UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria Elettrica

scuola di dottorato in ingegneria industriale
indirizzo ingegneria elettrotecnica
ciclo XXIV

Induction hardening of components for the aerospace industry

Direttore della Scuola:
Ch.mo Prof. Paolo Bariani .....

Coordinatore d'indirizzo:
Ch.mo Prof. Giovanni Martinelli .....

Supervisore:
Ch.mo Prof. Fabrizio Dughiero .....

Dottorando:
Alessandro Candeo .....

31 Gennaio 2012

1. Reviewer: prof. Mario Marchesoni

2. Reviewer: prof. Paolo Di Barba

3. Reviewer: ing. Michele Forzan

Day of the defense:

Signature from head of PhD committee:

Abstract

Induction heating has been widely used for heat treating and especially surface hardening in a broad variety of applications, ranging from the automotive to the renewable energy market. However, the lack of precise knowledge about the interrelation between all the concurrent physical phenomena occurring within the part during the heating cycle has restricted its use to mass-production items (mostly gears). The benefits of this technology, which is clean, repeatable, and cost-effective, could boost its introduction into more conservative industry sectors, such as aerospace, where furnace-based treatments (e.g., carburizing) represent the golden standard. The major limitation is related to the optimization of the induction hardening process, which usually requires significant material know-how and can thus be very long and expensive. Computer simulation could provide a general tool for understanding and improving the critical aspects of each step of the process, thus speeding up the spreading of the induction technology into new markets.

Sommario

Il riscaldamento a induzione è stato diffusamente impiegato nel settore dei trattamenti termici di componenti per i settori automobilistico ed eolico, in particolare per la tempra di indurimento in una varietà di applicazioni. I principali vantaggi legati all'utilizzo di questa tecnologia risiedono nell'elevato grado di ripetibilità ottenibile nel prodotto trattato, unitamente alla elevata velocità ed automazione di trattamento, fattori entrambi in grado di garantire una produzione efficiente e dal ridotto impatto ambientale. Tuttavia, a causa degli elevati requisiti di qualità del prodotto finito e dei ristretti quantitativi prodotti, alcuni settori industriali rimangono ancor oggi legati ai tradizionali processi di cementazione in forno, che risultano piuttosto dispendiosi per la lunga durata del trattamento termico ed il numero di operazioni di rettifica necessarie a valle dello stesso. L'impiego di metodologie di simulazione numerica al calcolatore permette un'accurata messa a punto del processo di tempra a induzione attraverso prove sperimentali mirate, che ne consentano una rapida implementazione a nuovi settori applicativi.

To my family.

Acknowledgements

This research work has been carried out in the period between January 2009 and December 2011 at the Electroheat Laboratory of the Department of Electrical Engineering (recently merged into the Department of Industrial Engineering) in the University of Padova. The experimental activity has been performed in the facilities of the Center for Thermal Technology of the Department of Mechanical Engineering in the University of Quebec (Ecole de Technologie Superieure, Montreal, Quebec, Canada).

With these few lines, I would like to acknowledge all the persons who have made this work possible, starting with my supervisor, Prof. Fabrizio Dughiero, who not only gave me the chance to work in this fascinating field, but also supported and motivated me with his challenging spirit. I would like to express my gratitude to Prof. Philippe Bocher of ETS, for offering me the opportunity to spend 6 months with his valuable team, which helped me focusing on the practical aspects of the subject. Among the many who also contributed to this work, I would like to mention Dr. Michele Forzan for sharing his huge experience and Dr. Florent Bridier of ETS for his great teaching attitude. I could not forget all the guys of the LEP and the ICIA: Angelo, Aristide, Cristian, Dario, Elisabetta, Fabio, Francesco, Marco and then Antoine, Benjamin, Christophe, Gregory, Mamy, Meysam and all those I might have unintentionally excluded.

Last but not least, a special thank goes to my family for supporting me over the years.

Contents

List of Figures

LIST OF FIGURES

LIST OF FIGURES

List of Tables

LIST OF TABLES

Glossary

BCC Body-Centered Cubic; crystal structure typical of Fe-α phase (ferrite).

CCT Continuous-Cooling-Transformation; non-isothermal diagrams illustrating phase decomposition at different cooling rates.

CHT Continuous-Heating-Transformation; non-isothermal diagrams illustrating austenite formation at different cooling rates.

FCC Face-Centered Cubic; crystal structure typical of Fe-γ phase (austenite).

FEM Finite Element Method; numerical technique for solving PDEs.

GPU Graphics Processing Unit; specialised circuit designed to do general purpose scientific and engineering computing.

HAZ Heat Affected Zone; area of base material which has had its micro-structure and properties altered by a heat treatment.

IGBT Insulated-Gate Bipolar Transistor; power semiconductor device primarily used as an electronic switch.

IT Isothermal-Transformation; same as TTT.

ITh Isothermal-Transformation (on heating); same as IT but for austenite formation.

LMP Larson-Miller Parameter; fitting parameter for describing the stress relieving behaviour of steels.

MEMS Micro-Electro-Mechanical Systems; very small mechanical devices driven by electricity.

MOSFET Metal-Oxide Semiconductor Field-Effect Transistor; transistor used for amplifying or switching electronic signals.

PAG Poly-Alkylene Glycol; class of polymer-based quenching fluids often used for hardening applications.

PDE Partial Differential Equation; type of differential equation.

GLOSSARY

SCR Silicon-Controlled Rectifier; four-layer solid state device that controls current.

TERSA Thermal Evaluation for Residual Stress Analysis; residual stress measurement technique.

TP Tempering Parameter; parameter used to describe the tempering behaviour of steels, according to the Hollomon-Jaffe equation.

TTT Time-Temperature-Transformation; isothermal diagrams illustrating characteristic transformation times at the given temperature (aka Isothermal-Transformation, IT diagrams).

XRD X-Ray Diffraction; residual stress measurement technique.

1

Introduction

> "... [When the steel pieces] are very hot and almost of a white colour because of the heat, in order that the heat may be quickly quenched, they are suddenly thrown into a current of water that is as cold as possible... In this way, the steel is made so hard that it surpasses almost every other hard thing..."
>
> Vannoccio Biringuccio, *de la Pirotechnia*, 1540.

1.1 Historical notes

The use of heat to affect steel properties, the so-called heat treatment, is one of the oldest of all technologies, dating back to the Iron Age. Though, it was not until the end of the 19^{th} and the begin of the 20^{th} century that this process came to be systematically investigated and its scientific backgrounds fully understood [1]. The first treatise on metalworking technology (see Figure 1.1) was published in the 16^{th} century by Vannoccio Biringuccio[1], an Italian metallurgist considered by some as the father of the foundry industry [2].

As people have been dealing with this technique for thousands of years, hardening high-carbon steel is rather simple: just heat the metal to a cherry-red colour and suddenly quench it into cold water, thus obtaining a considerable hardness increase. However, total control of the technique and full knowledge of the physical phenomena

[1] 1480-1539

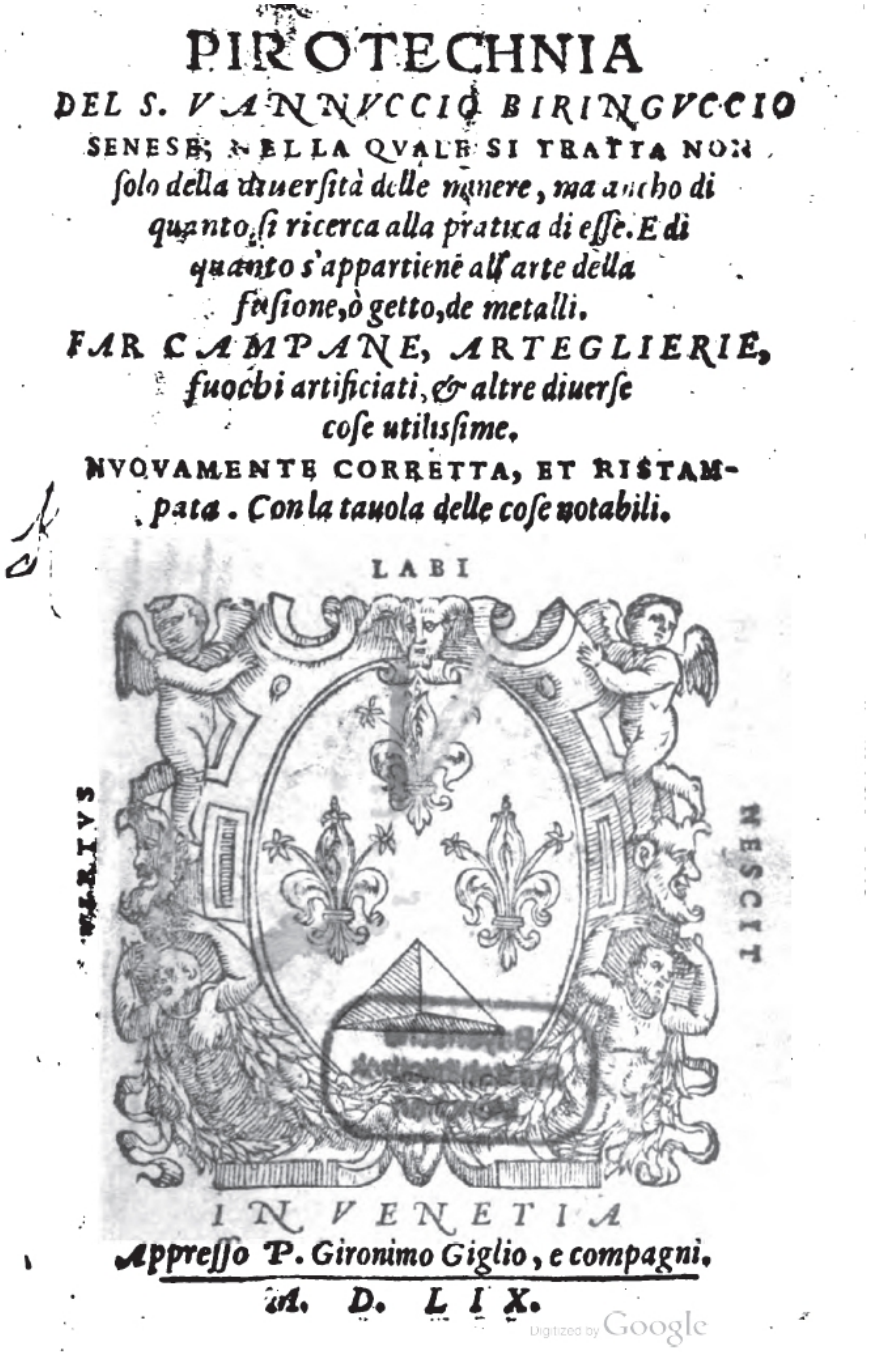

PIROTECHNIA
DEL S. VANNVCCIO BIRINGVCCIO
SENESE; NELLA QVALE SI TRATTA NON
solo della diuersità delle minere, ma anche di
quanto si ricerca alla pratica di esse. E di
quanto s'appartiene all'arte della
fusione, ò getto, de metalli.
FAR CAMPANE, ARTEGLIERIE,
fuochi artificiati, & altre diuerse
cose utilissime.
NVOVAMENTE CORRETTA, ET RISTAM-
pata. Con la tauola delle cose notabili.
LABI
NESCIT
IN VENETIA
Appresso P. Gironimo Giglio, e compagni.
M. D. L I X.

Figure 1.1: de la Pirotechnia - Front matter of the first treatise on metalworking technology by Vannoccio Biringuccio.

occurring during the process, have still not been achieved, particularly for the case of contact-less heating.

Heating technology without contact had been known for more than hundred years. It comprises a wide field of applications in metalworking and in the metal-casting industry because of its outstanding benefits. At present time induction heating and melting processes are commonly used all over the world. The progress in electro-heat became possible only after the discovery of the dynamo-electric principle and the availability of technically usable dynamo-machines in the second half of the 19^{th} century. At that time, the development of induction melting started, followed later by the high-frequency[1] heating of metallic bodies for hot-forming and heat-treating.

After the successful introduction of induction metal melting to the metallurgical industry and its general acceptance by the foundry industry, the attention of scientists and engineers was turned to induction heating technologies. Induction heat treating mainly includes the applications of surface hardening, through hardening, tempering, stress relieving, and annealing. Bar and billet heating for hot forming, reheating of strips, slabs, blooms, and plates and heating of wires, cables, rods as well as some other applications belong to induction mass heating. Although the possibility of using high frequencies for heating was already known, the absence of viable technical solutions able to support this idea caused a big delay in the practical implementation of such kind of applications. In fact, the introduction of surface hardening to the industry was initially slow because of a lack of powerful high frequency generators, insufficient knowledge about the metallurgical results of hardening and small number of successful industrial applications.

When appropriate medium and high frequency generators with acceptable power output became available, the activities in the various countries concentrated on induction heat treating and induction mass heating. Driven by the development of high frequency generators for radio communications, the situation dramatically changed in the early 1930s, when induction hardening of crankshafts was developed. Since this turning point, the number of experimental works and publications on induction hardening and other applications grew exponentially. In turn, this success stimulated production of new more powerful motor-generators and vacuum tube oscillators. Multiple surface

[1]The term "high frequency" has always been an unclear concept as it has been evolving with time and technology advancement.

hardening technologies were developed for a variety of parts including camshafts, gears, sprockets, railroad rails and engine cylinder liners [3].

2

The Heat Treatment of Steel

2.1 Principles of heat treating of steels

A steel is usually defined as an alloy of iron and carbon with the carbon content between a few hundreds of a percent up to about 2 wt%. Other alloying elements can amount in total to about 5 wt% in low-alloy steels and higher in more highly alloyed steels such as tool steels and stainless steels. Steels can exhibit a wide variety of properties depending on composition as well as the phases and micro-constituents present, which in turn depend on the heat treatment.

2.1.1 The Fe-C phase diagram

The basis for the understanding of the heat treatment of steels is the Fe-C phase diagram).

Figure 2.1 actually shows two diagrams; the stable iron-graphite diagram (dashed lines) and the metastable Fe-Fe_3C diagram. The stable condition usually takes a very long time to develop, especially in the low-temperature and low-carbon range, and therefore the metastable diagram is of more interest. The Fe-C diagram shows which phases are to be expected at equilibrium (or metastable equilibrium) for different combinations of carbon concentration and temperature.

We distinguish at the low-carbon end ferrite (α-iron), which can at most dissolve 0.028 wt% C at 727 °C (1341 °F) and austenite (γ-iron), which can dissolve 2.11 wt% C at 1148 °C (2098 °F). At the carbon-rich side we find cementite (Fe_3C). Of less interest, except for highly alloyed steels, is the δ-ferrite existing at the highest temperatures.

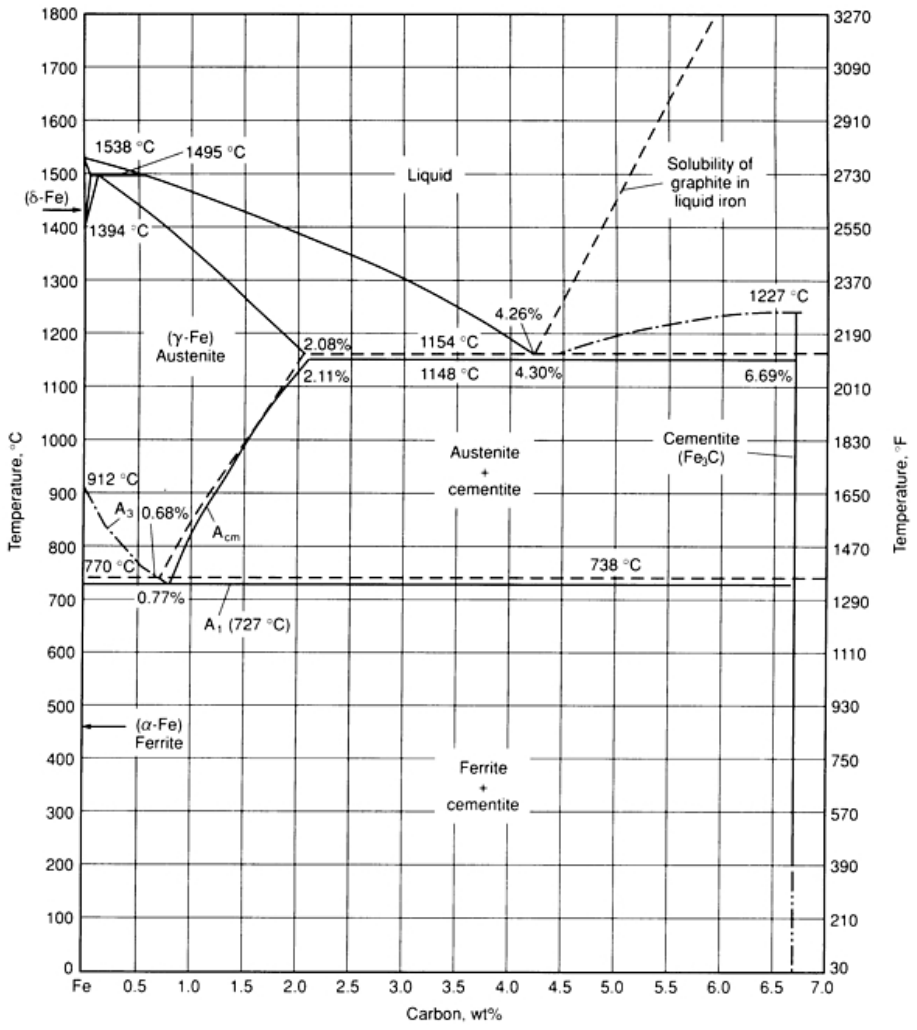


Figure 2.1: The Fe-C equilibrium diagram up to 6.67 wt% C - Solid lines indicate Fe-Fe_3C diagram; dashed lines indicate iron-graphite diagram [4].

Between the single-phase fields are found regions with mixtures of two phases, such as ferrite + cementite, austenite + cementite, and ferrite + austenite. At the highest temperatures, the liquid phase field can be found and below this are the two phase fields liquid + austenite, liquid + cementite, and liquid + δ-ferrite. In heat treating of steels, the liquid phase is always avoided. Some important boundaries at single-phase fields have been given special names that facilitate the discussion. These include:

- A_1, the so-called eutectoid temperature, which is the minimum temperature for austenite;
- A_3, the lower-temperature boundary of the austenite region at low carbon contents, that is, the $\gamma/\gamma + \alpha$ boundary;
- A_{cm}, the counterpart boundary for high carbon contents, that is, the $\gamma/\gamma + Fe_4C$ boundary.

Sometimes the letters c, e, or r are included.

The carbon content at which the minimum austenite temperature is attained is called the eutectoid carbon content (0.77 wt% C). The ferrite-cementite phase mixture of this composition formed during cooling has a characteristic appearance and is called pearlite and can be treated as a micro-structural entity or micro-constituent. It is an aggregate of alternating ferrite and cementite lamellæ that degenerates ("spheroidizes" or "coarsens") into cementite particles dispersed with a ferrite matrix after extended holding close to A_1.

The Fe-C equilibrium diagram is of experimental origin. The knowledge of the thermodynamic principles and modern thermodynamic data now permits very accurate calculations of this diagram. This is particularly useful when phase boundaries must be extrapolated and at low temperatures where the experimental equilibria are extremely slow to develop. If alloying elements are added to the iron-carbon alloy (steel), the position of the A_1, A_3, and A_{cm} boundaries and the eutectoid composition are changed.

2.1.2 Transformation diagrams

The kinetic aspects of phase transformations are as important as the equilibrium diagrams for the heat treatment of steels. The metastable phase martensite and the

morphologically metastable micro-constituent bainite, which are of extreme importance to the properties of steels, can generally form with comparatively rapid cooling to ambient temperature, that is, when the diffusion of carbon and alloying elements is suppressed or limited to a very short range. Bainite is a eutectoid decomposition that is a mixture of ferrite and cementite. Martensite, the hardest constituent, forms during severe quenches from supersaturated austenite by a shear transformation. Its hardness increases monotonically with carbon content up to about 0.7 wt%. If these unstable, metastable products are subsequently heated to a moderately elevated temperature, they decompose to more stable distributions of ferrite and carbide. The reheating process is sometimes known as tempering or annealing. The transformation of an ambient temperature structure like ferrite-pearlite or tempered martensite to the elevated temperature structure of austenite or austenite + carbide is also of importance in the heat treatment of steel. One can conveniently describe what is happening during transformation with transformation diagrams. Four different types of such diagrams can be distinguished. These include:

1. isothermal-transformation diagrams describing the formation of austenite, which will be referred to as ITh diagrams;
2. isothermal-transformation (IT) diagrams, also referred to as time-temperature-transformation (TTT) diagrams, describing the decomposition of austenite;
3. continuous-heating-transformation (CHT) diagrams;
4. continuous-cooling-transformation (CCT) diagrams.

Isothermal transformation diagrams

This type of diagram shows what happens when a steel is held at a constant temperature for a prolonged period. The development of the micro-structure with time can be followed by holding small specimens in a lead or salt bath and quenching them one at a time after increasing holding times and measuring the amount of phases formed in the micro-structure with the aid of a microscope. An alternative method involves using a single specimen and a dilatometer which records the elongation of the specimen as a function of time. The basis for the dilatometer method is that the micro-constituents undergo different volumetric changes, as listed in Table 2.1.

Table 2.1: Volume changes due to different transformations.

Transformation	Linear volume change
Spheroidized pearlite→Austenite	4.64-2.21×(%C)
Austenite→Martensite	4.64-0.53×(%C)
Austenite→Lower bainite	4.64-1.43×(%C)
Austenite→Upper bainite	4.64-2.21×(%C)

ITh diagrams (formation of austenite) During the formation of austenite from an original micro-structure of ferrite and pearlite or tempered martensite, the volume (and hence the length) decreases with the formation of the dense austenite phase. From the elongation curves, the start and finish times for austenite formation, usually defined as 1% and 99% transformation, respectively, can be derived. These times are then conveniently plotted on a temperature-log time diagram (see Figure 2.2). Also plotted in this diagram are the Ac_1 and Ac_3 temperatures. Below Ac_1 no austenite can form, and between Ac_1 and Ac_3 the end product is a mixture of ferrite and austenite. Notice that a considerable overheating is required to complete the transformation in a short time. The original micro-structure also plays a great role. A finely distributed structure like tempered martensite is more rapidly transformed to austenite than, for instance, a ferritic-pearlitic structure. This is particularly true for alloyed steels with carbide-forming alloying elements such as chromium and molybdenum. It is important that the heating rate to the hold temperature be very high if a true isothermal diagram is to be obtained.

IT diagrams (decomposition of austenite) The procedure starts at a high temperature, normally in the austenitic range after holding there long enough to obtain homogeneous austenite without undissolved carbides, followed by rapid cooling to the desired hold temperature. An example of an IT diagram is given in Figure 2.3. The cooling was started from 880 °C (1616 °F). The A_1 and A_3 temperatures are indicated as well as the hardness. Above A_3 no transformation can occur. Between A_1 and A_3 only ferrite can form from austenite. Despite some diagrams also depict intermediate fractions with dashed lines, normally only the 1% and 99% curves are reproduced. Notice that the curves are C-shaped. This is typical for transformation curves. A higher-temperature set of C-shaped curves shows the transformation to pearlite and

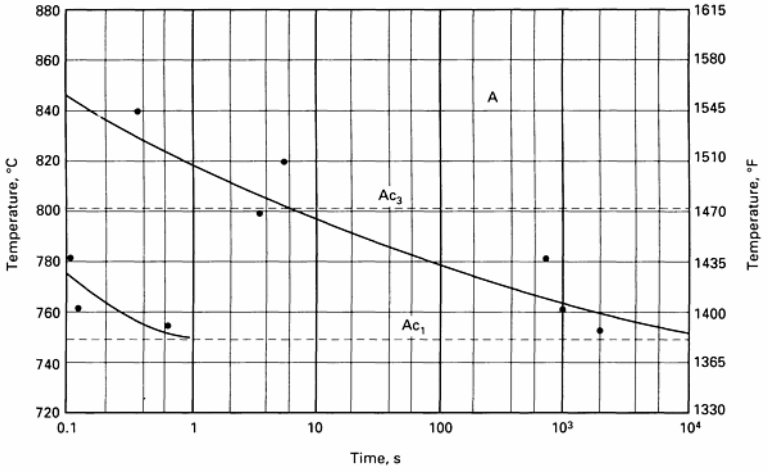


Figure 2.2: ITh diagram - Isothermal heating diagram for AISI 4140 steel. Heating rate to reach holding temperature is 1020 °C/s (1835 °F). Between Ac_3 and Ac_1, the final structure is a mixture of austenite and ferrite [5].

a lower temperature set indicates the transformation to bainite. In between is found a so-called austenite bay, common for certain low-alloy steels containing appreciable amounts of carbide-forming alloying elements such as chromium or molybdenum.

Continuous transformation diagrams

In practical heat treatment situations, a constant temperature is not required, but rather a continuous changing temperature during either cooling or heating. Therefore, more directly applicable information is obtained if the diagram is constructed from dilatometric data using a continuously increasing or decreasing temperature.

CHT diagrams Figure 2.4 shows an example for continuous heating of the same steel as in Figure 2.2. Indicated again are Ac_1 and Ac_3 and the same remarks as before are valid. The diagram was derived with constant heating rates and curves for 130 °C/s (265 °F/s), 10 °C/s (50 °F/s), and 0.2 °C/s (32 °F/s) are shown. Notice that the start and finish of the transformations are delayed relative to the isothermal diagram. This is generally true when one compares an isothermal and a continuous diagram, regardless of whether they are for heating or cooling. Like the ITh diagrams, the CHT diagrams are useful in predicting the effect of short-time austenitization that occurs in induction and laser hardening. One typical question is how high the maximum

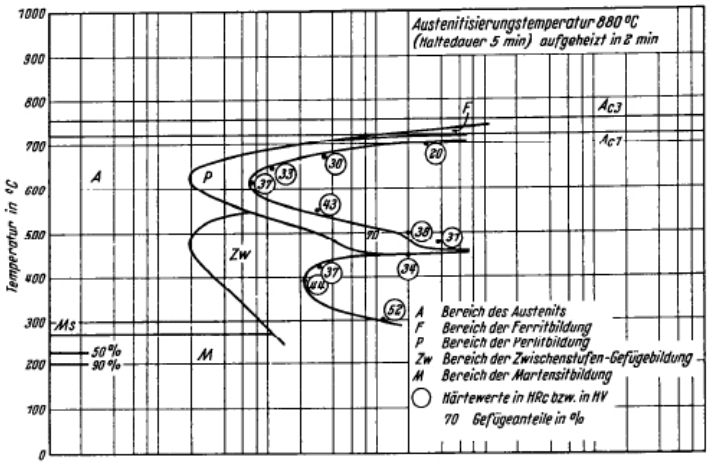


Figure 2.3: IT diagram - Isothermal cooling diagram for AISI 6150 steel austenitized at 880 °C for 2 min. The upper C-shaped curves describe transformation to pearlite; the lower C-shaped curves to bainite. Ferrite is not visible. [6].

surface temperature should be in order to achieve complete austenitization for a given heating rate. Too high a temperature may cause unwanted austenite grain growth, which produces a more brittle martensitic micro-structure.

CCT diagrams As for heating diagrams, it is important to clearly state what type of cooling curve the transformation diagram was derived from. Use of a constant cooling rate is very common in experimental practice. However, this regime rarely occurs in a practical situation. In Figure 2.5 is shown a CCT diagram for an AISI 6150 steel. Ferrite, pearlite, and bainite regions are indicated as well as the M_s temperature. Note that the M_s temperature is not constant when martensite formation is preceded by bainite formation, but typically decreases with longer times.

2.1.3 Hardenability concepts

The goal of heat treatment of steel is very often to attain a satisfactory hardness. The important micro-structural phase is then normally martensite, which is the hardest constituent in low-alloy steels. The hardness of martensite is primarily dependent on its carbon content as is shown in Figure 2.6. If the micro-structure is not fully martensitic, its hardness is lower. In practical heat treatment, it is important to achieve full hardness

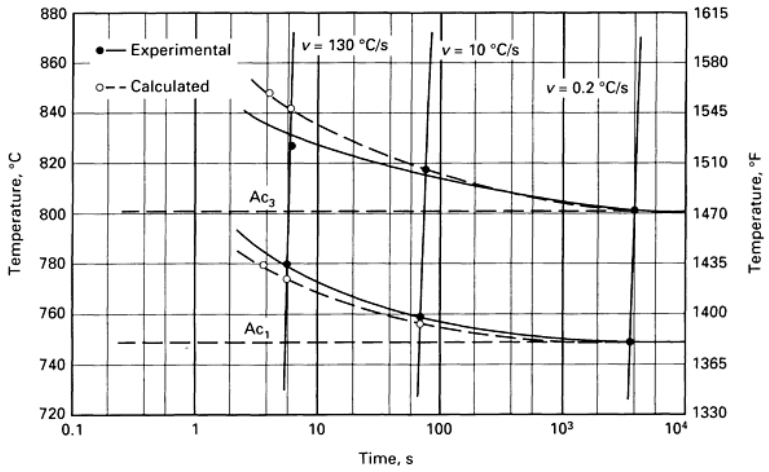


Figure 2.4: CHT diagram - Continuous heating transformation diagram for AISI 4140 steel. The phase being formed is austenite [5].

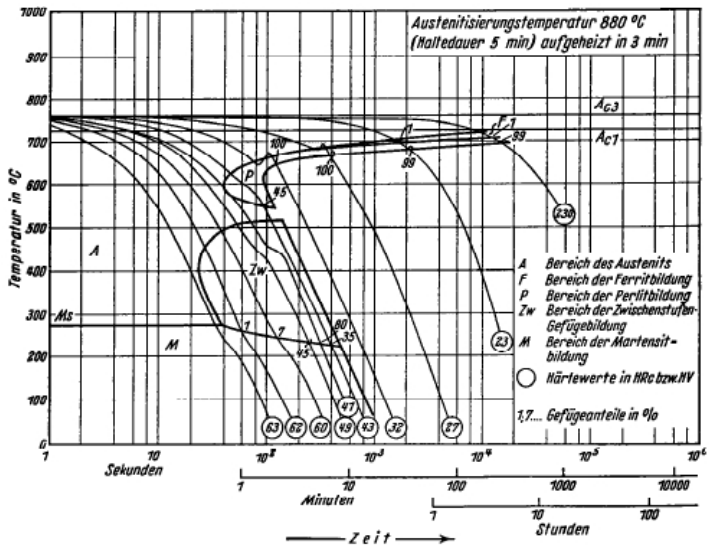


Figure 2.5: CCT diagram - Continuous cooling transformation diagram for AISI 6150 steel austenitized at 880 °C for 3 min [6].

to a certain minimum depth after cooling, that is, to obtain a fully martensitic microstructure to a certain minimum depth, which also represents a critical cooling rate. If a given steel does not permit a martensitic structure to be formed to this depth, one has to choose another steel with a higher hardenability (the possibility of increasing the cooling rate at the minimum depth will be discussed later). There are various ways to characterize the hardenability of a steel. The CCT diagram can serve this purpose if one knows the cooling rate at the minimum depth.

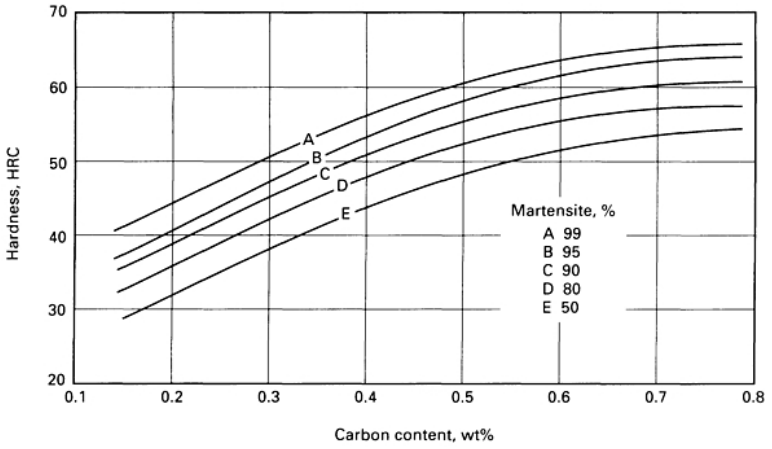


Figure 2.6: Hardenability of steel - Relationship between hardness, carbon content, and amount of martensite.

