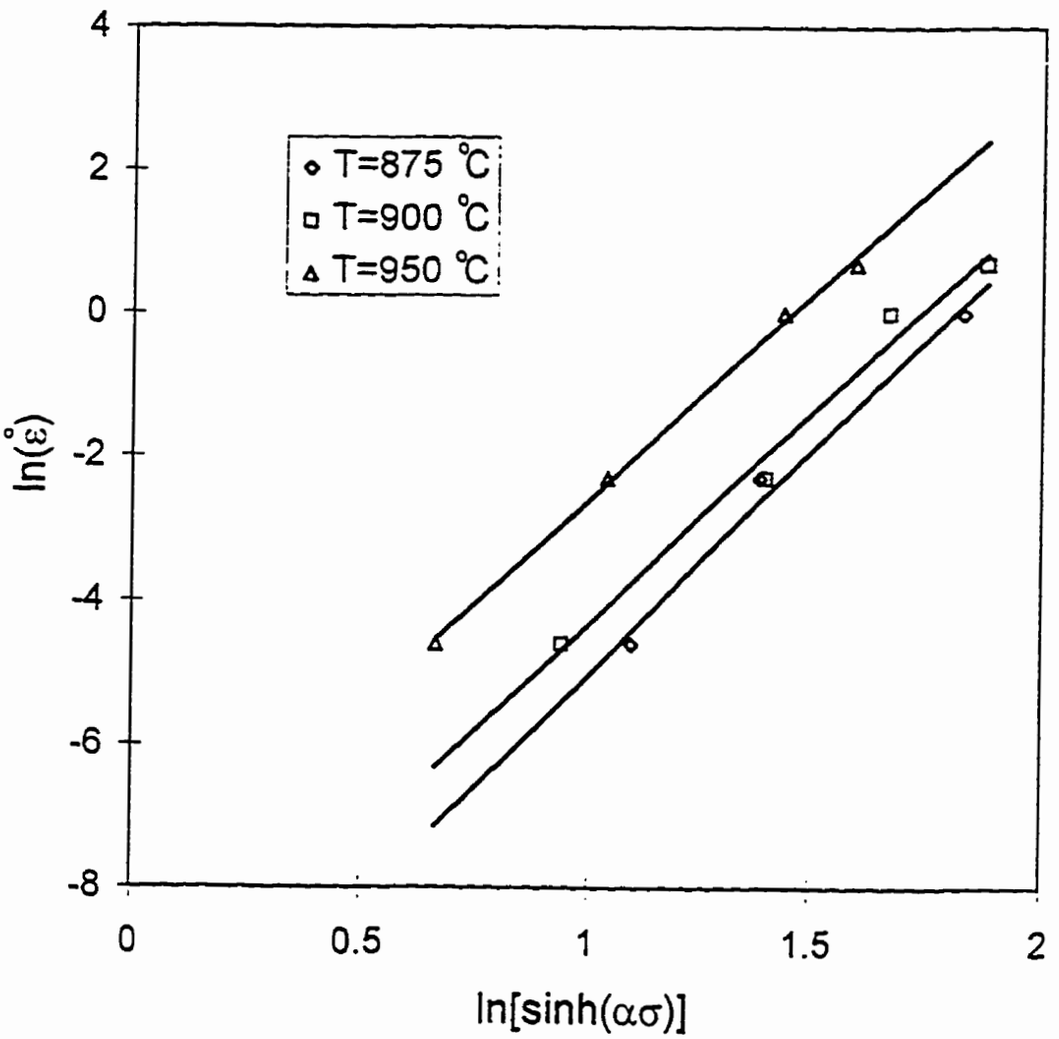

Figure 7-1 Plot for determination of strain rate sensitivity parameter, n.

The slope of a plot of $\ln[\sinh(\alpha\sigma)]$ versus (1/T) for a given strain and different strain rates should be equal, as another condition for applicability of the hyperbolic sine rate equation

and indicating that Q is also not a strong function of the temperature. A plot at the strain level of 0.3 is given in **Figure 7-2**, leading to activation energy values in the range of 330 kJ/mole.

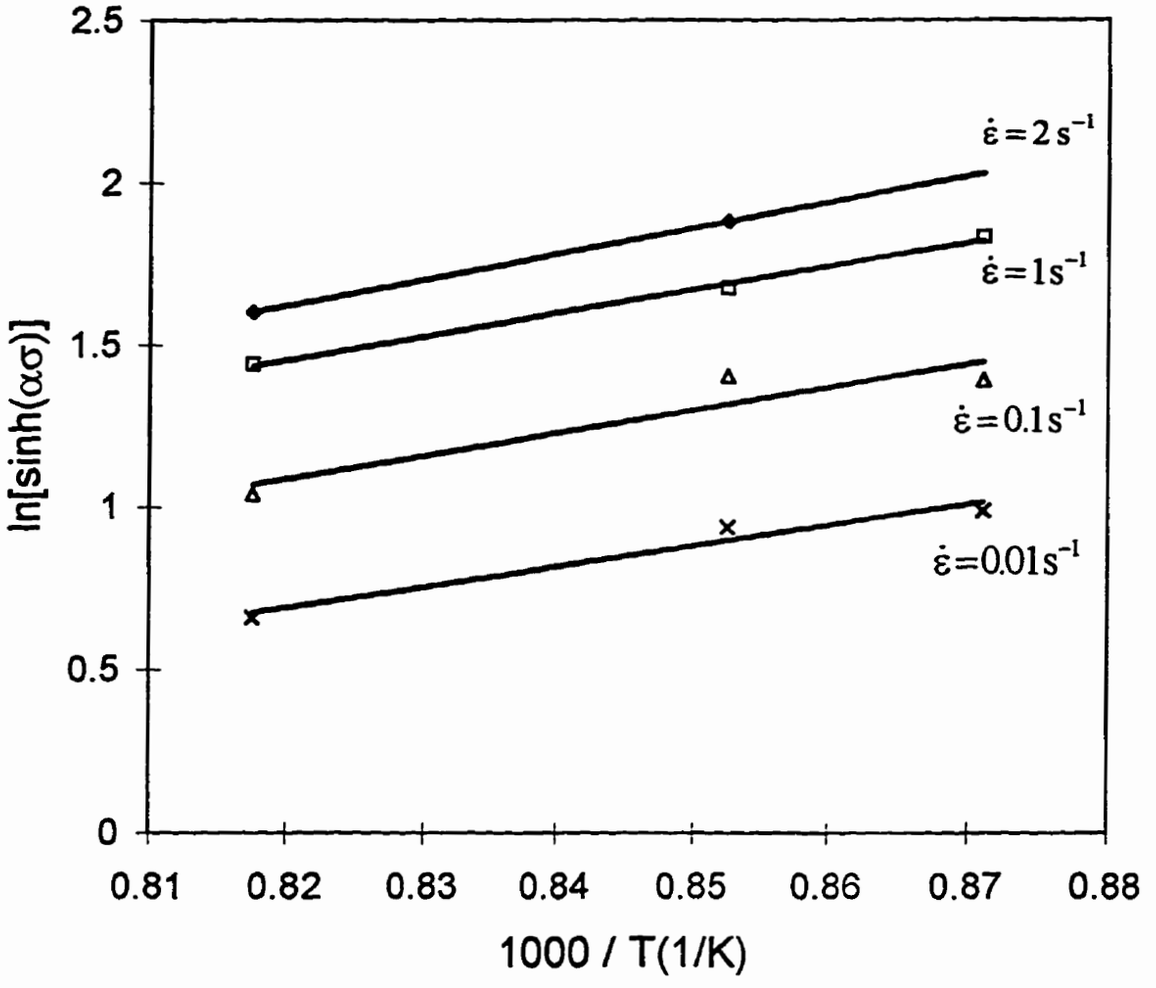

Figure 7-2 Plot for calculation of activation energy, Q.

The natural logarithm of Eq. (2.8) leads to

$$\ln(Z) = \ln(A) + n \ln[\sinh(\alpha\sigma)] \qquad \textbf{(7-3)}$$

A linear plot of ln(Z) versus $\ln[\sinh(\alpha\sigma)]$ at a given strain level for the whole range of strain rates and temperatures indicates the general applicability of the rate equation, as shown in **Figure 7-3**. The value of ln(A) and the strain rate sensitivity parameter can be obtained from these plots. This *n* value differs slightly from one obtained through Eq. (7.1), as shown in **Table 7-3**. The value of *n* appears to follow an inverse power relationship with strain, but the variations of ln(A) and the activation energy with strain are nearly linear for the present metal. The values of n, Q, and ln(A) for selected strain levels (0.1, 0.2, 0.3,...,0.9) obtained from **Figure 7-1** to **Figure 7-3** are given in **Table 7-3.** Rao et al.[141] reported an inverse power relationship with strain for all three parameters.

The values listed in **Table 7-3** were used to predict the flow stresses at given temperatures and strain rates. The predictions were compared with experimental values at 875 °C in **Figure 7-4**, also showing the results obtained using the neural networks, to be discussed in detail in the next section. A mean difference of 2.99%, defined as *100 x (experimental value-predicted value)/(experimental value)*, with a standard deviation of 0.68 was obtained. This level of accuracy is higher than the 3.35%, obtained on a low carbon steel by Rao et al[141]. This result was achieved by using an average *n* value. The *n* values found directly from Eqs. (7.1) and (7.3) produced mean differences of above 3.5%. Rao et al[141] used *n* values obtained from Equation (7.3) as it provided the minimum deviation for their data.

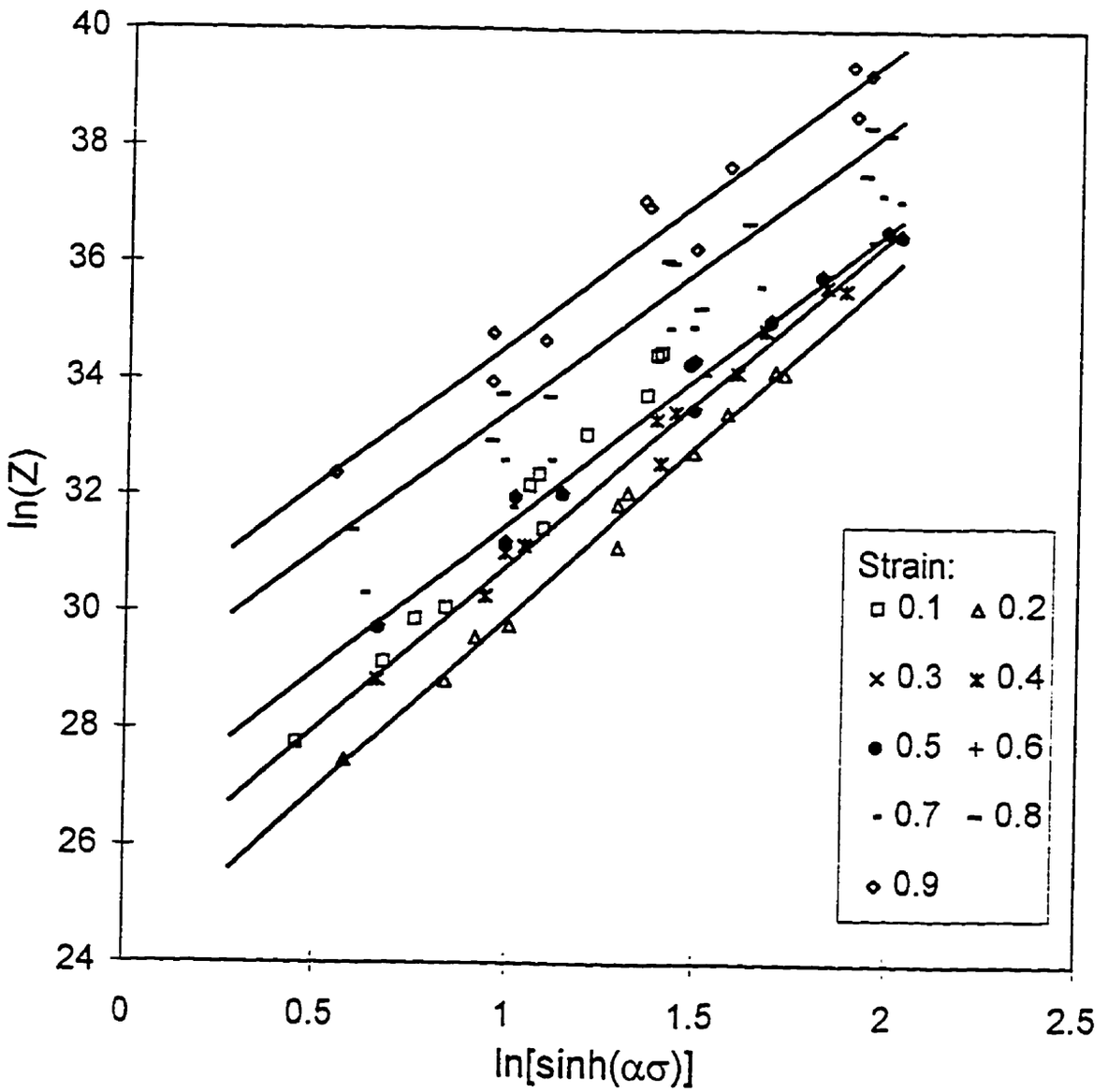

Figure 7-3 Zener-Hollomon plot for general validity test and calculation of the strength parameter, ln(A).

Table 7-3 The experimentally evaluated values for n, ln(A) and activation energy at different strain levels.

Strain	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
n (Eq. 7.1)	9.01	7.23	6.03	5.73	5.36	5.12	5.04	5.03	5.01	4.98
n (Eq. 7.3)	8.8	7.03	5.96	5.64	5.64	5.08	5.08	4.85	4.87	4.88
n (average)	8.91	7.13	5.99	5.68	5.5	5.1	5.06	4.94	4.94	4.93
ln(A)	28.05	24.43	23.93	25.17	25.17	26.41	26.44	27.39	28.58	29.7
Q (kJ/m)	330.7	320.2	322.0	336.2	357.3	346.7	351.8	341.9	351.8	352.5

7.1.1.3 Neural Network Modelling of Steel Flow Stress

A fully connected feedforward backpropagation network was designed to model the flow stress variation with strain, strain rate and temperature. From each compression test data, 9 data points were picked at equal strain intervals of 0.1, beginning at a strain of 0.1. Condition B2 of **Table 7-2** was set aside for network generalisation test and the rest of the data were used in training of the network. All input and output values were normalised into the range [0.1 to 0.9] to avoid premature saturation of the sigmoid function. The normalising scheme was described in Chapter 4.

Learning rates between 0.5 to 0.9 were attempted. The higher the learning rate, the faster the learning, however, too high learning rates may lead to oscillations in the network output error. One way to increase the learning rate without leading to oscillations is to modify the gradient descent algorithm to include a *momentum* term which takes into account the effects of past weight changes on the current direction of weight modification. Equation. (4.17) then changes to

$$\Delta W_{ij}(t+1) = \eta \delta_i^{\mu} V_j^{\mu} + \alpha \Delta W_{ij}(t) \qquad \textbf{(7-4)}$$

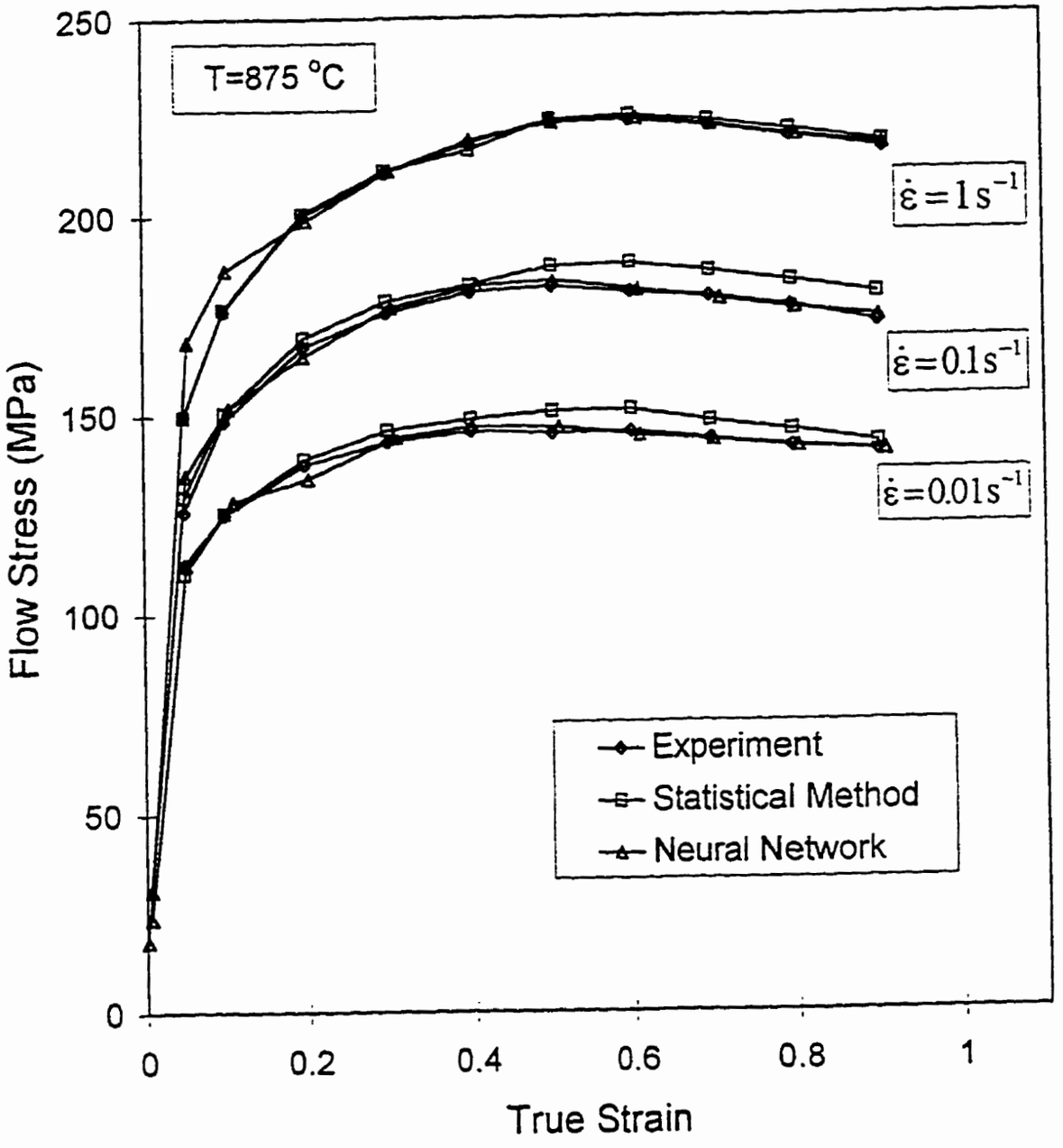

Figure 7-4 Comparison of the neural network predictions and statistical model predictions with the experimental flow stresses at the temperature of 875 °C.

where t indexes the iteration number and α, the momentum parameter, is a constant usually between 0.5 to 0.9. The learning rate of 0.7, with a momentum parameter of 0.9, provided the fastest convergence. The logistic sigmoid function with a constant steepness factor of 0.5 was chosen as the activation function. Both one-hidden layer and two-hidden layer networks were examined to investigate the effects of the extra hidden layers. The error measure for network performance evaluation was considered to be the average percent difference of network predictions and experimental values, as it was with the statistical model. It was found that the two-hidden layer topology had no advantage over the one-hidden layer for the equal number of total processing nodes in this particular application. Dynamic node generation scheme was used in order to find the optimum number of nodes. An increase in the number of hidden nodes up to 16 increased the accuracy. However, further increase in the number of hidden nodes had no considerable benefit.