Jominy end-quench test The most commonly used experimental method for hardenability is the well-known Jominy test [7]. For this test a round bar specimen that is 100 mm in length and 25 mm in diameter is used. The specimen is heated to the austenitizing temperature of the steel with a holding time of 20 min. One end face of the specimen is quenched by spraying it with a jet of water (see Figure 2.7). This causes the rate of cooling to decrease progressively from the quenched end along the length of the bar. When it is cool, two diametrically opposite flats that are 0.4 mm deep and parallel to the axis of the bar, are ground and the hardness is measured along the flats. The hardness values are plotted on a diagram at specified intervals from the quenched end (see Figure 2.8). To get good reproducibility, the time and temperature of austenitizing, the grinding of the flats to avoid burning, and the placement of the specimen in the hardness tester should be carefully controlled. It is also important to

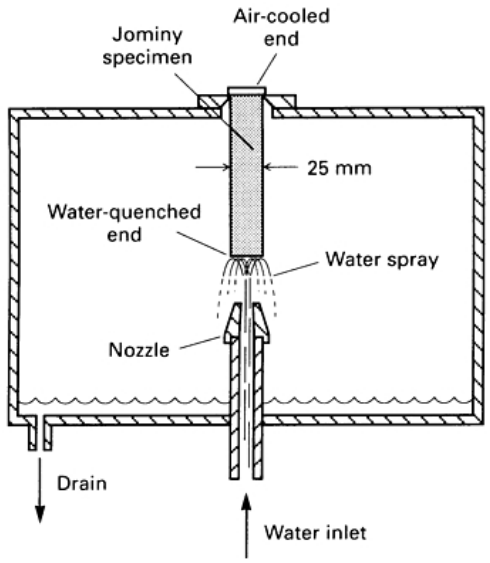


Figure 2.7: Jominy test layout - Hardening procedure for Jominy end-quench specimen.

protect the specimen against decarburization. Less critical are the water temperature, the diameter of the spraying nozzle, the height of the water jet, and the time to move the specimen from the furnace to the fixture. The rate of cooling at different dis-

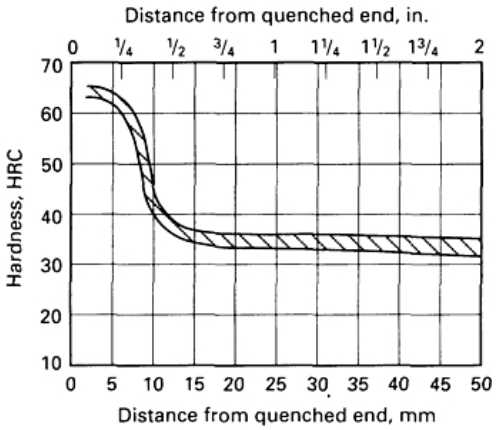


Figure 2.8: Jominy curves - Jominy hardenability curves as result of multiple testing test on one heat of AISI 4068 steel.

tances from the quenched end is approximately independent of the steel used because the thermal conductivity and heat capacity of hardenable steels do not vary very much and the heat transfer at the cooled end is steel independent. Therefore, the Jominy

bar presents a range of cooling curves that can be used to estimate a CCT diagram. The cooling rates are not linear but rather of the natural type according to Newton's law of cooling. The volume fractions of the various constituents of the micro-structure are evaluated quantitatively at different Jominy distances as well as hardness measurements. Conversely, the Jominy curve can be calculated from the TTT diagram using computer methods and expressions that give the hardness for the different phases [8]. The total hardness is then the weighted average. There is an economic trend to reduce the amount of experimental Jominy testing and replace it by the calculation of Jominy curves from the chemical composition and the grain size. Several regression formulas have been developed for different grades of steel which are quite accurate for a limited composition range.

2.2 General overview and classification

2.2.1 Stress-relief heat treating

Stress-relief heat treating is used to relieve stresses that remain locked in a structure as a consequence of a manufacturing sequence. This is performed by uniform heating of a structure to a suitable temperature below the transformation range (Ac_1 for ferritic steels), holding at this temperature for a predetermined period of time, followed by uniform cooling. Care must be taken to ensure uniform cooling, particularly in presence of variable section sizes. If the rate of cooling is not constant and uniform, new residual stresses can result that are equal to or greater than those that the heat-treating process was intended to relieve. Stress-relief heat treating can reduce distortion and high stresses from welding that can affect service performance.

Generally speaking, a number of factors influence the relief of residual stresses, including level of stress, permissible (or practicable) time for their relief, temperature, and metallurgical stability. When coming to Time-Temperature Factors. The relief of residual stresses is a time-temperature-related phenomenon, parametrically correlated by the Larson-Miller parameter (LMP):

$$LMP = T \cdot \left(C_{LMP} + \log_{10} \left(\frac{t}{3600} \right) \right) \tag{2.1}$$

where C_{LMP} is a material constant (C_{LMP}=20 for metals), T is temperature (K) and t is the time (s). Such parameter can be used practically in order to fit experimental

stress-to-rupture data with respect to time, thus obtaining the stress relief behaviour of the given steel (see Figure 2.9).

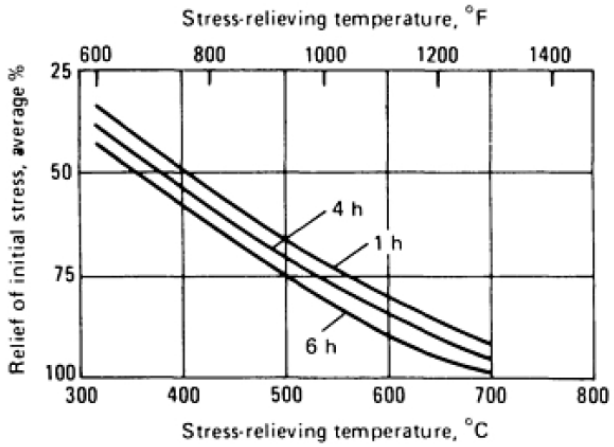
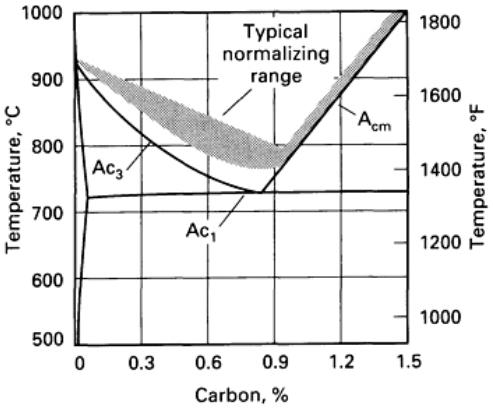


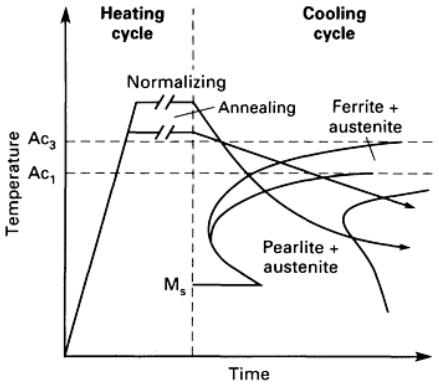
Figure 2.9: Larson-Miller relation - Relationship between time and temperature in the relief of residual stresses in steel.

2.2.2 Normalizing

Normalizing of steel is a heat-treating process that is often considered from both thermal and micro-structural standpoints. In the thermal sense, normalizing is an austenitizing heating cycle followed by cooling in still or slightly agitated air. Typically, the work-piece is heated to a temperature about 55 °C (100 °F) above the upper critical line of the iron-iron carbide phase diagram (i.e., above Ac_3 for hypo-eutectoid steels and above Ac_{cm} for hyper-eutectoid steels), as shown in Figure 2.10a. To be properly classed as a normalizing treatment, the heating portion of the process must produce a homogeneous austenitic phase (face-centered cubic, FCC, crystal structure) prior to cooling. Figure 2.10b compares the time-temperature cycle of normalizing to that of full annealing: the slower cooling of annealing results in higher temperature transformation to ferrite and pearlite and coarser micro-structures than does normalizing [9]. Normalizing is also frequently thought of in terms of micro-structure (see Figure 2.11). The areas of the micro-structure that contain about 0.8% C are pearlitic (lamellæ of ferrite and iron carbide). The areas that are low in carbon are ferritic (body-centered cubic, BCC, crystal structure). In hyper-eutectoid steels, pro-eutectoid



(a) Partial iron-iron carbide phase diagram showing typical normalizing range for plain carbon steels.



(b) Comparison of time-temperature cycles for normalizing and full annealing.

Figure 2.10: Normalizing - Typical range and cycle.

iron carbide first forms along austenite grain boundaries. This transformation continues until the carbon level in the austenite reaches approximately 0.8%, at which time a eutectoid reaction begins as indicated by the formation of pearlite. Air-hardening steels are excluded from the class of normalized steels because they do not exhibit the normal pearlitic micro-structure that characterizes normalized steels.

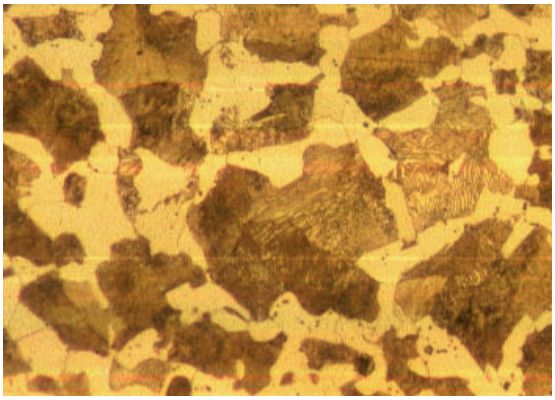


Figure 2.11: Normalized micro-structure - A typical medium carbon (0.3% C) engineering steel showing a micro-structure of ferrite (light areas) and pearlite (dark areas) [10].

The purpose of normalizing varies considerably. Normalization may increase or decrease the strength and hardness of a given steel in a given product form, depending on the thermal and mechanical history of the product. Actually, the functions of

normalizing may overlap with or be confused with those of annealing, hardening, and stress relieving. Improved machinability, grain-structure refinement, homogenization, and modification of residual stresses are among the reasons normalizing is done. Homogenization of castings by normalizing may be done in order to break up or refine the dendritic structure and facilitate a more even response to subsequent hardening. Similarly, for wrought products, normalization can help reduce banded grain structure due to hot rolling, as well as large grain size or mixed large and small grain size due to forging practice.

Depending on the mechanical properties required, normalizing may be substituted for conventional hardening when the size or shape of the part is such that liquid quenching might result in cracking, distortion, or excessive dimensional changes. Thus, parts that are of complex shape or that incorporate sharp changes in section may be normalized and tempered, provided that the properties obtained are acceptable. The rate of heating generally is not critical for normalizing; on an atomic scale, it is immaterial. In parts having great variations in section size, however, thermal stress can cause distortion. Time at temperature is critical only in that it must be sufficient to cause homogenization. Sufficient time must be allowed for solution of thermodynamically stable carbides, or for diffusion of constituent atoms. Generally, time sufficient for complete austenitization is all that is required. In cases where normalizing is done to homogenize segregated structures, longer times may be required. The rate of cooling significantly influences both the amount of pearlite and the size and spacing of the pearlite lamellæ. At higher cooling rates, more pearlite forms, and the lamellæ are finer and more closely spaced. The increased amount of fine-grained pearlite results in higher strength and higher hardness. Conversely, lower cooling rates result in softer parts. In any part having both thick and thin sections, the potential exists for variations in cooling rate, and thus for variations in strength and hardness as well. This can also increase the probability of distortion or even cracking. Cooling rate sometimes is enhanced with fans to increase strength and hardness of parts or to decrease the time required, following the furnace operation, for sufficient cooling of parts to permit convenient handling.

After parts have cooled uniformly through their cross section to black heat below Ar_1 (the parts are no longer red, as when they were removed from the furnace), they may be water or oil quenched to decrease the total cooling time. In heavy sections, cooling of the center material to black heat may require considerable time. Thermal

shock, residual thermally induced stress, and resultant distortions are factors to be considered. The micro-structure remains essentially unaffected by the increased cooling rate, provided that the entire mass is below the lower critical temperature Ar_1, although changes involving precipitates may occur.

2.2.3 Annealing

Annealing is a generic term denoting a treatment that consists of heating to and holding at a suitable temperature followed by cooling at an appropriate rate, primarily for the softening of metallic materials. Generally, in plain carbon steels, annealing produces a ferrite-pearlite micro-structure (see Figure 2.12). Steels may be annealed to facilitate cold working or machining, to improve mechanical or electrical properties, or to promote dimensional stability. The choice of an annealing treatment that will provide an adequate combination of such properties at minimum expense often involves a compromise. Terms used to denote specific types of annealing applied to steels are descriptive of the method used, the equipment used, or the condition of the material after treatment.

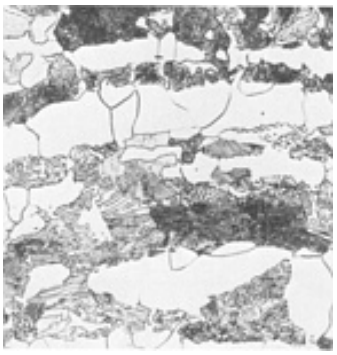


Figure 2.12: Annealed micro-structure - A fully annealed 1040 steel showing a ferrite-pearlite micro-structure [4].

Annealing cycles

In practice, specific thermal cycles of an almost infinite variety are used to achieve the various goals of annealing. These cycles fall into several broad categories that can be classified according to the temperature to which the steel is heated and the method of cooling used. The maximum temperature may be below the lower critical temperature,

A_1 (*subcritical* annealing); above A_1 but below the upper critical temperature, A_3 in hypo-eutectoid steels, or A_{cm} in hyper-eutectoid steels (*intercritical* annealing); or above A_3 (*full* annealing).

Because some austenite is present at temperatures above A_1, cooling practice through transformation is a crucial factor in achieving desired micro-structure and properties. Accordingly, steels heated above A_1 are subjected either to slow continuous cooling or to isothermal treatment at some temperature below A_1 at which transformation to the desired micro-structure can occur in a reasonable amount of time. Under certain conditions, two or more such cycles may be combined or used in succession to achieve the desired results. The success of any annealing operation depends on the proper choice and control of the thermal cycle, based on the metallurgical principles discussed in the following sections.

Subcritical Annealing Subcritical annealing does not involve formation of austenite. The prior condition of the steel is modified by such thermally activated processes as recovery, recrystallization, grain growth, and agglomeration of carbides. The prior history of the steel is, therefore, an important factor. In as-rolled or forged hypo-eutectoid steels containing ferrite and pearlite, subcritical annealing can adjust the hardnesses of both constituents, but excessively long times at temperature may be required for substantial softening. The subcritical treatment is most effective when applied to hardened or cold-worked steels, which recrystallize readily to form new ferrite grains. The rate of softening increases rapidly as the annealing temperature approaches A_1. Cooling practice from the subcritical annealing temperature has very little effect on the established micro-structure and resultant properties.

lntercritical Annealing Austenite begins to form when the temperature of the steel exceeds A_1. The solubility of carbon increases abruptly (nearly 1%) near the A_1 temperature. In hypo-eutectoid steels, the equilibrium structure in the intercritical range between A_1 and A_3 consists of ferrite and austenite, and above A_3 the structure becomes completely austenitic. However, the equilibrium mixture of ferrite and austenite is not achieved instantaneously. Undissolved carbides may persist, especially if the austenitizing time is short or the temperature is near A_1 causing the austenite to be

inhomogeneous. In hyper-eutectoid steels, carbide and austenite coexist in the intercritical range between A_1 and A_{cm}; and the homogeneity of the austenite depends on time and temperature. The degree of homogeneity in the structure at the austenitizing temperature is an important consideration in the development of annealed structures and properties. The more homogeneous structures developed at higher austenitizing temperatures tend to promote lamellar carbide structures on cooling, whereas lower austenitizing temperatures in the intercritical range result in less homogeneous austenite, which promotes formation of spheroidal carbides.

Supercritical or Full Annealing A common annealing practice is to heat hypo-eutectoid steels above the upper critical temperature (A_3) to attain full austenitization. The process is called full annealing. In hypo-eutectoid steels (under 0.77% C), supercritical annealing (i.e., above the A_3 temperature) takes place in the austenite region (the steel is fully austenitic at the annealing temperature). However, in hyper-eutectoid steels (above 0.77% C), the annealing takes place above the A_1 temperature, which is the dual-phase austenite-cementite region. In general, an annealing temperature 50 °C (90 °F) above the A_3 for hypo-eutectic steels and A_1 for hyper-eutectoid steels is adequate.

2.2.4 Quenching

Quenching refers to the process of rapidly cooling metal parts from the austenitizing or solution treating temperature, typically from within the range of 815 to 870 °C (1500 to 1600 °F) for steel. Stainless and high-alloy steels may be quenched to minimize the presence of grain boundary carbides or to improve the ferrite distribution but most steels including carbon, low-alloy, and tool steels, are quenched to produce controlled amounts of martensite in the micro-structure (see Figure 2.13). Successful hardening usually means achieving the required micro-structure, hardness, strength, or toughness while minimizing residual stress, distortion, and the possibility of cracking. The selection of a quenchant medium depends on the hardenability of the particular alloy, the section thickness and shape involved, and the cooling rates needed to achieve the desired micro-structure. The most common quenchant media are either liquids or gases. The liquid quenchants commonly used include:

- oil that may contain a variety of additives;

- water;
- aqueous polymer solutions;
- water that may contain salt or caustic additives.

The most common gaseous quenchants are inert gases including helium, argon, and nitrogen. These quenchants are sometimes used after austenitizing in a vacuum. The ability of a quenchant to harden steel depends on the cooling characteristics of the quenching medium. Quenching effectiveness is dependent on the steel composition, type of quenchant, or the quenchant use conditions. The design of the quenching system and the thoroughness with which the system is maintained also contribute to the success of the process.

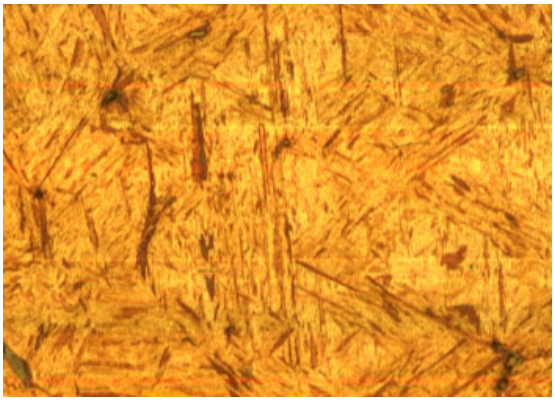


Figure 2.13: Quenched micro-structure - A water quenched steel showing a typical martensitic micro-structure [10].

Quenching processes

Fundamentally, the objective of the quenching process is to cool steel from the austenitizing temperature sufficiently quickly to form the desired micro-structural phases, sometimes bainite but more often martensite. The basic quenchant function is to control the rate of heat transfer from the surface of the part being quenched. The rate of heat extraction by a quenching medium and the way it is used substantially affects quenchant performance. Variations in quenching practices have resulted in the assignment of specific names to some quenching techniques:

- direct quenching;
- time quenching;
- selective quenching;
- spray quenching;
- fog quenching;
- interrupted quenching.

Direct quenching refers to quenching directly from the austenitizing temperature and is by far the most widely used practice. The term direct quenching is used to differentiate this type of cycle from more indirect practices which might involve carburizing, slow cooling, reheating, followed by quenching.

Time quenching is used when the cooling rate of the part being quenched needs to be abruptly changed during the cooling cycle. The change in cooling rate may consist of either an increase or a decrease in the cooling rate depending on which is needed to attain desired results. The usual practice is to lower the temperature of the part by quenching in a medium with high heat removal characteristics (e.g., water) until the part has cooled below the nose of the time-temperature-transformation (TTT) curve, and then to transfer the part to a second medium (e.g., oil), so that it cools more slowly through the martensite formation range. In some applications, the second medium may be air or an inert gas. Time quenching is most often used to minimize distortion, cracking, and dimensional changes.

Selective quenching is used when it is desirable for certain areas of a part to be relatively unaffected by the quenching medium. This can be accomplished by insulating an area to be more slowly cooled so the quenchant contacts only those areas of the part that are to be rapidly cooled.

Spray quenching involves directing high-pressure streams of quenching liquid onto areas of the work-piece where higher cooling rates are desired. The cooling rate is faster because the quenchant droplets formed by the high-intensity spray impact the part surface and remove heat very effectively. However, low-pressure spraying, in effect a flood-type flow, is preferred with certain polymer quenchants.

Fog quenching utilizes a fine fog or mist of liquid droplets in a gas carrier as the cooling agent. Although similar to spray quenching, fog quenching produces lower cooling rates because of the relatively low liquid content of the stream.

Interrupted quenching refers to the rapid cooling of the metal from the austenitizing temperature to a point above the M_s where it is held for a specified period of time, followed by cooling in air. There are three types of interrupted quenching: *austempering*, *marquenching* (martempering), and *isothermal quenching*. The temperature at which the quenching is interrupted, the length of time the steel is held at temperature, and the rate of cooling can vary depending on the type of steel and work-piece thickness. Comparisons of direct and interrupted quench cycles are shown in Figure 2.14.

Austempering consists of rapidly cooling the metal part from the austenitizing temperature to about 230 to 400 °C (450 to 750 °F) (depending on the transformation characteristics of the particular steel involved), holding at a constant temperature to allow isothermal transformation, followed by air cooling. Austempering is applicable to most medium-carbon steels and alloy steels. Low-alloy steels are usually restricted to 9.5 mm or thinner sections, while more hardenable steels can be austempered in sections up to 50 mm thick. Molten salt baths are usually the most practical for austempering applications. Oils have been developed that suffice in some cases, but molten salts possess better heat-transfer properties and eliminate the fire hazard.

Marquenching is similar to austempering in that the work-piece is quenched rapidly from the austenitizing range into an agitated bath held near the M_s temperature. It differs from austempering in that the work-piece remains at temperature only long enough for the temperature to be equalized throughout the work-piece.

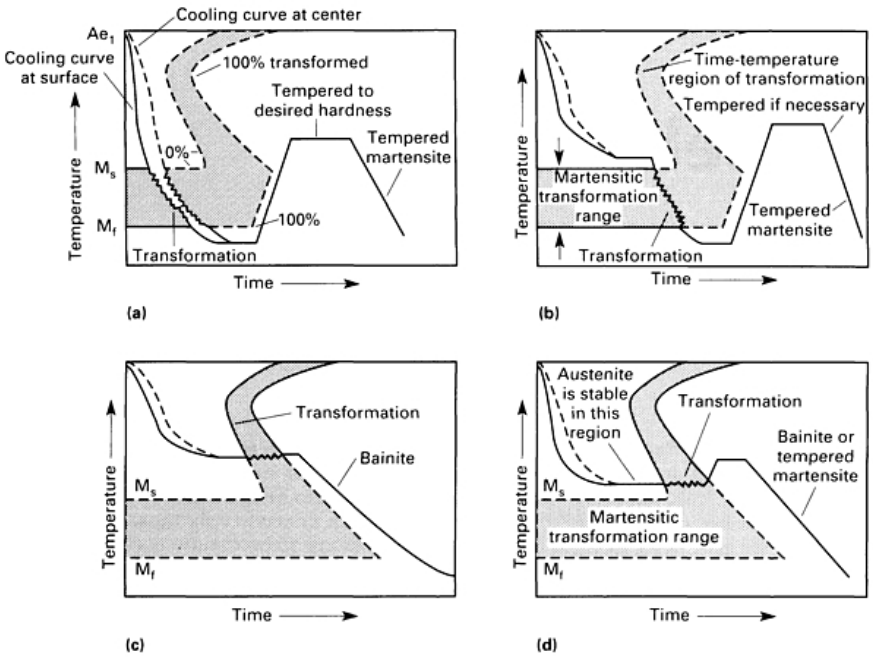


Figure 2.14: Quenching processes - Comparison of cooling rates and temperature gradients as work-pieces pass into and through martensite transformation range. (a) Conventional quenching and tempering processes that use oil, water, or polymer quenchants. (b) Marquenching, which uses either salt or hot oil as a quenchant. (c) Austempering, which uses a salt as a quenchant. (d) Isothermal quenching, which uses either salt or hot oil as a quenchant [11].

When the temperature has attained equilibrium but before transformation begins, the work-piece is removed from the salt bath and air cooled to room temperature. Oils are used successfully for marquenching, but molten salt is usually preferred because of its better heat-transfer properties. Cooling from the marquenching bath to room temperature is usually conducted in still air. Deeper hardening steels are susceptible to cracking while martensite forms if the cooling rate is too rapid. Alloy carburizing steels, which have a soft core, are insensitive to cracking during martensite formation, and the rate of cooling from the M_s temperature is not critical. Marquenching does not remove the necessity for subsequent tempering. The structure of the metal is essentially the same as that formed during direct quenching.

Isothermal quenching is also similar to austempering in that the steel is rapidly quenched through the ferrite and pearlite formation range to a temperature just above M_s. However, isothermal quenching differs from austempering in that two quench baths are employed. After the first quench, and before transformation has time to begin, the work-piece is transferred to a second bath at a somewhat higher temperature where it is isothermally transferred, followed by cooling in air.

2.2.5 Tempering

Tempering of steel is a process in which previously hardened or normalized steel is usually heated to a temperature below the lower critical temperature and cooled at a suitable rate, primarily to increase ductility and toughness, but also to increase the grain size of the matrix. Steels are tempered by re-heating after hardening to obtain specific values of mechanical properties and also to relieve quenching stresses and to ensure dimensional stability. Tempering usually follows quenching from above the upper critical temperature; however, tempering is also used to relieve the stresses and reduce the hardness developed during welding and to relieve stresses induced by forming and machining.

Variables associated with tempering that affect the micro-structure and the mechanical properties of a tempered steel include:

- tempering temperature;

- time at temperature;
- cooling rate from the tempering temperature;
- composition of the steel, including carbon content, alloy content, and residual elements.

In a steel quenched to a micro-structure consisting essentially of martensite, the iron lattice is strained by the carbon atoms, producing the high hardness of quenched steels. Upon heating, the carbon atoms diffuse and react in a series of distinct steps that eventually form Fe_3C or an alloy carbide in a ferrite matrix of gradually decreasing stress level (see Figure 2.15). The properties of the tempered steel are primarily determined by the size, shape, composition, and distribution of the carbides that form, with a relatively minor contribution from solid-solution hardening of the ferrite. These changes in micro-structure usually decrease hardness, tensile strength, and yield strength but increase ductility and toughness. Under certain conditions, hardness may remain unaffected by tempering or may even be increased as a result of it. For example, tempering a hardened steel at very low tempering temperatures may cause no change in hardness but may achieve a desired increase in yield strength. Also, those alloy steels that contain one or more of the carbide-forming elements (chromium, molybdenum, vanadium, and tungsten) are capable of secondary hardening: they may become somewhat harder as a result of tempering.

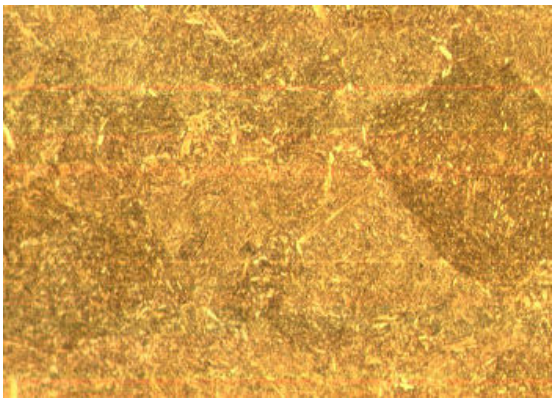


Figure 2.15: Tempered micro-structure - A quenched steel tempered by heating to 200 °C for 1 h [10].

Temperature and time are interdependent variables in the tempering process. Within limits, lowering temperature and increasing time can usually produce the same result as raising temperature and decreasing time. However, minor temperature changes have a far greater effect than minor time changes in typical tempering operations. With few exceptions, tempering is done at temperatures between 175 and 705 °C (350 and 1300 °F) and for times from 30 min to 4 h.

Tempering temperature Several empirical relationships have been made between the tensile strength and hardness of tempered steels such that the measurement of hardness is commonly used to evaluate the response of a steel to tempering. Figure 2.16 shows the effect of tempering temperature on hardness, tensile and yield strengths, elongation, and reduction in area of a low-alloy carbon steel (AISI 4340) held at temperature for 1 h. It can be seen that both room temperature hardness and strength decrease as the tempering temperature is increased. Ductility at ambient temperatures, measured by either elongation or reduction in area, increases with tempering temperature.

Tempering time The diffusion of carbon and alloying elements necessary for the formation of carbides is temperature and time dependent. The effect of tempering time on the hardness of a 0.82% C steel tempered at various temperatures is shown in Figure 2.17. The changes in hardness are approximately linear over a large portion of the time range when the time is presented on a logarithmic scale. Rapid changes in room-temperature hardness occur at the start of tempering in times less than 10 s. Less rapid, but still large, changes in hardness occur in times from 1 to 10 min, and smaller changes occur in times from 1 to 2 h. For consistency and less dependency on variations in time, components generally are tempered for 1 to 2 h. The levels of hardness produced by very short tempering cycles, such as in induction tempering, would be quite sensitive to both the temperature achieved and the time at temperature. The approximate hardnesses of quenched and tempered low- and medium-alloy steels can be predicted by means of an empirical tempering parameter (TP) developed by Holloman and Jaffe [13]:

$$TP = T \cdot (C_{TP} + \log_{10} t) \tag{2.2}$$

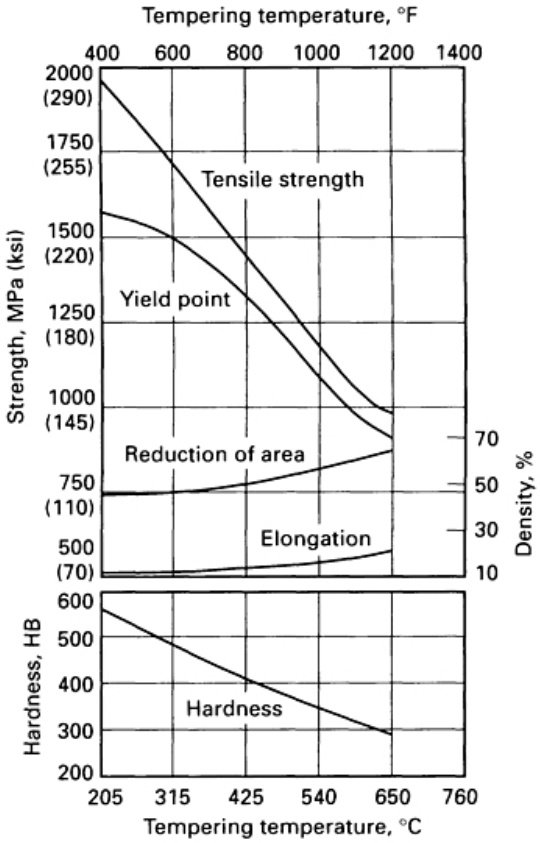


Figure 2.16: Mechanical properties on tempering - Effect of tempering temperature on the mechanical properties of oil-quenched AISI 4340 steel bar [12].

where T is temperature (K), t is time (s), and C_{TP} is a constant that depends on the carbon content of the steel (C_{TP}=10÷20). Reasonably good correlations are obtained except when significant amounts of retained austenite are present.

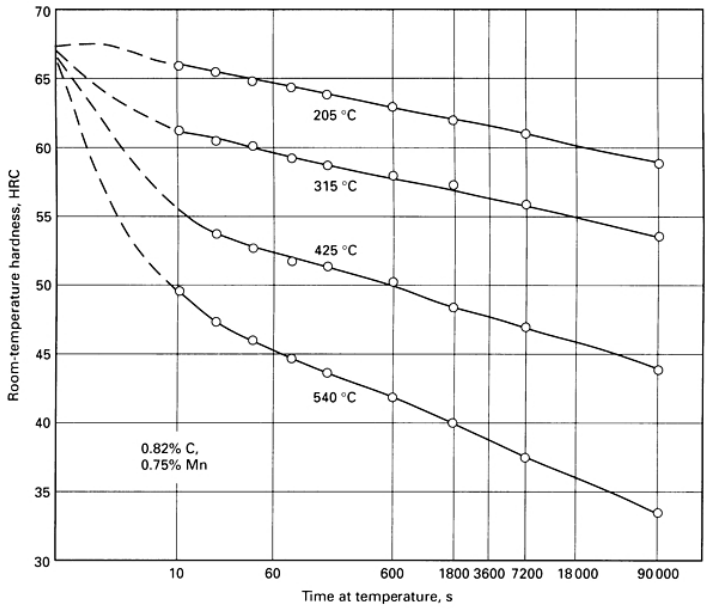


Figure 2.17: Effect of tempering time on hardness - Effect of time at four tempering temperatures on room-temperature hardness of quenched 0.82% C steel [14].

Cooling rate Another factor that can affect the properties of a steel is the cooling rate from the tempering temperature. Although tensile properties are not affected by cooling rate, toughness (as measured by notched-bar impact testing) can be decreased if the steel is cooled slowly through the temperature range from 375 to 575 °C (705 to 1065 °F), especially in steels that contain carbide-forming elements. Elongation and reduction in area may be affected also. This phenomenon is called temper embrittlement.

Carbon and alloy content The principal effect of carbon content is on as-quenched hardness. As it has already been shown in Figure 2.6, increasing the carbon content in steel alloys causes the material to become harder upon quenching. This is because the carbon sitting in the interstitial sites of the lattice structure hinders the movement of dislocation lines, thus increasing the strength of the material. The effect of temperature

and carbon content on the hardness of three carbon-molybdenum steels of different carbon content is shown in Figure 2.18, where the hardness curves after tempering for 1 h are plot as a function of temperature. The effect of carbon content is evident.

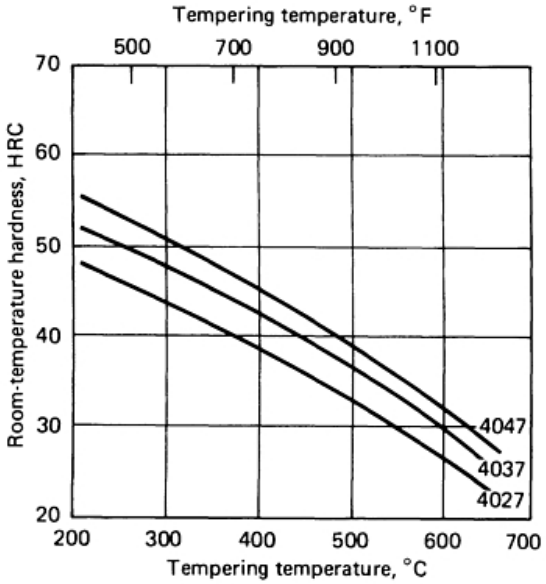


Figure 2.18: Effect of carbon content on tempering - Comparison of the tempering curves of three steels with different carbon content, tempered for 1 h. The relative difference in hardness compared with as-quenched hardness is retained after tempering.

The main purpose of adding alloying elements to steel is to increase the hardenability (i.e., the capability of the steel to form martensite upon quenching from above its critical temperature). The general effect of alloying elements on tempering is a retardation of the rate of softening, especially at the higher tempering temperatures. Thus, to reach a given hardness in a given period of time, alloy steels require higher tempering temperatures than do carbon steels. Alloying elements can be characterized as carbide forming or non-carbide forming. Elements such as nickel, silicon, aluminum, and manganese, which have little or no tendency to occur in the carbide phase, remain essentially in solution in the ferrite and have only a minor effect on tempered hardness. The carbide forming elements (chromium, molybdenum, tungsten, vanadium, tantalum, niobium, and titanium) retard the softening process by the formation of alloy carbides. The effect of the carbide-forming elements is minimal at low tempering temperatures where Fe_3C forms; however, at higher temperatures, alloy carbides are

formed, and hardness decreases slowly with tempering temperature.

Martempering

Martempering is a term used to describe an interrupted quench from the austenitizing temperature of certain alloy, cast, tool, and stainless steels. The purpose is to delay the cooling just above the martensitic transformation for a length of time to equalize the temperature throughout the piece. This will minimize the distortion, cracking, and residual stress. The term martempering is somewhat misleading and is better described as marquenching. The micro-structure after martempering is essentially primary martensitic that is untempered and brittle.
Martempering of steel (and of cast iron) consists of (see Figure 2.19):

- quenching from the austenitizing temperature into a hot fluid medium (hot oil, molten salt, molten metal, or a fluidized particle bed) at a temperature usually above the martensite range (M_s point);

- holding in the quenching medium until the temperature throughout the steel is substantially uniform;

- cooling (usually in air) at a moderate rate to prevent large differences in temperature between the outside and the center of the section.

Formation of martensite occurs fairly uniformly throughout the work-piece during cooling to room temperature, thereby avoiding formation of excessive amounts of residual stress. Straightening or forming is also easily accomplished upon removal from the marquenching bath while the part is still hot. The piece will hold its shape upon subsequent cooling in fixturing or in air cooling after removal from the forming die. The marquenching can be accomplished in a variety of baths including hot oil, molten salt, molten metal, or a fluidized particle bed. Martempered parts are tempered in the same manner as conventional quenched parts. The time lapse before tempering is not as critical because the stress is greatly reduced.

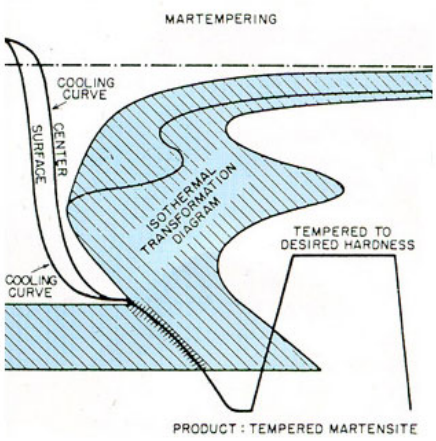


Figure 2.19: Martempering cycle - TTT diagram with superimposed cooling curves showing martempering.

Austempering

Austempering is the isothermal transformation of a ferrous alloy at a temperature below that of pearlite formation and above that of martensite formation. Austempering of steel offers several potential advantages:

- increased ductility, toughness, and strength at a given hardness;
- reduced distortion, which lessens subsequent machining time, stock removal, sorting, inspection, and scrap;
- the shortest overall time cycle to through-harden within the hardness range of 35 to 55 HRC, with resulting savings in energy and capital investment.

Steel is austempered by being (see Figure 2.20):

1. heated to a temperature within the austenitizing range, usually 790 to 915 °C (1450 to 1675 °F);
2. quenched in a bath maintained at a constant temperature, usually in the range of 260 to 400 °C (500 to 750 °F);
3. allowed to transform isothermally to bainite in this bath;
4. cooled to room temperature.

For true austempering, the metal must be cooled from the austenitizing temperature to the temperature of the austempering bath fast enough so that no transformation of austenite occurs during cooling, and then held at bath temperature long enough to ensure complete transformation of austenite to bainite.

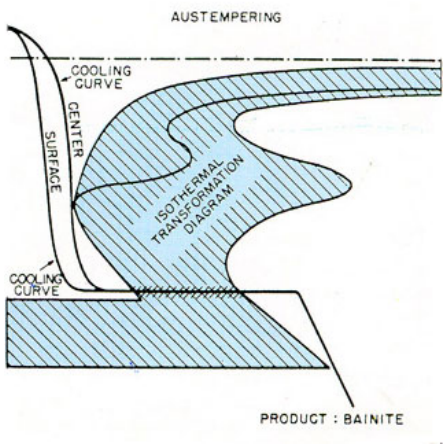


Figure 2.20: Austempering cycle - TTT diagram with superimposed cooling curves showing austempering.

2.2.6 Cold treating and cryogenic treatment

Cold treating of steel is widely accepted within the metallurgical profession as a supplemental treatment that can be used to enhance the transformation of austenite to martensite and to improve stress relief of castings and machined parts. Common practice identifies -84 °C (-120 °F) as the optimum temperature for cold treatment. There is evidence, however, that cryogenic treatment of steel, in which material is brought to a temperature of the order of -190 °C (-310 °F), improves certain properties beyond the improvement attained at cold-treatment temperatures.

Cold treating of steel

Cold treatment of steel consists of exposing the ferrous material to subzero temperatures to either impart or enhance specific conditions or properties of the material. Increased strength, greater dimensional or micro-structural stability, improved wear resistance, and relief of residual stress are among the benefits of the cold treatment of steel. All hardened steels are improved by a proper subzero treatment to the extent

that there will be less tendency to develop grinding cracks and therefore they will grind much more easily after the elimination of the retained austenite and the untempered martensite.

Hardening and retained austenite Whenever hardening is to be done during heat treating, complete transformation from austenite to martensite is generally desired prior to tempering. From a practical stand-point, however, conditions vary widely, and 100% transformation rarely, if ever, occurs. Cold treating may be useful in many instances for improving the percentage of transformation and thus for enhancing properties. During hardening, martensite develops as a continuous process from start (M_s) to finish (M_f) through the martensite formation range. Except in a few highly alloyed steels, martensite starts to form at well above room temperature (see Figure 2.21). In many instances, transformation is essentially complete at room temperature. Retained austenite tends to be present in varying amounts, however, and when considered excessive for a particular application, must be transformed to martensite and then tempered.

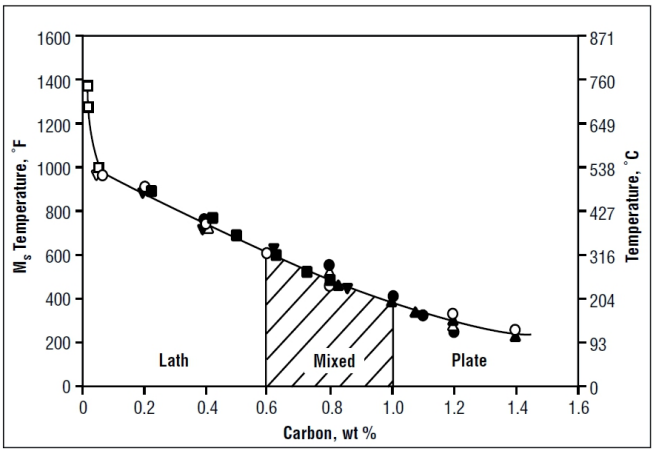


Figure 2.21: Effect of carbon content on M_s - Relationship between carbon content, martensite type, and start temperature [15].

Stress relief Residual stresses often contribute to part failure and frequently are the result of temperature changes that produce thermal expansion and phase changes, and consequently, volume changes. Under normal conditions, temperature gradients

produce non-uniform dimensional and volume changes. In castings, for example, compressive stresses develop in lower-volume areas, which cool first, and tensile stresses develop in areas of greater volume, which are last to cool. Shear stresses develop between the two areas. Even in large castings and machined parts of relatively uniform thickness, the surface cools first and the core last. In such cases, stresses develop as a result of the phase (volume) change between those layers that transform first and the center portion, which transforms last. When both volume and phase changes occur in pieces of uneven cross section, normal contractions due to cooling are opposed by transformation expansion. The resulting residual stresses will remain until a means of relief is applied. This type of stress develops most frequently in steels during quenching. The surface becomes martensitic before the interior does. Although the inner austenite can be strained to match this surface change, subsequent interior expansions place the surface martensite under tension when the inner austenite transforms. Cracks in high-carbon steels arise from such stresses.

Cryogenic treatment

Cryogenic treatment of steels consists of a slow cool-down (2.5 °C/min, or 4.5 °F/min) from ambient temperature to liquid nitrogen temperature. When the material reaches approximately 80 K (-315 °F), it is soaked for an appropriate time (generally 24 h). At the end of the soak period, the material is removed from the liquid nitrogen and allowed to warm to room temperature in ambient air. The temperature-time plot for this cryogenic treatment is shown in Figure 2.22. By conducting the cool-down cycle in gaseous nitrogen, temperature can be controlled accurately and thermal shock to the material is avoided. Single-cycle tempering is usually performed after cryogenic treatment to improve impact resistance, although double or triple tempering cycles are sometimes used.

2.3 Insight on surface hardening

Surface hardening, a process which includes a wide variety of techniques, is used to improve the wear resistance of parts without affecting the more soft, tough interior of the part. This combination of hard surface and resistance to breakage upon impact is useful in parts such as a cam or ring gear that must have a very hard surface to resist

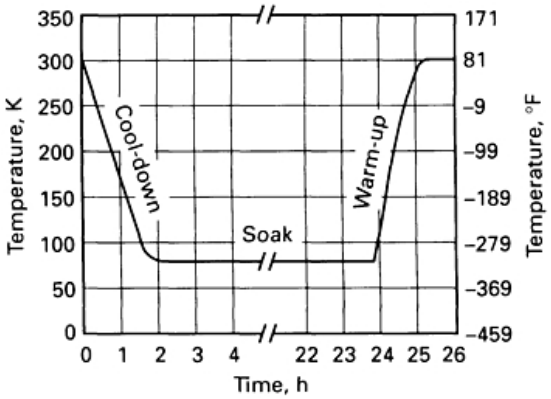


Figure 2.22: Cryogenic cycle - Plot of temperature versus time for the cryogenic treatment process [16].

wear, along with a tough interior to resist the impact that occurs during operation. Further, the surface hardening of steel has an advantage over through hardening because less expensive low-carbon and medium-carbon steels can be surface hardened without the problems of distortion and cracking associated with the through hardening of thick sections.

Leaving out the coverage of the methods involving an intentional addition of new layers to the original material in order to improve the overall performances of the part (e.g., hardfacing and coatings), the discussion will now focus on substrate treatments, which can be further classified into diffusion and selective hardening methods.

Diffusion methods modify the chemical composition of the surface with hardening species such as carbon, nitrogen, or boron. Diffusion methods allow effective hardening of the entire surface of a part and are generally used when a large number of parts are to be surface hardened.

Selective surface hardening methods allow, in contrast, localized hardening. Selective hardening generally involves transformation hardening (from heating and quenching), but some selective hardening methods (selective nitriding, ion implantation and ion beam mixing) are based solely on compositional modification.

2.3.1 Diffusion methods

Surface hardening by diffusion involves the chemical modification of a surface. The basic process used is thermochemical because some heat is needed to enhance the diffusion of hardening species into the surface and subsurface regions of a part. The depth of diffusion, or case depth (CD), exhibits a time-temperature dependence such that:

$$CD \propto K \cdot \sqrt{t} \tag{2.3}$$

where the diffusivity constant K depends on temperature, the chemical composition of the steel, and the concentration gradient of a given hardening species. In terms of temperature, the diffusivity constant increases exponentially as a function of absolute temperature. Concentration gradients depend on the surface kinetics and reactions of a particular process.

Methods of hardening by diffusion include several variations of hardening species (such as carbon, nitrogen, or boron) and of the process method used to handle and transport the hardening species to the surface of the part. Process methods for exposure involve the handling of hardening species in forms such as gas, liquid, or ions. These process variations naturally produce differences in typical case depth and hardness. Factors influencing the suitability of a particular diffusion method include the type of steel, the desired case hardness, and the case depth. It is important to distinguish between total case depth and effective case depth.

The term total case depth (TCD) refers to the depth of hardness where the hardened layer reaches the same hardness and properties as the base or core material. TCD is typically measured by sectioning the work-piece and polishing and etching with an acid solution to reveal the depth of the hardened layer. The measurements can then be taken visually and measured using a calibrated eyepiece or scale to qualify the total depth.

The term effective case depth (ECD) refers to the depth where a hardness measurement drops below a specified point. The hardness will then continue to decline until the TCD is reached. The hardness at the effective depth is specified based on the characteristics required and the hardenability of the material. For example, high carbon steel that may have a minimum surface hardness of 60 HRC may call for an ECD of 3 mm at 50 HRC. The method of determining ECD involves sectioning the piece and polishing

the surface. Measurements of the hardness are then taken at regular depth intervals until the hardness drops to the specified range. This distance from the surface is then measured to determine the effective depth.

Carburizing Carburizing is the addition of carbon to the surface of low-carbon steels at temperatures (generally between 850 and 950 °C, or 1560 and 1740 °F) at which austenite, with its high solubility for carbon, is the stable crystal structure. Hardening is accomplished when the high-carbon surface layer is quenched to form martensite so that a high-carbon martensitic case with good wear and fatigue resistance is superimposed on a tough, low-carbon steel core. Of the various diffusion methods, gas carburization is the most widely used, followed by gas nitriding and carbonitriding. Case hardness of carburized steels is primarily a function of carbon content. When the carbon content of the steel exceeds about 0.50%, additional carbon has no effect on hardness but does enhance hardenability. Carbon in excess of 0.50% may not be dissolved, which would thus require temperatures high enough to ensure carbon-austenite solid solution. Case depth of carburized steel is a function of carburizing time and the available carbon (carbon potential) at the surface. When prolonged carburizing times are used for deep case depths, a high carbon potential produces a high surface-carbon content, which may thus result in excessive retained austenite or free carbides. These two micro-structural elements both have adverse effects on the distribution of residual stress in the case-hardened part. Consequently, a high carbon potential may be suitable for short carburizing times but not for prolonged carburizing. Selection of carbon potential also depends on the carburizing response of a particular steel.

Nitriding Nitriding is a surface-hardening heat treatment that introduces nitrogen into the surface of steel at a temperature range (500 to 550 °C, or 930 to 1020 °F), while it is in the ferritic condition. Thus, nitriding is similar to carburizing in that surface composition is altered but different in that nitrogen is added into ferrite instead of austenite. Because nitriding does not involve heating into the austenite phase field and a subsequent quench to form martensite, nitriding can be accomplished with a minimum of distortion and with excellent dimensional control. The mechanism of nitriding is generally known, but the specific reactions that occur in different steels and with different nitriding media are not always known. Nitrogen has partial solubility

in iron. It can form a solid solution with ferrite at nitrogen contents up to about 6%. At about 6% N, a compound called gamma prime (γ'), with a composition of Fe_4N, is formed. At nitrogen contents greater than 8%, the equilibrium reaction product is ϵ compound, Fe_3N. Nitrided cases are stratified. The outermost surface can be all γ' and, if this is the case, it is referred to as the white layer (it etches white in metallographic preparation). Such a surface layer is undesirable and usually removed: it is very hard but is so brittle that it may spall in use. Special nitriding processes are used to reduce this layer or make it less brittle. The zone of the case is hardened by the formation of the Fe_3N compound, and below this layer there is some solid solution strengthening from the nitrogen in solid solution (see Figure 2.23).

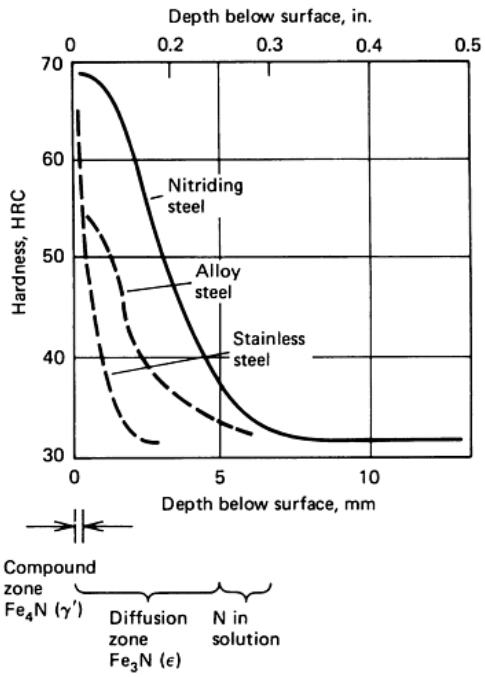


Figure 2.23: Nitriding - Nitride case profiles for various steels [17].

2.3.2 Selective surface hardening methods

The selective hardening of steel surfaces is typically achieved by localized heating and quenching, without any chemical modification of the surface. However, selective surface hardening can also include chemical modification by such techniques as ion implantation and selective carburization. The more common methods currently used

to harden the surface of steels include flame and induction hardening. However, each of these methods has shortcomings that can prevent its use in some applications. For example, the disadvantages of flame hardening include the possibility of part distortion, while induction hardening requires optimized induction coils, which must be carefully designed.

Flame hardening Flame hardening consists of austenitizing the surface of a steel by heating with an oxyacetylene or oxyhydrogen torch and immediately quenching with water. The result is a hard surface layer of martensite over a softer interior core with a ferrite-pearlite structure. There is no change in composition, and therefore the flame-hardened steel must have adequate carbon content for the desired surface hardness. The rate of heating and the conduction of heat into the interior appear to be more important in establishing case depth than the use of a steel of high hardenability. Flame-heating equipment may be a single torch with a specially designed head or an elaborate apparatus that automatically indexes, heats, and quenches parts (see Figure 2.24). Large parts such as gears and machine tool ways, with sizes or shapes that would make furnace heat treatment impractical, are easily flame hardened. With improvements in gas-mixing equipment, infrared temperature measurement and control, and burner design, flame hardening has been accepted as a reliable heat-treating process that is adaptable to general or localized surface hardening for small and medium-to-high production requirements.

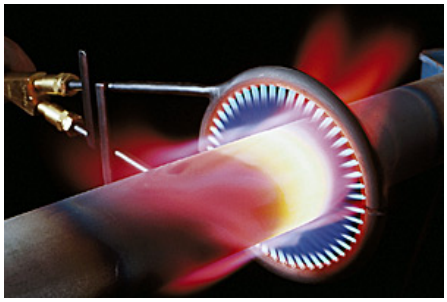


Figure 2.24: Flame hardening - Example of flame-heating equipment for tube hardening.

Induction heating Induction heating is an extremely versatile heating method that can perform uniform surface hardening, localized surface hardening, through hardening,

and tempering of hardened pieces. Heating is accomplished by placing a steel part in the magnetic field generated by high-frequency alternating current passing through an inductor, usually a water-cooled copper coil (see Figure 2.25). The depth of heating produced by induction is related to the frequency of the alternating current: the higher the frequency, the thinner or more shallow the heating. Therefore, deeper case depths and even through hardening are produced by using lower frequencies. The electrical considerations involve the phenomena of magnetic hysteresis and eddy currents.

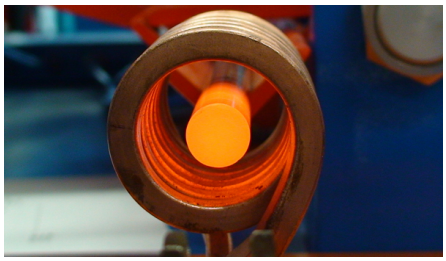


Figure 2.25: Induction hardening - Example of induction coil for heating billets.

Laser surface heat treatment Laser surface heat treatment is widely used to harden localized areas of steel and cast iron machine components. The heat generated by the absorption of the laser light is controlled to prevent melting and is therefore used in the selective austenitization of local surface regions, which transform to martensite as a result of rapid cooling (self-quenching) by the conduction of heat into the bulk of the work-piece. This process is sometimes referred to as laser transformation hardening to differentiate it from laser surface melting phenomena. There is no chemistry change produced by laser transformation hardening, and the process, like induction and flame hardening, provides an effective technique to harden ferrous materials selectively. Other methods of laser surface treatments include surface melting and surface alloying. Laser surface melting results in a refinement of the structure due to the rapid quenching from the melt. In surface alloying, alloying elements are added to the melt pool to change the composition of the surface. The novel structures produced by laser surface melting and alloying can exhibit improved electrochemical behaviour. Laser transformation hardening produces thin surface zones that are heated and cooled very rapidly, resulting in very fine martensitic micro-structures, even in steels with relatively low hardenability. High hardness and good wear resistance with less distortion result

from this process. The laser method (see Figure 2.26) differs from induction and flame heating in that the laser can be located at some distance from the work-pieces. Also, the laser light is reflected by mirrors to the focusing lens, which control the width of the heated spot or track.

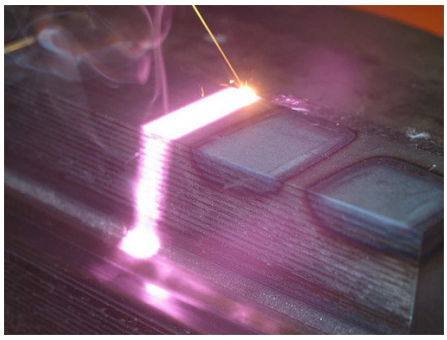


Figure 2.26: Laser hardening - Example of laser hardening.

2.4 Competing technologies

2.4.1 Case-hardening

Carburizing is a case-hardening process in which carbon is dissolved in the surface layers of a low-carbon steel part at a temperature sufficient to render the steel austenitic, followed by quenching and tempering to form a martensitic micro-structure. The resulting gradient in carbon content below the surface of the part causes a gradient in hardness, producing a strong, wear-resistant surface layer on a material, usually low-carbon steel, which is readily fabricated into parts. In gas carburizing, commercially the most important variant of carburizing, the source of carbon is a carbon-rich furnace atmosphere produced either from gaseous hydrocarbons, for example, methane (CH_4), propane (C_3H_3), and butane (C_4H_{10}), or from vaporized hydro-carbon liquids.

Carbon sources

Low-carbon steel parts exposed to carbon-rich atmospheres derived from a wide variety of sources will carburize at temperatures of 850 °C (1560 °F) and above. In the most primitive form of this process, the carbon source is so rich that the solubility limit of carbon in austenite is reached at the surface of the steel and some carbides

may form at the surface. Such atmospheres will also deposit soot on surfaces within the furnace, including the parts. While this mode of carburizing is still practised in parts of the world in which resources are limited, the goal of current practice in modern manufacturing plants is to control the carbon content of furnace atmospheres so that:

- the final carbon concentration at the surface of the parts is below the solubility limit in austenite;
- sooting of the furnace atmosphere is minimized.

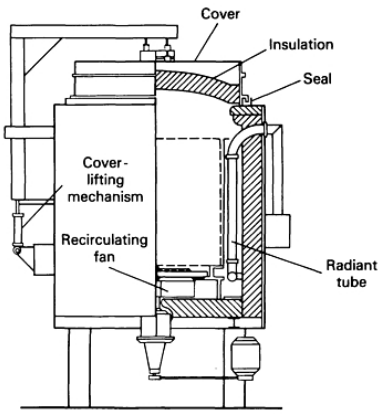
Controlled carburizing atmospheres are produced by blending a carrier gas with an enriching gas, which serves as the source of carbon. The usual carrier, endothermic gas, is not merely a diluent, but plays a role, described below, in accelerating the carburizing reaction at the surface of the parts. The amount of enriching gas required by the process depends primarily on the carbon demand, that is, the rate at which carbon is absorbed by the work load.

Carburizing equipment

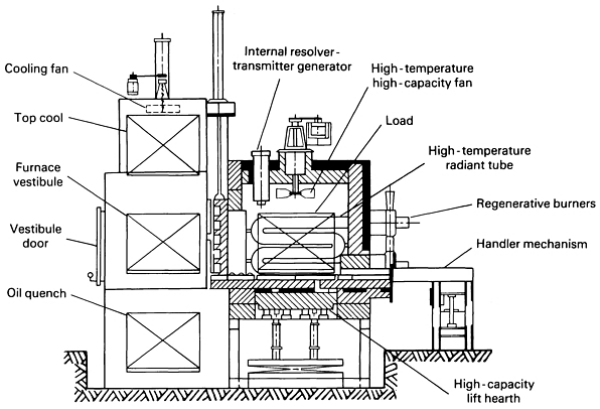
Gas carburizing furnaces vary widely in physical construction, but they can be divided into two major categories, batch and continuous furnaces. In a batch-type furnace, the work load is charged and discharged as a single unit or batch. In a continuous furnace, the work enters and leaves the furnace in a continuous stream. Continuous furnaces are favoured for the high-volume production of similar parts with total case depth requirements of less than 2 mm.

Batch furnaces The most common types of batch furnaces are pit furnaces and horizontal batch furnaces. Pit furnaces are usually placed in a pit with the cover or lid located just above floor level and are often loaded and unloaded with the aid of an overhead crane (see Figure 2.27a). Pit furnaces are frequently used for large parts requiring long processing times. If the work is to be direct quenched, the load must be moved through air before quenching. As a result, parts will be covered by an adherent black scale, which, depending on the needs of the application, may have to be removed by shot blasting or acid pickling.
Horizontal batch furnaces are frequently used for carburizing and direct quenching. Many of these furnaces are so-called sealed quench, or integral quench, furnaces; that

is, parts are discharged from the furnace into a vestibule that covers an oil quench tank (see Figure 2.27b). Because the furnace atmosphere also flows through the vestibule, parts can be kept free of oxidation prior to quenching. Sealed-quench batch furnaces are capable of processing many different types of loads with widely varying case depth requirements. Like pit furnaces, they can be made quite gas tight, with the result that positive furnace pressures are easily achieved.



(a) A pit batch carburizing furnace. Dashed lines represent the location of workload.



(b) A high-productivity gas-fired integral quench furnace.

Figure 2.27: Carburizing furnaces - Layout of typical batch furnaces for carburizing.

Continuous furnaces Types of continuous furnaces used for carburizing include mesh belt, shaker hearth, rotary retort, rotary hearth, roller hearth, and pusher designs. Many of these furnaces can be built with sealed oil quenching so that oxide-free parts can be produced. Most of these furnaces can be sealed well enough that positive furnace pressures can be maintained. Some continuous mesh belt furnaces, on the other hand, are open to the air at either end. Because air cannot be positively excluded, carburizing in these furnaces is often difficult to control.

Preparation of parts for carburizing

Parts, trays, and fixtures should be thoroughly cleaned before they are charged into a carburizing furnace. Often they are washed in a hot alkaline solution. Some users heat washed parts, trays, and fixtures in an oxidizing atmosphere at 400 °C (750 °F) before

carburizing to remove traces of organic contaminants [18]. Very thin oxide layers on parts (such as those produced by oxidation below 500 °C, or 900 °F) are reduced by the carburizing atmosphere. Heavy oxide layers, such as forging scale, will be reduced to iron flakes, which are not adherent to the part. Residues from alkaline washer solutions deposited on parts, particularly those with silicates, can cause spotty carburizing, as well as give the parts a blotchy appearance. In addition, alkaline residues can adversely affect the life of heat-resistant furnace alloys. Quenching salts remaining on trays and fixtures can also damage furnace hardware (for example, silicon carbide rails in pusher furnaces). Chlorine- or sulfur-containing residues on parts will release gases that can react with brickwork, the protective oxide films on heat-resistant alloy fixtures, or the work load.

To function properly, some parts must be selectively carburized, that is, carburized only on certain surfaces. Some gears are carburized only on teeth, splines, and bearing surfaces. In addition to satisfying the performance requirements of a part, selective carburizing may facilitate the machining or welding of non-carburized surfaces in the hardened condition. Surfaces that are not to be carburized must be protected by a coating or shield that is impervious to the carburizing atmosphere. Various means are employed to protect or stop off selected surfaces from the atmosphere.

Copper plating is widely used for this purpose because it is relatively easy to apply, is machinable, and does not contaminate furnace atmospheres. Surfaces that are not to be copper plated can be coated with a chemical-resistant lacquer, which is removed prior to carburizing. After carburizing, the copper can be chemically stripped from the part or removed in subsequent machining operations.

Ceramic coatings in the form of paint can also protect selected surfaces from carburizing. Surfaces must be thoroughly cleaned before ceramic paint is applied, and the first coat is allowed to dry before a second coat is applied. Ceramic paint coatings must adhere tightly in order to be impervious to the carburizing atmosphere.

Blind holes can be *stopped-off* by inserting copper plugs or by filling them with clay. If air is entrapped by the plug, a means of venting must be provided to relieve the pressure build-up during heating. Through holes may be plugged at either end to limit access by the atmosphere, thereby minimizing carburization. Internal threads can be protected by the insertion of a copper screw; external threads by capping with a copper nut. If a steel screw or nut is used, the threads should be coated with a stopoff

material to facilitate removal. The success of all stopoff methods depends largely on the care used in their applications. If parts are cooled slowly after carburizing, they will be soft enough to permit the removal of the case in selected areas by machining. After subsequent reheating and quenching, these areas will remain lower in carbon and relatively soft.
Operations can also be planned so that the case on a hardened part can be removed selectively by *grinding*. This practice is usually confined to small areas.