The plot in **Figure 7-4** shows a comparison of the predicted and experimental flow stresses at the temperature of 875 °C. It is found that the predicted flow curves follow the experimental flow curves very closely. The average percent difference between the predictions and experimental points for all temperatures and strain rates, was calculated to be 2.17% with a standard deviation of 0.48, compared to the average difference of 2.99% and standard deviation of 0.68 obtained through statistical analysis. The accuracy of Neural Network predictions is about 27% higher than the Statistical method.

Generalisation Test

During learning, the outputs of a supervised neural network attempt to approximate the target values, given the inputs in the training set. This ability is useful in itself, but the purpose of using a neural net or any other model is to approximate target values given inputs that are not in the training set. This feature is called generalization, which has been discussed in detail in Section 4.2.10.

In the present work condition B2 was set aside to be used for testing the generalization ability of the network. The network predictions for this condition (900 °C

and 0.1 s^{-1}) are compared to the experimental values in **Figure 7-5**. The average difference between network predictions and experimental values for this unseen data was found to be 3.3%.

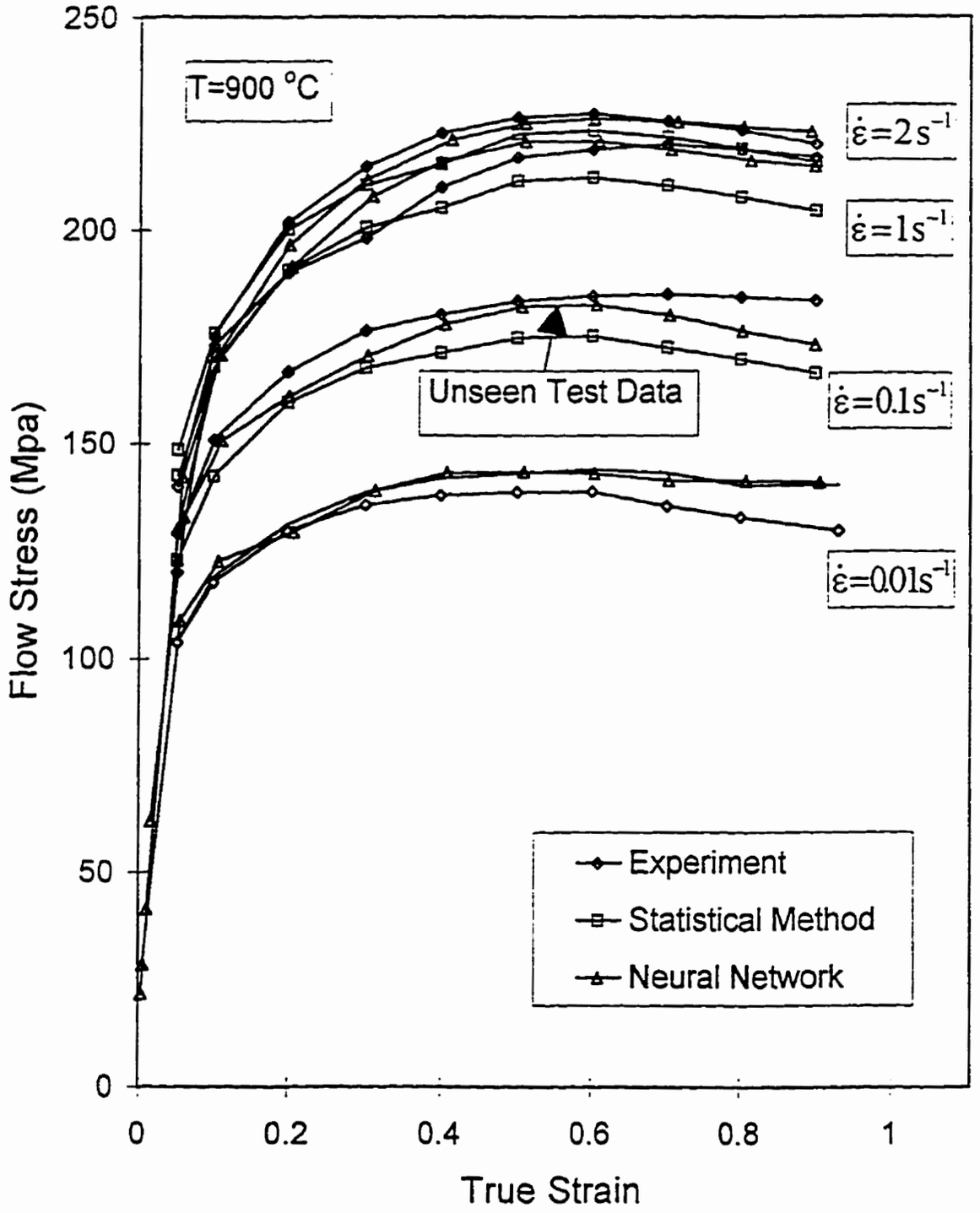

Figure 7-5 Comparison of the neural network predictions, statistical model predictions, and the experimental flow stresses of the B2 test condition.

7.1.2 Flow Stress Modelling of Hot Aluminium

Artificial Neural Network based models were developed to predict the hot flow stresses of an aluminium alloy.

7.1.2.1 Experimental Procedure and Data

Compression test samples of the Al 1100-H14 alloy (Si=1, Zn=0.1, Cu=0.05, Mn=0.05 and remainder aluminium) were machined from aluminium plates with the longitudinal direction parallel to the rolling direction. The samples were sized to 20 mm in diameter and 30 mm in height. Boron nitride was used as a lubricant to minimise interfacial friction. Recesses of 0.1~0.2 mm depth were machined at each end to retain the lubricant. A type-K thermocouple was embedded centrally in each specimen. The compression tests were carried out on a servohydraulic computer controlled constant strain rate Instron machine, as described in section 5.3. The test conditions and their designations are presented in **Table 7-4**. The temperature was set to 400, 450, and 500 °C and the strain rate ranged from 0.97 to 11.53 s^{-1}.

Table 7-4 Experimental matrix used to develop a database of aluminum hot flow stresses at different temperature and strain rates.

Temperature⇩ - Strain Rate ⇨	0.97	2.97	5.04	7.58	11.53
400 °C	A1	A2	A3	A4	A5
450 °C	B1	B2	B3	B4	B5
500 °C	C1	C2	C3	C4	C5

7.1.2.2 Flow Stress Prediction Results

A fully connected feedforward backpropagation network was designed to predict aluminium flow stresses at different temperatures and strain rates. The purpose of modelling was to develop an effective representation of material behaviour at high temperatures.

From each set of compression test data, 15 data points were picked in equal strain intervals of 0.05. Conditions B2 and B4 were set aside as network generalisation test data and the rest of the data was used in training of the network. All input and output values were normalised into the range [0.1 to 0.9], to avoid premature saturation of the sigmoid function. The learning rates and momentum rates were varied between 0.5 to 0.9. The learning rate of 0.9, with a momentum rate of 0.7, resulted in the fastest convergence. The logistic sigmoid function with constant steepness factor of 0.5 was chosen as the activation function. Both one hidden layer and two hidden layer networks were examined to investigate the effects of extra hidden layers. It was found that the two hidden layer topology had no advantage over one hidden layer for equal numbers of total processing nodes. Increasing the number of hidden nodes up to eight nodes increased the accuracy. However, a further increase in the number of hidden nodes had no considerable benefit. To avoid overfitting and enhance model generalisation ability, it is desired to use the minimum possible number of processing nodes. Therefore, an eight-node network was concluded to be the optimum network topology.

The plots in **Figure 7-6** show the comparison of the experimental values of flow stress with those predicted by the Neural Network at the strain rate of 5.04 s^{-1}. It is clearly observed that the predicted flow curves follow experimental flow curves very closely. The average percent difference was calculated to be less than 1.9%. A similar accuracy was found with other strain rates as well. This clearly indicates that the network was able to accurately learn the training data set.

The main quality indicator of a neural network, however, is its generalisation ability, i.e., its ability to accurately predict the output for a given unseen test data. The network predictions for unseen B4 condition is plotted along with B4 experimental values in **Figure 7-7**. The average percent difference was found to be 2.83%.

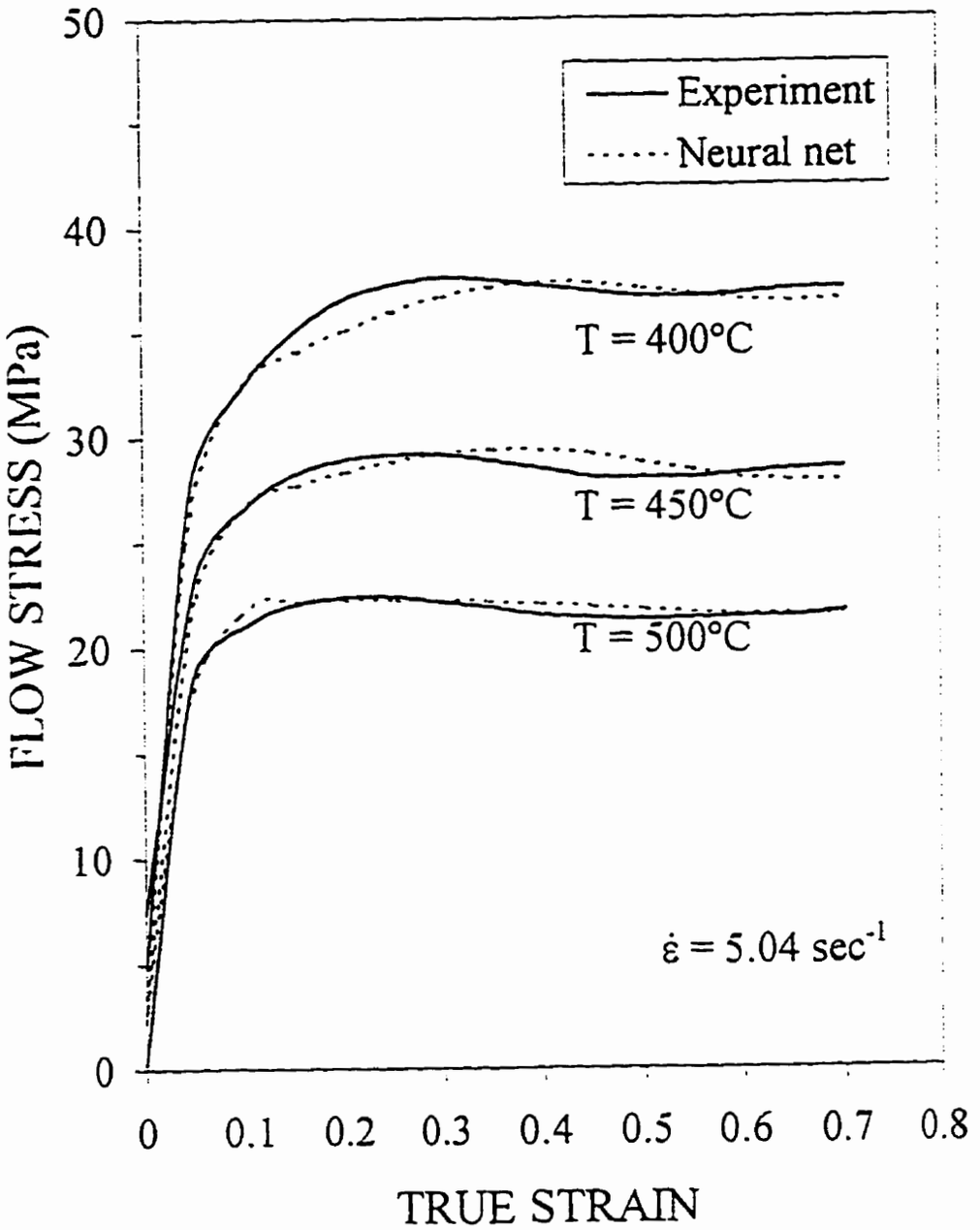

Figure 7-6 Experimental stress-strain curves of an aluminum alloy at high temperatures and performance of a Neural Network model.

This error is still less than the errors that usually arise in flow stress measurements due to the errors in the temperature, strain rate, and stress measurement.

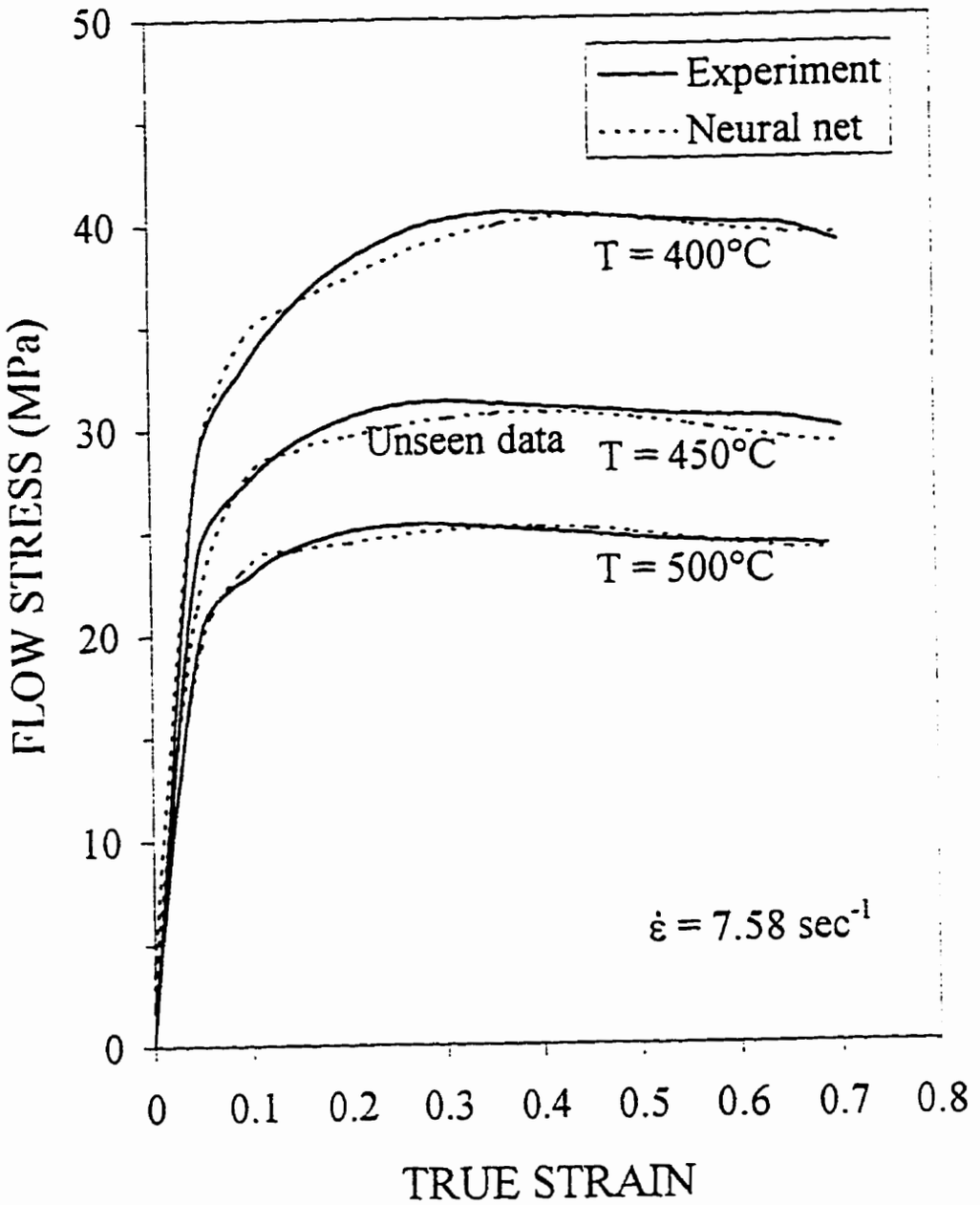

Figure 7-7 Performance of a Neural Network model on prediction of stress-strain curve of an unseen test data.

7.2 Experimental Rolling Force Predictions

Neural Network models were developed and tested to predict the rolling forces during hot rolling and cold rolling of an aluminium alloy.

7.2.1 Aluminium Hot Rolling Force Prediction

7.2.1.1 Experimental procedure

Strips of a 3000 aluminium alloy were cut along the rolling direction and sized to 6.12~6.16mm thickness, 50~52 mm width and 310 mm length. A type-K (chromel-alumel) thermocouple was embedded to a depth of 15 mm in each strip's tail end. Rolling experiments were carried out on a two-high experimental mill with the roll diameters of 250 mm. Two industrial hot-air guns were used to heat the rolls to approximately 90 °C. The roll forces were measured by two load cells located under the bearing blocks of the lower roll. Two torque transducers were mounted on the roll spindles to measure the roll torque applied to both upper and lower rolls. The rolling speed was measured by a digital shaft encoder, installed in the top drive spindle. In order to measure the forward slip, two photo-sensors were installed at the mill exit a known distance apart. The signals of these sensors are used by a clock to measures the time elapsed for strip to travel between these sensors. The strip exit velocity was then calculated. Rolling reductions were set to 15% and 35% . The rolling speed was varied from 20 rpm to 160 rpm, giving surface velocities of 0.26 to 2.1 m/s, the higher value of which is close to industrial operating speeds.

Four different lubricants (referred to as natural A, natural B, semi-synthetic A and semi-synthetic B) with different emulsion concentrations were used in the hot rolling tests. In the case of the natural oils, the emulsion ratios (by volume) were 1% and 3%, and for semi-synthetic oil emulsion ratios were 1% and 10%. Lubricant A is designed for low friction applications and lubricant B for higher friction conditions. Both lubricants are based on synthetic esters. The viscosities of these lubricants at 40 °C and 100 °C are given in the **Table 7-5**.