Carburizing process variables

The successful operation of the gas carburizing process depends on the control of three principal variables:

- temperature;
- time;
- atmosphere composition.

Other variables that affect the amount of carbon transferred to parts include the degree of atmosphere circulation and the alloy content of the parts.

Temperature The maximum rate at which carbon can be added to steel is limited by the rate of diffusion of carbon in austenite. This diffusion rate increases greatly with increasing temperature; the rate of carbon addition at 925 °C (1700 °F) is about 40% greater than at 870 °C (1600 °F). The temperature most commonly used for carburizing is 925 °C (1700 °F). This temperature permits a reasonably rapid carburizing rate without excessively rapid deterioration of furnace equipment, particularly the alloy trays and fixtures. The carburizing temperature is sometimes raised to 955 °C (1750 °F) or 980 °C (1800 °F) to shorten the time of carburizing for parts requiring deep cases. Conversely, shallow case carburizing is frequently done at lower temperatures because case depth can be controlled more accurately with the slower rate of carburizing obtained at lower temperatures.
For consistent results in carburizing, the temperature must be uniform throughout the work load. Temperature gradients through the work load will persist for a substantial period of time while the work is being heated to the carburizing temperature. Because

parts at the exterior of the load reach the furnace temperature first, they will begin carburizing well before parts at the interior of the load. The consequence is variability in case depth from part to part and within a single part. In addition, soot can be deposited on cold parts exposed to a carburizing atmosphere. Therefore, for best results, the work load should be heated to the carburizing temperature in a near-neutral furnace atmosphere. In batch furnaces, parts can be heated until they reach the furnace temperature; then carburizing can commence with the addition of the enriching gas. Many new continuous furnaces are being built with separate preheat chambers to ensure that the load is at a uniform temperature before entering the carburizing zone.

Time The effect of time and temperature on total case depth is shown in Figure 2.28. The data given are computed assuming saturated austenite at the surface of the work-pieces. When the surface carbon content is controlled so that it is less than the saturation value, case depths will be less than they otherwise would be. Moreover, for a given case depth, increasing the carburizing temperature will decrease the carburizing time. In addition to the time at the carburizing temperature, several hours may be required to bring large work-pieces or heavy loads of smaller parts to operating temperature. For work quenched directly from the carburizing furnace, the cycle may be lengthened further by allowing time for the work to cool from the carburizing temperature to about 843 °C (1550 °F) prior to quenching. If the workload is exposed to the carburizing atmosphere during heating, some carburizing will occur before the nominal start of carburizing. Similarly, additional diffusion and interchange of carbon with the atmosphere will occur during cooling prior to quenching. More complex mathematical models that allow for variations in temperature and atmosphere carbon potential with time can be constructed to allow a better prediction of case depth.

Atmosphere composition The carbon potential of a furnace atmosphere at a specified temperature is defined as the carbon content of pure iron that is in thermodynamic equilibrium with the atmosphere. The carbon potential of the furnace atmosphere must be greater than the carbon potential of the surface of the work-pieces in order for carburizing to occur. It is the difference in carbon potential that provides the driving force for carbon transfer to the parts. The combined effects of time, temperature, and carbon concentration on the diffusion of carbon in austenite can be expressed by Fick's

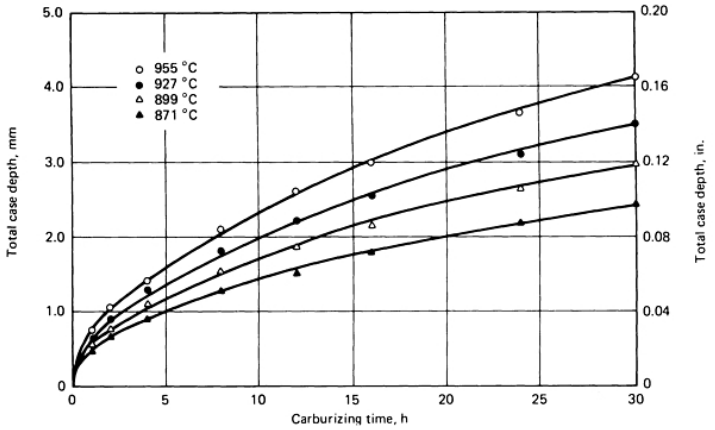


Figure 2.28: Carburizing time - Plot of total case depth versus carburizing time at four selected temperatures [19].

laws of diffusion. Fick's first law states that the flux of the diffusing substance perpendicular to a plane of unit cross-sectional area is proportional to the local carbon gradient perpendicular to the plane:

$$J = -D\frac{\partial \phi}{\partial x} = -D\nabla\phi \tag{2.4}$$

where J is the diffusion flux (mol/m^2·s), D is the diffusion coefficient (m^2/s), ϕ is the particle concentration (mol/m^3), and x is the particle position (m).
Fick's second law is a material balance within an elemental volume of the system; the flux of carbon into an elemental volume of iron minus the flux of carbon out of the elemental volume equals the rate of accumulation of carbon within the volume:

$$\frac{\partial \phi}{\partial t} = D\frac{\partial^2 \phi}{\partial x^2} = \nabla \cdot (D\nabla\phi) \tag{2.5}$$

where t is the diffusion time (s). Combining the two laws (2.4) and (2.5) leads to a partial differential equation that describes the diffusion process and provides solutions to the diffusion equation for a variety of boundary conditions and part configurations, for example, plate, rod, sphere, and so on [20]. With these solutions and values of the diffusion coefficient, it is possible to predict the carbon gradient and depth of penetration occurring for any combination of time, temperature, and surface carbon concentration.

2.4.2 Induction hardening

Induction heating is a method of heating electrically conductive materials by the application of a varying magnetic field whose lines of force enter the work-piece. In this process, the varying magnetic field induces an electric potential (voltage), which can then create an electric current depending on the shape and the electrical characteristics of the work-piece. These so-called eddy currents dissipate energy and produce heat by flowing against the resistance of an imperfect conductor. Because all metals are fair electrical conductors, induction heating is applicable to several types of metal processing operations such as melting, welding, brazing, heat treating, stress inducement, zone refining, and heating prior to hot working. The technique also lends itself to a variety of non-metal applications including adhesive bonding, graphitizing carbon, drying, curing, and super-heating glass. Of these applications, heat treating predominates in terms of the number of units used, with surface hardening of steel and cast iron being the most prevalent use. Compared to furnace techniques, induction heating can often provide energy savings and much higher heating rates than convection, radiation, conduction, or even flame-impingement processes. Other advantages of induction heat treating, which stem from this non-contact method and its generation of heat within the work-piece, are:

- ease of automation and control;
- reduced floor space requirements;
- quiet and clean working conditions;
- suitability for integration in a production line or general work area;
- self-monitoring capability.

Principles of induction heating

The basic components of an induction heating system are an induction coil, an alternating-current (AC) power supply, and the work-piece itself. The coil, which may take different shapes depending on the required heating pattern, is connected to the power supply so that a magnetic field is generated from the current flow (see Figure 2.29). The magnitude of the field depends upon the strength of the current and

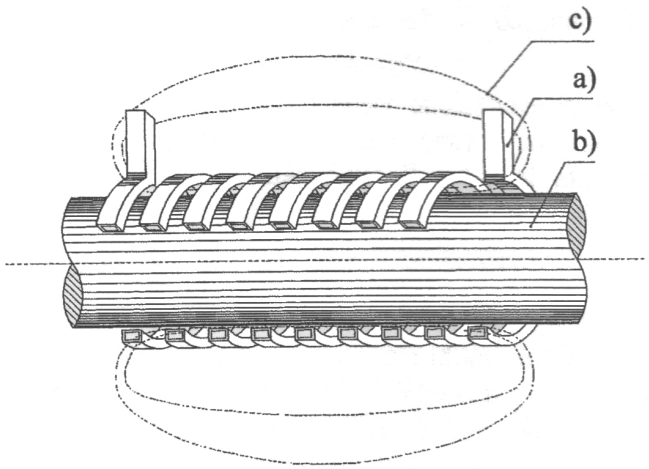


Figure 2.29: Solenoid coil - Pattern of currents and magnetic field in a solenoid coil. (a) inductor; (b) load; (c) magnetic field lines [21].

the number of turns in the coil. If an electrically conductive object is placed inside a coil with a varying current, the eddy currents are generated from the varying magnetic field within the object. These eddy currents are induced as a result of Faraday's law of electromagnetic induction:

$$e = -N\frac{d\Phi}{dt} \tag{2.6}$$

which relates an induced voltage e (V) to the time rate of change of the magnetic field Φ (A/m^2) and the number of turns N in the coil. In effect, the coil acts in much the same manner as a primary winding of a transformer, with the work-piece acting as a single-turn (or short-circuited) secondary winding.

At any moment the direction of the induced current in the work-piece is approximately opposite to that in the inductor coil, and in general its flow pattern will describe a kind of "shadow image" of the coil conductors. The induced currents also generate their own magnetic fields, which are in opposition to the field generated by the coil and thereby prevent the field from penetrating to the center of the object. Therefore, the eddy currents are more concentrated at the surface and decrease in strength toward the center of the object. This phenomenon of the eddy currents travelling closer to the surface of a conductor is called the "skin effect."

The mathematics needed to explain the skin effect involve a differential equation that has solutions in the form of Bessel functions. These solutions demonstrate that the in-

duced current in a large planar object (that is, a plate much thicker than the expected eddy current penetration) decreases exponentially from the surface into the workpiece, or electrical load. This allows the definition of an "effective" depth of eddy currents known as the "reference depth" or "penetration depth," where the current density has dropped by 1/e (or 37%) of its surface value. The formula for this depth is given by:

$$\delta = \sqrt{\frac{2\rho}{\omega\mu\mu_0}} = \sqrt{\frac{\rho}{\pi\mu\mu_0 f}} \tag{2.7}$$

where δ is the reference depth (m), ρ is the resistivity of the work-piece (Ω·m), μ_0 is the magnetic permeability of the vacuum (μ_0=$4\pi\cdot10^{-7}$ H/m), μ is the relative magnetic permeability of the work-piece, ω is the pulsation (rad/s), and f is the frequency (Hz) of the alternating magnetic field of the work coil. If the current is squared, a curve that relates to the heating effect of the eddy currents is obtained. At the depth of penetration δ, the heating effect therefore drops to $1/e^2$ (or 14%) of its value at the surface (see Figure 2.30).

Figure 2.31 presents plots of the penetration depth at various frequencies as a function of

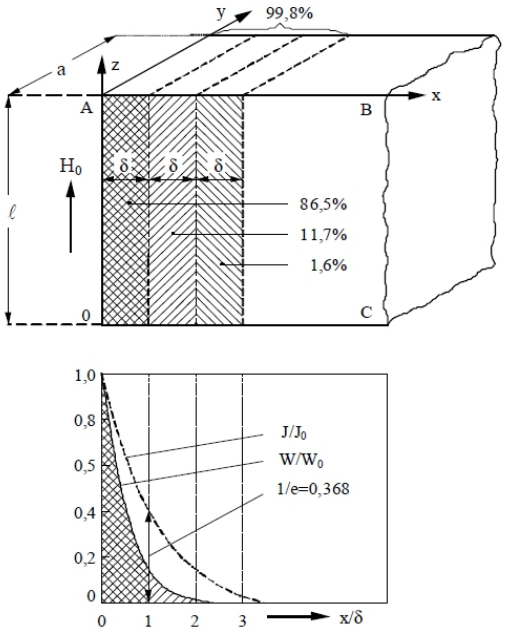


Figure 2.30: Penetration depth - Power and current density distribution for flat slab arrangement [21].

the power density. These variations occur for any given material, because the resistivity

of conductors increases with temperature. Moreover, for ferromagnetic metals, the relative permeability varies with temperature, decreasing to a value of one at the so-called Curie temperature. Also, as the power density is increased, steels may become magnetically saturated, leading to decreased permeability and hence increased reference depth. Because of these effects, the reference depth in non-magnetic materials may vary by a factor of two or three over a heat-treatment temperature range, whereas for magnetic steels it can vary by a factor of 20.

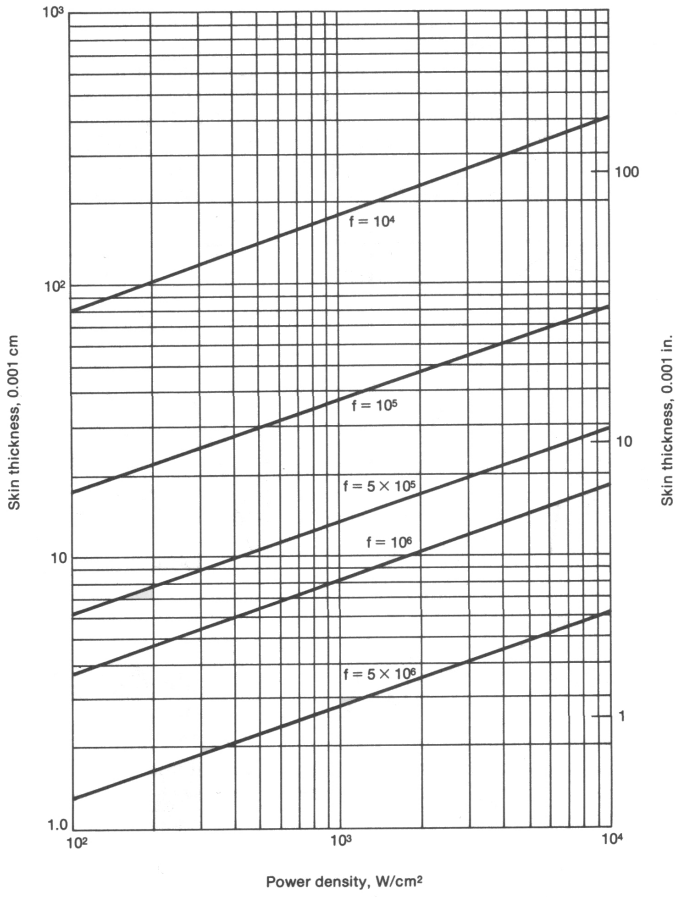


Figure 2.31: Effect of power and frequency on the penetration depth - Reference skin depth for magnetic steel as a function of power density and frequency [1].

Heat generation The primary mechanism of heat generation occurs from the flow of eddy currents into the work-piece. A second mechanism of heat generation occurs in ferromagnetic materials when energy is dissipated during the reversal of magnetic

domains. This effect in ferromagnetic materials is usually ignored in induction heating calculations, because the effect is small for the typical strength levels of magnetic fields for induction heating. Therefore, the rate of heating in units of power is largely a function of I^2R, or Joule heating. The heating rate is affected by several variables such as:

- field strength;
- proximity of the work-piece to the coil conductors;
- the electrical and magnetic properties of the material such as resistivity (ρ) and magnetic permeability (μ).

Heating rates are typically controlled by the coil voltage, because the field strength (or coil current) is roughly proportional to coil voltage. The desired rate of heating varies with the application, and the mathematical analysis of induction heating processes can be quite complex for all but the simplest of work-piece geometries. This is because of the effects of non-uniform heat generation through the work-piece, heat transfer, and the fact that the electrical, thermal, and metallurgical properties of most materials exhibit a strong dependence on temperature. For this reason, quantitative solutions exist for the most part only for the heating of round bars or tubes and rectangular slabs and sheets. Nevertheless, such treatments, if properly applied, may provide useful insights into the effects of coil design and equipment characteristics on heating patterns in irregularly shaped parts. This information, coupled with knowledge generated through years of experimentation in both laboratory and production environments, serves as the basis for the practical design of induction heating processes.

Coupling and flux shaping The degree of coupling between the work-piece and the magnetic field of the coil is determined by the number of magnetic flux lines that enter the work-piece itself. This flux density (number of lines of force) is proportional to the coil current, and the amount of energy transferred is proportional to the square of the number of flux lines intercepted by the work-piece. The frequency of coil currents also influences the pattern of induced currents, because flux lines tend to stay closer to the coil conductors as frequency is increased. Flux leakage refers to the flux that does not thread through the heated material. The degree of flux leakage depends, in part, upon

the electrical gap between the coil conductors and the work-piece. This so-called air gap is of greater concern as frequency increases, because flux lines tend to stay closer to the surface of coil conductors as frequency is increased.

Flux leakage can also be influenced by the use of flux concentrators, which are ferromagnetic materials purposely placed on or near the coil assembly. These devices, provide a path of low magnetic reluctance and facilitate the concentration of flux lines in desired regions.

Induction heat-treating equipment

An induction heating system typically consists of a power supply, a workstation, an inductor (heating) coil, controls, and work-piece handling units. When a steel or cast iron is being hardened, the system may also include a quenching system, depending on the hardenability of the particular grade. Some annealing practices may also be enhanced by accelerated cooling. The basic architecture of an induction heating system includes a workstation (or heat station), which contains load matching components such as output transformers and capacitors, plus high-frequency contactors, protective devices, cooling-water manifolds, and quench valves. The principal function of the workstation is to provide the proper electrical impedance match between the output of the power supply and the inductor (induction heating coil) for optimum power transfer into the heated load. The coil is normally mounted on the front of and close to the workstation. Workstation components may sometimes be installed in the power supply enclosure, as is often the case in small solid-state, medium-frequency heaters, and in many of the vacuum-tube oscillator units.

Power supplies Besides the induction coil and work-piece, the power supply is probably the most important component of an overall induction heating system. The function of the power supply is to accept power from a line at (usually 50 or 60 Hz) and to deliver AC electrical energy through the load matching circuit to the inductor coil circuit at a useful voltage and frequency. Over the years, three basic frequency ranges have evolved for the classification of AC induction heating systems:

- low-frequency systems (50 or 60 Hz line-frequency systems);
- medium-frequency systems (converters in the range 5-30 kHz);

- high-frequency systems (100-450 kHz).

Medium- and high-frequency power supplies suggest the need for conversion of line-frequency power, preferably three phase, to a single phase of higher frequency. A very basic block diagram that applies to nearly all induction heating power supplies is shown in Figure 2.32. The first block represents the AC to DC converter or *rectifier*. This section may provide a fixed DC voltage, a variable DC voltage, or a variable DC current. The second block represents the *inverter* or oscillator section, which switches the DC to produce a single-phase AC output. The third block represents the *load-matching* components, which adapt the output of the inverter to the level required by the induction coil. The control section compares the output of the system to the command signal and adjusts the DC output of the converter, the phase or frequency of the inverter, or both to provide the desired heating.

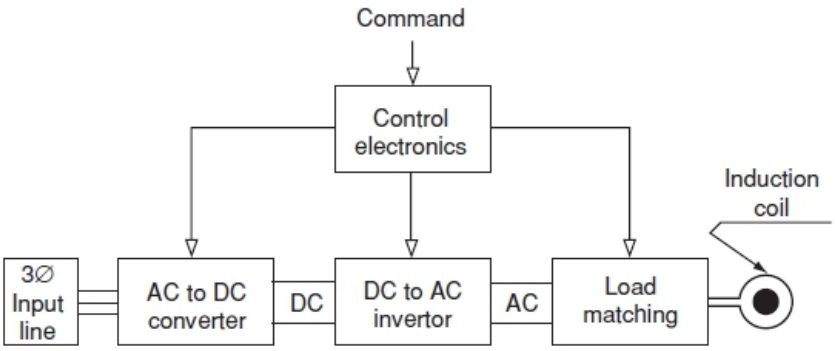


Figure 2.32: Induction heat treatment system - Basic block diagram of an induction heat treatment power supply [22].

The choice of a power source depends primarily on the desired frequency, although there is often considerable overlap of application capability for any given unit in its frequency range. The frequency required for efficient induction heating is determined by the material properties (i.e., resistivity and relative magnetic permeability) and the work-piece cross-sectional size and shape. In a great variety of cases, a suitable frequency may also be impacted by the application requirements such as case depth in surface hardening of steels or the allowable temperature gradient for through hardening (see Figure 2.33). In general, however, equipment cost per kilowatt increases with frequency, so a power supply of the lowest possible frequency that will accomplish the job efficiently is usually the best place to start when considering equipment.

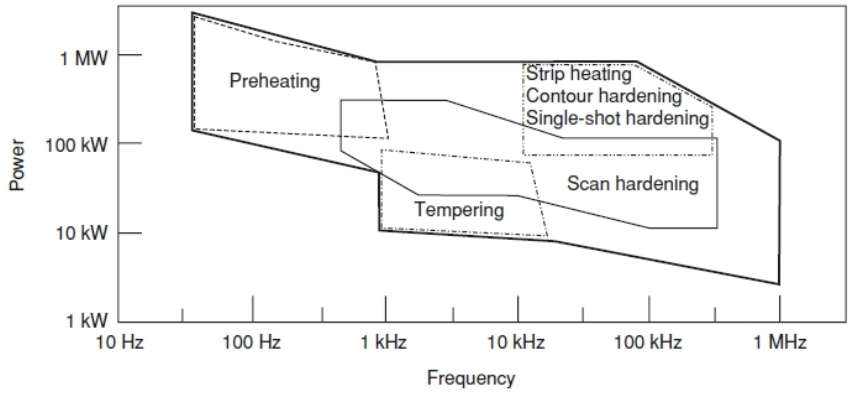


Figure 2.33: Induction heat treatment applications - Typical power and frequency range for typical induction heat treatment applications [22].

Power requirements for induction heating are also dependent on the specific application. For through heating applications, the power needed is generally based upon the amount of material processed per unit of time, the peak average temperature, and the material's heat capacity at this temperature. A power determination for localized heating and other operations, such as surface hardening of steel, is not as simple because of the effects of heat conduction to the adjacent metal and/or unhardened core.

Medium-frequency power supplies using *solid-state* frequency conversions make up the vast majority of induction heating and melting power supplies in use today. The inverter circuits that convert DC to AC use solid-state switching devices such as thyristors and transistors. For high power and lower frequencies, large thyristors are commonly used. These systems, which utilize high-power silicon-controlled rectifiers (SCRs) having rapid turn-on/turn-off times, have largely replaced motor-generator sets for medium frequency conversion. The basic advantages of solid-state power supplies are their improved efficiency (see Figure 2.34), low initial cost and maintenance, and availability in a multitude of sizes and frequencies. With the exception of an AC-AC solid-state converter, which does not have wide usage in heat-treating applications, all of the commercial medium-frequency solid-state inverters for induction heating use a two-step conversion in which 50 Hz AC is changed to DC and the DC changed back to AC at the desired frequency. Diodes or thyristors first rectify three-phase line frequency to produce direct current, and solid-state logic then converts DC into AC by either a swept-frequency or load-resonant inverter. In both types of solid-state inverters, the

system frequency is tuned to the load, whereas in motor-generator circuits the load is tuned to the fixed frequency of the generator. This greatly simplifies load matching when the electrical characteristics of the load change during a heating cycle.

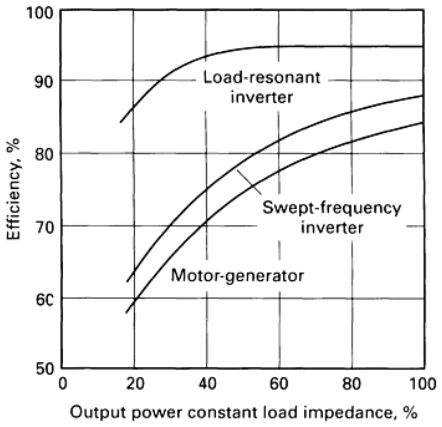


Figure 2.34: Inverter efficiency - Comparison of the overall system efficiency of a typical motor-generator, swept-frequency inverter, and load-resonant inverter.

For frequencies above 150 kHz, frequency conversion can be performed by either vacuum-tube or transistor systems (see Figure 2.35).
Vacuum-tube systems have been widely used in the past, thanks to their flexibility and capability of attaining frequencies up to about 2 MHz for induction heating. However, the low efficiency (50-70% depending on vacuum tube type), limited life time (4000-6000 operating hours) and high voltages required (10-12 kV), made the vacuum tube a rather inefficient equipment [23].
Recently, transistor systems are more and more used because of their ability to be turned on and off very fast with low-switching losses. Presently, transistor systems like the so-called Insulated-Gate Bipolar Transistor (IGBT), produce RF heating energy in the frequency span 20 to 100 kHz and a wide range of power outputs. Metal-Oxide Semiconductor Field-Effect Transistor (MOSFET) allow to extend the frequency range up to 500 kHz with acceptable power levels (>100 kW). The major difference of the two switches can be found within their switching capability and robustness. The MOSFET is capable of reaching higher frequencies but is quite susceptible to high voltages. On the contrary, the IGBT is rather slower but very robust. In addition, the frequency range of the IGBT can be increased by the so-called multiplexing technique which means that the frequency is split, therefore reducing the switching losses to an acceptable level.

For high frequency applications, the switching losses will become the determining factor [24].

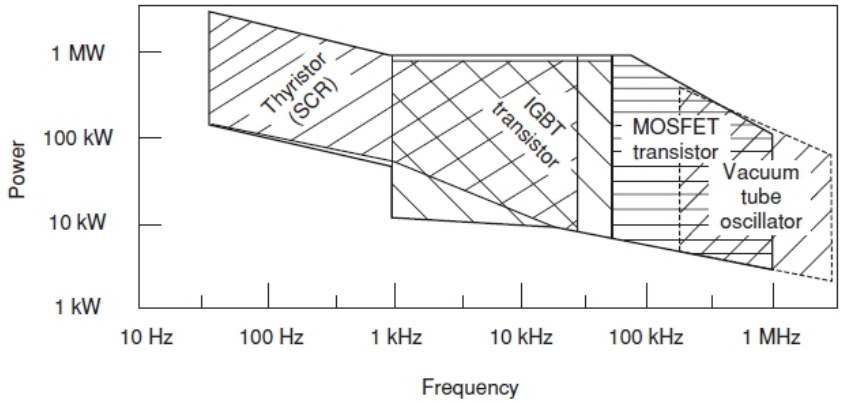


Figure 2.35: Induction heat treatment inverters - Modern inverter types and power-frequency range for induction heat treatment [22].

The first development in solid-state frequency conversion for induction heaters was based on the *swept-frequency* or *variable-frequency* inverter. With this system, line voltage is converted to direct current and then is applied by means of a capacitive voltage divider to an SCR (DC-AC) inverter circuit. The system utilizes internal commutation to turn off the SCRs in the inverter circuit. A feedback circuit from the resonant tank, together with power- and voltage-control signals, is fed to a local oscillator whose output signal then controls the inverter firing rate. The system then operates at a frequency along the straight portion of the tank's resonant frequency curve, but not at the resonant frequency. By offsetting the firing rate of the oscillator from the resonant frequency of the tank, the unit can be made to operate along the power curve (see Figure 2.36). The frequency from the local oscillator can thus be shifted by the resonant frequency of the tank circuit (or heat station), and the power supply (if set for constant output at the load) can shift its frequency as the load is heated. Nevertheless, it is necessary first to tune the system so that the tank frequency is within the frequency band of the local oscillator. Once the tank is tuned within the range of the local oscillator, it too will provide a constant output as the resonant frequency of the tank shifts, especially when the temperature passes through the Curie point during heating of magnetic steels.

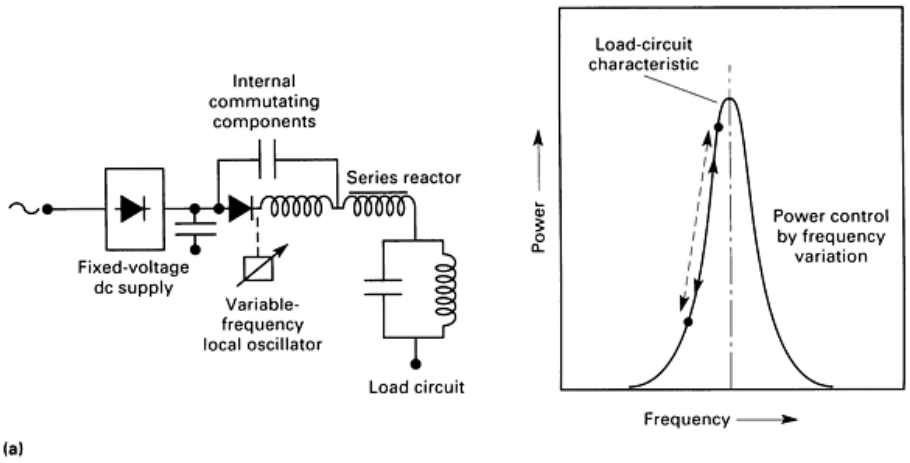


Figure 2.36: Swept-frequency inverter - Circuit schematics for a swept-frequency inverter [4].

The other common solid-state inverter, known as the *load-resonant*, also utilizes a DC power source. However, this inverter is provided with an SCR for voltage control. The SCR inverter uses no local oscillator but derives its commutating effect directly from the resonant tank circuit. Thus, it is load-resonant. In actuality, for reasons based on the operational characteristics of the SCR, the system is usually operated with a slightly leading power factor and thus an operating frequency slightly higher than the load-resonant frequency (see Figure 2.37). Because the frequency is dependent on the tank-resonant conditions, changes in the coil inductance (such as when a magnetic material passes through the Curie temperature) will shift the frequency of operation without any change in power output. If it is desired to select a specific frequency, depending on the application, capacitor taps can be adjusted to meet this requirement. Because the system always operates at the resonant frequency of the tank circuit, its power factor is usually close to unity, and therefore impedance matching by means of a transformer-tap change is the only routine adjustment that need be made to secure optimum power output. This procedure is the same as that for tuning a fixed-frequency motor-generator system.

Full-bridge inverter The most common inverter configuration is the full bridge as shown in Figure 2.38. Often referred to as an H-bridge, it has four legs that each

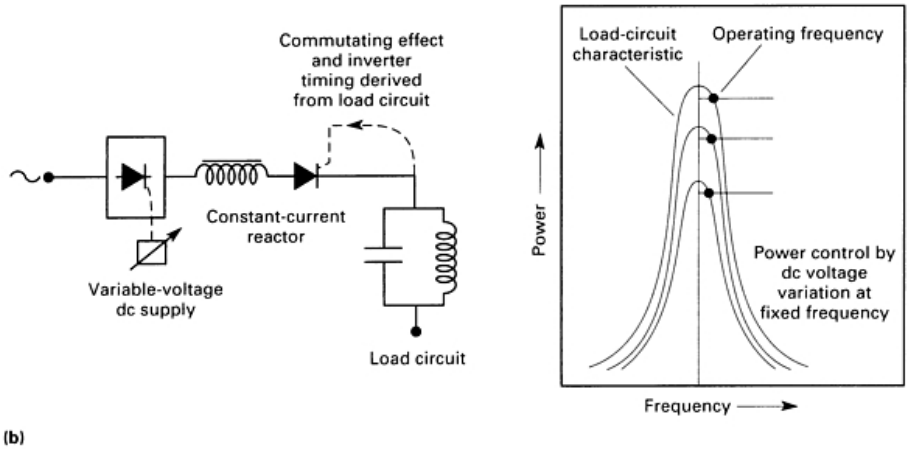


Figure 2.37: Load-resonant inverter - Circuit schematics for a load-resonant inverter [4].

contains a switch. The output is located in the center of the H so that when switches S1 and S2 are closed, current flows from the DC supply through the output circuit from left to right. When switches S1 and S2 are opened and switches S3 and S4 are closed, current flows in the opposite direction, from right to left. As this process is repeated, an AC is generated at a frequency determined by the rate at which the switches are opened and closed.

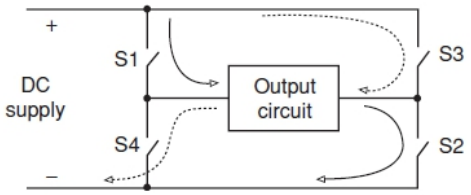


Figure 2.38: Full-bridge inverter - Basic scheme of a full-bridge inverter [22].

Half-bridge inverter The half-bridge inverter, as its name implies, requires only two switches and two-filter capacitors to provide a neutral connection for one side of the output circuit as shown in Figure 2.39. The other side of the output circuit is then switched between positive DC supply by S1 and the negative supply by S2, thus generating an AC voltage across the output. This configuration is used in place of the full bridge where lower output voltage or output power is desired.