Table 7-5. Specifications of lubricants used during aluminum hot rolling

Lubricant	Density (g/mm^3)	Viscosity at 40 °C (mm^2/s)	Viscosity at 100 °C (mm^2/s)
Natural A	N/A	37.9	5.5
Natural B	N/A	76.3	8.6
Semi-synthetic A	0.866	28.5	5.5
Semi-synthetic B	0.883	29.6	5.7

7.2.1.2 Aluminium Hot Rolling Force Prediction Results

A dataset of the force variation during the aluminium hot rolling process was developed as explained in the previous section. In the first step of Neural Network modelling, the effects of the lubricant type was excluded, and a network was trained to predict roll forces for each lubricant. The model inputs were: reduction, roll speed, strip temperature, and emulsion concentration. The network was trained using all 22 data points and after final training the force variation was satisfactorily modelled. The maximum percent difference between model predictions and experimental values was found to be 7.5% with an average percent difference of 2.5%. This level of accuracy is quite satisfactory for this application and dataset. This error value is smaller than errors that normally arise due to experimental variations and the accuracy of instrumentation.

After successful development of a force model for a given lubricant, the lubricant type was also incorporated into the force prediction model. The experimental matrix included 66 data sets, of which 59 were used for network training. The rest was set aside to test performance of the trained neural network. A configuration of one hidden layer with 8 nodes, five inputs and one output node, with a learning rate of 0.7 and momentum rate of 0.7 was found to provide optimum accuracy. After 10,000 iterations, the network converged to a solution and further iterations had an insignificant effect on error reduction. The relative error values of the force model for both the training set and the test set are shown in **Figure 7-8**. The trained network predicted the roll force for 66% of the conditions within 5% relative error

band, 95% within 10% error band , and only three of the conditions were predicted with errors up to a maximum of 13%. The network also generalised well on the test data with errors less than 10%. The network roll force predictions for the 15% and 35% reductions have been plotted along with the experimental values in **Figure 7-9**.

7.2.1.3 Roll Torque Prediction

A network similar to the one used for roll force predictions was trained to predict roll torques. The experimental matrix, training set and test set were also the same as the previous model. The training data and predictions on the test data after 15,000 iterations are shown in **Figure 7-10**.

The relative errors of most of the training data are within 10% error band except two points. The network also predicted well on the test data with errors less than 10%. The network predictions for the 15% and 35% reductions have been plotted along with the experimental values in **Figure 7-11**.

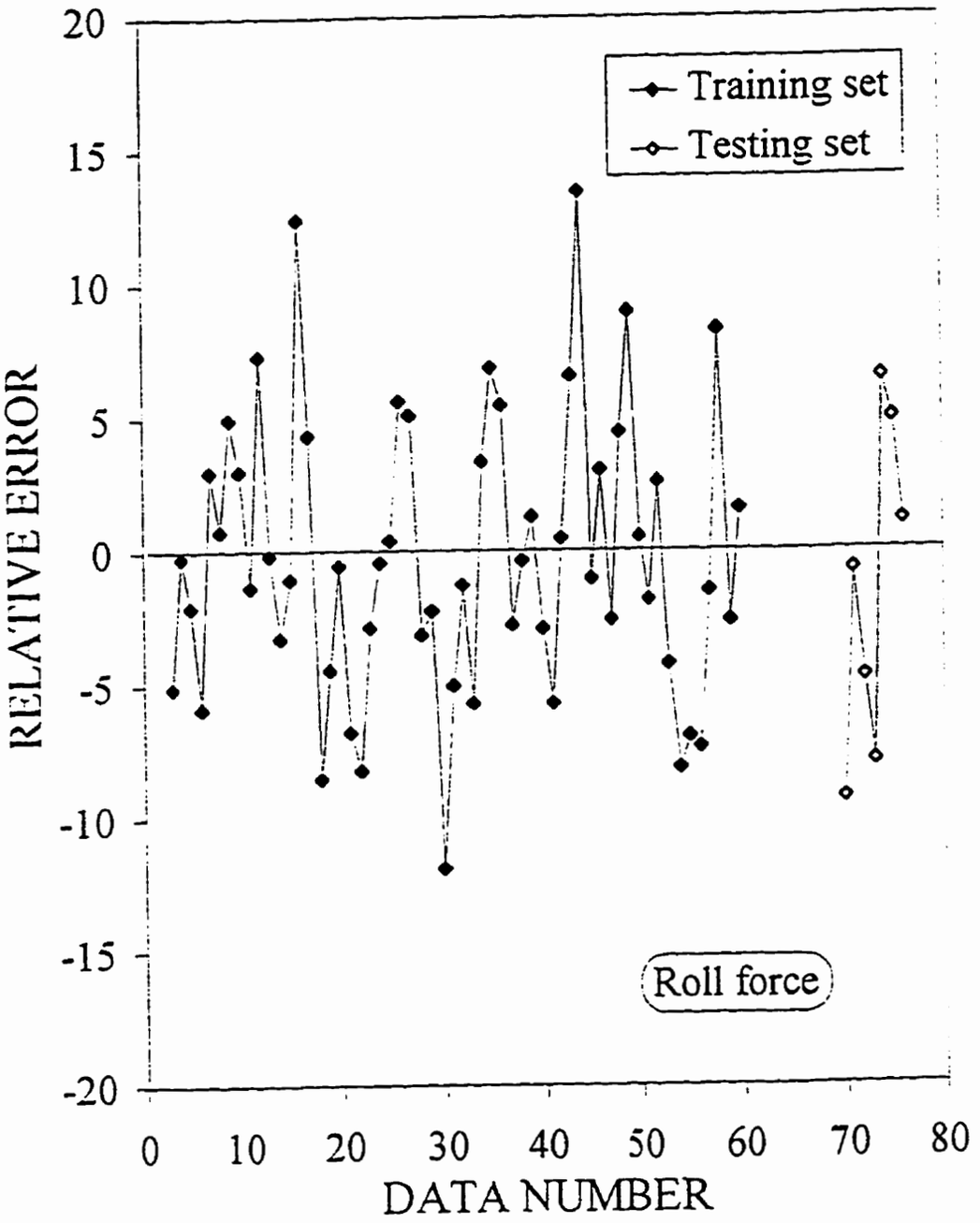

Figure 7-8 Percent difference of roll force predictions of a Neural Network model from experimental roll force values during aluminum hot rolling.

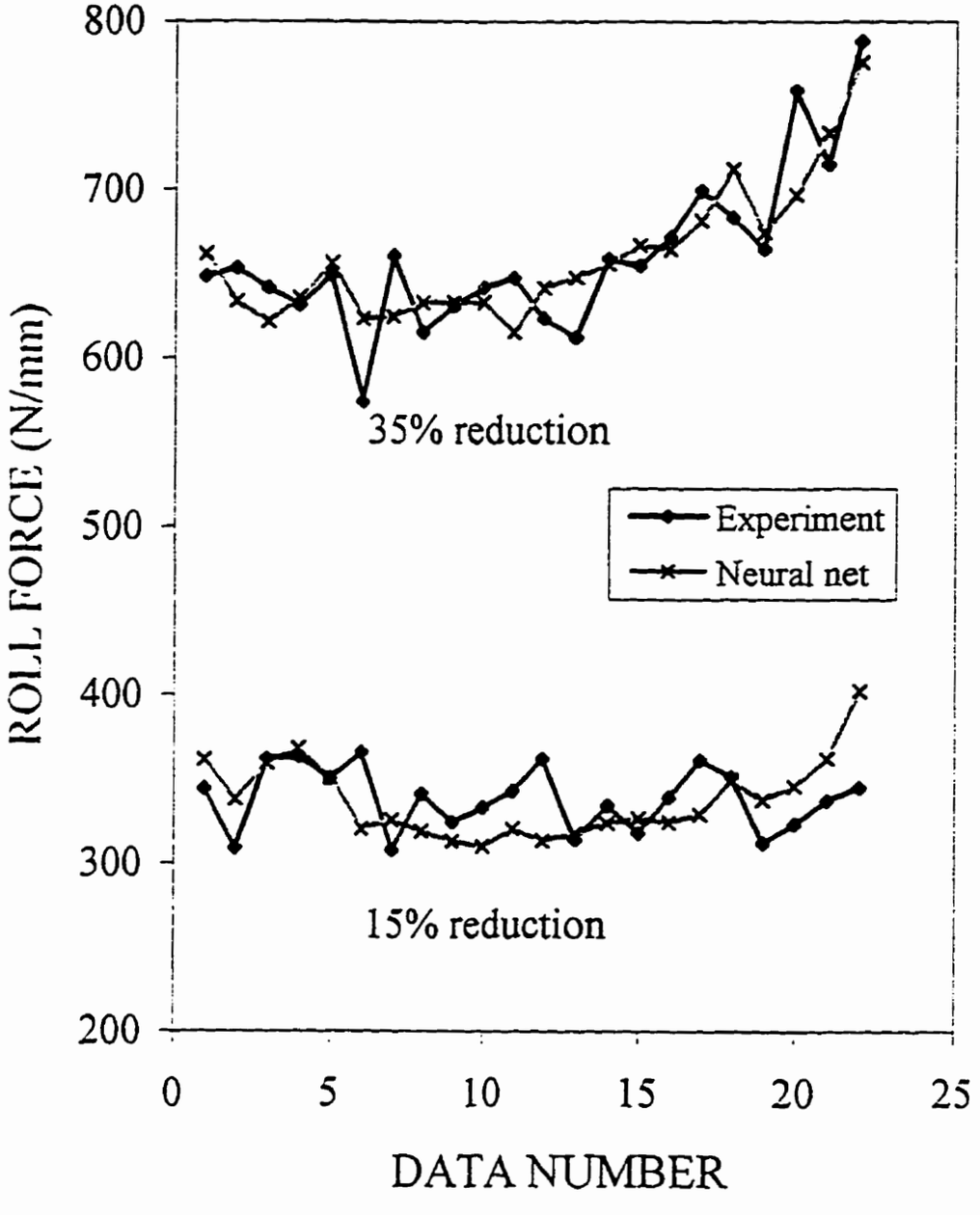

Figure 7-9 Neural Network roll force predictions and experimental values for two different reductions during hot rolling of aluminum strips.

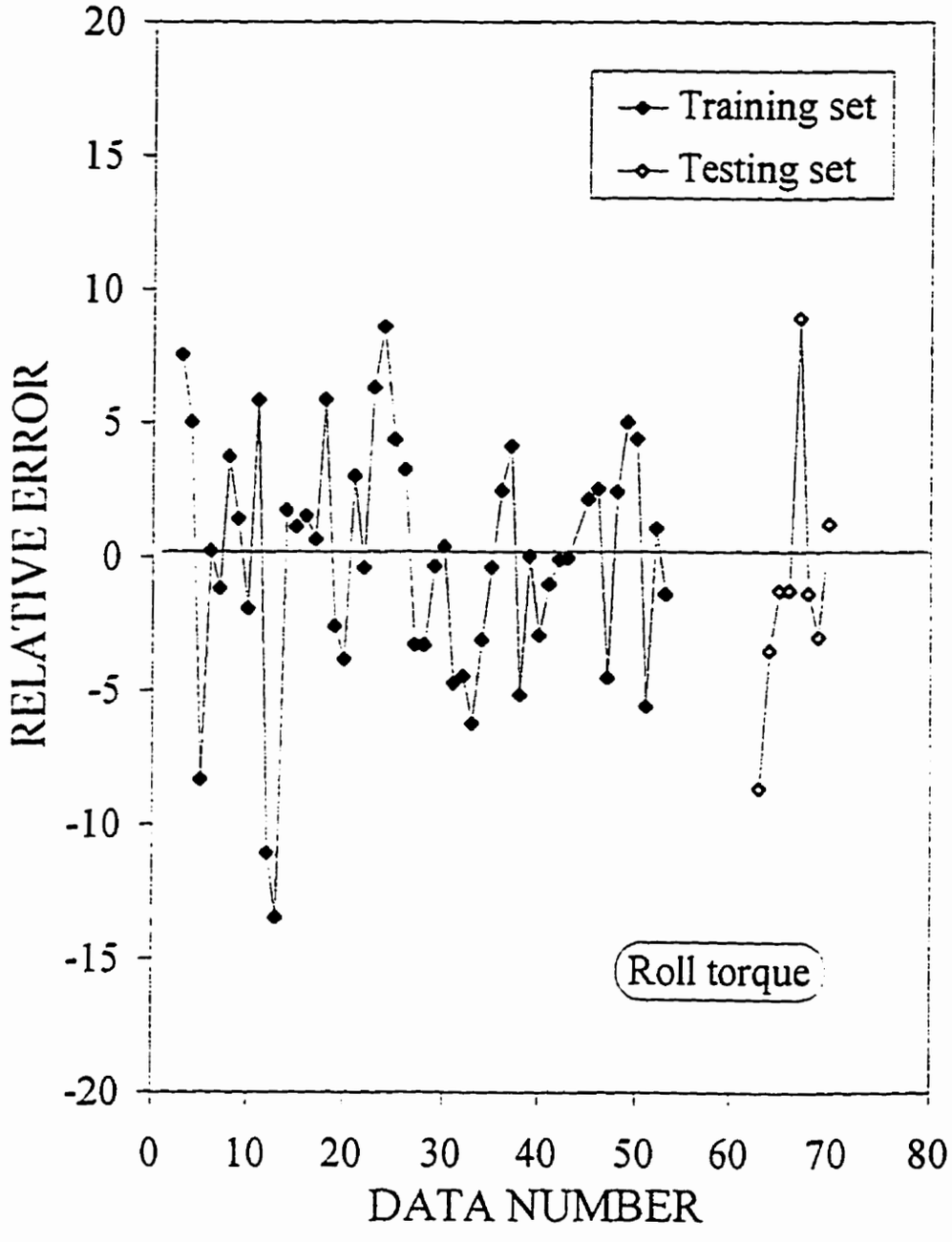

Figure 7-10 Percent difference of roll torque predictions of a Neural Network model from experimental roll torque values during aluminum hot rolling.

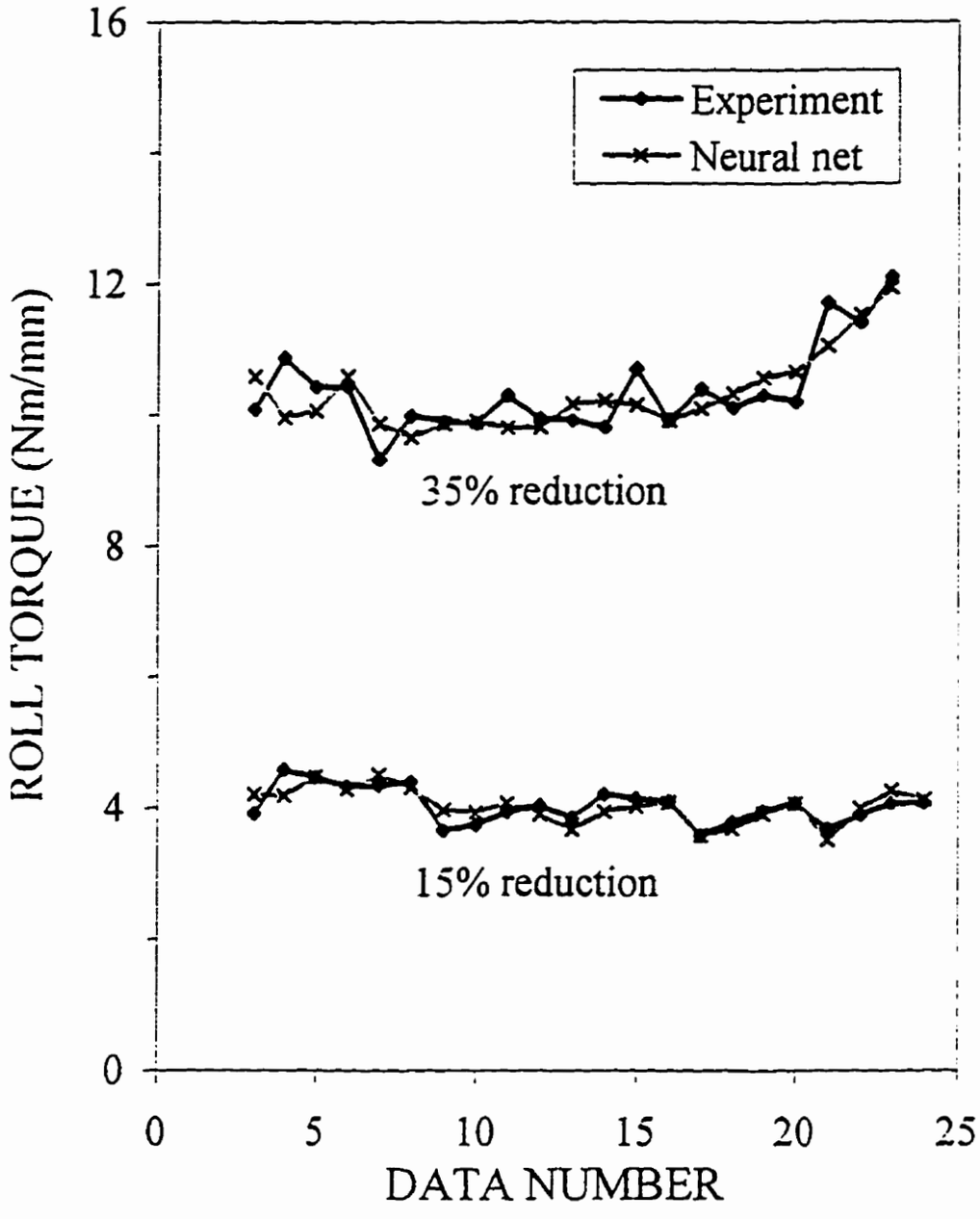

Figure 7-11 Neural Network roll torque predictions and experimental values for two different reductions during hot rolling of aluminum strips.

7.2.2 Aluminium Cold Rolling Force Prediction

The 25 mm wide and 1 mm thick strips of Al1100-H14 aluminium were rolled in the mill described in Section 7.2.1. The rolling speeds used were 20, 60, 100, 140, and 180 RPM. The reductions were aimed at 10, 20, 30, 40, and 50 percent. Mineral Seal oil, with no additive and with 1, 3, 5, and 7 percent Lauryl alcohol was used as lubricant. The average value of roll forces and roll torques for these 125 tests were collected. The first objective of the Neural Network modelling was to develop a predictive model to estimate the roll force and torque for any rolling conditions within the experimental domain used in these tests. The model inputs were rolling speed, reduction and percent of Lauryl alcohol additive in the lubricant. Every other ten data point (all together 13 for this data set) was set aside for network generalisation test and the rest of the data was used to train the network. A one hidden layer network with 12 processing nodes, a learning rate of 0.5, and a momentum rate of 0.7 provided the most accurate solution. Upon the convergence of the network to a solution with a minimum error, the network was tested through its prediction of roll forces for the 13 unseen test data. Roll force network predictions on these test data and the actual experimental values are compared in **Figure 7-12**. The results, in spite of an average percent difference of 8%, can be considered satisfactory, and more accurate than conventional roll force models. The average percent difference is in the same range as experimental errors with the existing equipment.