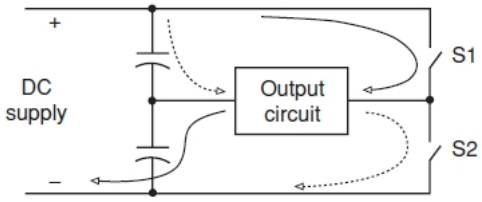


Figure 2.39: Half-bridge inverter - Basic scheme of a half-bridge inverter [22].

Figure 2.40 shows the principal design features of the inverter configurations most commonly used in induction heating power supplies. The two major types are the voltage-fed and the current-fed. The figure further subdivides each of these by the DC source (fixed or variable), the mode of inverter control, and the load-circuit connection (series or parallel).

Voltage-fed inverters with simple series load Voltage-fed inverters are distinguished by the use of a filter capacitor at the input of the inverter and a series-connected output circuit as shown in the simplified power circuit of Figure 2.41. The voltage-fed inverter is used in induction heating to generate frequencies from 90 Hz to as high as 1 MHz. Thyristors, which are also called silicon-controlled rectifiers (SCRs), can be used to switch the current at frequencies below 10 kHz. Below 50 kHz, insulated gate bipolar transistors (IGBTs) are commonly used. Above 50 kHz power, metal-oxide semiconductor field-effect transistors (MOSFETs) are chosen for their very fast switching speeds.

Voltage-fed inverter with series connection to a parallel load A popular variation of the voltage-fed inverter for induction heating has an internal series connected inductor and capacitor that couple power to a parallel resonant output or tank circuit as shown in Figure 2.42. The values of the internal series inductor and capacitor are selected to be resonant above the operating or "firing" frequency of the inverter with an impedance at this firing frequency that will allow sufficient current to flow from the bridge to permit full-power operation. A very important feature of this style of inverter is that the internal series circuit isolates the bridge from the load. This protects the inverter from load faults caused by shorting or arcing and from badly tuned loads, making it one of the most robust thyristor-based induction power supplies available for heat

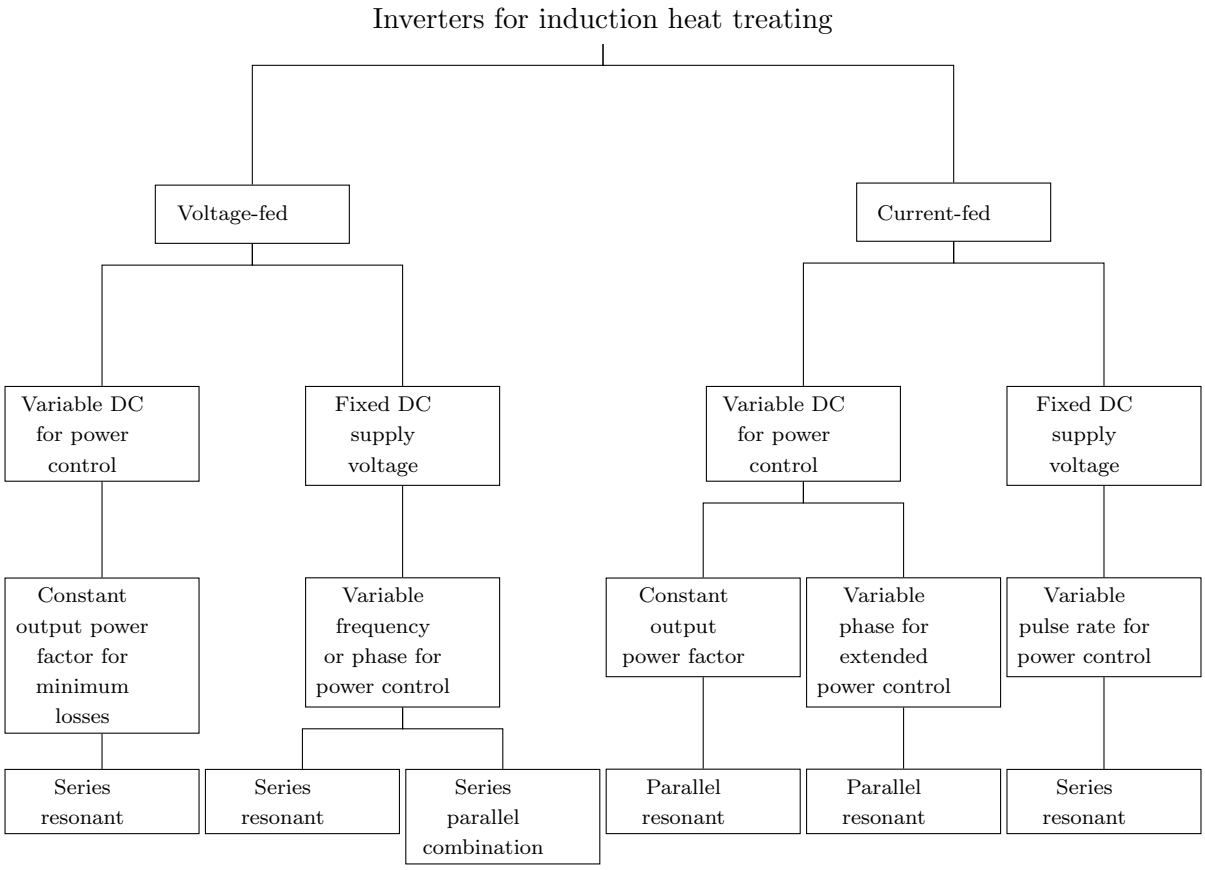


Figure 2.40: Inverter classifications - Tree chart of induction heat treatment inverters [22].

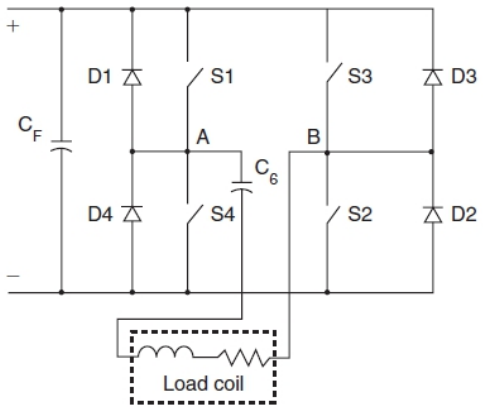


Figure 2.41: Voltage-fed inverter - Voltage-fed series-connected output inverter type [22].

treatment. A second feature of this series-parallel configuration is realized when the internal series circuit is tuned to the third harmonic of the firing frequency. The power supply is then capable of developing full power into the parallel tank circuit tuned to either the fundamental firing frequency or the third harmonic. Because load current is not used for commutation, this system can be operated with the output shorted for easy troubleshooting. The voltage-fed inverter with series connection to a parallel load commonly uses thyristors for power switching in the bridge and has an unregulated DC input supply. The regulation of output power is accomplished by varying the firing frequency relative to the parallel load resonant frequency.

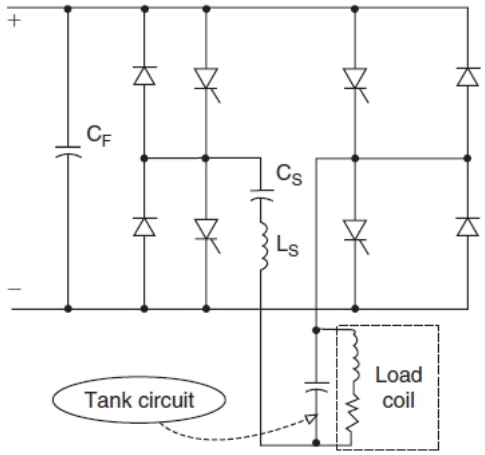


Figure 2.42: Voltage-fed inverter (parallel load) - Voltage-fed inverter with series connection to parallel load [22].

Current fed inverters Current-fed inverters are distinguished by the use of a variable voltage DC source followed by a large inductor at the input of the inverter bridge and a parallel resonant load circuit at the output as shown in the simplified power circuit of Figure 2.43. Current-fed inverters are available in models that cover the entire 90 Hz to 1 MHz range of frequencies used for induction heat treatment. Thyristors are commonly used below 10 kHz, whereas transistors are chosen for the higher frequencies.

Load matching A very important facet of induction heat treatment that is often overlooked in the initial design stages is the ability to successfully deliver to the work-

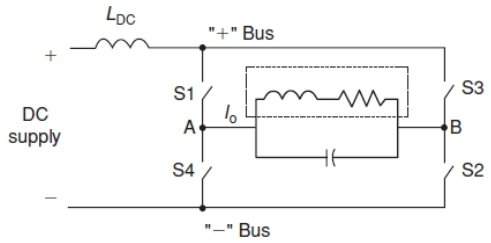


Figure 2.43: Current-fed inverter - Current-fed full bridge inverter [22].

piece the maximum available power from a given power supply at the minimum cost. Circumstances do not always allow for an optimal design of a complete induction heat treatment system in which the power supply design is based on the application including the specific induction coil parameters. Quite often, the induction coil is designed to achieve the desired induction heat treatment pattern without regard for the power supply that will be used. When this is the case, a flexible interface is required to match the output characteristics of the power supply to the input characteristics of the induction coil and work-piece combination [25, 26]. If this match is not provided, the power supply will not be able to develop its rated power if the coil requires more voltage or current than the supply can deliver. There are many factors involved, any of which can cause complications in arriving at the stated goal. Variable ratio transformers, capacitors, and sometimes inductors are connected between the output of the power supply and the induction coil. The adjustment of these components is commonly referred to as load matching or load tuning.

Understanding load matching for solid-state power supplies The most common example of matching a power source and load would be a simple lighting circuit application, where a 6 V light bulb is available for use on a 120 V_{AC} power line (see Figure 2.44). Obviously, there is a need for some type of interface hardware to prevent 120 V_{AC} from destroying the light bulb. This would commonly be accomplished by inserting a transformer between the light bulb and the power line. Induction heating circuits have not only a resistive element but also considerable inductance. As part of the electric circuit, any inductor can be introduced as a combination of resistance and reactance (inductance). Both the resistance and the reactance of the inductor are non-linear functions of several parameters such as coilto-work-piece geometry, material

properties, and frequency. Furthermore, the electrical resistivity and magnetic permeability of metals are non-linear functions of the temperature and vary during the heating cycle. In addition, modern metalworking processes require that work-pieces of different sizes are heated in the same inductor. The combinations of production mix and variations of material properties result in changing coil resistance and reactance, which affects the tuning and performance of the power supply. Generally speaking, a change in coil resistance and reactance results in a change of the phase angle between the coil voltage and coil current of a given circuit. Such a change can be characterized by the coil power factor, which refers to the cosine of the phase angle ($\cos\varphi$). Power factors of different types of inductors are affected differently by the various factors. Along with the power factor, the quality factor (or "Q-factor") is widely used as an induction circuit characteristic, as it refers to the ratio of the inductive reactance to the resistance of the coil (in other words, it represents the ratio of the energy stored to the energy dissipated in the system). In addition to these factors, the process itself usually requires that the part be heated at some frequency other than the line frequency. In conventional heat treatment, the applied frequency typically ranges from 200 Hz to 400 kHz. Since a relatively large current is required to successfully heat a work-piece, it is necessary to build power sources with extremely high output current capability or to use a simple resonant circuit to minimize the actual current or voltage requirement of the frequency converter.

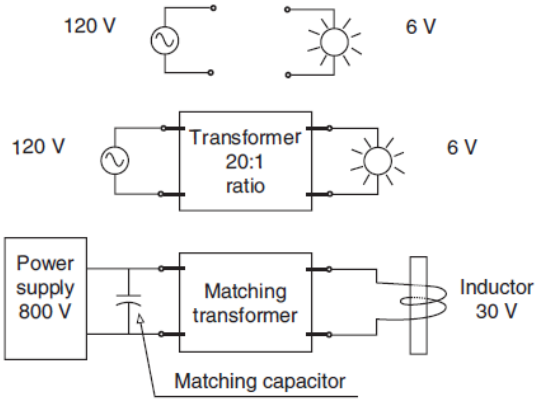


Figure 2.44: Load matching - Example of load tuning and impedance matching techniques [22].

As previously shown, there are two basic types of resonant frequency converters

that use parallel and series resonant circuits (see Figure 2.45).

Looking first at the parallel circuit in Figure 2.46b, it is easy to see that if the capacitor value is equal to zero, then a given voltage applied to the circuit at a fixed frequency will result in a specific amount of power dependent on the circuit impedance. When sufficient capacitance is added to the circuit to tune the load circuit near resonance, the circuit impedance rises dramatically and the amount of current drawn from the power source falls off dramatically. The circuit voltage required to achieve a specific power level is the same as it was in the initial case of zero capacitance, but now the higher current required by the load is supplied by the capacitors rather than the power source. In a parallel-tuned load circuit, we have a Q rise in current in the tank circuit compared with the input line from the power source.
This analogy can be repeated for the case of the series circuit in Figure 2.46a. With the calculated change in circuit impedance, the circuit current will be much higher for a given input voltage when the circuit is tuned near the resonant frequency because the impedance is approaching zero. The load coil current required for a given power is the same for the given load circuit regardless of whether the connection is series or parallel, but as the overall impedance has fallen and the required current is fixed, the required driving voltage is approximately a factor of Q lower than the coil voltage.

Hence, we have a Q rise in current in the parallel circuit and a Q rise in voltage with the series-connected circuit. It is therefore imperative to have an understanding of what type of circuit connection exists in order to understand the effect that tuning changes will have on the power source and workstation components.

A solid-state RF power supply is similar to an SCR inverter in that line voltage enters the system through an isolation transformer and is then changed to a DC voltage, the level of which is adjustable through thyristor rectifiers. This filtered DC output is the input for the solid-state RF supply's power boards. The MOSFETs on each board and the boards themselves are connected in parallel. Frequency sensing from the tank circuit feeds the tank resonant-frequency signal back to the microprocessor. The MOSFETs in turn are driven at this frequency by the microprocessor. The software will also change the system operating frequency in response to changes in tank resonant frequency. As does the load-resonant inverter, the solid-state RF system operates at the peak of the resonant curve, and its efficiency is therefore similar. Similarly, the software for these microprocessor-driven systems can be modified to provide constant

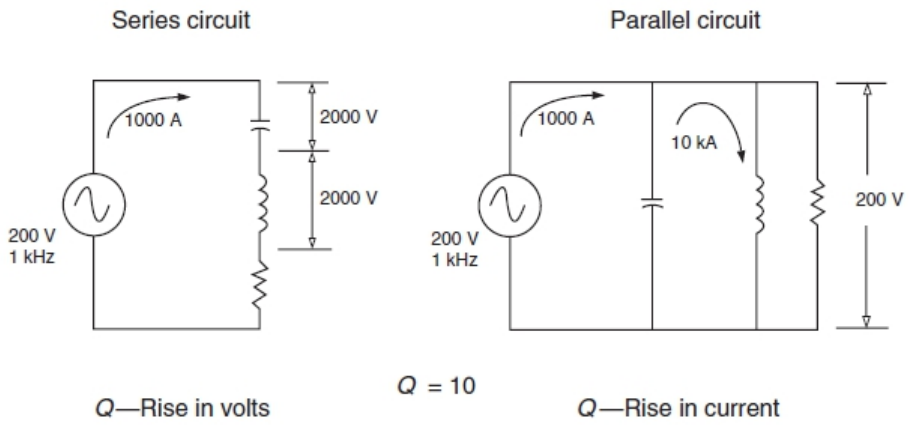
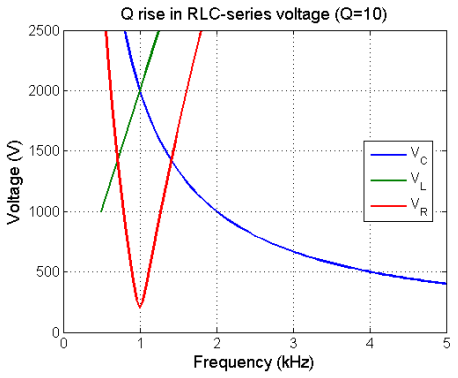
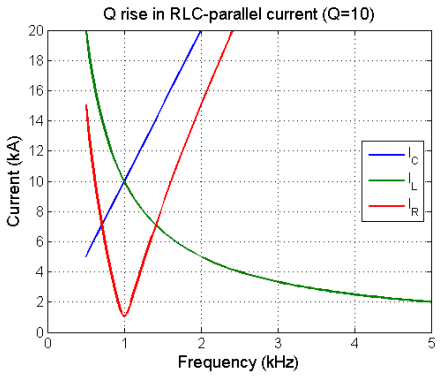


Figure 2.45: Types of resonant circuits - Series and parallel resonant circuits [22].



(a) Voltage rise in RLC-series resonant circuit.



(b) Current rise in RLC-parallel resonant circuit.

Figure 2.46: Behaviour at resonance - Voltage and current rise for RLC resonant circuits at the frequency of resonance.

current, constant power, or constant voltage, depending on the conditions required for the particular application. The software also permits interrogation of the system through a computer-keyboard interface.

3

Induction Hardening of Gears

3.1 Mechanical gears

Gears are machine elements that transmit rotary motion and power by the successive engagements of teeth on their periphery. They constitute an economical method for such transmission, particularly if power levels or accuracy requirements are demanding. Gears have been used for more than three thousand years and they are now used as an important element in many types of machinery used in current times. Gears range in size from few micrometers (e.g., micro-electro-mechanical systems, MEMS) to many meters (e.g., wind power generators) and can be made out of many different materials, from plastics to ultra high-strength heat-treated steels [27].

A mechanical gear (no matter what type) is essentially a toothed wheel or cylinder that works in tandem with another gear (or gears) to transmit motion, change speed or direction. When transmitting power from a source to the point of application requires a compact, efficient, and high-speed drive, gear trains offer a competitive and suitable solution, featuring reversibility, flexibility, and robustness [28].

3.1.1 Gear classification

A variety of different geometries of gear can be found, mainly depending on the specifications and requirements of the selected application. A rough classification is presented in the following.

Spur gears

The *spur gear* is the most basic mechanical power transmission product. Basically, in every application where a shaft turns, a spur gear exists. In a spur gear, the teeth, which are on the outer surface of the cylinder, are straight and parallel to the hole (or axis) so when two come together, or *mesh*, they do so in the same plane (see Figure 3.1) As a result of how they meet, spur gears can increase or decrease the speed or torque of whatever they are moving. A number of variations of the basic spur gear exist, including pinion wire, stem pinions, rack and internal gears.

Figure 3.1: Spur gear - Example of spur gear set.

Pinion wire is a long wire or rod that has been drawn through a die so that gear teeth are cut into its surface. It can be made into small gears with different face widths, hubs, and bores (see Figure 3.2).

Figure 3.2: Pinion wires - Example of pinion wires.

Stem pinions are bore-less spur gears with small numbers of teeth cut on the end of a ground piece of shaft. They are especially suited as pinions when large reductions are desired (see Figure 3.3).

Racks are yet another type of spur gear. Unlike the basic spur gear, racks have their teeth cut into the surface of a straight bar instead of on the surface of a cylindrical blank (see Figure 3.4).

Figure 3.3: Stem pinion - Example of stem pinion.

Figure 3.4: Racks - Example of racks.

Internal gears have their teeth cut parallel to their shafts like spur gears, but they are cut on the inside of the gear blank (see Figure 3.5).

Figure 3.5: Internal gear - Example of internal gear.

Helical gears

Helical gears are similar to spur gears except that their teeth are cut at an angle (known as the helix angle) to the hole (axis) rather than straight and parallel to the hole like the teeth of a spur gear. Helical gears are used to connect non-intersecting shafts. Standard helical gears with 45-degree helix angles are used to connect parallel shafts or shafts at right (90-degree) angles. Helical gears are manufactured as both right and left-hand gears. The teeth of a left-hand helical gear lean to the left when the gear is placed on a flat surface. The teeth of a right-hand helical gear lean to the right when placed on a flat surface. Opposite hand helical gears run on parallel shafts.

Gears of the same hand operate with shafts at 90-degrees (see Figure 3.6).

Figure 3.6: Helical gear - Example of right-hand helical gear set.

Worms and worm gears

Worms are a type of gear with one or more cylindrical, screw-like threads (also referred to as "start") and a face width that is usually greater than its diameter. A worm gear has a center hole (bore) for mounting the worm on a shaft.
Worm gears, like worms, also are usually cylindrical and bored for mounting on a shaft. The diameter of a worm gear, however, is usually much larger than the width of its face. Worm gears differ from spur gears in that their teeth are somewhat different in shape and they are always formed on an angle to the axis to enable them to mate with worms (see Figure 3.7). Worms and worm gears work in sets, rotating on shafts at right angles to each other, in order to transmit motion and power at various speeds and speed ratios. The worm usually drives the worm gear. Accordingly, the worm gear is usually the driven member. In worm and worm gear sets, both the worm and worm gear are of the same hand. Right-hand gearing are considered standard, thus right-hand sets will always be assumed unless otherwise specified.

Figure 3.7: Worm gear - Example of worm and worm gear set.

Bevel and miter gears

When an application calls for the transmission of motion and/or power between shafts that intersect at right angles (90-degrees), bevel gears and miter gears are often the reference choice. A *bevel gear* is shaped like a section of a cone, its teeth being straight or spiral. If they are spiral, the pinion and gear must be of opposite hand in order for them to run together. Since bevel gears provide a ratio and they are used to reduce speed, the pinion always has fewer teeth (see Figure 3.8).
Miter gears are identical to bevel gears except that in a miter gear set, both gears always have the same number of teeth. Their ratio, therefore, is always 1:1. As a result, miter gears are not used when an application calls for a change of speed.

Figure 3.8: Bevel gear - Example of bevel gear set.

3.1.2 Gear nomenclature

This section will review some of the most common terms used in the gear industry to describe gears design and geometry. The reader can refer to Figures 3.9-3.10 for more understanding.

Active profile The part of the gear tooth profile that actually comes in contact with the profile of the mating gear while in mesh.

Addendum The height of the tooth above the pitch circle.

Backlash The amount by which the width of a tooth space exceeds the thickness of the engaging tooth on the operating pitch circle.

Base circle The circle from which the involute tooth profiles are generated.

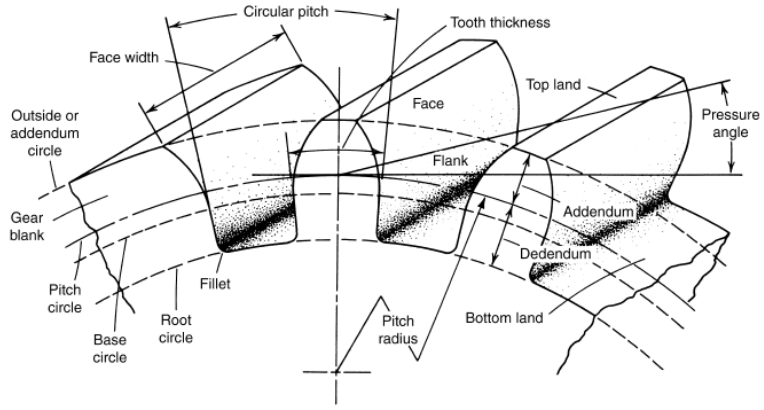


Figure 3.9: Gear nomenclature - Schematic of typical gear nomenclature [27].

Bottom land The surface at the bottom of a tooth space adjoining the fillet.

Center distance The distance between the axes of rotation between two mating gears.

Circular pitch Length of arc of the pitch circle between corresponding points on adjacent teeth.

Circular thickness The length of arc between the two sides of a gear tooth at the pitch circle.

Dedendum The depth of the tooth below the pitch circle.

Diametral pitch (DP) A measure of tooth size in the English system. In units, it is the number of teeth per inch of pitch diameter. As the tooth size increases, the diametral pitch decreases. Diametral pitches range from 0.5 to 200. Coarse pitch gears are those with DP$\leq$20. Fine pitch gears are those with a DP$>$20.

Face width The length of the gear teeth in an axial plane.

Fillet radius The radius of the fillet curve at the base of the gear tooth.

Gear A geometric shape that has teeth uniformly spaced around the circumference. In general, a gear is made to mesh its teeth with another gear.

Gear blank The work-piece used for the manufacture of a gear, prior to machining the gear teeth.

Gear pinion When two gears mesh together, the smaller of the two is called the pinion; the larger is called the gear.

Gear ratio The ratio of the larger to the smaller number of teeth in a pair of gears.

Helix angle The angle between any helix and an element of its cylinder. In helical gears and worms, it is at the standard pitch circle unless otherwise specified.

Involute gear tooth A gear tooth whose profile is established by an involute curve outward from the base circle.

Normal section A section through a gear that is perpendicular to the tooth at the pitch circle.

Pitch circle The circumference of a gear measured at the point of contact with the mating gear.

Pitch diameter The diameter of the pitch circle.

Pitch line In a cross section of a rack, the pitch line corresponds to the pitch circle in the cross section of the gear.

Pitch point The tangency point of the pitch circles of two mating gears.

Pitch radius The radius of the pitch circle in a cross section of gear teeth in any plane other than a plane of rotation.

Pressure angle The angle between a tooth profile and a radial line at its pitch point. The pressure angle of an involute gear tooth is determined by the size ratio between the base circle and the pitch circle. Common pressure angles used by the gear industry are 14.5, 20, and 25°.

Rack A rack is a gear having a pitch circle of infinite radius. Its teeth lie along a straight line on a plane. The teeth may be at right angles to the edge of the rack and mesh with a spur gear, or the teeth on the rack may be at some other angle and engage a helical gear.

Root circle The root circle coincides with the bottoms of the tooth spaces.

Tooth thickness The thickness of the tooth measured at the pitch circle.

Top land The surface of the top of a tooth.

Transverse section A section through a gear perpendicular to the axis of the gear.

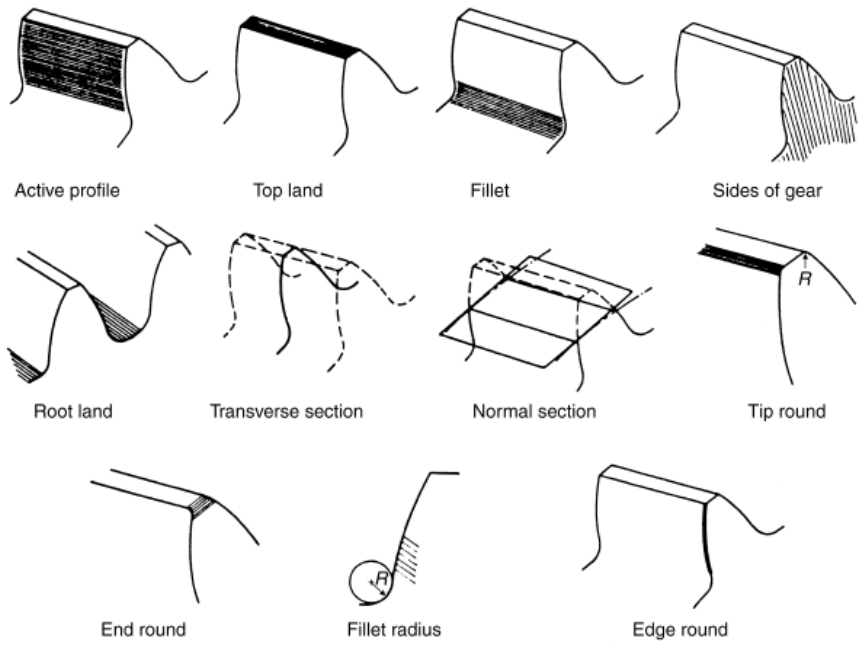


Figure 3.10: Tooth nomenclature - Schematic of typical tooth nomenclature [27].

3.2 Manufacturing process of gears

Gear manufacturing has been one of the most complicated of the metal cutting processes. In fact, manufacturing of gears presents a demanding challenge for metallurgists in heat treatment, for supervisors in machining and gear cutting, and for quality engineers in keeping the quality to the required standards. In the last decades, gear manufacturing process dynamics are undergoing a major breakthrough, aiming at cutting the number of operations required to obtain a work gear within the final specifications (in terms of dimensions, tooth form quality, operational noise, etc.).
In *soft gear* process dynamic, the gear teeth are generated by gear hobbing or shaping, depending on the component design constraints. Soft finishing of gear teeth is carried out by gear shaving, rolling or grinding to attain the gear quality grade. Even after the heat treatment deterioration, the quality specification remains well within the desired

final specification to meet product final performance requirements.
In *hard gear* process dynamic, hobbed and/or shaped, or warm forged/rolled gears after heat treatment undergo final finishing operation, such as hard finishing, honing, or grinding.
Obviously, overall economy becomes the deciding factor for selection of the process dynamic: the most competitive manufacturing choice will be the one which allows the optimum combination of the production related aspects with the operation related ones. The earlier accounts for the manufacturing costs, whereas the later for the product quality [29].

3.2.1 Raw material for transmission gears

Gears are generally designed for a finite life. Alloy steels are most favoured gear material. Case-hardening steels provide the ideal features required for gear material. For gear teeth, the surface is to be hard with soft and tough core to provide wear and fatigue resistance. Case-hardening steels do have varying chemical composition and are named accordingly (e.g., chrome steel, low-molybdenum steel, chrome-molybdenum steel, nickel-chrome-molybdenum steel). Basic requirements of good gear materials may be summarised as follows:

- well controlled hardenability[1], that helps in getting consistent and predictable result after heat treatment;
- least non-metallic inclusions, especially oxides that generally present machining difficulties;
- good formability for better forge die life and consistency of forge quality;
- high and consistent machinability;
- low and stabilised quenching distortion;
- no grain growth during present practice of high temperature carburizing, which can cause higher quenching distortion and lower toughness.

[1]Hardenability is the property of a steel that determines the depth and distribution of the hardness induced by quenching.

Among the through-hardening steels in wide use are AISI 1040, 1060, 4140, and 4340. These steels can also be effectively case hardened by induction heating. Among the carburizing steels used in gears are AISI 1018, 1524, 4026, 4118, 4320, 4620, 4820, 8620, and 9310 (AMS 6260).

Most gears are made of carbon and low-alloy steels, including carburizing steels and the limited number of low-alloy steels that respond favorably to nitriding. In general, the steels selected for gear applications must satisfy two basic sets of requirements that are not always compatible: those involving fabrication and processing and those involving service. Fabrication and processing requirements include machinability, forgeability, and response to heat treatment as it affects fabrication and processing. Service requirements are related to the ability of the gear to perform satisfactorily under the conditions of loading for which it was designed and thus encompass all mechanical-property requirements, including fatigue strength and response to heat treatment.

In all gears the choice of material must be made only after careful consideration of the performance demanded by the end-use application and total manufactured cost, taking into consideration such issues as machining economics. Key design considerations require an analysis of the type of applied load, whether gradual or instantaneous, and the desired mechanical properties, such as bending fatigue strength or wear resistance, all of which define core strength and heat treating requirements.

Different areas in the gear tooth profile see different service demands. Consideration must be given to the forces that will act on the gear teeth, with tooth bending and contact stress, resistance to scoring and wear, and fatigue issues being paramount. For example, in the root area, good surface hardness and high residual compressive stress are desired to improve endurance, or bending fatigue life. At the pitch diameter, a combination of high hardness and adequate subsurface subsurface strength are necessary to handle contact stress and wear and to prevent spalling. Numerous factors influence fatigue strength, including:

- hardness distribution, as a function of case hardness, case depth, core hardness;
- micro-structure, as a function of retained austenite percentage, grain size, carbides (size, type, distribution), non-martensitic phases;
- defect control, as a function of residual compressive stress, surface finish, geometry, inter-granular toughness.

Surface hardening steels

Carburizing or nitriding grades of steel are commonly specified where maximum wear resistance is required for bearing surfaces. Such grades usually correspond to plain carbon or alloy steels, with low carbon content. The progressive diffusion of induction heating as a viable option for surface hardening applications has broadened the material choice range, which now includes also medium carbon steels. Case-hardened gears are best suited for heavy-duty service, for example, transmission gears, and offer high resistance to wear, pitting, and fatigue. Surfaces must be sufficiently hard to resist wear and of sufficient depth to prevent case crushing. A rough rule for case depth is that it shall not exceed one-sixth of the base thickness of the tooth. A case-hardened gear provides maximum surface hardness and wear resistance and at the same time provides interior toughness to resist shock. In general, case-hardened gears can withstand higher loads than through-hardened gears, although the latter are quieter and less expensive because of the simpler heat treatment required [30].