The main advantage of the neural network modelling compared to the classical roll force models is that it does not require any explicit knowledge of the coefficient of friction. Other models require the value of friction coefficient, which is a function of almost all process parameters, i.e. rolling speed, reduction, temperature, roll material and surface roughness, strip material and surface roughness, and type of lubricant. Furthermore, there is no direct way for the measurement of coefficient of friction in an industrial setting. That is why in the control algorithm of industrial strip mills, the coefficient of friction and all the roll/material interface variables are lumped into a single factor called the "geometric term". The geometric term is back-calculated from measured roll forces after each pass and saved in a database. The geometric term is used in roll force prediction of new slabs for initial roll gap setup purposes.

The value of geometric term for different family of grades and different gauge families is stored in a lookup table. This table is updated after each pass.

Outlier Detection

In all experimental or even industrial databases there is a possibility for the existence of outlier data points. The inclusion of outliers in a database is deleterious to the accuracy and generalisation of any model developed based on this kind of databases. Hence, it is important to employ proper techniques to detect the outliers and exclude them from further analysis. Neural Network modelling is still in early stages of development and author is not aware of any systematic approach to deal with this issue. However, there are some intuitive ways of detecting outliers in simple applications with only a few variables. For example, considering the performance of the roll force model of aluminium strips, depicted on **Figure 7-12,** it is observed that the network prediction on test number 8 is significantly worse than the rest of the test points. This could be due to the deficiency of the network in learning the roll force variation trend. However, it is also possible that this data point is actually an outlier. To investigate this possibility, this data point along with its neighbouring points were plotted in a roll force versus reduction graph for the same kind of lubricant and different rolling speeds, as depicted in **Figure 7-13**. It is clearly observed that the data for test number 8, denoted as outlier in **Figure 7-13,** is not in line with the rest, due to excessive experimental errors. After excluding this data point from modelling process, the average percent difference of the model predictions from experimental values dropped to 6.7%.

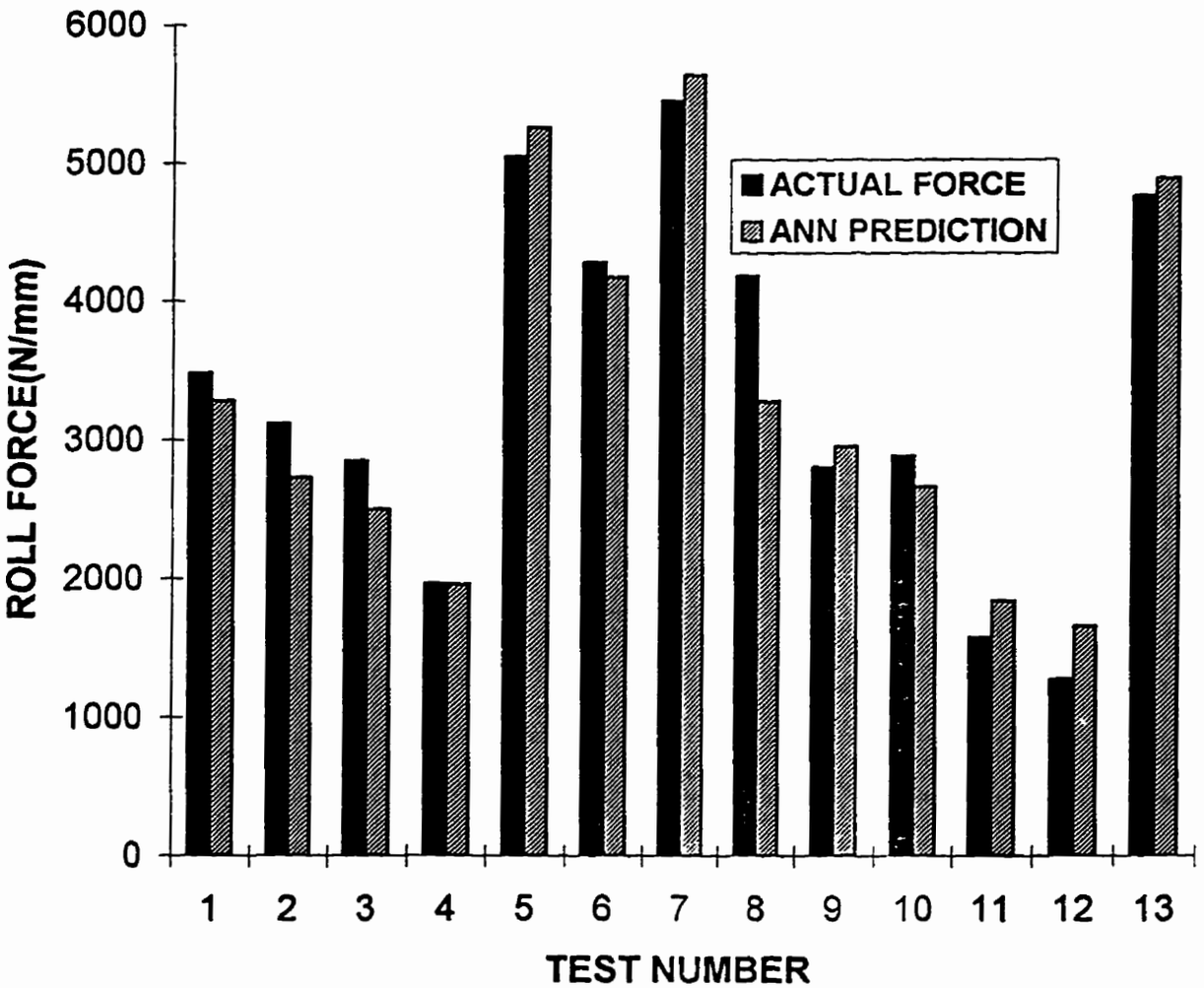

Figure 7-12 Neural Network model performance in predicting rolling forces during experimental cold rolling of aluminum strips.

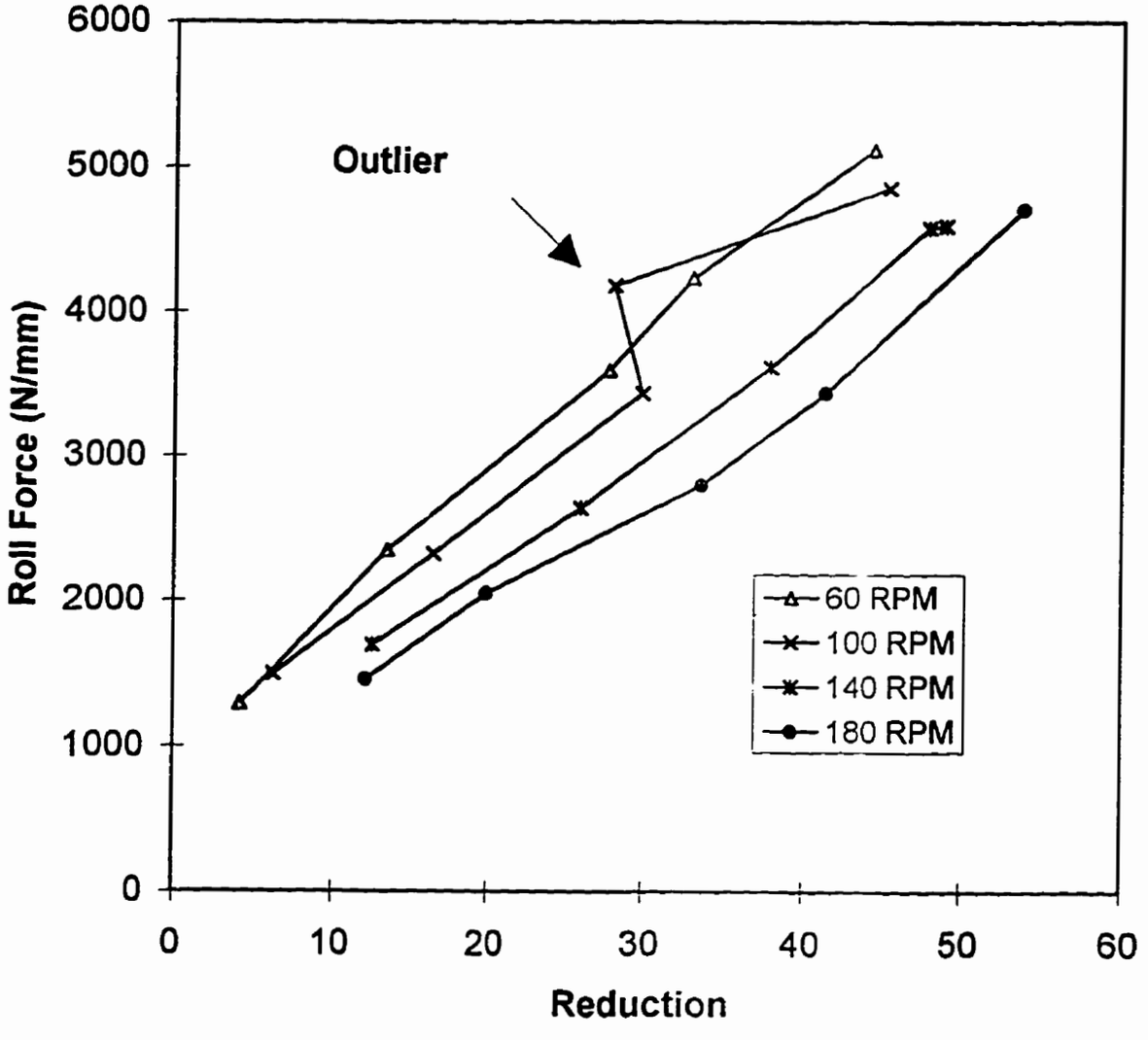

Figure 7-13 Investigation of the state of the test number 8 and visual observation of its deviation from the general trend of the rest of the data

Inverse Modelling :

In the modelling of industrial processes, the usual objective is to find the dependence of the process outputs on a given set of process inputs. However, in many applications we need to find out the conditions (process inputs) under which a desired process output can be obtained. For example, it could be of interest to use roll force models and try to find out for a given set of process conditions (roll speed and lubricant type) how we may set the roll force value in order to achieve a desired reduction. This is called static inverse modelling. Inverse models are useful in the process control applications. A common technique to develop feedback control algorithms of industrial processes is to integrate the variable of "time" into these models and develop dynamic inverse models.

Since Neural Networks do not rely on any closed form equations, lookup tables, or any physical understanding of the process, it appears that it is quite easy to convert Neural Network models to function as an inverse model. As an example, the database of aluminium cold rolling was used here to develop a model that predicts the reduction for any given set of roll speed, lubricant type, and measured roll force. The same network topology and learning parameters that were used for the force model, are used for reduction prediction model. The performance of the network on the 13 unseen test points is demonstrated in **Figure 7-14**. The average percent difference of the predictions from experimental values was calculated to be 7.6% (without removing the test number 8, which was an outlier).

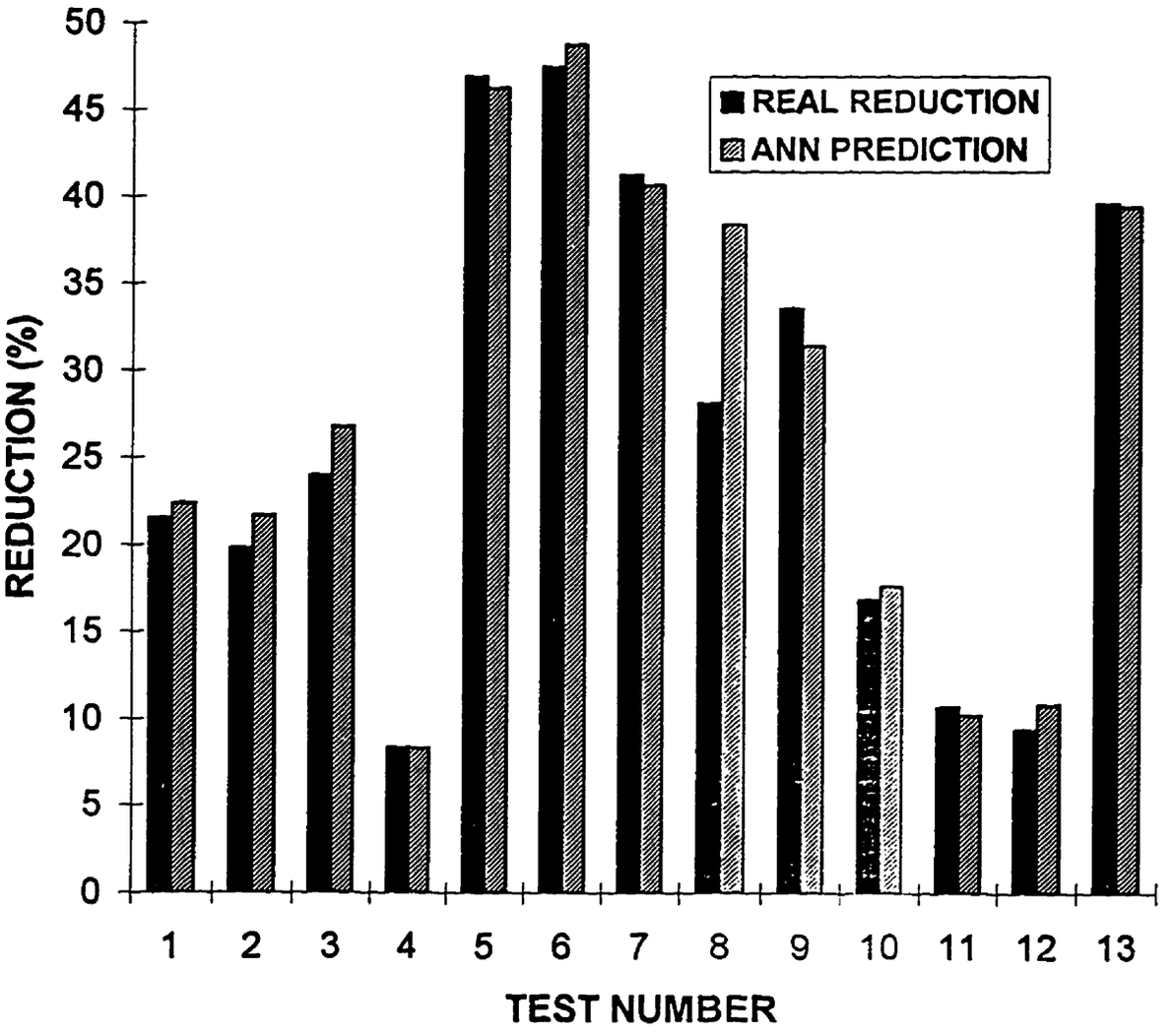

Figure 7-14 Neural Network model performance in predicting the reduction for given roll speed, lubricant and rolling force.

7.3 Hot Strip Mill Roll Force Prediction

7.3.1 Industrial Importance and Applications

One of the main quality indicators of steel strips, especially those which require stamping into complex shapes, is the gauge uniformity and consistency throughout the coil. The accuracy is achieved through an accurate roll gap setup, which is a direct function of the predicted rolling forces at each stand. Furthermore, a large error in the roll force prediction and consequently wrong roll gap setup leads to the occurrence of cobbles (failure of strip to pass through all stands). Even in the modern strip mills there is about half a percent yield loss due to occasional cobbles. This half a percent yield loss for a typical strip mill can sum to couple of million dollars of loss in production each year.

A main focus of this research was to utilise Neural Networks for hot strip mill roll force predictions in an industrial setting. In this section, the results of the application of this technique to the hot strip mill of Dofasco Inc. will be presented. The results demonstrated superior performance of Neural Networks which are readily applicable to industrial practice.

7.3.2 Process Database

A database of 831 coils of a high niobium HSLA steel (steel number 1, in Table 5.1) was extracted from a seven-stand hot strip mill. The data for each coil included material chemical composition, entry temperature to finisher, and for each stand: roll diameter and speed, strip entry and exit thickness, entry and exit width, predicted temperature, looper tensions, predicted force and actual measured force. Based on these data, the strain, strain rate, and L/h ratio (the ratio of projected length of arc of contact to the average strip thickness through the stand) for each stand were calculated. This database included three different types of measured and predicted roll forces: measured head-end, measured steady-state, and predicted head-end. Measured head-end force is the force measured after two seconds of rolling. Predicted head-end force is the prediction of the existing on-line roll force model for the first two seconds of rolling (before control action for roll gap correction takes place) based

on the predicted values for pass strains, strain rates, and temperatures for any given rolling schedule. This value at each stand is used to setup the roll gap at that stand. Steady-state roll force is the measured rolling force after 10 seconds of rolling. This is enough time for the feed-back controller to adjust the roll gaps and reach steady-state rolling. This value is used to update hardness factors in the corresponding lookup tables.

There were three reasons for this choice of material. First, this grade of high niobium strip is amongst the newly developed grades which are gaining a good portion of high strength strip market and they are not as well-understood as other grades. Second, there are speculations regarding the occurrence of dynamic recrystallization during rolling of these grades. Third, this grade is considered as a bad-behaved and difficult-to-roll grade due to the complex metallurgical events taking place during its hot rolling.

The coefficients and exponents of empirical roll force models being used by steel mills are usually stored in lookup tables specific to a grade and a gauge family. This is because the accuracy of the predictions will be enhanced if the range of variation of input variables is smaller. For example, the strips are divided into output gauge families of : less than 2.7 mm, between 2.7 and 4.5 mm, between 4.5 to 7.2 mm, and above 7.2 mm. However, it is very desirable to have more general models applicable to wider variety of grades and output gauges. The main reason for this is not the difficulties associated with managing a large number of lookup tables. The bigger drawback of breaking products into a larger number of grade or gauge families is that, if a grade is not rolled for a while, then due to the continuous changes in the mill conditions over the time the saved model and its coefficients for this grade will not be able to provide accurate predictions. By the time that models' coefficients are adapted to new conditions there would be some off-gauge rolling or the possibility of cobble. Taking this into account, it was desired to develop Neural Network models that are able to predict rolling forces for a wider range of products. This would reduce the accuracy of the model. However, it would keep the model up-to-date all the time. Hence, two series of models were developed. First, a general model being able to predict roll forces for all output gauges, and the second series which was specific to the same output gauge families that Dofasco Inc. uses.

7.3.3 Network Architecture

The first step in the modelling of a large industrial process is to decide or analyse which input variables or groups of variables have a significant effect on the process outputs. There are a number of statistical techniques that can be used for this purpose. The simplest one is the analysis of linear correlations between each possible input and the process output. However, most of these techniques are able to verify only linear dependencies.