When selecting a case-hardened gear, the following factors must be considered :

- high tooth pressures will crack a thin case;
- too soft a core will not provide proper backing for a hard case; compressive stresses in the case improve fatigue durability, and a high case hardness increases wear resistance;
- if the ratio of case depth to core thickness is too small, excessive stresses in subsurface layers can produce poor fatigue life;
- residual tensile stresses are highest with low core hardness and increase with increasing case depth; these stresses can be relieved by tempering.

Grain size variations have an important effect on core properties. These variations are influenced by the type of steel and the method of heat treatment used subsequent to carburizing. Section thickness also influences core properties. A tough tooth core may not be required in applications where a gear will not be subjected to impact loading. In these applications, core properties are relatively unimportant, provided the core is sufficiently hard to support the case. Considering the case alone, it is important that the surface resist wear and fatigue bending, because bending stresses vary from a maximum at the surface to zero near the tooth center.

3.2.2 Gear manufacturing process

Gears can be made by a variety of manufacturing processes, which will be briefly reviewed in this section. The type of gear being machined (spur, helical, bevel, or other) is usually the major factor in the selection of machining process, although one or more of the following factors usually must be considered in the final choice of the method:

- size of the gear;
- configuration of integral sections (flanges or other);
- quantity requirements;
- accuracy requirements;
- gear-to-pinion ratio;
- cost.

For a more detailed description considering the type of gear as the major variable, the reader may refer to [27].

Metal removal processes

As shown in Figure 3.11, gear blanks can be shaped by a number of cutting (machining) and finishing processes. Often blanks are processed by a series of rough cutting and finishing operations. Machining and finishing processes make up about 60% of gear manufacturing costs (see Figure 3.12).

Broaching is a machining operation which rapidly forms a desired contour in a work-piece by moving a cutter, called a broach, entirely past the work-piece. The broach has a long series of cutting teeth that gradually increase in height. The broach can be made in many different shapes to produce a variety of contours. The last few teeth of the broach are designed to finish the cut rather than to remove considerably more metal. Broaches are often used to cut internal gear teeth, racks, and gear segments on small gears, and usually are designed to cut all teeth at the same time.

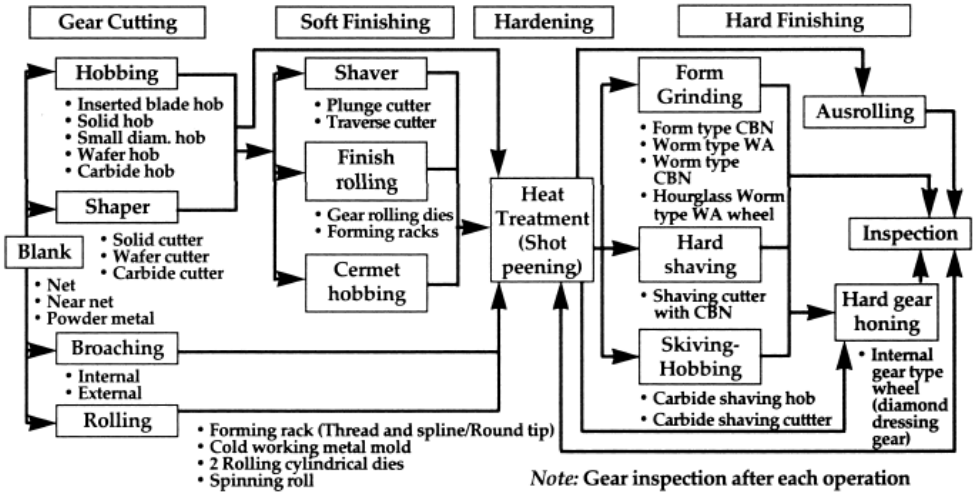


Figure 3.11: Gear manufacturing - Overview of various gear manufacturing processes [31].

Grinding is a process that shapes the surface by passes with a rotating abrasive wheel. Grinding is not a practical way to remove large amounts of metal, so it is used to make very fine-pitch teeth, or to remove heat treat distortion from large gears that have been cut and then fully hardened. Many different kinds of grinding operations are used in gear manufacture.

Hobbing is a gear cutting method that uses a tool resembling a worm gear in appearance, having helically spaced cutting teeth. In a single-pitch hob, the rows of teeth advance exactly one pitch as the hob makes one revolution. With only one hob, it is possible to cut interchangeable gears of a given pitch of any number of teeth within the range of the hobbing machine.

Honing is a low-speed finishing process used chiefly to produce uniform high dimensional accuracy and fine finish. In honing, very thin layers of stock are removed by simultaneously rotating and reciprocating a bonded abrasive stone or stick that is pressed against the surface being honed with lighter force than is typical of grinding.

Lapping is a polishing operation that uses abrasive pastes to finish the surfaces of gear teeth. Generally a toothed, cast iron lap is rolled with the gear being finished.

Milling is a machining operation which removes the metal between two gear teeth by passing a rotating cutting wheel across the gear blank.

Shaping is a gear cutting method in which the cutting tool is shaped like a pinion. The shaper cuts while traversing across the face width and rolling with the gear blank at the same time.

Shaving is a finishing operation that uses a serrated gear-shaped or rack-shaped cutter to shave off small amounts of metal as the gear and cutter are meshed at an angle to one another. The crossed axes create a sliding motion which enables the shaving cutter to cut.

Skiving is a machining operation in which the cut is made with a form tool with its face so angled that the cutting edge progresses from one end of the work-piece to the other as the tool feeds tangentially past the rotating work-piece.

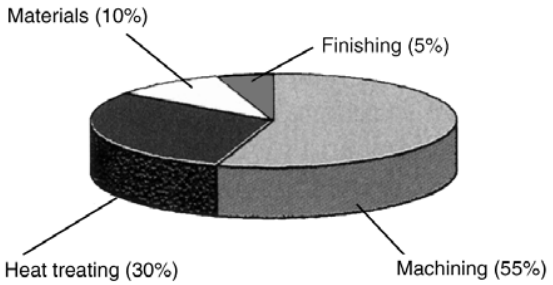


Figure 3.12: Gear manufacturing costs - Typical partitioning of gear manufacturing costs; manufacturing represents about 55% of the total, heat treatment about 30% [27].

Casting, forming, and forging processes

Casting is a process of pouring or injecting molten metal into a mould so that the metal solidifies and hardens into the desired shape. Casting is often used to make gear blanks that will have cut teeth. Small gears are frequently cast complete with teeth by the die casting process, which uses a precision mould of tool steel and low-melting-point alloys for the gears.

Stamping is a fast inexpensive method of producing small gears from thin sheets of metal. The metal is sheared by a punching die which stamps through the sheet stock into a mating hole.

Gear rolling is a process which rapidly shapes fine gear teeth or worm threads by high pressure rolling with a toothed die.

Powder metallurgy (PM) processing consists of pressing a powder to the desired shape, followed by heating (sintering) at an elevated temperature below its melting point. There are a number of variations of the PM process that are applicable to gears. The PM process is suitable for high volume production of small gears. It is not economical for low-to-medium volume production.

Injection moulding is a method of forming a plastic to the desired shape by forcing the heat softened plastic into a relatively cool cavity (die) under pressure. It is widely used for high volume production of thermoplastic resin (e.g., acetals and nylons) gears. Often lubricants are added to the thermoplastic material to further improve the inherent lubricity of the material.

Forging has long been used to create blanks that will be subsequently shaped into gears by metal removal methods. However, it is increasingly being used for the production of near-net shape and net shape gears for demanding applications where great strength and durability are required. The forging process is carried out hot with metal preheated to a desired temperature under intense pressure until it fills the die cavity. The resultant grain flow which smoothly follows gear tooth contours makes forged gears stronger than those made by other processes.

Forming processes, including forging and blank machining, are preliminary to the real manufacturing chain, which is dedicated to the creation of the desired tooth profile as well as other compulsory geometrical details.
Forging involves the shaping of metal using localized compressive forces, which deform the internal grains maintaining their continuity throughout the part and giving rise to a piece with improved strength characteristics. Uniform grain flow also reduces distortion during subsequent heat treatments. Hot forging is a manufacturing process most commonly used for gears. Cold/warm forgings are high production methods used to produce gear blanks with much better dimensional control and considerable material saving. New cold forging methods produce a neat finished gear profile combining forming with rolling. A good forging is a necessity: a faulty forging may compromise any effort put in excellent design and careful manufacturing of gears, even with the best

available material. Machinability, ultimate strength, final quenching distortion, and surface finish will all be affected by the forging practices.
Quality of gear manufacturing starts with *blank machining*. Accuracy in blank machining is a necessity for attaining the desired quality standard of finished gears.

Whether the manufacturing of gears consists of many different phases, the rather general flow chart shown in Figure 3.15 shall give an idea of the production steps involved. In practice, gear manufacturing processes can be grouped in two categories.

1. Category one relates to teeth cutting, finishing and all necessary operations related to gear tooth profiles (e.g., hobbing, shaping, shaving, honing, etc.).

2. Category two relates to the rest of the conventional machining (e.g., drilling, milling, grinding, etc.).

Gear cutting

Gear hobbing and shaping are the most commonly used cutting processes used for generating the gear teeth. Basis for selection of either of the two depends on application.
Hobbing is a machining process for making gears, splines, and sprockets on a hobbing machine, which is a special type of milling machine. The teeth or splines are progressively cut into the work-piece by a series of radial cuts made by a cutting tool called a hob. Compared to other gear forming processes it is relatively inexpensive and quick, especially for gears with larger face width or when stacking multiple gears with narrow face width. Hobbing cannot be used for internal gears.
Shaping produces a series of straight lines parallel to the axis of the gear. This technique is generally slower and more accurate than hobbing, also providing a better surface finishing. Shaping can be used for internal gears.

Gear shaping In the process of *gear shaping* (see Figure 3.13), a gear of desired tooth profile with cutting capability can generate the similar tooth profile in a blank and produce a gear suitable for meshing with any gear of interchangeable series. The cutter with a particular number of teeth on its periphery rotates in the correct ratio required for generating the desired number of teeth on the rotating blank. The rotations of the cutter and the work gear are in opposite direction for external gear and in same

direction for internal gear. The cutter simultaneously reciprocates parallel to the tooth profile of the work gear. The distance between the cutter and work gear axes gradually reduces till the final size (pitch circle diameter) of the gear being generated is reached. Cutting may occur in downward stroke or in upward stroke. When cutting occurs in downward stroke, it is called down cutting or push shaping. When cutting occurs in upward stroke, it is termed as up cutting or pull shaping.

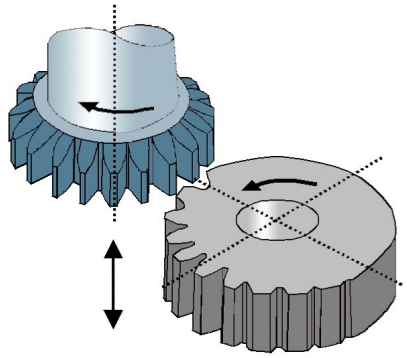


Figure 3.13: Gear shaping - Schematic illustration.

Gear hobbing In *gear hobbing*, a worm like cutter known as hob with cutting teeth having the basic reference profile of a rack cuts teeth on a cylindrical blank. Successive hob teeth come in contact with each tooth in the gear blank and generate gear tooth by producing a large number of flats that envelop the tooth profiles (see Figure 3.14). Hob is tilted according to the hob thread angle and helix angle on the gear teeth, to align the hob teeth with the teeth of the gear to be cut. A single thread hob generates one tooth space in one turn of its rotation. The hob and the blank rotate in a constant timed relation to each other that depends on the number of thread of the hob and the number of teeth on the work gear. The hob moves radially to the desired depth of the teeth a little clear from the blank and then feeds axially along the width of the gear teeth.

Edge finishing

During shaping and hobbing, burrs appear at the exit side of the cutter. Process sequence, tool design, tool replacement frequency and cutting parameters are some of

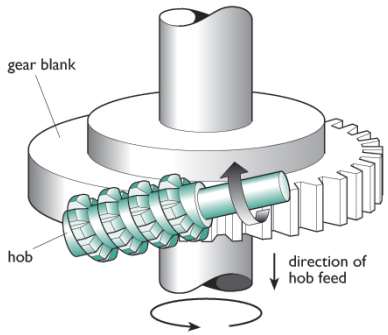


Figure 3.14: Gear hobbing - Schematic illustration.

the major factors that decide the type and strength of the burr produced. Minimisation of burr and placing them in the best position for easy removal is the main target of a manufacturing engineer. *Deburring* may seem to be a simple operation to start with, but it is really difficult, particularly so in gear manufacturing. When the burr is being removed from one side it tries to move in a different direction. As a principle, deburring must be carried out just after the operation where the burrs have been generated.
Along with the chances of injury in manual handling, the presence of burrs may damage the tooth surfaces during transportation and produce nicks. Mostly these nicks are created on the gear teeth by the sharp edges of gears themselves from accidental hits during in-process handling. These nicks cause meshing defects and are observed as sudden and very high deviation of profile or lead during double flank roll testing. Moreover, the sharp edges tend to become super-carburized during heat treatment. Because of excessive brittleness, these edges break off and cause further harmful damages (or noise problems) during actual operation in transmission. Besides removing burrs, deburring also produces a control-sized chamfer. Chamfering the gear tooth represents an essential procedure to protect the active flanks of the gear from damages.

Soft gear finishing

Soft finishing process is planned to achieve the desired quality grade, surface finish and tooth modifications of gear teeth to meet the requirements of specific application. Again, surface roughness produced by hobbing/shaping processes will cause noise and also defective rolling. So gear finishing operation becomes necessary to ensure better and consistent surface finish on the profiles.

Tooth modifications are essential for many reasons. True involute trace for teeth may not be ideal for gears running at high speeds. Undesirable noise is caused by impact load during tooth meshing even for normal no-load condition. Tooth profile is to be modified to reduce the impact loading. Again, in case of heavy loading, the teeth tend to bend. The bent tooth enters mesh before its mating tooth is in the proper location to receive it, and produces interference and noise. The entering tooth is to be modified to the same amount that the loaded tooth has deflected out of position. Again, the gear teeth may require certain modification at soft stage to compensate for distortion during heat treatment. The amount of modification required is established by continuous monitoring of heat treatment effect on lead and profile of gears.
Gear shaving, gear roll finishing, and gear grinding are the processes applied by various manufacturers for soft gear finishing. Gear shaving is the most extensively used process. Shaving is used for the gears of up to 40 HRC. It has been established that shaved gears perform frequently with lesser noise problems than the ground gears. Shaving is also superior in respect of macro-geometric and micro-geometric characteristics of the resulting gear profiles. It does not produce heat checks that are generally produced in grinding.

Gear shaving *Gear shaving* is basically a low pressure, free-cutting process. A helical gear-like cutter with closely spaced grooves extending from the tip to the root of each tooth, rotates with gear in close mesh in both directions during the shaving cycle. The centre distance between the gear and the cutter is reduced in small controlled steps to remove metal from the gear tooth surfaces till the final required size is achieved. The helix of the cutter is different from that of the gear to be shaved. Shaving removes the cutter marks, waviness and surface irregularities of the pre-shave gear generating process. Surface finish of shaved gears may be even as good as that after grinding. Tooth size is maintained as specified within a closer tolerance. Tooth quality is improved depending on the nature of the gear tooth error. Profile and lead accuracy are remarkably improved. Base pitch error and the difference between the adjacent pitches are reduced greatly. However, the gear may still have a greater cumulative pitch error, if the concentricity during gear cutting has not been controlled carefully. The profile corrections, such as tip relief or root relief, are obtained by modifying the tooth profile of the shaving cutter according to the requirement. For certain methods of

shaving, all modifications are attained through the modifications of shaving cutter only. These modifications compensate for misalignment in final transmission assembly and for heat treatment distortions as well as produce the desired tooth bearing for uniform load distribution. Gear noise is reduced and load carrying capacity is increased. Under favourable conditions, shaved gears take four times as much load as hobbed gears in high speed transmissions.

Gear roll finishing In *gear roll finishing*, a soft gear is meshed with the rolling die and rotated under pressure for finishing by plastic deformation occurring simultaneously along contact line. As a gear rolling die tooth engages the approach side of a gear tooth, sliding action occurs along the line of action in the arc of approach in a direction from the top of the gear tooth toward the pitch point where instantaneous rolling action is achieved. When the contact leaves the pitch point, sliding occurs now in the opposite direction towards the pitch point in the arc of recession. The contact between the teeth of die and gear on the trail side produces exactly the opposite direction of sliding to that on approach side. As a result, the material is compressed towards the pitch point on the approach side and is extended away from the pitch point on the trail side. This action causes a greater quantity of material to be displaced on the trail side than on the approach side. On the approach side, the tendency is to trap the material. While on the trail side, the tendency is to permit it to flow towards the top and root of the teeth. Obviously, the material stock allowance for finish rolling and hardness of the material is extremely critical and influences the accuracy and quality of gear roll finishing to a very large extent. In roll finishing the metal flows and smoothens the surface. There is no metal removal as in gear shaving. Some obvious advantages of this roll finishing process are:

- extremely short cycle time (5-8 s);
- excellent surface finish;
- improved tooth strength;
- better and uniform dimensional control.

Heat treatment

Heat treating is one of the most important steps in the manufacture of precision gearing. Its contribution is vitally important for cost control, durability, and reliability. As shown in Figure 3.12, heat treating represents about 30% of a typical gear manufacturing cost. If not properly understood and controlled, it can have a significant impact on all aspects of the gear manufacturing process. For transmission gears, the core of the tooth is to be soft and ductile for impact absorption without breakage during actual running whereas the surface of the tooth should be hard enough to resist wear. Surface hardening process includes:

- carburizing or induction heating;
- quenching;
- tempering.

Hard gear finishing

Heat treatment distorts the bore, deforms the outside diameter, causes deviations from the soft finished dimensions and geometry. Tooth flank becomes rough because of scale formation in heat treatment, shot blasting and shot peening. Carburizing usually results in a drop in profile and increase in pressure angle. Amount of drop varies with pitch and case depth. Carburizing also causes unwinding or decrease in the helix angle. For thicker gears, the unwinding will be less. All these variations produced during heat treatment generate the necessity of finishing the hardened gear to bring back the dimensional accuracy as well as tooth quality up to the required specifications. Finishing of hardened gears comprises of:

- grinding of external or internal bearing surfaces;
- finishing of gear tooth flank.

Superfinishing

Superfinishing processes are those that can achieve a surface roughness average (R_a) of few hundreds of nanometres [32]. Superfinishing can be achieved under carefully

controlled conditions by grinding, honing/micro-honing, lapping, polishing/electro-polishing, or combinations of finishing processes (e.g., grinding, honing, and lapping). In the past, such processing was rarely carried out on gears because of the time and expense involved. Recently, however, there have been a number of studies that show that the reduction of surface roughness improves the lubricating condition of gears and offers the possibility of increasing their surface fatigue lives [33].

3.3 Induction surface hardening

Induction surface hardening is a method of hardening the surfaces of components, usually in selected areas, by the short time application of high-intensity heating followed by quenching. The heating and hardening effects are localized, and the depth of hardening is controllable. Unlike thermochemical case hardening treatments (carburizing, nitriding, and carbonitriding) applied to steels, induction hardening do not promote chemical enrichment of the surface with carbon or nitrogen but rely on the presence of an adequate carbon content already in the material to achieve the hardness level required. The properties of the core remain unaffected and depend on material composition and prior heat treatment.

Generally speaking, the goal of surface hardening of components made of steels or cast irons is to create a martensitic layer with specific depth in the work-piece. The martensitic layer provides a hard and wear-resistant surface while the core material is unaffected and remains relatively soft and ductile. Since the electromagnetic skin depth in a given material varies with its electrical resistivity, magnetic permeability and applied frequency, it is almost possible to heat specific areas of a piece of metal without affecting neighbouring regions.

Induction surface hardening has been successfully used on most gear types (spur, helical, bevel, etc.) when gear teeth require high hardness and bending strength, but either size and configuration of the part or process requirements (e.g., cost, efficiency, repeatability, automation, etc.) does not lend itself to carburizing and quenching the entire part.

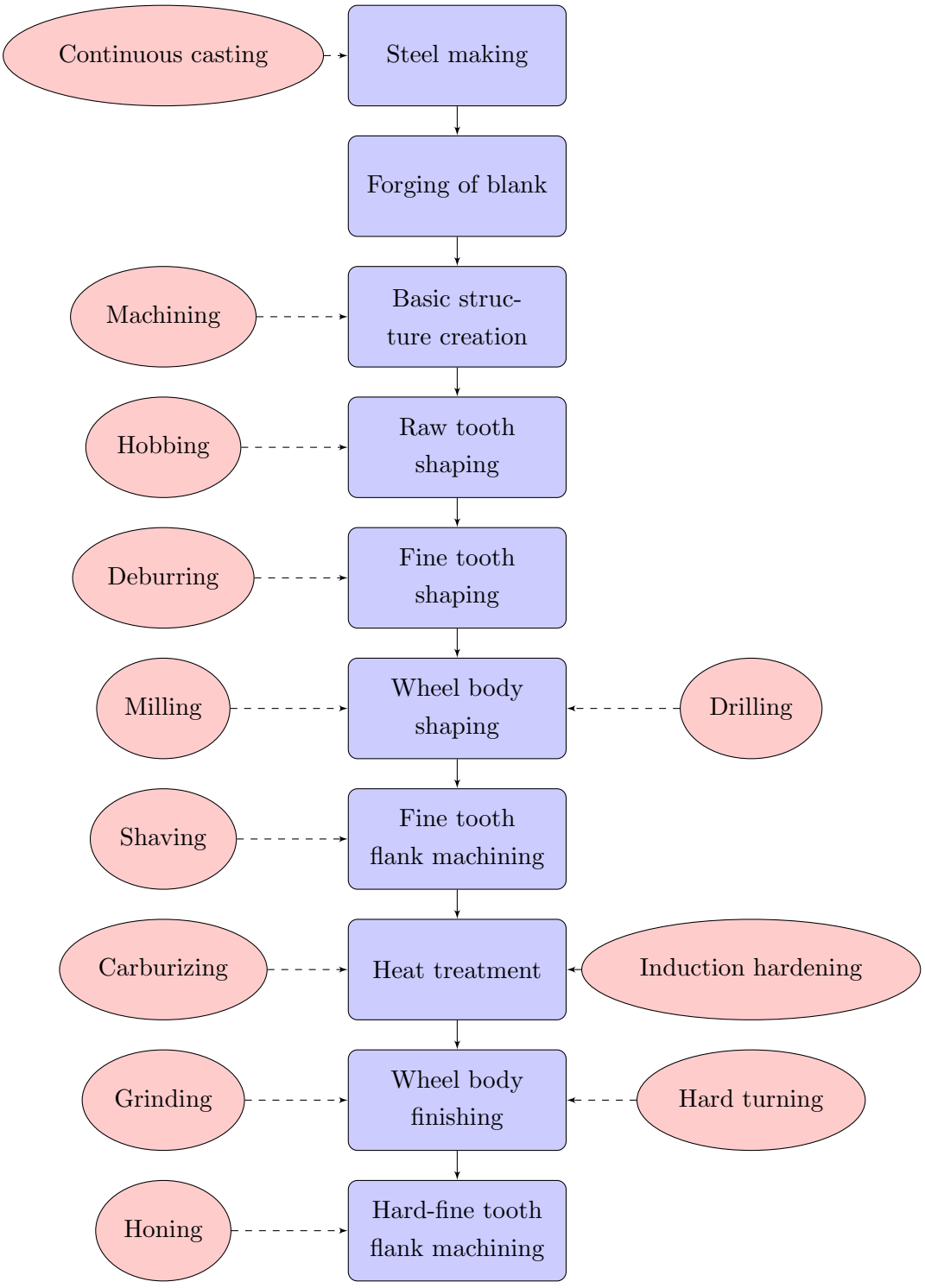


Figure 3.15: Gear process chain - Flow chart of the manufacturing chain of gears [34].

3.3.1 Hardening of gears

Together with crankshafts, gears have been among the parts earlier considered for induction heat treatment, thanks to the intrinsic advantages of this technology, namely short heating times, flexible adaptation to the hardening tasks, strict process control, and compatibility with production line implementation and automation. However, the lack of precise knowledge about the interrelation between all the concurrent physical phenomena occurring within the part during the heating cycle has almost restricted its use to mass-production items. For years, the main demand has come from the automotive industry, where many components made from steels and cast irons require surface hardening at high production rate (e.g., crankshafts, camshafts, axle shafts, gear wheels, hubs, etc.). More recently, the benefits of this technology, which is clean, repeatable, and cost-effective, have pushed its introduction into more conservative industry sectors, such as aerospace, where furnace-based treatments (e.g., carburizing) have always represented the golden standard. In this case, the limiting factor is related to the optimization of the induction hardening process, which usually requires significant material know-how and can thus be very long and expensive [35]. Furthermore, due to their complex geometry and variety of materials, types, and sizes, gear hardening represents one of the most challenging tasks to be performed in practice.

Though different patterns of hardened layer may be acceptable, depending on particular application, it was found relatively early that the best service properties correspond to a hardness layer more or less uniform along all the surface of the gear, repeating its contour. However, breakage in the hardened layer may be accepted on the tips of the teeth, where mechanical loading is very low. On the flanks and roots of the teeth the hardened layer must be continuous in order to withstand big contact pressure and tensile stress from loading.

Contour hardening of the whole gear by using an encircling inductor may be achieved only in a narrow range of process parameters, namely specific power, frequency, and heating time. The optimal frequency value depends upon the gear geometry and material, hardness layer depth and pattern, and tooth profile: a too high value of frequency will lead to hardening of tooth tips only, whereas a too low value will mainly affect tooth roots (see Figure 3.16). In addition to frequency, heating time and specific power have to be accurately selected: in general, the smaller the gear module is, the greater

the amount of energy to be provided is (i.e., higher power densities with lower heating times). This is because the small mass of the tooth tends to be heated up very fast and thermal diffusion from the surface have to be prevented in order to achieve a satisfactory contour heating.

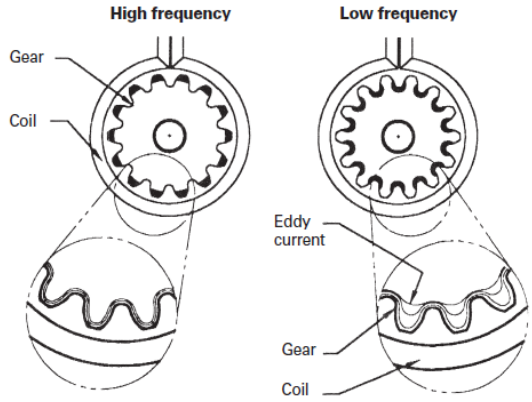


Figure 3.16: Frequency selection for gear surface hardening - Influence of frequency on induced current density distribution: higher frequencies mainly affect the tooth tip, lower frequencies focus on the root [36].

Single shot hardening

The simplest arrangement for gear hardening consists of using a simple inductor encircling the outer radius of the gear. Thanks to its very easy design, such kind of inductor represents an efficient and flexible mean for providing a single shot heat treatment of gears. In order to achieve a uniform heating pattern, the gear is rotated during the treatment, thus minimising the effect of magnetic field inhomogeneities in the area of the coil leads. Small gears (M=2) are generally treated with high frequencies in the range 300-500 kHz; however, such gears present a hardened layer which is deeper at the tooth tips than in the root area. Additional flexibility may be achieved by means of power delivery profiling techniques, such as pulse single frequency or dual frequency hardening.

Pulse single frequency hardening consists of a pre-heating stage at lower power followed by one or more higher power pulses. During the first stage, a moderate rise in temperature is produced on the whole tooth, because of the low power and relatively long time. Once the temperature has been raised up to around 500 °C (932 °F), high power

pulses are provided during the second stage, inducing the austenite transformation of the surface layers only, thanks to the high frequency used.

Dual frequency hardening takes advantage of two different generators, each one providing a power source at a different frequency, usually in the range 5-30 kHz (medium frequency, MF) and 100-450 kHz (high frequency, HF). The system may consist either of two separate inductors connected to each generator (this is referred to as "double frequency" or "DF" technology) or a single inductor simultaneously fed by both generators ("simultaneous dual frequency" or "SDF" technology).

In the DF approach, the gear is first moved into the MF inductor, where roots are heated to a temperature close to the Curie point, then quickly transferred into the HF inductor, where flanks and tips reach the required transformation temperature, and finally into a quenching device, typically a spray ring, where cooling is performed. It should be pointed out that the second heating cycle is much shorter than the first one, mainly due to the very high power densities needed to further increase the steel temperature above the Curie point. DF heating may also be carried out without transferring the gear from one inductor to another. In this case, the inductor is made of two half-rings, connected to a MF power source on the one side and to a HF power source on the other. Initial heating takes place at medium frequency with HF contact short-circuited. Then high frequency is applied with short-circuited MF connection.

The SDF systems use both medium and high frequency flowing simultaneously in the same coil. Two power supplies are connected in parallel via a system of filters and feed the coil. The main advantages of this method are its high flexibility, short heating cycles, high production rates, and true contour surface hardening of complex shaped gears. The success of the SDF technology is mainly related to the development of special high power generators (up to 3 MW combined MF/HF power) and optimal power delivery to the part.

Dual frequency heating is more flexible than heating at optimal frequency. In fact, varying times of heating, MF/HF power ratios and delays in application of the process make it possible to treat gears of different modules on the same installation. Moreover, with correct selection of frequencies and power level, optimum hardening results with minimum power consumption may be achieved.

Tooth-by-tooth hardening

Tooth-by-tooth (or gap-by-gap) hardening is a process where each gear tooth or valley is hardened individually. These technique can be realised by applying a single-shot or scanning mode and is typically not very suitable for small and fine-pitch gears (i.e., M<6). In case of big non-heavy loaded gears, it is acceptable to harden only tooth flanks. Coil geometry depends upon the shape of the teeth and the required hardness pattern and is required to be symmetrical between the flanks of two adjacent teeth. Such an inductor can be designed to heat only the root and/or flank of the tooth, leaving the tip and tooth core soft, tough and ductile. Simple loop inductors have always been used for this purpose, where magnetic flux concentrators made of laminations are employed to increase efficiency and avoid overheating of the tooth corner on the ends of the gear.

In order to produce "contour" hardening of each valley, other types of inductors must be used. When the production rate is not high, it is possible to use a scanning inductor made of two copper plates profiled in the form of a tooth valley and a central longitudinal conductor connecting these plates. A magnetic concentrator on the central conductor is essential for this type of inductor, so as to increase the power density induced in the root area. Inductors of this type produce a "heat print" in a form of a butterfly (see Figure 3.17). Thus, in order to distribute the temperature uniformly within gear roots and flanks, scanning mode is used. This strategy is particularly beneficial for treating large face width gears, where a uniform contour hardness layer is produced by moving the inductor along the thickness of the part. Particular attention should be paid to electromagnetic end/edge effects and the ability to provide the required pattern in the gear face areas (gear ends) as well as along the tooth perimeter. In fact, pattern uniformity is quite sensitive to coil positioning and local under or overheating (if not even arcing) is possible with a misplaced inductor.

After hardening of adjacent valleys, a narrow non-hardened strip remains along the tooth tips (see Figure 3.18). The main concern when applying tooth-by-tooth hardening techniques is undesirable heating of the areas adjacent to the hardened zone, the so-called *back tempering*. The main reason why undesirable tempering occur is related to thermal conductivity effects: when heating one side of the tooth to temperatures above Ac_3, there is a danger that the opposite side of the gear tooth will be heated by

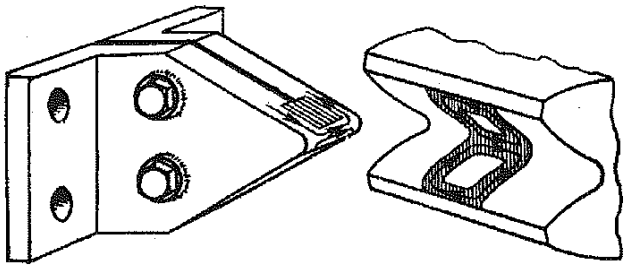


Figure 3.17: Butterfly inductor - Original layout of the butterfly inductor for tooth-by-tooth hardening of gears; coil current flow direction and induced power deposition on the part [3].

conduction to an inappropriately high temperature, resulting in undesirable tempering back of previously hardened areas (see Figure 3.19). Whether a hardened side of the tooth will be softened due to back tempering depends upon several factors, including the applied frequency, gear module, tooth shape, heat time, case depth and other pattern features. In the case of shallow and moderate case depths and large teeth, the root of the tooth, its fillet, and the bottom of the tooth flank are typically not overheated, since the massive area below the tooth root serves as a heat sink, which helps to conduct excessive heat and protects the hardened side of the tooth from back tempering. Conversely, the tooth tip and top of the tooth flank can be considered a troublesome area as back tempering takes place due to the relatively small mass of metal at the tooth tip, which make heat diffuse very fast from the heating side to the already hardened side of the tooth.