A quite interesting feature of Neural Network models is their capability to give strong indications regarding the significance of the variation of each process output with any given process input. This feature is overlooked in the literature, however, it is of significant importance for real industrial modelling where the number of possible input variables can easily be over ten or twenty. A simple procedure for the detection of important variables is as follows. A network is developed which includes all of the possibly important variables. All of the available data points are normalised and used for training. The network is trained for a large number of epochs for a number of learning rates and momentum parameters. The number of processing nodes is kept as small as possible. Then the weight matrix of the network with the lowest final error is analysed. The sum of the weights of the connections of any given input variable and the hidden layer is normalised to the [0-100] range. This normalised sum value is an indication of the covariance of the network output with any given input variable. They are analogous to the normalised eigenvalues of the covariance matrix.

This technique or other statistical methods can identify important variables, however, they cannot replace the need for some physical understanding of the process for the optimal selection of input variables. For example, in the case of roll force modelling, it is easily understood that the roll diameter, input thickness, and output thickness have some effects on the rolling forces. However, through some understanding of the process, we may group these variables to one single variable, reducing the number of input variables, yet developing a more

accurate model. This variable, also used in other models, is the L/h ratio, the ratio of projected arc of contact to the average of the input and output thickness.

Based on the above technique and considering the existing roll force models, a number of different input variables were tried. The results of the analysis indicated that the most accurate roll force model is developed once the following input variables were used: strain, strain rate, temperature, looper tension, and L/h ratio.

Based on previous experience with aluminium rolling force models, it was decided not to try two-hidden layer networks. They did not produce any better results for the same number of processing nodes (CPU time). The dynamic node generation technique was employed to decide the optimum number of the processing nodes. The results indicated that 24 processing nodes are able to capture the variation of the data adequately.

The early stopping method was used to decide about the number of epochs (iterations) for the best results. Every other ten data point was set aside for testing and early stopping purpose. This method stops the iterations once the average error in the network predictions on the test data starts to increase with the number of iterations.

Both logistic sigmoid and hyperbolic tangent functions were tried. The results were identical. The scheme explained in Chapter 4 was used for data normalising.

7.3.4 Final Roll Force Model

An individual backpropagation Neural Network model was developed and trained to learn and predict the rolling force at each of the seven stands. Upon completion of training, the performance of these networks was tested on the testing data. For confidentiality reasons, no numerical data regarding the performance of the existing online model and Neural Network model is provided here. However, to appreciate the significant improvement in the prediction accuracy by using Neural Networks, the predicted force values for unseen slabs are plotted

with the measured values in **Figure 7-15**. For better visualisation, the same values for 50 slabs are plotted in the **Figure 7-16.**

This level of accuracy is quite satisfactory. The model easily outperforms the existing online roll force model. The improvement in the roll force prediction will translate to a marked reduction in the strip head-end off-gauge reduction. The model can readily replace the Set-up Module for roll gap set-up purposes. For a better appreciation of the model performance, the average percent difference of the Neural Network model predictions from measured values are compared to those of the existing roll force model for all seven stands in **Figure 7-17.** This level of accuracy is still improved further once the data is broken down to the Dofasco Inc. gauge families and individual models were developed for each family. This was expected, since the variation of some of the input variables, i.e. strain, strain rate, and L/h ratio are reduced.

The accuracy of the roll force predictions in experimental cold and hot rolling of aluminium strips is not as close to the accuracy achieved here. The obvious reason is the significantly lower level of noise in the database of industrial rolling.

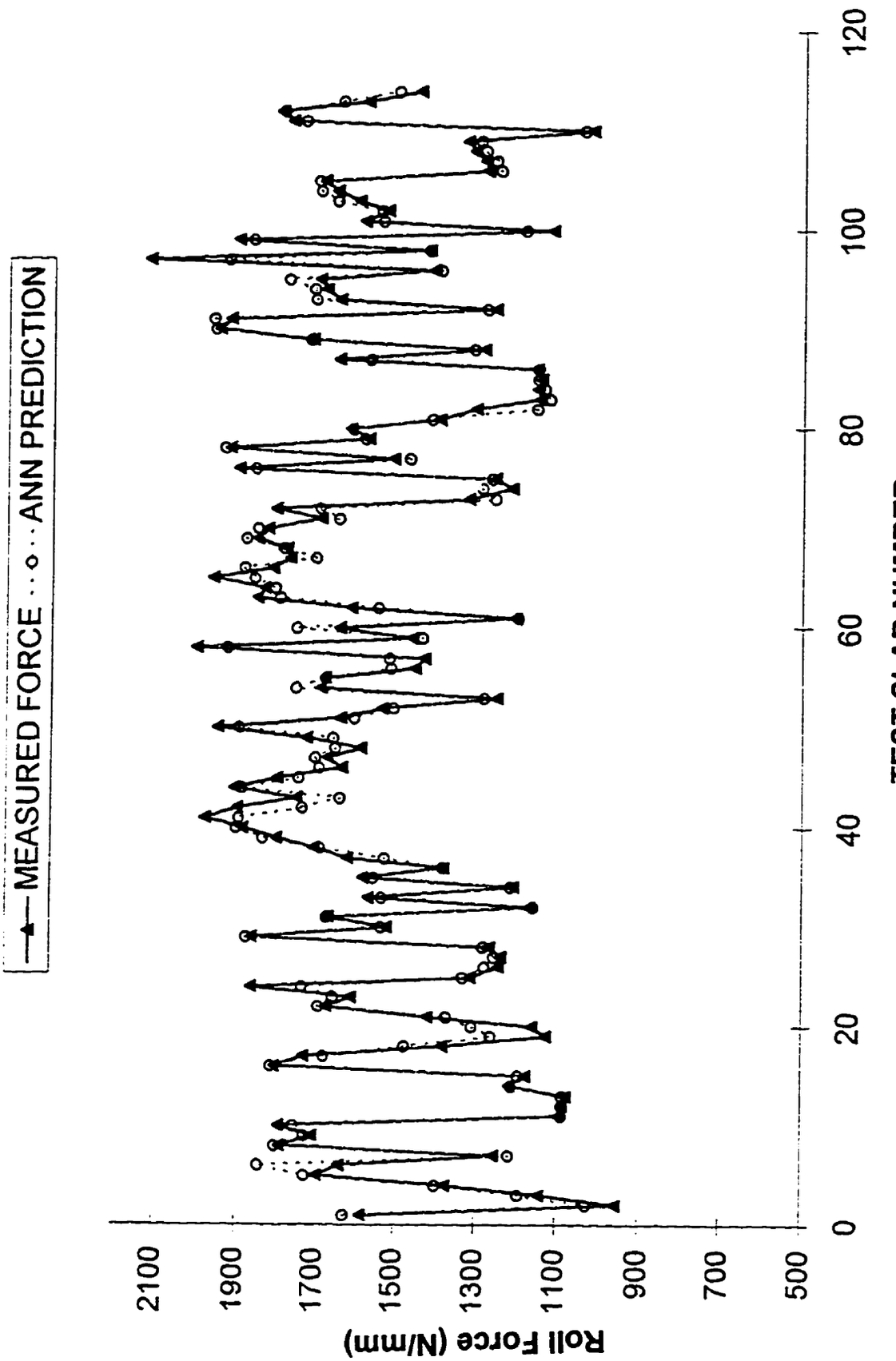

Figure7-15 Stand one roll force predictions by Neural Network model and measured oll force values

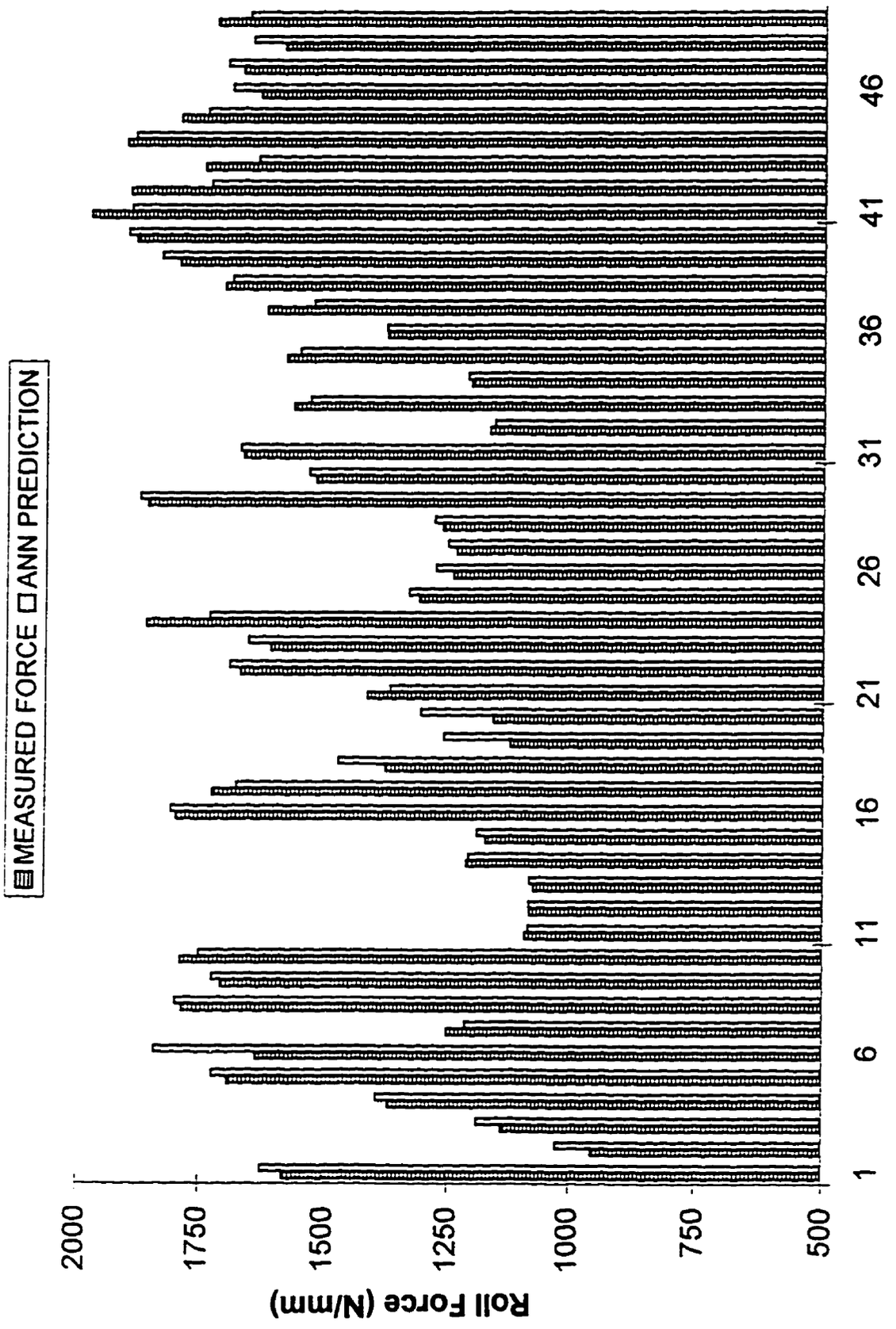

Figure7-16 Bar graph presentation of predicted and measured roll forces of 50 test slabs.

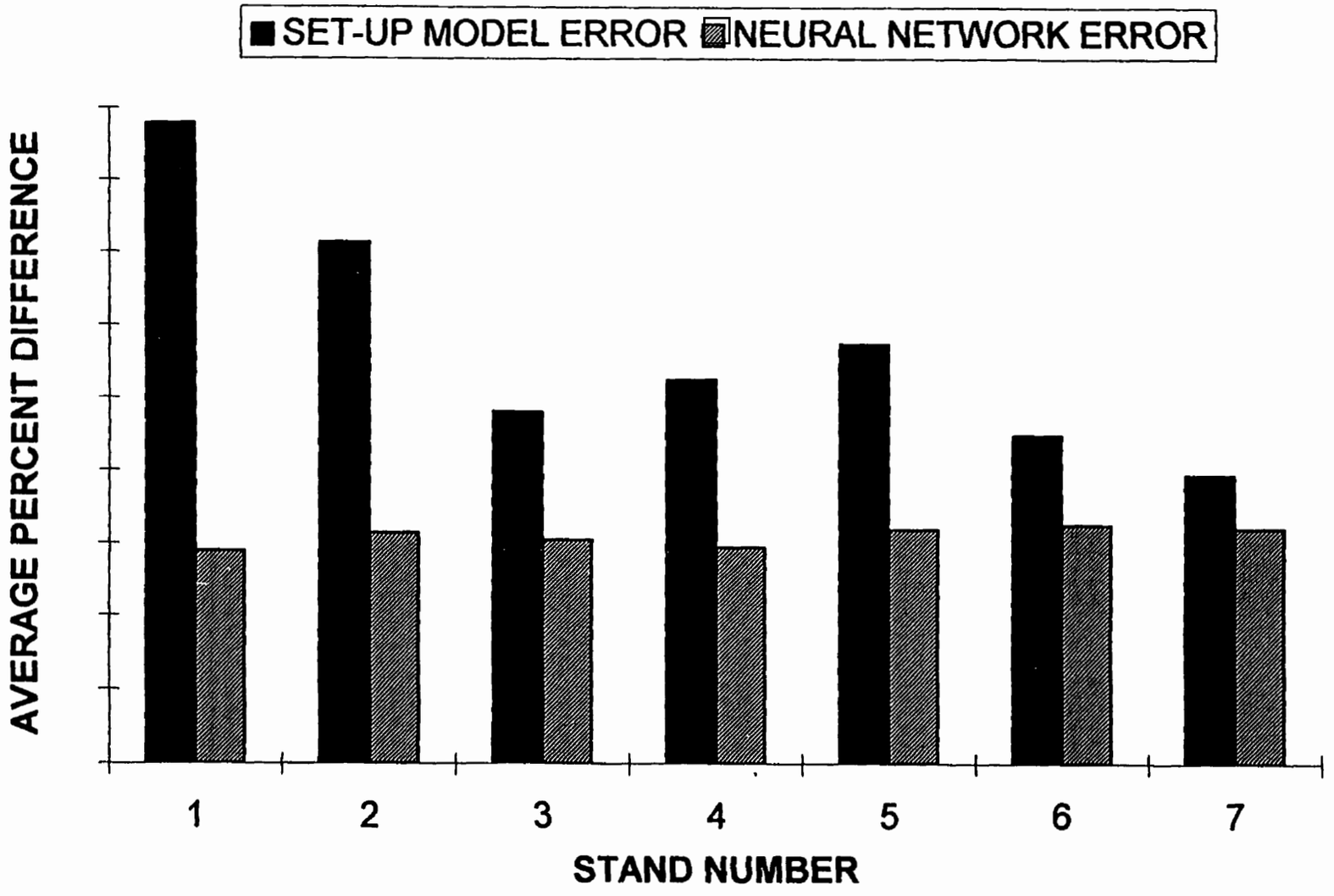

Figure 7-17 Percent differences of measured roll force values from roll force predictions by Neural Network model and online roll force model (Actual values are not shown due to the confidentiality)

The superior performance of this model demonstrates the great potential of Neural Network modelling for the roll force prediction applications and many other industrial processes. Moreover, Neural Networks provide a number of other advantages. These include: ease of development, adaptation and maintenance, recall speed, and ease of integration to the control algorithms. These benefits are briefly explained here.

Once the technique is understood and a Neural Network simulator code is developed, the processes of model development, training, testing, and information recall are quite routine. With the future development of commercial codes, it can be argued that the ease of model development will be one of the main advantages of Neural Network modelling technique.

Another very important characteristic of each modelling technique is model maintenance and adaptability. Neural Network models rely neither on lookup tables nor do they need any tuning, hence are easily maintainable. Model adaptation to the new process conditions and new experiential knowledge is done through retraining. However, this retraining is quite fast as the network initialisation is not random any more. Once new data are added to the database, the network is retrained with the initial weights and thresholds equal to those of previous trained network. The training is off-line and can be done as often as needed. However, retraining takes some time depending on the size of the network and database.

The Neural Network recall process is quite fast. Recall involves a one-shot non-iterative process of propagating input values through the network to output nodes. This feature renders the technique suitable for online control applications. Actually most of the industrial research in the Neural Network field is concentrated in control applications.

7.4 Dynamic Recrystallization Modelling

7.4.1 Background

The possibility and occurrence of dynamic recrystallization during hot rolling of HSLA strips still remains controversial. Torsion simulations presented in Chapter 6 revealed some

softening after the second interpass time which could be attributed to dynamic recrystallization. Mill log analysis also confirmed the results of torsion simulations in more than 80% of the slabs. However, there was no softening or there was some work hardening in the rest of the slabs. Independent attempts were made by the author and Dofasco's researchers to model the occurrence and the quantitative amount of this softening, using linear and non-linear regression techniques. However, none of the different forms of the regression models resulted in an R-squared value above 0.5. Hence, it was decided to employ Neural Networks to tackle this modelling task. The Neural Network technique managed to model this softening with satisfactory accuracy. The fact that Neural Network was able to perform this task indicates that there is a definite trend and physical basis for this softening and it is not a random occurrence. Neural Networks cannot learn anything from a random process. The accuracy of the results were further improved by integration of Principal Component Analysis to Neural Network modelling.

7.4.2 Mill Log Analysis and Modelling

Mill logs of 380 slabs of steel number 1 with an output gauge range of less then 2.7 mm were analysed. The purpose of the study was to determine the occurrence of unaccounted softening during strip rolling. The average softening of all six interstand intervals for all 380 slabs were calculated and plotted in **Figure 7-18**. This trend, repeated in other gauge families, clearly demonstrates a significant softening in the second interpass time.

This softening is automatically accounted for in the Finishing Set-up Module (FSU, control module for roll force calculations and roll gap setup), because the hardness modifiers are saved for each gauge range and stand. However, in the lookup table used by the FSU module there is only one average hardness modifier value stored for each grade and stand. If the variation of the softening amount is small, this scheme could be sufficiently versatile. However, it is not so and this interpass softening is not occurring consistently. This fact has been shown in **Figure 7-19**. The average value of this softening saved in the lookup table is about 5%, however, there are softening in the range of -10% to 20%.