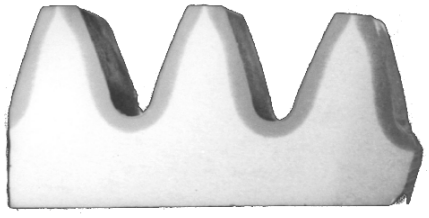


Figure 3.18: Tooth-by-tooth hardening pattern - Hardening of tooth flanks usually leaves a narrow non-hardened strip on tooth tips [36].

In order to overcome the problem of back tempering, additional cooling blocks can be used. For gears of big modules, water spray chambers are used as well as special

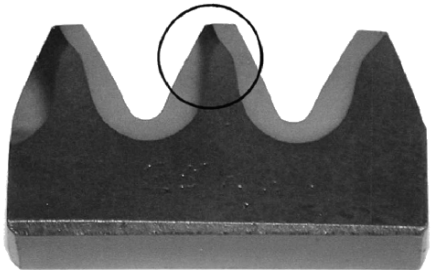


Figure 3.19: Back tempering problem - Back-tempered hardening pattern due to processing of consecutive teeth [36].

cooling devices to prevent back tempering of the previously hardened zone on the back side of the tooth. Extra cooling protects already-hardened areas while heating unhardened areas of the gear (see Figure 3.20). For gears of small module and diameter, the process takes place in water or oil, which serves as quench and as cooling media in order to prevent back tempering. In this case, both the gear and the inductor are completely submerged into liquid and a steam blanket forms between the coil and gear surface, thus reducing thermal losses to acceptable values.

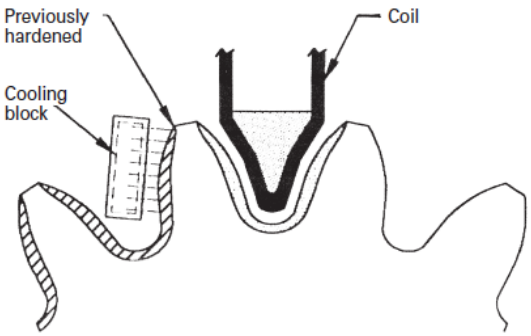


Figure 3.20: Back tempering solution - Hardening scheme for overcoming back tempering by using specific cooling blocks on previously-treated teeth [36].

4

The Physics behind Induction Hardening

4.1 Eddy current heating

4.1.1 Electromagnetic problem

The technique of calculating electromagnetic fields depends on the ability to solve Maxwell's equations, which are a set of equations that describe the interrelated nature of electric and magnetic fields. Maxwell's theory, first proposed in 1864 and then published in 1873 with the famous "A treatise on electricity and magnetism," is a generalization of the work done by Faraday, Ampère and Gauss and represents the second great unification in physics, after the first one realised by Newton. For general time-varying electromagnetic fields Maxwell's equations in differential form are:

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \tag{4.1a}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \tag{4.1b}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{4.1c}$$

$$\nabla \cdot \mathbf{D} = \rho_c \tag{4.1d}$$

where $\mathbf{E}$ is the electric field intensity (V/m), $\mathbf{D}$ is the electric flux density (C/m^2), $\mathbf{H}$ is the magnetic field intensity (A/m), $\mathbf{B}$ is the magnetic flux density (Wb/m^2), $\mathbf{J}$ is the elctric current density (A/m^2), and ρ_c is the electric charge density (C/m^3).

Maxwell's equations (4.1) are in indefinite form, because the number of equations is less than the number of unknowns. When considering a uniform, isotropic, conductive, and linear medium, the following additional constitutive relations hold true:

$$\mathbf{D} = \varepsilon_0 \varepsilon_r \mathbf{E} \tag{4.2a}$$

$$\mathbf{B} = \mu_0 \mu_r \mathbf{B} \tag{4.2b}$$

$$\mathbf{J} = \sigma \mathbf{E} \tag{4.2c}$$

where ε_r is the relative permittivity, ε_0 is the permittivity of free space ($\approx 1.257 \cdot 10^{-6}$ H/m), μ_r is the relative magnetic permeability (H/m), μ_0 is the magnetic permeability of free space ($\approx 8.854 \cdot 10^{-12}$ F/m), and σ is the electrical conductivity (S/m) which is the inverse of the electrical resistivity ρ (Ω·m). Relative permittivity and relative permeability values respectively indicate the ability of a given material to conduct electric and magnetic flux better than air or vacuum.

For induction heating applications, where the typical frequency range is below 1 MHz, the contribution of the electric flux density variation $\frac{\partial \mathbf{D}}{\partial t}$ to the total current is negligible compared to that of the conduction current $\mathbf{J}$. Thus, also the relative impact of such displacement currents on total heat generation within the part can be neglected in (4.1a).

When several regions with different magnetic and/or electric properties are exposed to an electromagnetic field, certain components of such field can be continuous or discontinuous on different sides of the interfaces between each region. In the simple case of two different and contiguous media, as could be the case of a ferromagnetic billet exposed to a magnetic field propagating in air, the following boundary conditions are satisfied:

$$(\mathbf{B}_2 - \mathbf{B}_1) \cdot \mathbf{n} = 0 \tag{4.3a}$$

$$(\mathbf{D}_2 - \mathbf{D}_1) \cdot \mathbf{n} = \rho_{cs} \tag{4.3b}$$

$$(\mathbf{H}_2 - \mathbf{H}_1) \times \mathbf{n} = \mathbf{J}_s \tag{4.3c}$$

$$(\mathbf{E}_2 - \mathbf{E}_1) \times \mathbf{n} = 0 \tag{4.3d}$$

where $\mathbf{n}$ is the normal vector to the boundary in each point and $\mathbf{J}_s$ is the surface current density. Equations (4.3) can be interpreted as follows:

1. the normal component of the magnetic flux density is continuous across the interface between the two media;

2. the normal component of the electric flux density is in general discontinuous across the interface by a factor proportional to the superficial charge density ρ_{cs}; if no superficial charge density is present, then continuity of the electric flux density's normal component is assured;

3. the tangential component of the magnetic field strength are in general discontinuous across the interface by a factor proportional to the superficial current density $\mathbf{J}_s$; if no surface current density is present, then continuity of the magnetic field's tangential component is assured;

4. the tangential components of the electric field intensity are continuous across the interface between the two media.

Because the magnetic flux density $\mathbf{B}$ satisfies a zero divergence condition (i.e., the magnetic field lines always form continuous loops without having any origin or termination points), it can be expressed in terms of the so-called magnetic vector potential $\mathbf{A}$, as follows:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{4.4}$$

By substituting (4.2) and (4.4) in (4.5), it is possible to derive the following equation:

$$\nabla \times \mathbf{E} = -\nabla \times \frac{\partial \mathbf{A}}{\partial t} \tag{4.5}$$

which, after integration, provides the expression for the electric field intensity:

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \tag{4.6}$$

where V (V) represents the electric scalar potential. In a similar way, the expression for the electric current density can be derived from (4.2c):

$$\mathbf{J} = \sigma \mathbf{E} = -\sigma \frac{\partial \mathbf{A}}{\partial t} + \mathbf{J}_{source} \tag{4.7}$$

where $\mathbf{J}_{source}$ (A/m^2) represents the source current density (e.g., coil current). In this way it is possible to combine Maxwell's equation (4.1) and constitutive relations (4.2) into a single expression which is a function of the magnetic vector potential only. In the

simple case of linear, isotropic material, such general formulation takes the following form:

$$\frac{1}{\mu_0\mu_r}(\nabla \times \nabla \times \mathbf{A}) = \mathbf{J}_{source} - \sigma\frac{\partial \mathbf{A}}{\partial t} \tag{4.8}$$

Assuming a quasi-stationary (time-harmonic) nature for the applied electromagnetic field oscillating at the angular frequency ω (rad/s), the above equations become, after some vector algebra:

$$\frac{1}{\sigma}\nabla^2\mathbf{H} = j\omega\mu_0\mu_r\mathbf{H} \tag{4.9a}$$

$$\frac{1}{\mu_r}\nabla^2\mathbf{E} = j\omega\sigma\mu_0\mathbf{E} \tag{4.9b}$$

$$\frac{1}{\mu_0\mu_r}\nabla^2\mathbf{A} = \mathbf{J}_{source} + j\omega\sigma\mathbf{A} \tag{4.9c}$$

4.1.2 Thermal problem

As described in the previous section, the steady-state assumption of the electromagnetic field quasi-stationary (time-harmonic) nature with an angular frequency ω is widely used for describing the phenomena related to eddy currents. Thanks to its comprehensive formulation, it allows to simulate the great majority of induction heat treating applications without appreciably compromising the accuracy of the computation and minimizing the computation time. However, the same assumption cannot be

The transient heat transfer process in a metal work-piece can be described by means of the Fourier equation:

$$c\gamma\frac{\partial T}{\partial t} + \nabla \cdot (-k\nabla T) = Q \tag{4.10}$$

where T is temperature (°C), γ is the density of the metal (kg/m^3), c is the specific heat (J/kg·°C), k is the thermal conductivity (W/m·°C), and Q is the heat source density produced by eddy currents per unit time in a unit volume (W/m^3). Such heat source density is obtained as a result of solving the electromagnetic problem and assumes the form $Q = \rho J^2$. Both k and c are generally non-linear functions of temperature, so assuming a constant value for any of them would result in appreciable errors in predicting the right induced power density pattern and predicting the correct final heating distribution.

Equation (4.10), with suitable initial and boundary conditions, is able to predict 3-D temperature distribution at any time and at any point within the work-piece. The *initial*

temperature distribution is usually assumed uniform and set to ambient temperature value; however, there are some cases where the effect of previous technological processes have to be taken into account as a non-uniform, residual heat pattern within the work-piece. Thermal *boundary* conditions represent the combined effect of surface conduction and the heat losses due to both convection and radiation effects. This is described by the following equation:

$$-k\frac{\partial T}{\partial n} = \alpha(T_s - T_a) + c_s(T_s^4 - T_a^4) + Q_s \tag{4.11}$$

which expresses the point-wise thermal gradient normal to the boundary as a function of the convective heat transfer coefficient α (W/m^2·°C), the radiative heat loss coefficient c_s (W/m^2·°C^4), and the surface conductive losses Q_s (W/m^2) resulting from the contact with cooler or hotter parts. In equation (4.11) the surface and ambient temperature values are associated to T_s and T_a, respectively. It should be pointed out that while tabulated values for α can be easily found in literature for many different materials, even in function of temperature, the coefficient c_s is usually associated with the emissivity of the given material, which is expressed as the ratio c_s/k_B, where k_B is the Stefan-Boltzmann constant ($k_B \approx 5.670 \cdot 10^{-8}$ W/m^2·°C^4).

4.1.3 Coupling of electromagnetic and thermal problems

One of the major features of induction heat treating computation deals with the fact that both the electromagnetic and heat transfer phenomena are tightly coupled due to the interrelated non-linear nature of the material properties. This feature requires special computational algorithms, able to deal with such interrelated effects, to be developed. It is important to note that the time constants typical of each one of these two phenomena are quite different: electromagnetic processes are very fast, with time constants ranging between milliseconds and microseconds; conversely, thermal processes are much longer, typically in the range from seconds to a few minutes. Specific applications dealing with very high frequencies (i.e., 100-450 kHz) and short heating times (i.e., 1-5 s), like gear contour hardening, may even scale by several orders of magnitude such difference [37].

Two-step coupling

There are several ways to couple the electromagnetic and heat transfer problems. The simplest method is called the two-step coupling approach (see Figure 4.1). According to this technique, the electromagnetic problem is solved during the first computational step; then both the obtained current and power density distributions are used to determine the heat sources to be used for solving the thermal problem. The main assumption of this method is that the electromagnetic physical properties of the material (i.e., electrical resistivity and magnetic permeability) are constant during the whole heating phase. The main advantages of the two-step procedure are its short execution time and moderate memory requirements, together with a straight-forward implementation. The principal limitation is its inability to take into account the modifications of material properties in the course of heating, which is a crucial point in those applications where a transition across the Curie temperature is needed. Therefore, the two-step approach can only be used for low-temperature (i.e., in the range 100-300 °C) heating of aluminium or copper strips, rods, or wires.

Indirect coupling

The most common approach to coupled electromagnetic and heat transfer problems is called the indirect coupling method (see Figure 4.2). This technique calls for an iterative process which consists of multiple elementary solving sequences for the electromagnetic and heat transfer problems. Practically, for each time-step of the solution, Maxwell's and Fourier's equations are solved multiple times, iteratively exchanging the results of calculations and accordingly updating material physical properties. This solving loop is repeated until a given precision of results (i.e., a sufficiently small variation of the residual) is reached. The main assumption here is that temperature variations are not significant during a certain interval of time predetermined by the user. Depending on the application, at least five iterations are typically required to make corrections in the electromagnetic field distribution via recalculation of the heat sources. For the great majority of induction heating applications, the indirect coupling is highly effective. Of course, the accuracy of the results are greatly affected by the chosen time-stepping for the solution and the required precision for the convergence of each step.

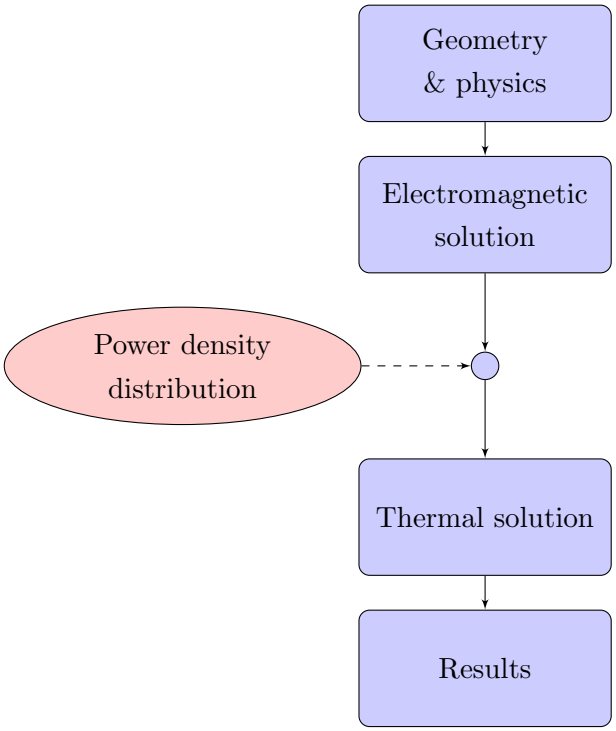


Figure 4.1: Two-step coupling - Schematic chart of the two-step coupling approach for simulating induction heating problems [38].

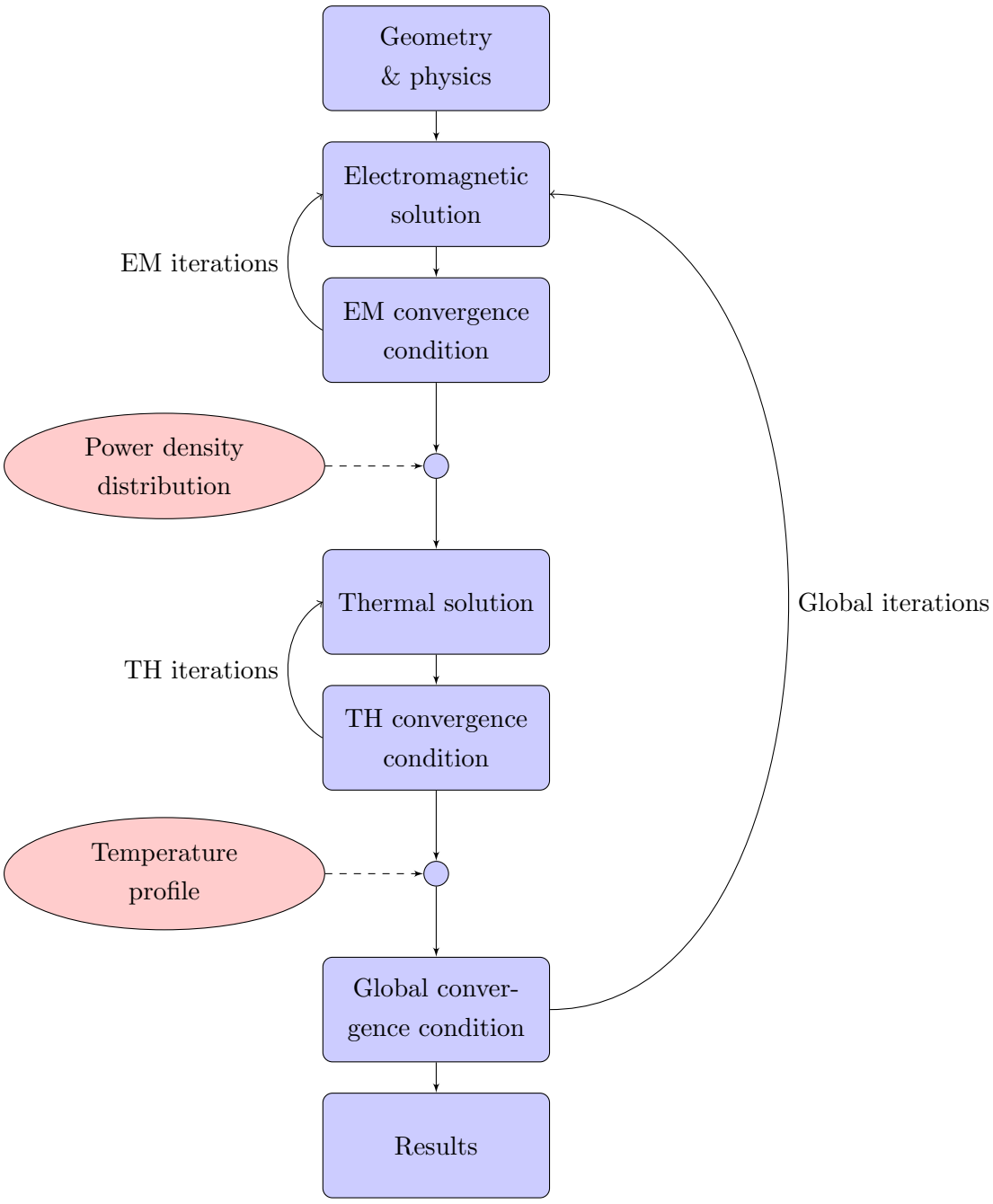


Figure 4.2: Indirect coupling - Schematic chart of the indirect coupling approach for simulating induction heating problems [38].

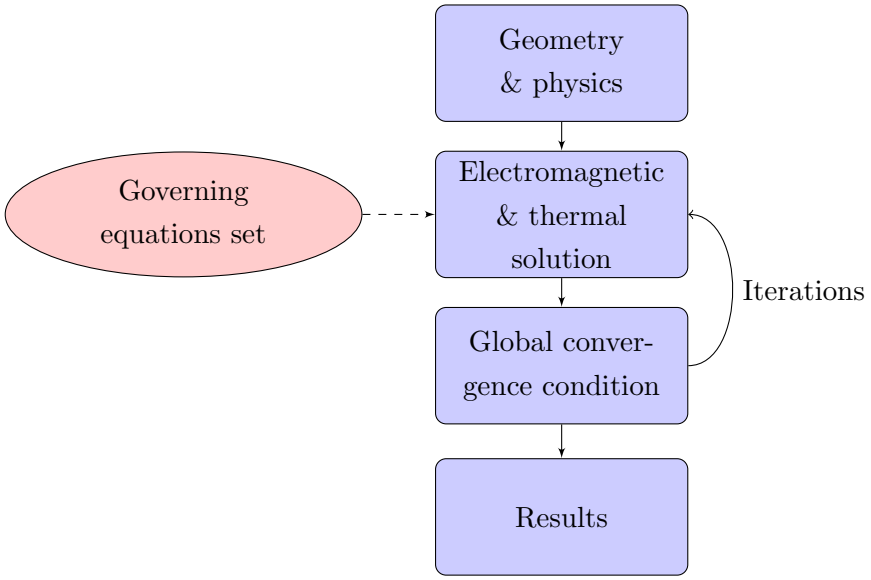


Figure 4.3: Direct coupling - Schematic chart of the direct coupling approach for simulating induction heating problems [38].

Direct coupling

The direct coupling of the electromagnetic and heat transfer problems (see Figure 4.3) could represent the final approach for the accurate simulation of induction heating problems as it would allow the simultaneous solution of the governing equations on a real multi-physics basis. The main issue however, is represented by the necessity of formalizing a consistent and compact set of global equations comprising both the electromagnetic and thermal problem unknowns in a single notation. At present, such a kind of approach has not yet been implemented in a general form, mainly due to its intrinsic complexity and extremely high computational cost.

4.2 Phase transitions in steel

Surface hardening of a steel part consists of raising a surface layer above the transformation temperature (denoted by Ac_3 on the iron-iron carbide phase diagram) at which it will be transformed to austenite and rapidly cooling the part to produce a hard martensitic structure in this region. Design of surface-hardening treatments demands consideration of the work-piece material and its starting condition, the effect of rapid heating on Ac_3 or Ac_{cm} temperature, property requirements, and equipment

selection. Induction surface hardening is applied mostly to hardenable grades of steel, although some carburized and slow-cooled parts are often reheated in selected areas by induction heating. Some typical induction surface hardened steels are:

- medium-carbon steels, such as 1030 and 1045, used for automotive drive shafts, gears, and so forth;
- high-carbon steels, such as 1070, used for drill and rock bits, hand tools, and so forth;
- alloy steels used for bearings, automotive valves, and machine-tool components.

4.2.1 Critical temperatures

The critical temperatures that must be considered to understand the principles of the different heat treatments are those that define the onset and completion of the transformation to or from austenite. For a given steel, the critical temperatures depend on whether the steel is being heated or cooled. Critical temperatures for the start and completion of the transformation to austenite during heating are denoted, respectively, by Ac_1 and Ac_3 for hypo-eutectoid steels and by Ac_1 and Ac_{cm} for hyper-eutectoid steels. These temperatures are higher than the corresponding critical temperatures for the start and completion of the transformation from austenite during cooling, which are denoted, respectively, by Ar_3 and Ar_1 for hypo-eutectoid steels and by Ar_{cm} and Ar_1 for hyper-eutectoid steels[1]. These critical temperatures converge to the equilibrium values Ae_1, Ae_3, and Ae_{cm} as the rates of heating or cooling become infinitely slow. The positions of the Ae_1, Ae_3, and Ae_{cm} lines are close to the more general (i.e., near equilibrium) A_1, A_3, and A_{cm} lines on the iron-iron carbide binary phase diagram shown in Figure 4.4. Various alloying elements markedly affect these critical temperatures. For example, chromium raises the eutectoid temperature, A_1, and manganese lowers it. According to [39], it is possible to calculate upper and lower critical temperatures using the actual chemical composition of the steel. The following equations will give an approximate critical temperature for a hypo-eutectoid steel (the presence of other alloying elements will also have marked effects on these critical temperatures):

$$Ac_1(^\circ\mathrm{C}) = 723 - 20.7(\%Mn) - 16.9(\%Ni) + 29.1(\%Si) - 16.9(\%Cr) \qquad (4.12a)$$

[1]The "c" and "r" in the symbols are derived from the French words *chauffage* for heating and *refroidissement* for cooling.

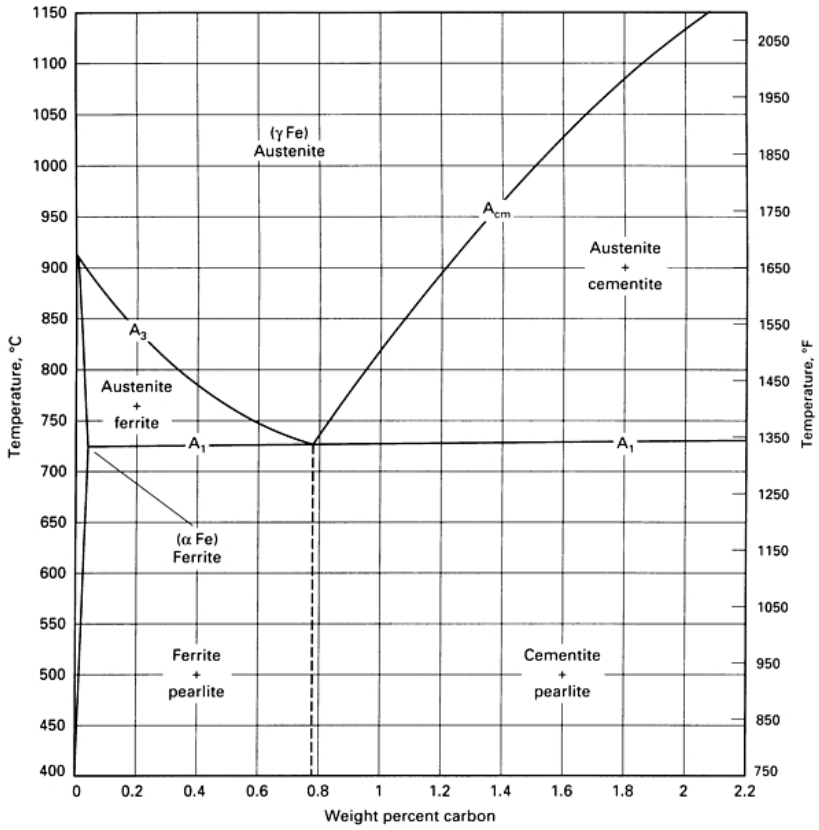


Figure 4.4: The iron-iron carbide binary phase diagram - Equilibrium diagram for combinations of carbon in a solid solution of iron (close-up in the steel region).

$$Ac_3(^\circ\mathrm{C}) = 910 - 203(\sqrt{\%C}) - 15.2(\%Ni) + 44.7(\%Si) + 104(\%V) + 31.5(\%Mo) \quad (4.12b)$$

In a similar way, also the M_s temperature can be approximated by using the following equation [40]:

$$M_S(^\circ\mathrm{C}) = 521 - 353(\%C) - 225(\%Si) - 24.3(\%Mn) - 27.4(\%Ni) - 17.7(\%Cr) - 25.8(\%Mo) \quad (4.13)$$

4.2.2 Hardness and hardenability

Hardenability refers to the ability of a steel to be transformed partially or completely from austenite to martensite at a given depth when cooled under prescribed conditions. This definition reflects the empirical nature of steel hardenability, and many types of experiments have been devised to measure or describe the hardenability of various kinds of steel. Martensite is the micro-structure usually desired in quenched carbon and low-alloy steels. The cooling rate must be fast enough to avoid the nose of the continuous-cooling-transformation (CCT) curve of the steel being quenched, as indicated by curve A in Figure 4.5, in order to obtain the maximum amount of martensite. If the cooling rate is not fast enough to miss the nose of the CCT curve (curves B, C, and D), some transformation to bainite, pearlite, or ferrite will take place, with a corresponding decrease in the amount of martensite formed and a decrease in the as-quenched hardness developed. The cooling rate in quenched parts should be fast enough so that a high percentage of martensite is produced in critically stressed areas of the part. Higher percentages of martensite result in higher fatigue and impact properties after tempering. The maximum hardness obtainable in a steel depends almost exclusively on the carbon concentration. The relation of carbon concentration and percentage martensite to the approximate maximum as-quenched hardness is shown in Figure 4.6.

4.2.3 TTT and CCT diagrams

The cooling rates required to achieve particular micro-structures are often represented by time-temperature-transformation (TTT) diagrams and continuous-cooling-transformation (CCT) diagrams. In order to understand how to read and interpret these diagrams, the example of an AISI 4340 low-alloy steel will be used in the following.

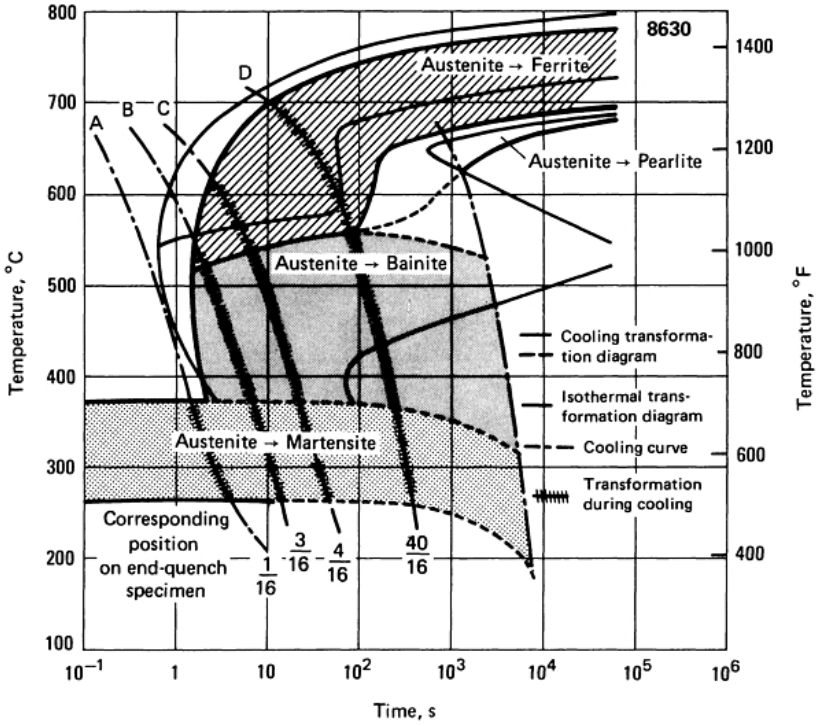


Figure 4.5: CCT diagram - Transformation diagrams and cooling curves for 8630 steel, indicating the transformation of austenite to other constituents as a function of the cooling rate [4].

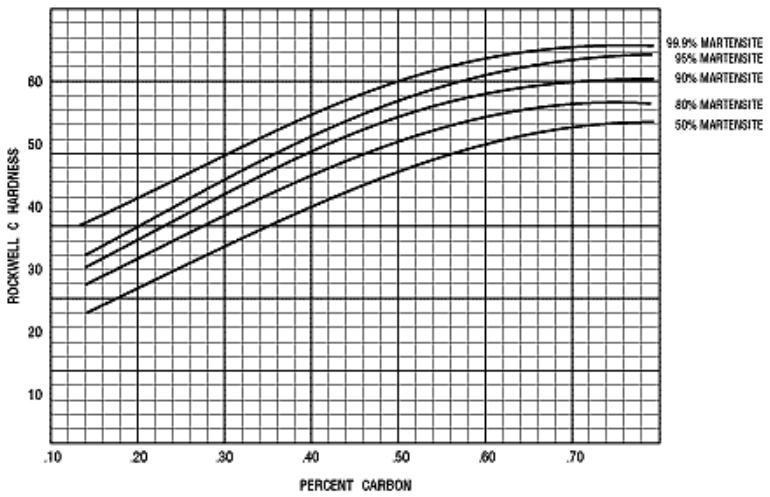


Figure 4.6: Hardenability of steel - Average relationships between carbon content, hardness, and percentage of martensite in quenching [41].

TTT curves are usually produced by solution treating (austenitizing) small samples of steel at the appropriate temperature for the alloy, quickly transferring samples to a lead or salt bath, holding for selected periods of time, and water quenching. The micro-structure of each sample after quenching is examined to determine the point in time when the transformation to ferrite, pearlite, or bainite began and the rate at which the transformation progressed with increasing isothermal holding time. The start of transformation in TTT curves is usually defined as the time required to produce 0.1% transformation at the specified holding temperature.