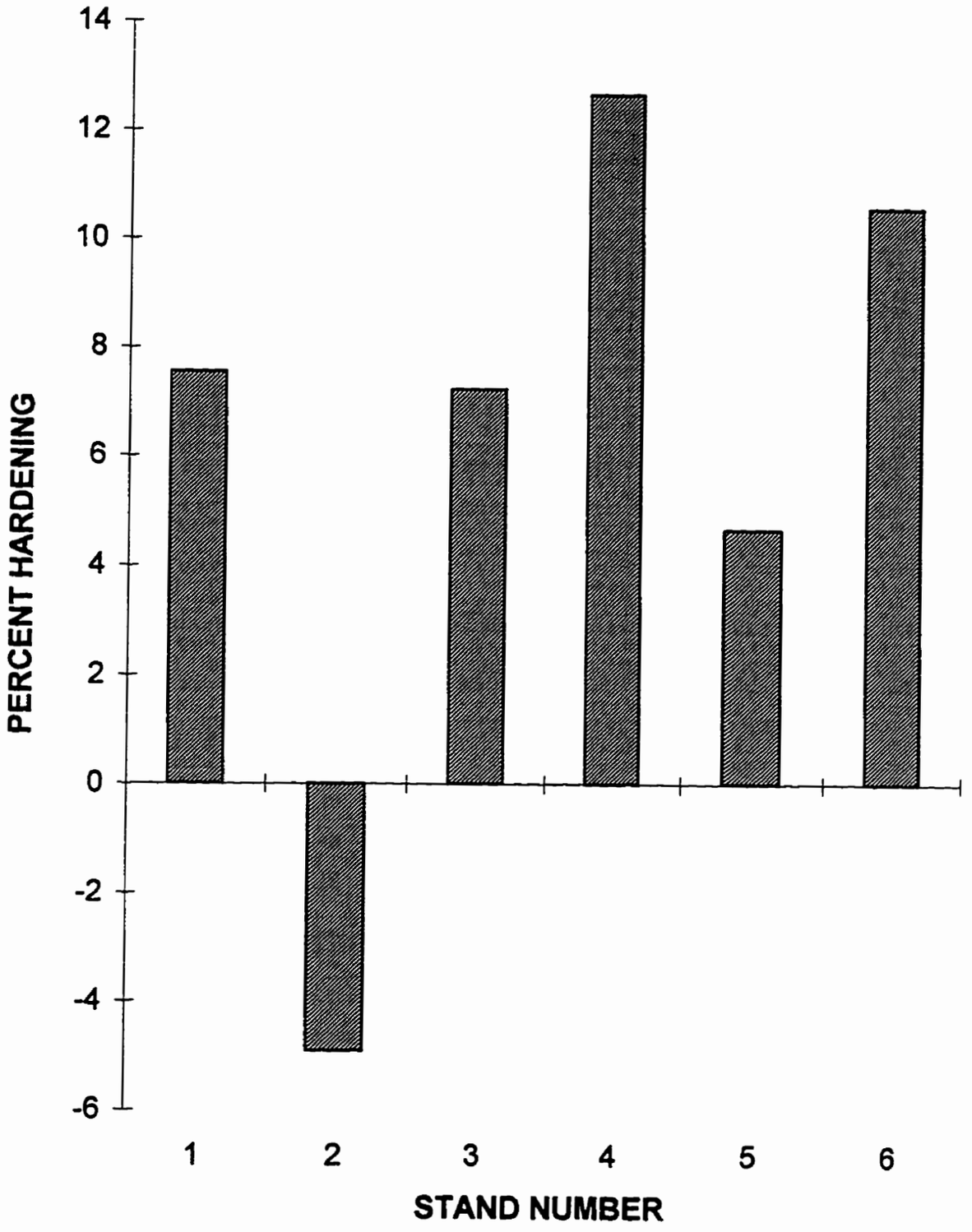

Figure 7-18 Average interpass hardening during finish rolling of steel number 1

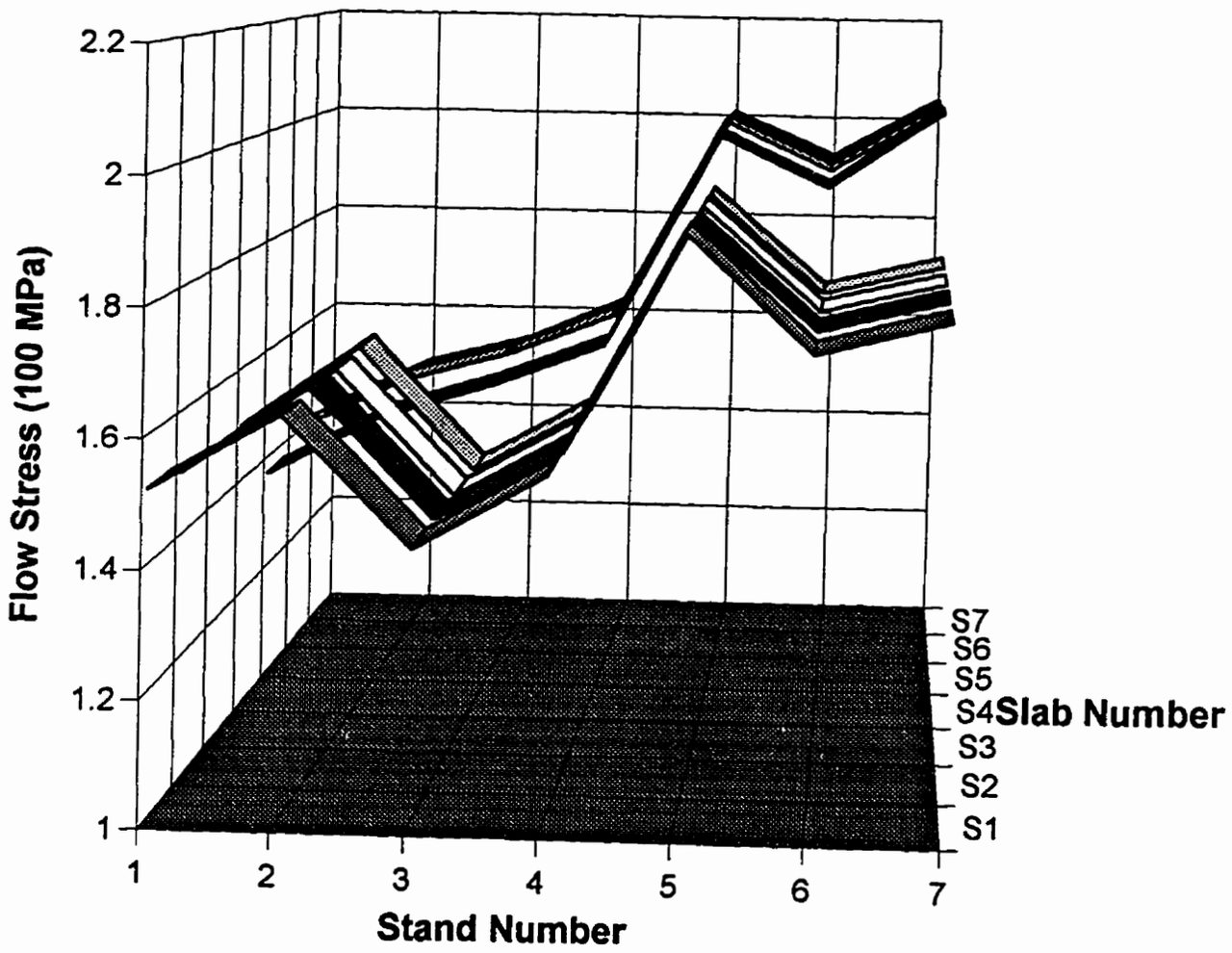

Figure 7-19 Inconsistent interpass softening in the second interpass time.

This much variation has to be modelled and accounted for during mill set-up, otherwise there would be consequential errors in the roll force prediction.

In the first stage of Neural Network modelling, the strain, strain rate, and temperature at the first two stands were chosen as network inputs. The interpass times in the first and second passes are linear combinations of the strains and strain rates in the first two stands, so they were not included in the inputs. A network with these six inputs and 16 processing nodes was trained to learn the variation of this interpass softening. The predictions of this network have been plotted along with the real softening in **Figure 7-20**. Careful consideration of this plot reveals that the Neural Network model follows the general trend of the data well. However, the average absolute difference between the model predictions and the real values was calculated to be 3.40 MPa. The amount of this softening was varied from -10 MPa (i.e. hardening) to 20 MPa in this database. The arrows in the plot point to the worst predictions.

In the next stage, the processing conditions in the third pass (strain, strain rate, and temperature) were included in the inputs of the model. After training of the network with the same number of the processing nodes as the pervious network, the average absolute difference of the model predictions and real softening was reduced to 2.73 MPa. This improvement was an indication of the effects of processing condition in the third pass on the interpass softening. The network predictions are plotted with real softening in the **Figure 7-21**.

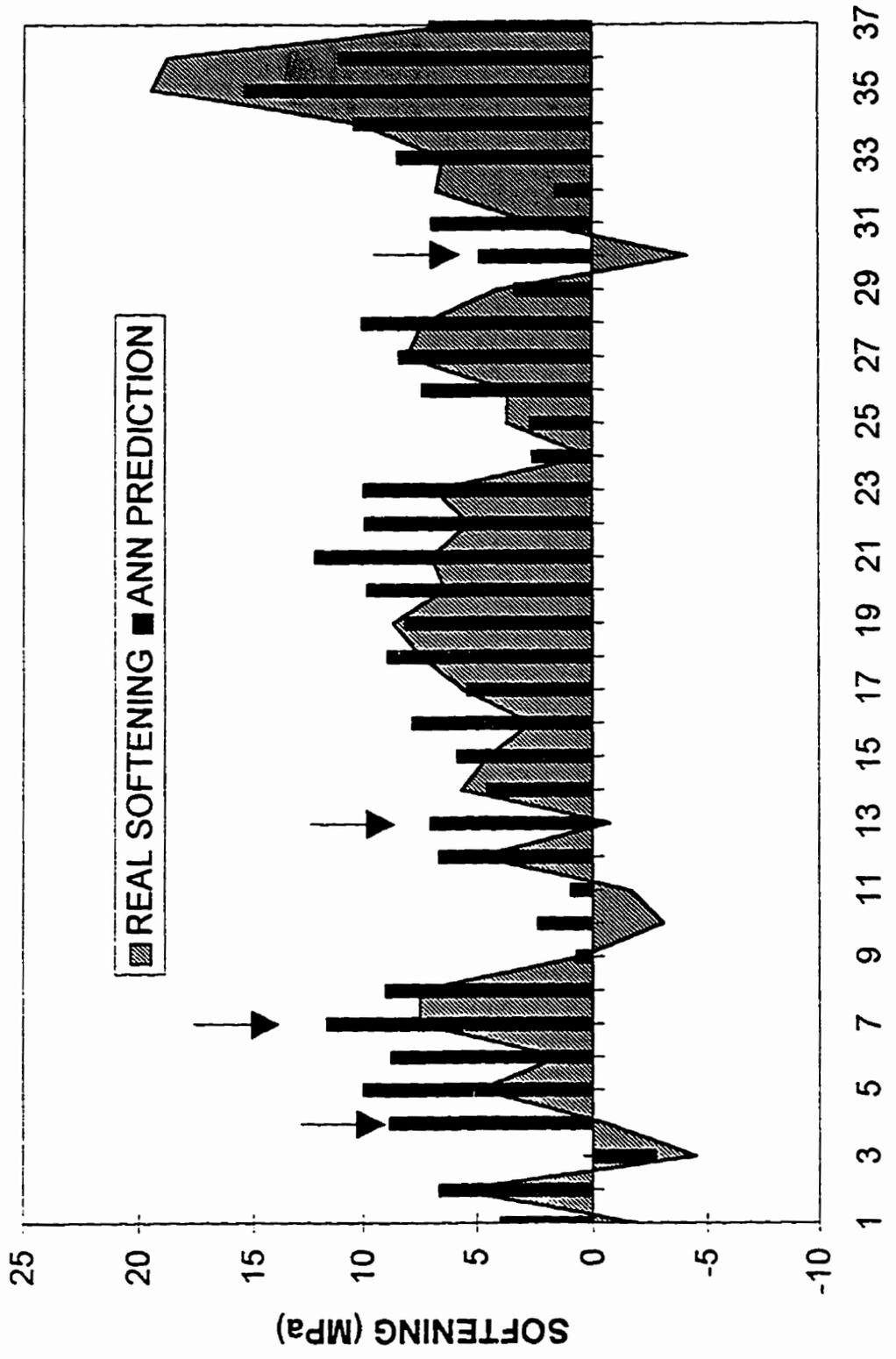

Figure 7-20 Neural Network prediction of interpass softening once six input variables were used.

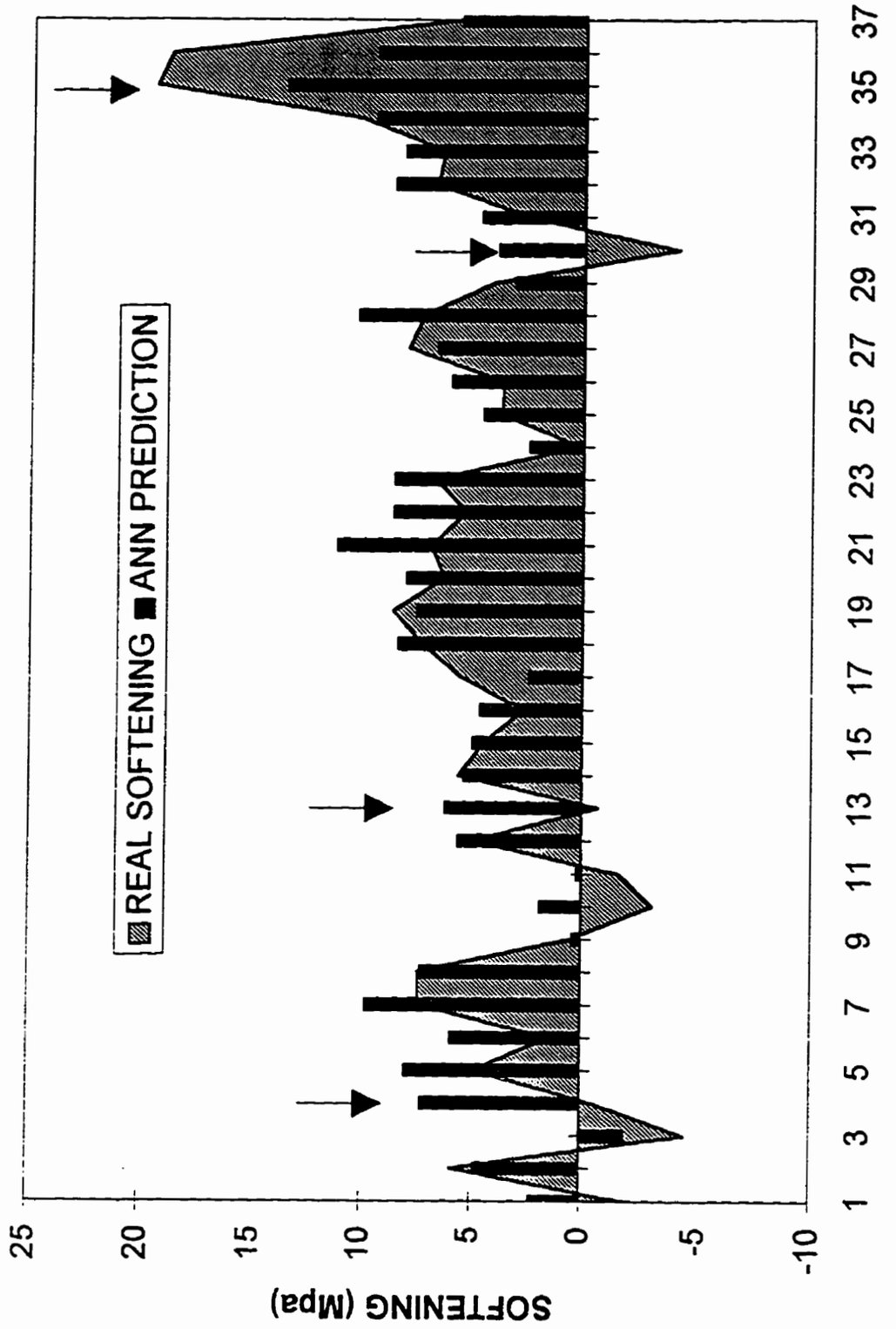

Figure 7-21 Neural Network prediction of interpass softening once nine input variables were used.

7.4.3 Analysis of the Correlation Between Input Variables

As discussed in Chapters 3 and 4, the linear colinearities of input variables in any given database create vagueness and instability in any model developed based on this kind of database. In the modelling of the dynamic recrystallization it is expected that there is some degree of linear correlation between input variables. To verify this, the correlation matrix of input variables was calculated. This matrix is demonstrated in the **Figure 7-22**. Looking at this matrix, there are some strong linear correlations between some of the input variables. A systematic way of looking at these correlations is to calculate normalised eigenvalues of this matrix. If the relative magnitude of eigenvalues are about the same, it means that the variation of these inputs are independent of each other. However, if some of the eigenvalues are much smaller than the rest, it means that there are some strong correlations. The eigenvalues of the correlation matrix of input variables are given in the **Table 7-6**. These normalised eigenvalues add up to 9, the number of variables.

Table 7-6 Eigenvalues of the correlation matrix of the original variables

Eigen 1	Eigen 2	Eigen 3	Eigen 4	Eigen 5	Eigen 6	Eigen 7	Eigen 8	Eigen 9
4.510	1.776	1.384	0.716	0.509	0.077	0.016	0.006	0.002

It is observed that the first eigenvalue is one order of magnitude larger than the fifth eigenvalue and two orders of magnitude larger than eighth eigenvalue. This means we can expect some model performance improvement if these correlated inputs are decoupled through transformation into the principal component space.

	Strain 1	Strain Rate 1	Temp 1	Strain 2	Strain Rate 2	Temp 2	Strain 3	Strain Rate 3	Temp 3
Strain 1	1	0.38	0.395	0.55	0.513	0.566	0.402	0.314	0.172
Strain Rate 1	0.38	1	0.489	-0.14	0.705	0.626	0.358	0.246	0.205
Temp 1	0.395	0.489	1	-0.051	0.379	0.833	0.417	0.276	0.15
Strain 2	0.55	-0.14	-0.051	1	0.533	0.365	0.108	0.245	0.277
Strain Rate 2	0.513	0.705	0.379	0.533	1	0.805	0.367	0.396	0.415
Temp 2	0.566	0.626	0.833	0.365	0.805	1	0.462	0.404	0.34
Strain 3	0.402	0.358	0.417	0.108	0.367	0.462	1	0.941	0.814
Strain Rate 3	0.314	0.246	0.276	0.245	0.396	0.404	0.941	1	0.955
Temp 3	0.172	0.205	0.15	0.277	0.415	0.34	0.814	0.955	1

Figure 7-22 Correlation matrix of the original data (nine variables)

7.4.4 Principal Component Analysis pre-processing

A computer code was developed to read in the original database of a process, form the correlation matrix, calculate eigenvalues and eigenvectors of the correlation matrix, and transform the original data into principal components. The details of this development is explained in Chapter 3. This code was integrated into the Neural Network simulator code as a pre-processing module and was applied to the above database for dynamic recrystallization modelling.

Considering that the eigenvalues of the correlation matrix of original variables are the variances of principal components, then the sum of the variations exhibited by the last three principal components is only 0.26% of the total variation of data (sum of the last three eigenvalues divided by nine). This is better understood if we look at the standardised covariance matrix of principal components. The code was used to form this matrix which is given in the **Table 7-7** .

	Prin. 1	*Prin. 2*	*Prin. 3*	*Prin. 4*	*Prin. 5*	*Prin. 6*	*Prin. 7*	*Prin. 8*	*Prin. 9*
Prin. 1	**4.510**	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 2	≅ 0.0	**1.776**	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 3	≅ 0.0	≅ 0.0	**1.384**	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 4	≅ 0.0	≅ 0.0	≅ 0.0	**0.716**	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 5	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	**0.509**	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 6	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	**0.077**	≅ 0.0	≅ 0.0	≅ 0.0
Prin. 7	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	**0.016**	≅ 0.0	≅ 0.0
Prin. 8	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	**0.006**	≅ 0.0
Prin. 9	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	≅ 0.0	**0.002**

Table 7-7 Standardized covariance matrix (correlation matrix) of the nine principal components corresponding to nine input variables.

The diagonal elements in this matrix are variances of principal components. The first diagonal element is the variance of the first principal component, the second element is the variance of the second principal component and so on. These variances which are the eigenvalues of the correlation matrix of the original variables, are ordered in descending order. If there are strong correlations between the original variables, then the variation of the last principal components are small. If so, the last principal components can be ignored without loosing much of the variation of the data. In this way, the problem can be analysed in a lower dimensional space. Moreover, the principal components do not have any linear correlations with each other. This is evident from the covariance matrix of principal component, in which

all of the off-diagonal elements are zero (Theoretically off-diagonal elements are zero, however, here they have very small values due to machine round-off errors).

7.4.5 Integration of Principal Component Analysis and Neural Networks

Based on the values of the nine principal components above, corresponding to the nine original input variables, the last three principal components were ignored and only the first six principal components were used in the analysis. A six-input Neural Network model was developed and trained based not on the original variables, but on these six principal components. Upon completion of the training, the Neural Network learned the variation of the softening based on the first six principal components. The recall phase (calculation of model output for any given input) changed to a two-step procedure. First, the principal components corresponding to the input variables of the data point under consideration were calculated. Then these principal components were propagated through the network to determine the network prediction.

The average absolute difference between network predictions and the measured softening for all the 130 unseen test data points was calculated to be 2.17 MPa. This means a 36% performance improvement compared to a six-nodded network with process conditions in the first two stands, and a 20% improvement compared to the nine-nodded network using original variables. The predictions of this network is plotted along with the measured interpass softening in the **Figure 7-23.**

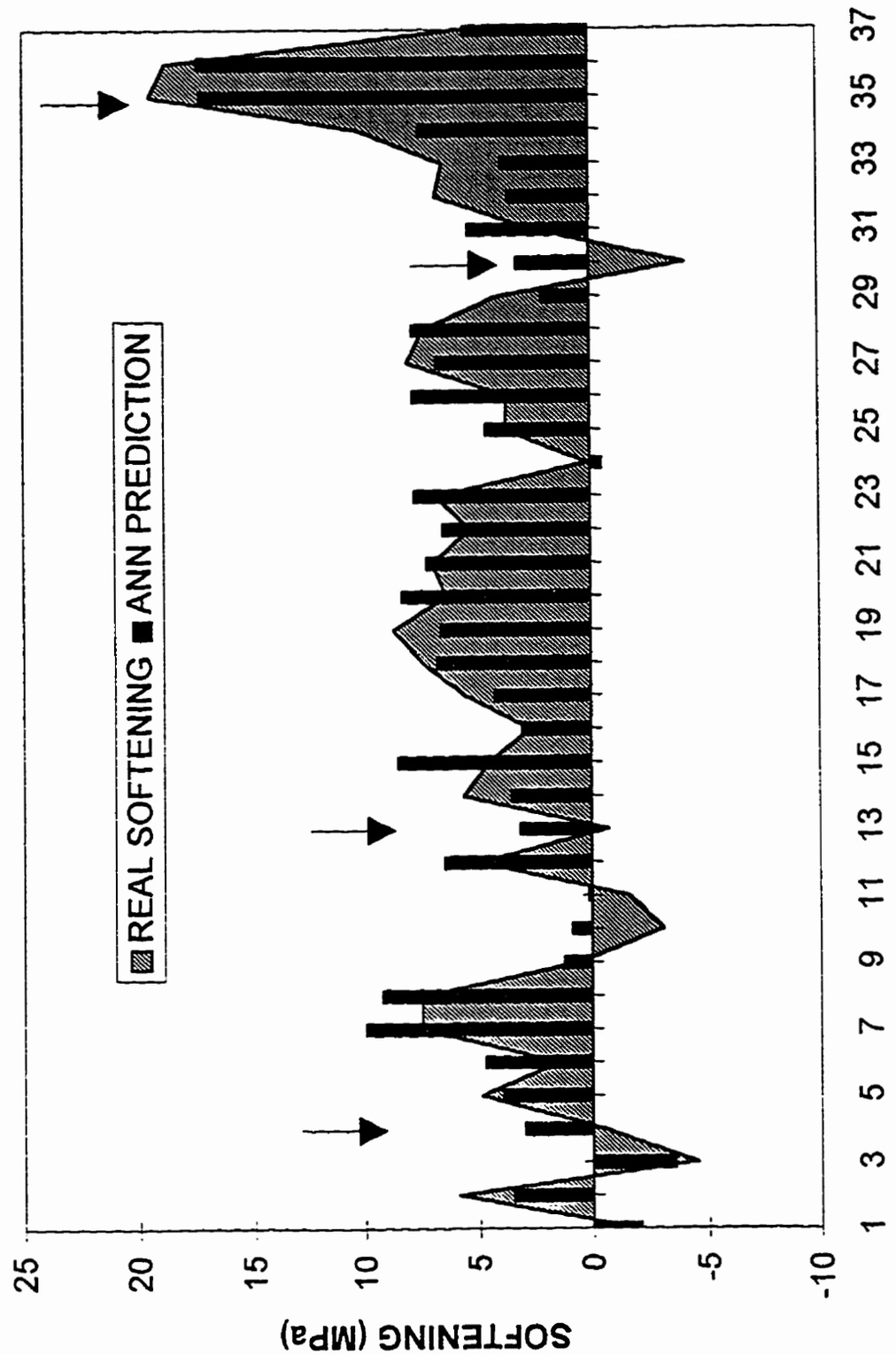

Figure 7-23 Neural Network predictions of interpass softening once the first six principal components were used.

This improvement in the performance compared to the nine-nodded network is solely attributed to the decoupling of the input variables. Another added benefit of this kind of data pre-processing prior to the Neural Network modelling is savings in CPU time. Principal Component Analysis is a non-iterative procedure, which takes a minimal computing time compared to the iterative network training procedure. Principal Component Analysis pre-processing, through a reduction of the number of input nodes, reduces the network size and training time significantly. Of course this is only possible once there are considerable linear correlations between input variables, which is the case for most of the industrial processes such as multi-stand rolling, machining, and welding.

7.5 Comparison of Neural Net Models with Statistical Models

The results discussed in this chapter provide another case-specific comparative study of the Neural Network techniques versus statistical methods. The Neural Network based model of steel hot flow stress prediction outperforms the statistical method in terms of both accuracy (mean difference of 2.17% versus 2.99%) and consistency (standard deviation of 0.48 versus 0.68). Another advantage of Neural Network modelling is ease of development, maintenance and recall. In hyperbolic sine function modelling the individual values of n, Q, and ln(A) at each strain level should be stored and retrieved from a data bank. Handling of such data banks, in the cases of multidimensional processes becomes cumbersome and requires some computational time to search and interpolate. However, the recall mechanism of backpropagation networks is very fast since all that is necessary is a one-step forward propagation of inputs through the layers. This recall speed, along with its accuracy, renders the technique promising for on-line control applications. Another major advantage of neural network models is their easy adaptability to new data. Upon presentation of a new data point, the network has to be trained again to learn this new pattern. However, the convergence to an acceptable solution is achieved with minimal iterations, as the previous weight matrix is used to initialise the network.

There are also some drawbacks to this technique. The network does not represent the knowledge in the familiar form of equation-based models. Thus, reasoning about the material or process behaviour and desired changes would not be as intuitive as they are in the regression methods. Moreover, network training relies on a comprehensive database of the given problem. However, the criteria to judge what constitutes the comprehensiveness of the data is neither well defined nor obvious.

8. Conclusions

A number of conclusions may be deduced from the results of physical simulations, mill data analysis, applications of Neural Networks, and Principal Component Analysis. These conclusions are listed under the above headings and are presented in the following sections.

8.1 Compression Testing

1. The type of pre-test solution treatment of compression samples prior to reheating to the test temperature has a significant effect on both microstructure and mechanical behaviour of HSLA steel specimens. When specimens are solution treated, quenched, then reheated to the test temperature, it results in finer grains and much higher super-saturation and potential for the Nb(C,N) precipitation upon reheating compared to the specimens which are solution treated and cooled to the test temperature.
2. The above increased potential for precipitation also affects the recrystallization kinetics substantially. The recrystallization kinetics results produced from the tests with different pre-test solution treatments should not be mixed together.
3. Interpass times experienced in hot strip rolling of HSLA steels are not long enough for complete static recrystallization. In the absence of full static recrystallization, the pass-to-pass strain is accumulated and can reach the critical strain for the initiation of dynamic recrystallization.

8.2 Torsion Testing

1. Kinetics of static recrystallization during interpass times of HSLA strip rolling is not fast enough to relax the strain during the first passes, even in a seven stand mill which has lower strain per pass compared to the six or five stand mills. Strain accumulation leads to the nucleation of new strain-free grains during deformation. During interpass times after the nucleation of new grains, kinetics of recrystallization is accelerated drastically. This accelerated recrystallization is due to the surpassing of the grain nucleation phase of the recrystallization process.
2. The accelerated softening after dynamic formation of new grains relaxes all accumulated strain and softens the material. This complete relaxation translates to lower flow stresses in the intermediate stands of the hot strip mill. If this softening is not taken into account, erroneous roll force predictions will result.
3. Once appropriate corrections have been incorporated, torsion based physical simulations of hot rolling processes are able to simulate the real process quite closely.

8.3 Mill Data Analysis

1. A portion of the errors in roll force predictions in the industrial rolling of HSLA steels is due to the unaccounted accelerated softening (metadynamic recrystallization) in the intermediate stands. The existing method of storing material flow stress characteristics in terms of one single value of hardness factor for each grade at each stand is not appropriate for HSLA strips.
2. The amount of this softening is between 20 MPa and -10 MPa (hardening). So far, there is no mathematical model able to predict the amount of this softening based on process input variables.
3. The existing strip rolling conditions of high Nb HSLA steels lie in the border of the domain of processing conditions that initiate dynamic recrystallization. If the change from one softening mechanism to another is not properly modelled and included in the control algorithm, then there would be a consequential unaccounted variation in the material flow

stress which will cause roll gap setup problems. This adds to the importance of such a model and explains why these grades are called "bad-behaved" grades.

8.4 Artificial Neural Network Applications

1. Artificial Neural Networks have evolved as powerful modelling techniques. They outperform statistical models of steel flow stress variation at high temperatures. Neural Networks are equally applicable to aluminium and other materials, as they do not rely on any mathematical equation derived from the material's physics.
2. Neural Network based models for roll force predictions during strip rolling of HSLA steels have a significant advantage in terms of prediction accuracy over the existing roll force models used in the industry. Their increased accuracy can be utilised in order to minimise the head-end off-gauge problems, reduce the occurrence of cobbles, and increase gauge consistency of the rolled strips.
3. The most advantageous applications of Neural Networks are in the modelling of those processes where there is no mathematical equation available to describe the input-output relationships. This is the case with the modelling of the amount of softening in the intermediate stands of hot strip rolling of HSLA steels. There is no other model available able to predict the amount of this softening. However, Neural Networks are able to do so with an acceptable accuracy.
4. Neural Network modelling is a potential candidate for the adaptive control of multi-variate industrial processes. It is expected that in the next millennium, Neural Network based control algorithms will be the dominant choice for the control systems of the industrial processes in general and the rolling processes in specific.

8.5 Integration of Principal Component Analysis to Neural Networks

1. Neural Network training algorithms, similar to statistical models, are poorly behaved when linear correlations exist between input variables. Principal Component Analysis is an effective and reliable multi-variate statistical technique which can transform linearly correlated input variables to uncorrelated principal components space. Principal Component Analysis was integrated into the Neural Network training algorithm as a data

pre-processor. Decoupling of the data using Principal Component Analysis increases the prediction accuracy of Neural Network based models significantly, for a minimal computational cost.

REFERENCES

1. D. Schmitter, "Automatic grain size determination and classification of iron carbides with neural nets", Steel Research, No. 10, 1995, pp. 449-453
2. T H. Courtney, "Mechanical Behaviour of Materials", McGraw-Hill, 1990, ISBN 0-07-0123265-8
3. R.W. Hertzberg, "Deformation and Fracture Mechanics of Engineering Materials:, J. Wiley & Sons, 1996, ISBN 0-47-1012149
4. H. Mecking and U.F. Kocks, Acta Metallu., vol. 29, 1981, p. 1865
5. A. Seeger, "Dislocations and Mechanical Properties of Crystals", ed. J.C. Fisher, Wiley, New York, 1957, p. 243.
6. H. Mecking and G. Gottstein, "Recrystallization of Metallic Materials", ed. F. Haessner, Dr. Reideree, Verlag, Stuttgart, 1978, p. 1865.
7. P. Cotterill and P.r. Mould, "Recrystallization and Grain Growth in Metals", Surrey University Press, 1976, p2.
8. U.F. Kocks, "Dislocations and Properties of Real Materials", Proc. Conf. Ins. Metals, Dec 84, London, p.125.
9. W. Roberts, "Deformation, Processing and Structure", 1982 ASM Materials Science Seminar, ed. G. Krauss, ASM, Metals Park, Ohio, 1984, p. 109.
10. H.J. McQueen and J.J. Jonas, J. Applied Metal Working, vol. 3, 1984, p.233.
11. C.M. Sellars, Metals Forum, vol. 4, 1981, p.75.
12. W. Roberts and B.Ahlblom, Acta Metall., vol. 26, 1978, p. 801.
13. C.M. Sellars, Proc. 7th Int. Symp. on Metallurgy and Materials Science, eds. N. Hansen et al., Riso National Laboratory, Roskilde, Denmark, 1986, p. 167.
14. L. Blaz, T. Sakai, and J.J.Jonas, Metal Science, vol. 17, 1983, p.609.
15. A. Laasraoui and J.J. Jonas, Metall. Trans., vol. 22A, 1991, p. 1545.
16. N.D. Ryan and M.J. McQueen, Can. Met. Quart., vol. 29, 1990, p. 147.

18. John. D. Verhoeven, "Fundamentals of Physical Metallurgy", John Wiley & Sons, 1975, p.328.
19. P. D. Hodgson, "Mathematical Modelling of Recrystallization Processes During the Hot Rolling of Steel", Thesis desertation, University of Queensland, 1990, p. 35.
20. O. Kwon and A. J. DeArdo, Acta Metallu. and Mater., vol. 39, 1991, p.526
21. R.W.K. Honeycomb, "The Plastic Deformation of Metals", ed. E. Arnold, ASM, Metals Park, Ohio, 1984, p.290.
22. R.A. Petkovic, M.J. Luton, and J.J. Jonas, Can. Met. Quart. vol. 14, 1975, p. 137
23. K.Lucke and K. Deret, Acta Metall, vol. 5, 1957, p.628
24. I. Weiss and J.J.Jonas, Metall. Trans., vol. 10A, 1979, p.831
25. Barr, W and Tipper., C.F., Journal of Iron and Steel Institute, 157, 223 (1947)
26. Morrison, W.B., Journal of Iron and Steel Institute, 201, 317 (1964)
27. Philips, R., Duckworth, W.E., and Copley, P.E., Journal of Iron and Steel Institute, 201, 593 (1964)
28. Kubota, H. Kozasu, T., Nippon Kokan Technical Report No 46, P205, 1969
29. Sekine, H. and Muruyama, T., Trans. Iron and Steel Inst. Japan, 16, 427 (1976)
30. D. Hodgson, Ph. D. Thesis dissertation, University of Queensland, Australia, 1990
31. C.M. Sellars and J.J. Jonas, "Physical and Computer Modelling of the Thermomechanical Processing of Steels", 10th PTD Conference Proceedings, p. 447, 1992
32. C. Roucoules, P.D. Hodgson, S. Yue, and J.J. Jonas, "Softening and Microstructural Changes Following the Dynamic Recrystallization of Austenite", Metallurgical and Materials Transactions A, volume 25A, February 1994, pp. 389- 400
33. A. Laasrouoi, Doctoral Thesis Dissertation, McGill University, 1990.
34. E.C. Sarmento and J.F. Evans, "Effects of strain accumulation and dynamic recrystallization on the flow stresses of HSLA steels during flat rolling", Proc. Int. Conf. on the Processing, Microstructure and Properties of Microalloyed and other Modern HSLA Steels, ed. A.J. DeArdo, Iron and Steel Society of Metallurgical Society of AIME, Warrendale, PA, 1992, pp105-112

35. T. Sakai and J.J. Jonas, "Dynamic Recrystallization: Mechanical and Microstructural Considerations", Acta metall., vol. 32, 1984, 189-209.
36. F.H. Samuel, S. Yue, J.J. Jonas and B.A. Zbinden, "Modelling of Flow Stress and Rolling Load of a Hot Strip Mill by Torsion Testing", ISIJ Int., vol. 29, No 10, 1989, pp. 878-886
37. F. H. Samuel, S. Yue, J.J. Jonas and K.R. Barnes, "Effect of Dynamic Recrystallization on Microstructural Evolution during Strip Rolling", ISIJ, Int. vol. 30 No. 3, 1990, pp. 216-225
38. A.J. DeArdo, Proc. Int. Symp. Mathematical Modelling of Hot Rolling of Steel, Hamilton, Canada, 1990, ed. S. Yue, CIMM, (1990), p. 220
39. N. Hatta, J. Kokado, S. Kikuchi, and H. Takuda, "Modelling on Flow Stress of Plain Carbon Steel at Elevated Temperature", Steel Research, vol 56, 1985, pp. 575-583.
40. C.M. Sellars and W.J. McG. Tegart, Mem. Sci. Rev. Metall. , vol.63, No. 9, pp. 731-746.
41. F. Wang and J.G. Lenard, "The Constitutive Behaviour of a Nb-V HSLA Steel in the Temperature Range of 900 to 975 oC", Steel Research, No.2, 1991, pp. 66-71.
42. A. Laasraoui and J.J. Jonas, "Prediction of Steel Flow Stresses at High Temperatures and Strain Rates", Metall. Trans. A., vol. 22, July 1991, pp. 1545-1558.
43. K. Minami, F. Siciliano, T.M. Maccagno and J.J. Jonas, "Mathematical Modelling of Mean Flow Stress during the Hot Strip Rolling of Nb Steels", ISIJ Int. , No 12, 1996, pp. 1507-1514
44. F. Garofalo, Trans. Metall. Soc. AIME 227, No. 2, 1963, pp. 351-355.
45. K.P. Rao and E.B. Hawbolt, "Development of Constitutive Relationships Using Compression Testing of a Medium Carbon Steel", J. Eng. Mat. Tech., ASME Trans. Vol. 114, Jan. 1992, pp. 116-123.
46. T. Sakai and M. Ohashi, "Tetsu to Hagane 67, No.11, 1981, pp. 2000-2009.
47. J.H. Beynon and C.M. Sellars, "Modeling Microstructure and its Effects during Multipass Hot Rolling, ISIJ Int., vol 32, No. 2, 1992, pp. 359-367.
48. J.G. Lenard and M. Tajima, "Thermo-mechanical Treatment of a High Nb-High V Bearing Micro-alloyed Steel", ISIJ Int. vol.35, No. 12, 1995, pp. 1509-1517.

49. S. Shida, "Effects of Carbon Content, Temperature, and Strain Rate on Compression Flow Stress of Carbon Steels", Hitachi Res. Rep., 1974, pp. 1-9.
50. A. Murthy and J.G. Lenard, "Statistical Evaluation of Some Hot Rolling Theories", J. Eng. Mat. Tech., vol. 104, 1982, pp. 47-51.
51. E. Orowan, "The Calculation of Roll Pressure in Hot and Cold Flat Rolling", Proc. Inst. Mech. Eng., vol.150, No.4, 1943, pp. 140-167.
52. H. Ford and J.M. Alexander, "Simplified Hot-Rolling Calculations", J. Inst. Metals, vol. 92, 1964, pp. 397-406.
53. G. Nadkarni, F. Wang and J.G. Lenard, "The role of Constitutive Formulations in the Analysis of Hot Rolling", Journal of Engineering Materials and Technology, vol. 109, Oct. 1987, pp. 343-350.
54. A.H. Tsoi, "Application of Neural Network Methodology to the Modelling of the Yield Strength in a Steel Rolling Plate Mill", Advances in Neural Information Processing, ed. Houson and Lippman, vol. 4, pp. 698-705.
55. Y.J. Hwu, Y.T. Pan, and J.G. Lenard, "A comparative Study of Artificial Neural Networks for the Prediction of Constitutive Behaviour of HSLA and Carbon Steels", Steel Research, vol. 67, No.2, 1996, pp. 59-66.
56. K.P. Rao and Y.K. Prasad, "Neural Network Approach to Flow Stress Evaluation in Hot Deformation", Journal of Materials Processing Technology, vol. 53, 1995, pp. 552-566.
57. Using NWorks, NeuralWork Professional II/Plus User Manual, Neural Ware Inc., USA.
58. Siemens quartely News Magazine, May, 1997
59. C. Chatfield and A.J. Collins, "Introduction to Multivariate Analysis", Chapman and Hall, London, 1980.
60. R.F. Gunst and R.L. Mason, "Regression Analysis and its Applications: A Data-Oriented Approach", Dekker, New York, 1980
61. I.T. Jollife, "Principal Component Analysis", Springer-Verlag, New York, 1986
62. J.A. Biglou, "GTA Weld Process Modelling using Principal Component Analysis", Master's Thesis, Mechanical Engineering Department, University of Waterloo, 1990.

63. S. C. Chapra and R.P. Canale, "Numerical Methods for Engineers", McGraw-Hill, New York, 1988.
64. W.H. Press, "Numerical Recipes in C: The Art of Scientific Computing", Cambridge University Press, 1992
65. K. Hornik, M. Stinchcombe, H. White, "Multilayer Feed-Forward Networks are Universal Approximators", Neural Networks, vol. 2, 1989, pp. 359-366
66. J. Hertz, A. Krogh, and R.G. Palmer, "Introduction to the Theory of Neural Computation", Addison Wesley, 1991.
67. D.A. Stacey, "Artificial Neural Networks, SD 733 Lecture Notes", University of Waterloo, System Design Engineering Dept., 1995.
68. F. Rosenblatt, "Principles of Neurodynamics", Spartan, New York, 1962
69. W.S. McCulloch and W. Pitts, "A logical Calculus of Ideas Immanent in Nervous Activity", Bulletin of Mathematical Biophysics, vol. 5, 1943, pp115-133
70. A.N. Burkitt, "Optimisation of the Architecture of Feedforward Neural Networks with Hidden Layers by Unit Elimination", Complex Systems, vol. 5, 1991, pp. 371-381
71. S. Tamura, M. Tateishi, M. Matumoto, S. Akita, "Determination of the Number of Redundant Hidden Units in a Three-layered Feedforward Neural Network", Proc. Int. Joint Conf. on Neural Networks, 1993, pp. 335-338
72. D.B. Parker, "Optimal Algorithms for Adaptive Networks", IEEE 1st Int. Conf on Neural Networks, San Diego, eds. M. Caudill and C. Butler, vol. II, 1987, pp593-600
73. D.E. Rumelhartm, G.E. Hinton, and R.J. Williams, "Learning Representation by Backpropagating Errors", Nature, vol. 323, 1986, pp533-536
74. D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition", vol. 1:Foundation, Cambridge, MIT Press, 1986
75. Matlab Neural Network Toolbox, The MathWorks Inc.,
76. A.S. Weigend, "On Overfitting and Effective Number of Hidden Units", Proc. of "Connectionist Models Summer School", Ed. M. Mozer et al, 1993, pp. 335-342

77. J.E. Moody, "The Effective Number of Parameters: An Analysis of Generalisation and Regularisation in Non-linear Systems", Advances in Neural Information Processing Systems 4, ed. J.E. Moody et al, Morgan Kaufmann, 1992, pp. 847-854
78. W.S. Sarle, "Neural Networks and Statistical Models", Proc of 19th SAS Users Group Int. Conf., Cary, NC: SAS Institute, pp. 1538-1550
79. C.M. Bishop, "Neural Networks for Pattern Recognition", Oxford, Oxford University Press
80. D. Michie, D.J. Speigelhalter and C.C. Taylor, "Machine Learning, Neural and Statistical Classification", Ellis Horwood, 1994
81. Y.J. Hwu, Y.T. Pan, and J.G. Lenard, "A comparative Study of Artificial Neural Networks for the Prediction of Constitutive Behaviour of HSLA and Carbon Steels", Steel Research, vol. 67, No.2, 1996, pp. 59-66.
82. K.P. Rao and Y.K. Prasad, "Neural Network Approach to Flow Stress Evaluation in Hot Deformation", Journal of Materials Processing Technology, vol. 53, 1995, pp. 552-566.
83. Using NWorks, NeuralWork Professional II/Plus User Manual, Neural Ware Inc., USA.
84. A.H. Tsoi, "Application of Neural Network Methodology to the Modelling of the Yield Strength in a Steel Rolling Plate Mill", Advances in Neural Information Processing, ed. Houson and Lippman, vol. 4, pp. 698-705.
85. E.R. Schmitter, "Automatic Grain Size Determination and Classification of Iron Carbide with Neural Nets", Steel research, vol. 66, No. 10, 1995, pp. 449-453
86. S.Y. Yun and K.S. Chang, "Dynamic Prediction Using Neural Network for Automation of BOF Process in Steel Industry", Iron and Steel Maker, August 1996, pp. 37-42
87. M. Roscheisen, R. Hofman, and V. Tresp, "Neural Control for Rolling Mills: Incorporating Domain Theories to Overcome Data Deficiency", Advances in Neural Information Processing System 4, ed. J. Moody, pp. 659-666
88. N.F. Portman, D. Lindhoff, G. Sorgel, Application of Neural Networks in Rolling Mill Automation", Iron and Steel Engineer, February 1995, pp 33-36
89. X. Wu, J.H. Garrett, and J. Ghaboussi, "Representation of Material Behaviour: Neural Network Approach,

90. M. Armand, Master Thesis, University of Waterloo, 1992

91. J.G. Lenard and A.N. Karagiozis, "Accuracy of High-Temperature, Constant Rate of Strain Flow Curves, In Factors That Affect the Precision of Mechanical Tests, eds. Papirno and Weiss, H.C., STP 1025, ASTM, Bal Harbour, 1987, 206-216.

92. H. Yada et al., "Strength and structural changes under high strain rate hot deformation of C steels", Trans. ISIJ, vol. 23, 1983, 100-109.

93. Ch. Roucoules, Ph. D. thesis, McGill University, 1992.

94. H. Ford; JIM; 96, (1968)

95. C. M. Sellars et al., "Plane strain compression testing at elevated temperatures", Report on research work supported by science research council grant B/RG/1481. University of Sheffield, October 1976

96. J.J. Jonas and T. Sakai; "A new approach to dynamic recrystallization", Presented at the 1982 ASM materials science seminar of 'Deformation, Processing and Structure', G. Krauss, ed., October 1982

97. F. Korber, "The plastic Deformation of Metals", J. of Inst. of Metals, vol. 48, p. 317, 1932.

98. J.J. Jonas and C.M. Sellars; "Physical and computer modelling of the thermomechanical processing of steel", 10th PTD conference proceedings, 1992, pp447-451

99. W.J. Liu and J.J. Jonas, "Calculation of Ti(CyN1-y)-Ti4C2S2-MnS Austenite Equilibrium in Ti-Bearing Steels", Mat. Sci. and Tech., 3, 1987, pp.197-206

100. D. S. Fields, Jr. and W.A. Backofen, "Determination of strain hardening characteristics by torsion testing", Proc. ASTM, 1957, vol. 57, pp. 1259-1272

101. G. R. Canova et al, "The use of torsion testing to assess material formability", Formability of metallic materials -2000 A.D., ASTM STP 753, J.R. Newby and B.A. Niemeir eds., ASTM, 1982, pp. 189-210.

102. J. Gil Selillano et al, "Large strain work hardening and textures, Progress in Materials Science, 1981, vol. 25, pp. 69-412

103. E. Anelli, "Application of Mathematical Modelling to Hot Rolling", ISIJ Int. vol. 32, No. 3, pp. 440-449

104. T.M. Maccagno and J.J. Jonas, "Correcting for the Effects of Static and Metadynamic Recrystallization during the Laboratory Simulation of Rod Rolling", ISIJ Int. vol.34, No. 7, 1994, pp. 607-614

105. C. Roucoules, P.D. Hodgson, S. Yue, and J.J. Jonas, "Softening and Microstructural Changes Following the Dynamic Recrystallization of Austenite", Metallurgical and Materials Transactions A, volume 25A, February 1994, pp. 389-400

106. H. Yada et al., "Strength and structural changes under high strain rate hot deformation of C steels", Trans. ISIJ, 23, 1983, pp. 100-109

107. P. Choquet, A. LeBon, C. Rossard, C. Perdrix, and G. Joannes, "The hot Torsion Testing at IRSID: Application to the Simulation and Modelling of Hot Forming Processes, THERMEC-88, 1988, ed. I. Tamura

108. H. Saito, K. Ushioda, T. Senuma, T. Nakamura, and K. Esaks, "Structure and Texture Evolution During Subsequent Annealing of Steel Hot-Rolled in α Phase, THERMEC-88, ed. I. Tamura, 1988, pp.628-635

109. W. Robertrs, A. Sandburg, T. Siwecki, and T. Werlefors, "Prediction of Microstructure Development during Recrystallization Hot Rolling of Ti-V Steels", HSLA Steels: Technology and Applications, ASM, 1983, pp. 67-84.

110. C. M. Sellars and W.J. McTeggart, "Hot Workability", Int. Metall. Rev., vol. 17, 1972, pp. 1-24

111. C.M. Sellars, "Modelling Microstructural Development during Hot Rolling", Material Science and Technology", vol. 6, Nov. 1990, pp. 1072-1081

112. C.M. Sellars, "The physical Metallurgy of Hot Working", Hot Working and Forming Processes, ed. C.M. Sellars and G.J. Davis, Metal Society, London, 1979, pp.3-15

113. C. Roucoules. P.D. Hodgson, S. Yue, and J.J. Jonas, "Softening and Microstructural Changes Following Dynamic Recrystallization of Austenite", Metall. And Metals Trans. A, vol. 25A, Feb 1994, pp. 389-400

114. B. Dutta and C.M. Sellars, "Effect of Composition and Process Variables on Nb(C,N) Precipitation in Niobium Microalloyed Austenite", Materials Science and Technology, vol. 3, March 1987, pp. 197-206

115. J.G. Lenard and A.N. Karagiozis, "Accuracy of High-Temperature, Constant Rate of Strain Flow Curves, In Factors That Affect the Precision of Mechanical Tests, eds. Papirno and Weiss, H.C., STP 1025, ASTM, Bal Harbour, 1987, 206-216.

116. D. Hall and J. Worbec, "Torsion Simulation of the Strip Rolling Process", Phase Transformations during the Thermal/Mechanical Processing of Steel, Proc. of the International Symp., Ed. E.B. Howbott and S. Yue, 1992.

117. E.C. Sarmento and J.F. Evans, "Effects of Strain Accumulation and Dynamic Recrystallization on the Flow Stresses of HSLA Steels during Flat Rolling", Proc. Int. Conf. on the Processing, Microstructure and Properties of Microalloyed and other Modern HSLA Steels, ed. A.J. DeArdo, Iron and Steel Society of Metallurgical Society of AIME, Warrendale, PA, 1992, pp105-112.

118. F.H. Samuel, S. Yue, J.J. Jonas and K.R. Barns, "Effect of Dynamic Recrystallization on Microstructural Evolution during Strip Rolling", ISIJ Int. vol. 30, No. 3, 1990, pp216-225.

119. F.H. Samuel, S. Yue, J.J. Jonas & K.R. Barnes, "Effect of Dynamic Recrystallization on Microstructural Evolution during Strip Rolling", ISIJ Int. vol. 29, No. 3, 1989, pp. 216-225.

120. T. Siwecki, Proc. of the Int. Conf. on Physical Metallurgy of Thermomechanical Processing of Steel and Other Metals, I, I. Tamura, ed., ISIJ, Tokyo, 1988, pp. 232-240.

121. C.M. Sellars, "Physical Metallurgy of Hot Working", Proc. of Int. Conf. on Hot Working and Forming Processes, July 1979, Sheffield, C.M. Sellars & G.J. Davis, Ed., The Metal Society, London, pp. 3-15.

122. T. Tanaka, N. Tabata, T. Hatomura and C. Shiga, "Modelling of Flow Stress and Rolling Load of a Hot Strip Mill by Torsion Testing", Proc. Int. Symp. on Microalloying 75, Union Carbide Corp., New York, 1977, pp. 107-115.

123. L.J. Cuddy, J.J. Bawin, and J.C. Raley, "Recrystallization of Austenite", Met. Trans. A., Vol. 11A, 1980, pp. 381-389.

124. L.J. Cuddy, " Thermomechanical Processing of Microalloyed Austenite", TMS-AIME, Warrendale, PA, 1982, pp. 129-136.

125. S. Yue and J.J. Jonas, "The Three Critical Temperatures of Steel Rolling and Their Experimental Determination", Materials Forum, 14, 1990, pp. 245-252.

126. F.H. Samuel, S. Yue, J.J. Jonas & B.A. Zbinden, "Modelling of Flow Stress and Rolling Load of a Hot Strip Mill by Torsion Testing", ISIJ Int., vol. 29, No. 10, 1989, pp. 878-886.

127. I. Weiss and J.J. Jonas, "Interaction Between Recrystallization and Precipitation During the High Temperature Deformation of HSLA Steels", Met. Trans., A, vol. 10a, 1979, pp. 831-840.

128. R.A. Petkovic, M.J. Luton and J.J. Jonas, "Recovery and Recrystallization of Carbon Steels between Intervals of Hot Working" Canadian Metallurgical Quarterly, 14, 1975, 137-145.

129. C. Ouchi, T. Sampet and I. Kozasu, Trans. ISIJ, vol. 22, 1982, 214.

130. T. Sakai and J.J. Jonas, "Dynamic Recrystallization: Mechanical and Microstructural Considerations", Acta metall., vol. 32, 1984, pp. 189-209.

131. F.H. Samuel, S. Yue, J.J. Jonas and K.R. Barns, "Effect of Dynamic Recrystallization on Microstructural Evolution during Strip Rolling", ISIJ Int. vol. 30, No. 3, 1990, pp. 216-225.

132. J.A. Biglou and J.G. Lenard, "Physical Simulation of Hot Rolling Processes", Proc. 36th Mechanical Working and Steel Processing Conference, ISS, vol. 32, Oct. 1994, Baltimore, MD, pp. 439-445

133. T.M. Maccagno and J.J. Jonas, "Correcting for the Effects of Static and Metadynamic Recrystallization during the Laboratory Simulation of Rod Rolling", ISIJ Int., vol. 34, No. 7, 1994, pp. 607-614

134. S. Shida, "Effects of Carbon Content, Temperature, and Strain Rate on Compression Flow Stress of Carbon Steel", Hitachi Res. Lab. Report, 1974, pp1-9.

135. C.M. Sellars, "Modelling Microstructural Development during Hot Rolling", Material Science and technology, vol. 6, (1990), pp. 1072-1081

136. C. Rossard, "Microstructure and Design of Alloys", Metals and Iron Steel Ins. London, vol. II,, 1974, pp. 175-182

137. J. Alexander and H. Ford, "Simplified Hot-Rolling Calculations", Journal of the Institute of Metals, vol. 83, 1954, pp 80-f.

138. R. B. Sims, Proc. Inst. Mech. Engrs, vol. 168, 1954, pp. 191-205.

139. Tajima and J.G. Lenard, "Thermo-mechanical Treatment of a High Nb-High V Bearing Micro_alloyed Steel", ISIJ Int., vol. 35, No. 12, 1995, pp. 1509-1517.

140. F. Garofalo, Trans. Metall. Soc. AIME 227, No. 2, 1963, pp. 351-355.

141. K.P. Rao and E.B. Hawbolt, "Development of Constitutive Relationships Using Compression Testing of a Medium Carbon Steel", J. Eng. Mat. Tech., ASME Trans. Vol. 114, Jan. 1992, pp. 116-123.