The TTT diagram for an AISI 4340 steel is shown in Figure 4.7. Starting from an austenitizing temperature of about 745 °C (1373 °F), the steel must be cooled past 500 °C (932 °F) and to the start of the martensitic transformation (M_s) in less than about 20 s to produce a fully martensitic structure and avoid the bainitic area . Slower cooling rates may result in lower hardening fractions. Many TTT curves have been published and provide valuable assistance to metallurgists regarding the relative hardenability of different alloys [6, 42, 43]. However, it is difficult to quantitatively determine from a TTT curve whether a steel of a particular composition and section size can be hardened in a particular quenchant.

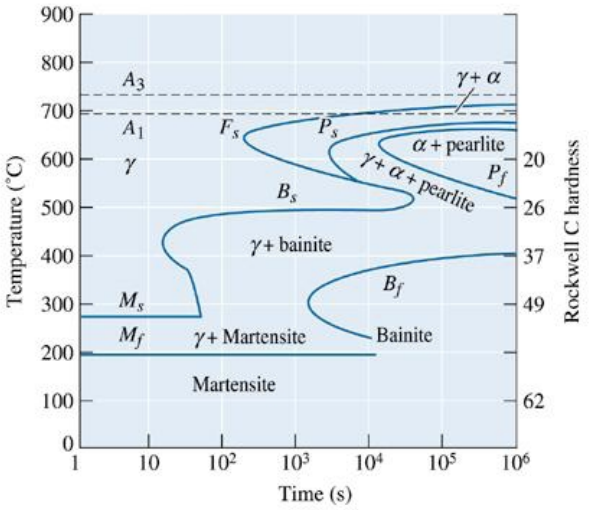


Figure 4.7: Example of TTT diagram - Time-Temperature-Transformation diagram for an AISI 4340 low-alloy steel.

CCT diagrams are constructed by cooling samples of alloys and examining the resulting micro-structures to determine the phases associated with particular cooling

rates. CCT diagrams for a wide variety of alloys have also been published [44, 45, 46]. The CCT diagram for an AISI 4340 steel is shown in Figure 4.8. This diagram indicates that a fully martensitic micro-structure in this alloy requires a cooling rate above 8 °C/s (46 °F/s) at 745 °C (1373 °F) . Slower cooling rates result in spurious amounts of other phases which reduce the hardening fraction.

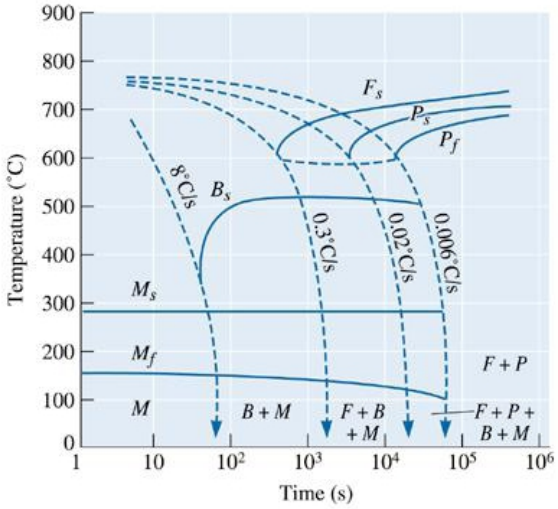


Figure 4.8: Example of CCT diagram - Continuous-Cooling-Transformation diagram for an AISI 4340 low-alloy steel.

4.2.4 Mathematical modelling

In order to perform an accurate analysis of the hardening process, the mathematical modelling of phase transformations is a very important step. Over the years, various approaches have been developed. Johnson, Mehl, and Avrami [47, 48, 49] proposed an analytical equation for diffusional transformation under the isothermal condition. Together with this equation, Scheil's additive rule [50] has been widely adopted to describe non-isothermal cooling processes like quenching, by subdividing the cooling curve into various small isothermal steps. For handling a diffusion-less transformation like an austenitemartensite transformation, Koistinen and Marburger [51] proposed an empirical model which can be used to predict the volume fraction of the martensite after rapid quenching.

The so-called Johnson-Mehl-Avrami-Kolmogorov (JMAK) model is commonly applied for the prediction of the diffusional transformations, both on heating (i.e., $\alpha \rightarrow \gamma$)

and on cooling (i.e., $\gamma \rightarrow \alpha$). Diffusion-less transformation instead, are described by means of the Koistinen-Marburger (KM) model.

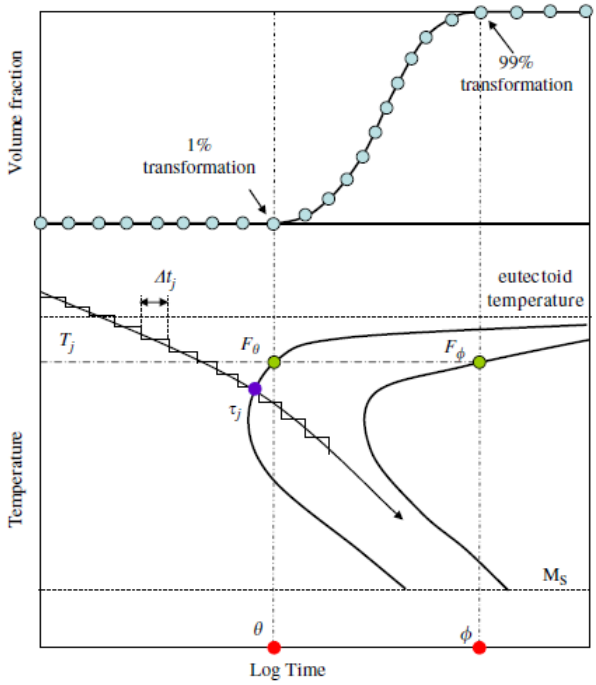


Figure 4.9: Scheil's additivity rule - Example of step-wise isothermal discretization of the cooling curve based on the Scheil's additivity rule [52].

The schematic TTT diagram of eutectoid steel is shown in Figure 4.9. Under the isothermal condition, kinetics of the diffusional transformation can be expressed as the following JMAK equation:

$$F_i = 1 - \exp\left(-a(T) \cdot t_j^{n(T)}\right) \tag{4.14}$$

where F_i indicates the volume fraction of the i-th phase during the phase transformation, a is the diffusional coefficient (s^{-1}) and n is the transformation exponent. The quantities a and n are temperature-dependent material parameters directly obtained from the TTT diagram through the following equations:

$$a = -\frac{\ln(1 - F_\theta)}{\theta^n} \tag{4.15a}$$

$$n = \frac{\ln\left[\ln(1 - F_\theta)/\ln(1 - F_\phi)\right]}{\ln(\theta/\phi)} \tag{4.15b}$$

where F_θ and F_ϕ are respectively the initial and final volume fractions of the considered phase (typically F_θ=0.01 and F_ϕ=0.99), while θ and ϕ are respectively the times at which the transformation begins and completes (s).

The transformation time t_j is taken as the sum of the time interval Δt_j and the fictitious time obtained by considering the volume fraction of the considered phase at the previous time-step F^{j-1}, as follows:

$$t_j = \Delta t_j + \left\{ -\frac{\ln\left(1 - F^{j-1}\right)}{a(T)} \right\}^{1/n(T)} \tag{4.16}$$

Using this transformation time, the volume fraction of each phase can be calculated using equation (4.14). However, since the JMAK equation cannot be applied to a non-isothermal transformation process like quenching, the so-called Scheil's additivity rule is adopted to describe the non-isothermal transformation behaviour. According to this rule, the cooling curve is subdivided into a number of small isothermal time-steps. Then, a nucleation factor $\Delta t_j/\tau_j$ is calculated and summed for each time-step until it reaches unity and the transformation is assumed to begin. The nucleation condition reads as follows:

$$\sum_{j=1}^{m} \frac{\Delta t_j}{\tau_j} = 1 \tag{4.17}$$

where τ_j is the transformation starting time at the j-th time-step.

For the case of diffusion-less transformations, the volume fraction of martensite is calculated using the empirical model established from experiments by Koistinen and Marburger. The amount of martensite can be totally expressed as a function of the temperature as follows:

$$F_m = \left[1 - \exp\left(-b(T) \cdot (M_s - T)\right)\right] \left(1 - \sum\nolimits_i F_i\right) \tag{4.18}$$

where F_m and F_i are the volume fractions of martensite and other phases, respectively, while M_s and T are the temperature at which martensite transformation begins and current temperature (°C), respectively. The parameter b represents the temperature-dependent diffusion-less coefficient ($°C^{-1}$), commonly assumed as constant (b=0.011). Since ferrite and pearlite cannot transform into martensite, once the austenite is transformed into one of these two product phases, equation 4.18 is multiplied by the adjustment term $(1 - \sum_i F_i)$.

4.3 Fundamentals of quenching

4.3.1 Cooling curves

The examination of quenching performance by cooling curve analysis is becoming increasingly popular and perhaps the most informative method of characterizing a quenchant medium. Cooling curves are obtained by quenching a test piece containing one or more thermocouples into a test sample of the quenching fluid, in a laboratory quenching bath, or in the production bath itself. The test piece (probe) may be constructed from the alloy of interest, from an austenitic stainless steel, or a nickel-base alloy such as Inconel. The use of austenitic steel and nickel alloy specimens reduces or eliminates the need for a protective atmosphere while the test pieces are being solution treated. Silver has also been used as the probe material. The test probes are heated to an elevated temperature and then quenched into the medium of interest. A high-speed recorder is used to record temperature changes with respect to time. The resulting time-temperature curves reflect the heat removal characteristics of the quenching fluid and the mass and surface area of the test probe. The resulting cooling curves provide information about the cooling rates achieved in the part. Most metallurgical transformation data is presented in terms of the cooling rate needed to achieve a specific micro-structure and these rates can often be related to the cooling rates obtained from quenchant cooling curve analysis. In a given grade of steel, low cooling rates usually produce ferritic micro-structures, while progressively higher rates produce pearlite, bainite, and finally, martensite. Cooling rates produced by quenchants may be related to specific micro-structures, as will be subsequently discussed.

There are generally considered to be three stages of heat removal, referred to as A, B, and C stages, associated with quenching in liquids. A temperature-time cooling curve illustrating the three stages is shown in Figure 4.10. This curve was obtained with stainless steel probe solution treated at 845 °C (1550 °F) and quenched in non-agitated water. The probe had a thermocouple located at the geometric center. A cooling rate curve, which is obtained by taking the first derivative of the time-temperature curve, is also shown. Figures 4.11 and 4.12 illustrate the three stages of cooling: vapour blanket (A-stage), nucleate boiling (B-stage), and convective (C-stage). These cooling stages are typically obtained when an austenitized steel rod is quenched into an aqueous poly-

alkylene glycol (PAG) polymer solution. Similar studies have also been performed with other quenchants including water, oil, and other aqueous polymer quenchants.

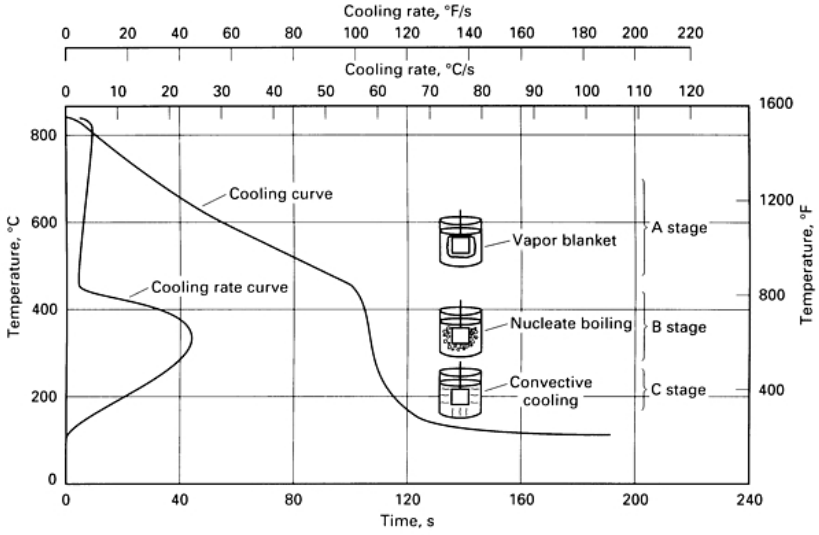


Figure 4.10: Cooling curves during quenching - Cooling curve and cooling rate curve at the center of a 25 mm diameter probe quenched with 95 °C (200 °F) water flowing at 15 m/min.

Stage A: vapour blanket phase ("gas") The first stage of cooling is characterized by a quenchant vapour blanket around the part. This is also referred to as the vapour blanket cooling stage and is characterized by the Leidenfrost effect, a phenomenon in which a liquid, in near contact with a mass significantly hotter than the liquid's boiling point, produces an insulating vapour layer which keeps that liquid from boiling rapidly [53]. The vapour blanket develops and is maintained while the supply of heat from the interior of the part to the surface exceeds the amount of heat needed to evaporate the quenchant and maintain the vapour phase itself. Relatively slow cooling occurs during this period because the vapour envelope acts as an insulator, and cooling occurs mainly by radiation through the vapour film. The temperature above which a total vapour blanket is maintained is called the characteristic temperature of the liquid and is also known as the Leidenfrost temperature. The A-stage of cooling is not usually present in parts quenched in aqueous solutions containing more than about 5 wt% of an ionic material such as potassium chloride, lithium chloride, sodium hydroxide, or sulfuric acid. Cooling curves for these solutions start immediately with stage B. The

presence of the salts at the hot metal quenchant interface initiates nucleate boiling almost immediately. A-stage cooling is not observed when quenching in non-volatile quenchant media such as molten salt baths. Conversely, heat transfer in gas quenchants such as air and inert gases occurs exclusively by a vapour blanket mechanism.

Stage B: nucleate boiling phase ("bubbles") The highest cooling rates occur in the so-called nucleate boiling stage. During this period, the vapour envelope collapses, and high heat extraction rates are achieved that are associated with nucleate boiling of the quenchant on the metal surface. Heat is rapidly removed from the surface as liquid quenchant contacts the metal surface and is vaporized.

Stage C: convective cooling phase ("waves") Stage C is called the liquid cooling stage and begins when the temperature of the metal surface is reduced below the boiling point of the quenching liquid. Below this temperature, boiling stops and cooling takes place by conduction and convection into the quenchant. C-stage cooling rates are dependent on the viscosity of the quenchant: all other factors being equal, cooling rates decrease with increasing viscosity.

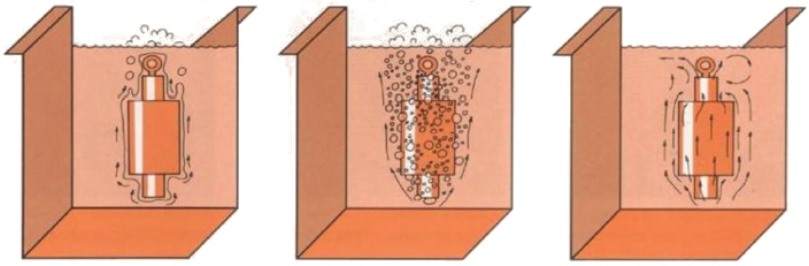


Figure 4.11: Cooling stages during quenching - Schematic representation of the three stages of heat removal during quenching [54].

4.3.2 Factors affecting cooling

The actual cooling rates and temperature ranges associated with the three stages of cooling vary with the type of quenchant and the mass and surface area of the part being quenched. The highest cooling rates are generally obtained with brine solutions, followed by water, synthetic polymer quenchants, oils, salt baths, fluidized beds, and gases. There are, of course, substantial variations in the attainable cooling rates within

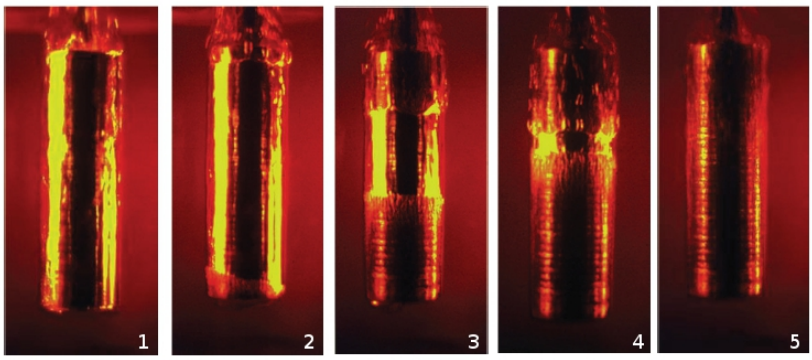


Figure 4.12: Cooling frames during quenching - Photo sequence of a hot steel rod being quenched in a PAG polymer in water solution. (1) Moment of immersion, vapour film forms around the probe. (2) After 5 s, boiling commences at corners. (3) After 10 s, boiling front moves along the probe. (4) After 15 s, boiling collapses and convection stage begins. (5) After 30 s, polymer film has completely redissolved and heat removal is achieved entirely by convection [54].

particular classes of quenchants depending on the temperature, extent of agitation, viscosity, molecular weight, wetting characteristics, polymer or oil additive concentration, gas pressure, and gas velocity. The cooling rates are also functions of the thickness and geometry of the part, and characteristics of the quenching facility including the extent of agitation, racking procedure, extent of surface oxidation, and the effective heat transfer coefficient between the part and the quenchant. The same heat transfer mechanisms involved in cooling test pieces and instrumented probes are involved in quenching parts during a heat-treating operation. Although a particular cooling curve is strictly related only to the size and material of the test piece, thermocouple location in the test piece, and conditions of the quenching liquid under which a test was performed, cooling curves developed under one set of conditions may be translated to other conditions by applying appropriate heat transfer formulas.

The effect of irregular configuration on heat flow from a gear to the quenching area is illustrated in Figure 4.13. High temperatures persist near the surface at the roots of the teeth where large vapour bubbles are trapped. If the gear were induction or flame heated, and thus had a uniformly thin heated layer conforming to the contour, quenching would progress more rapidly and uniformly because heat also would flow simultaneously to the cold metal underlying the heated exterior and the quenchant.

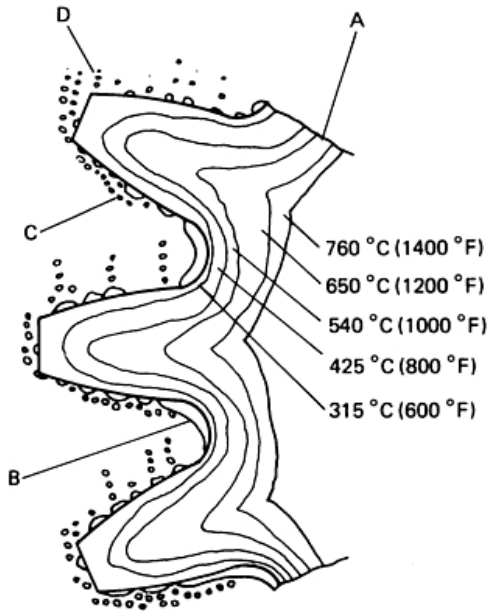


Figure 4.13: Cooling of a geared part - Temperature gradients and other factors affecting the edgewise quenching of a gear in a quiescent volatile liquid. A, flow of heat from hot core of gear; B, vapour blanket stage still exists due to large source of heat and poor agitation; C, trapped vapour bubbles condensing slowly; D, vapour bubbles escaping and condensing.

Agitation refers to the movement of the quenching fluid relative to the part. Agitation is usually obtained by stirring the liquid, but in some cases it is obtained by moving the part in the liquid. Agitation has an extremely important influence on the heat transfer. Agitation causes mechanical disruption of the vapour blanket in Stage A and a faster transition to B-stage cooling. Increasing agitation usually produces a shorter A-stage cooling time and faster cooling rates in all three regions. Conversely, higher quench bath temperatures typically produce longer A-stage cooling times and slower cooling rates in the B- and C-stage regions. However, modestly higher temperatures in oil quench baths can improve the heat removal characteristics by reducing oil viscosity. An improvement in quench oil wetting characteristics may be obtained with increasing bath temperature, which results in higher cooling rates. Therefore, the cooling performance of many quench oils is often independent of modest variations in bath temperature. Non-uniform quenching may result if agitation is not used because of localized hot spots resulting from uneven heat removal from the metal surface. This may lead to spotty hardness, increased surface cracking, distortion, and higher residual stresses.

Heat transfer during quenching

Heat removal from parts during quenching can be mathematically described in terms of the effective interface heat transfer coefficient. A quenchant must impart a sufficiently high interface heat transfer coefficient to produce a cooling rate that will minimize transformation of austenite to ferrite or pearlite and yield the desired amount of martensite or bainite. The interface heat transfer coefficient (HEC) is defined as follows:

$$h = \frac{q}{A(T - T_{ext})} \tag{4.19}$$

where h is the heat transfer coefficient (W/m^2·°C), q is the heat flow from the part to the quenchant (W), A is the surface area of the part in contact with the fluid (m^2), T is the part surface temperature, and T_{ext} is the fluid temperature. A similar ratio, more widely used in steel quenching, is the Grossmann number H (m^{-1}), which is defined by the following equation:

$$H = \frac{h}{2k} \tag{4.20}$$

where k is the thermal conductivity of the metal (W/m·°C).

4.3.3 Distortion and cracking

Distortion and cracking during quenching limits the severity of the quenchant and equipment that may be used. A more severe quench produces martensite to a greater depth (with a steel of given hardenability), but it also increases the likelihood of distortion and cracking.

Distortion during quenching can be understood by remembering that:

- steel has a higher strength when cold than when hot;
- steel shrinks while cooling and expands while hardening.

Linear dimensional changes occurring during cooling and transformation are shown in Figure 4.14 for both slow cooling and fast quenching conditions. Quenching to form martensite results in an expansion of the material compared to that achieved with a slow cooled pearlitic matrix. However, it is recognized that both pearlitic and martensitic materials contract over 1% during cooling from the austenite temperature, but martensitic materials have a lower net contraction. Distortion is a result of three phenomena: warping, thermally induced deformation, and martensite formation.

Warping is the result of non-uniform heating or non-uniform support of a part during heating.

Thermal deformation is a result of non-uniform contraction during cooling.

Martensite formation also induces stresses that cause distortion.

Warping during non-uniform cooling is schematically illustrated in Figure 4.15. Assume that the bar was initially at a uniformly high temperature. If the bar were quenched on one side, the more rapidly cooled side would contract earlier and at a higher rate than the opposite side. The rapidly cooled side becomes shorter and stronger as it cools and causes plastic deformation in the hot side. The deformation is followed by cooling and contraction of the more slowly cooled side. When the part has cooled, it will then be warped (if the stress exceeds the yield strength of the bar) with the slowly cooled side being shorter and warped concave. The plastic deformation on the hot side results in compressive stresses on the rapidly cooled side and tensile stresses on the more slowly cooled side.

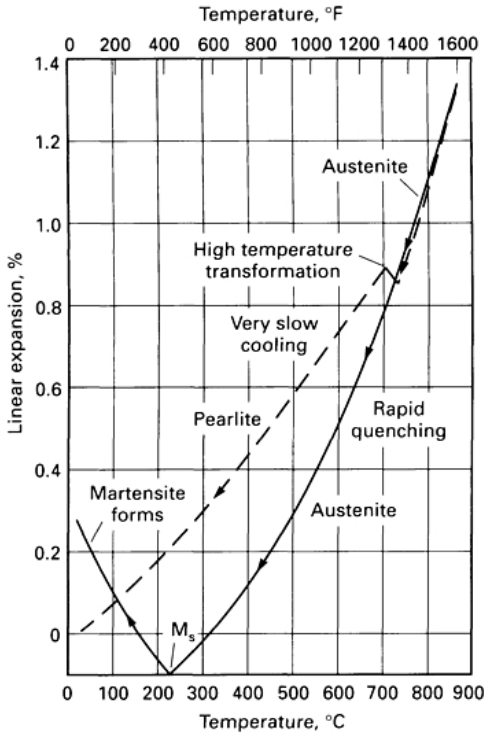


Figure 4.14: Dimensional changes during cooling - Dilatometric study of steel cooled from a high temperature to show that the steel undergoes reversals of its dimensional contractions depending on variations in its quenching rate [55].

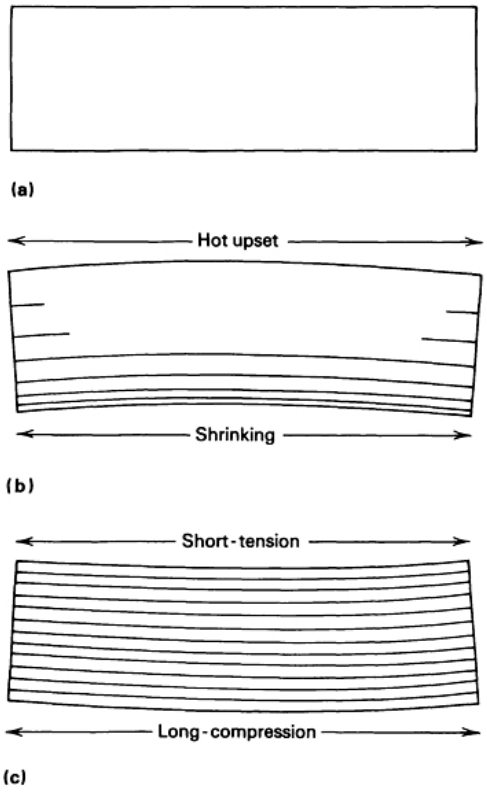


Figure 4.15: Warping - Warpage caused by non-uniform quenching of a steel block [56]. (a) Uniformly hot. (b) Non-uniform cooling. (c) Uniformly cold.

Thermally induced deformation resulting from thermal gradients produced during cooling is illustrated in Figure 4.16. If a part is initially uniformly hot, and is rapidly quenched, the outer surface shrinks while the center is still relatively hot. This process puts the outer surface into tension and the inside into compression, causing internal plastic flow. As the center of the part cools and the temperature reaches a uniformly low value, thermal contraction in the center of the bar occurs, which reverses the stress state and places the center in tension. Finish machining of the outer shell removes the tensile stress present in the core.

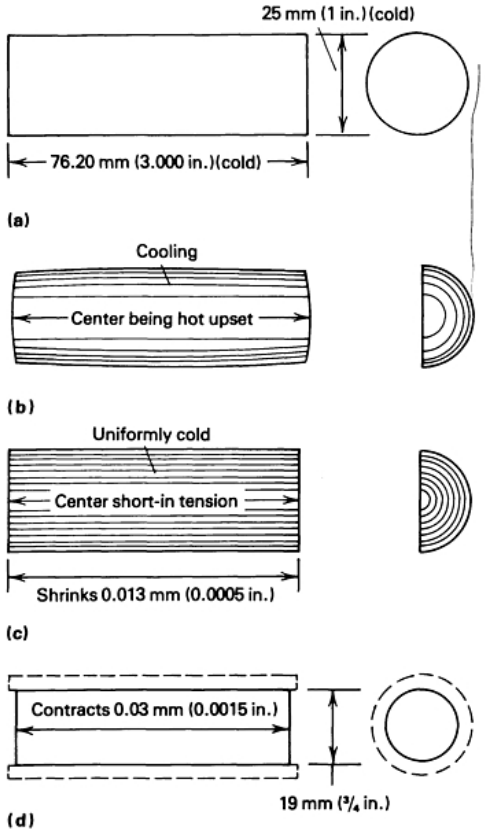


Figure 4.16: Thermally induced deformation - Effect of quenching on a steel cylinder that was quenched in water from 775 °C (1425 °F) [56]. (a) Uniformly hot. (b) Center portion of cylinder is hot upset during quenching. (c) Center portion of cylinder is short in tension when uniformly cold. (d) Machining of the shell removes tensile stress from core to allow cylinder to decrease in length.

The expansion of a steel part being hardened associated with martensite formation is a similar deformation process. The martensitic transformation occurs with a

net shrinkage of about 1% during cooling from the austenitizing temperature to room temperature. During rapid cooling, the outer surface is quenched to martensite and the subsequent contraction of the core produces surface compression and center tension (see Figure 4.17).

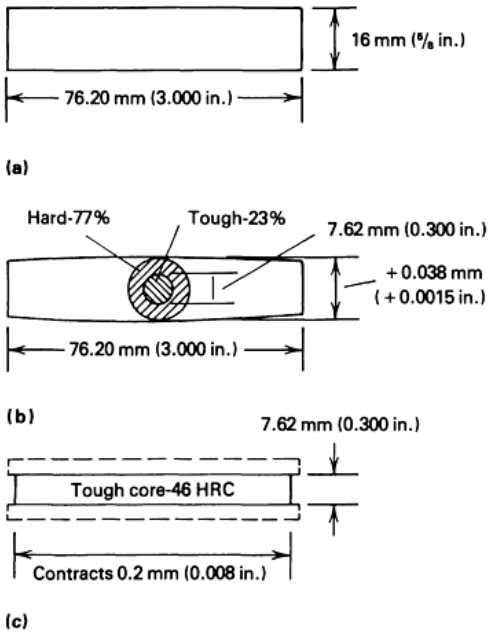


Figure 4.17: Martensite-related deformation - Strains produced by hardening a cylinder of 1.10% C tool steel [56]. (a) Dimensions of an annealed bar before quenching. (b) Hardness values at selected locations inside cylinder after water quenching at 775 °C (1425 °F) based on surface hardness of 68 HRC. (c) Core hardness after water-quenched cylinder diameter reduced from 16 mm to 7.62 mm by wet grinding.

More generally speaking, the distortion occurring during quenching depends on the size and shape of the part, its composition, and the characteristics of the quenchant employed. Also, parts with large section size variations, very thin cross sections, containing slots, keyholes, drilled holes, or grooves cause problems because of the difficulty in achieving uniform heating and cooling rates. The steel composition can also have a significant effect on warping and distortion. High-carbon alloys and those with high hardenability are prone to more distortion than low-carbon with low-hardenability alloys. Similarly, high austenitizing temperatures and low martensite start (M_s) temperatures tend to aggravate distortion. The characteristics of the quench procedure, particularly conditions promoting high rates of cooling through the transformation

range, aggravate distortion and cracking. Rapid quenching to a temperature just above the M_s temperature, followed by slow cooling through the martensite formation range, minimizes distortion. Uniform rates of heat removal per unit of surface area also tend to minimize distortion.

Cracking occurs for the same reasons as distortion, but cracks form when the localized strain exceeds the failure strain in the material. Usually, the cracks follow austenite grain boundaries, but cracking does not appear to be related to the austenite grain size. The tendency for cracking typically decreases as the M_s temperature increases.

4.4 Deformations and residual stresses

Heat treatment often causes stress- and strain-related problems such as residual stress, quench cracks, and deformation and/or distortion. The residual stress may be defined as the self-equilibrating internal or locked-in stress remaining within a body with no applied (external) force, external constraint, or temperature gradient [57]. Whether talking about their tensile or compressive nature, depending on the location within the material, residual stresses can be categorised in two big families:

Macro- or long-range residual stress, is a first-order stress that represents an average of body stresses over all the phases in poly-phase materials. Macro-residual stress act over large regions as compared to the grain size of the material and is traditionally considered when designing mechanical parts.

Micro- or short-range residual stress, is a second-order or texture stress, which is associated with lattice defects (e.g., vacancies, dislocations, pile-up of dislocations) and fine precipitates (e.g., martensite) [58, 59]. Micro-residual is the average stress across one grain or part of the grain of the material and is representative of the essential behaviour of material deformation.

4.4.1 Sources of residual stress

Residual stress can be originated from many sources, mainly during *processing* of the material (rolling, casting, forging), *forming* operations (shearing, bending, drawing, machining), and *fabrication* of a structure (welding). Whenever a component is stressed

beyond its elastic limit and plastic flow occurs, residual stresses are present. Examples can be found in cold forming, quenching, grinding, and welding; in cold forming global bending loads create stresses on corners, raising local fatigue strength to critical conditions [60]; quenching of thick sections results in high residual compressive stresses on the surface of the material, which are balanced by residual tensile stresses in the internal areas of the section [61]; depending on the grinding operation, compressive or tensile residual stress can be introduced very superficially, causing warping of thin parts [62]; welding usually results in localized residual stresses that approach levels equal to or greater than the yield strength of the material at room temperature [63].

4.4.2 Effects of residual stress

The major effects of residual stress include dimensional changes and resistance to crack initiation. Dimensional changes occur when the residual stress (or a portion of it) in a body is eliminated. In terms of crack initiation, residual stresses can be either beneficial or detrimental, depending on whether the stress is of tensile or compressive nature. In fact, it should be noticed that since residual stresses are algebraically summed with the applied stress generated by an external load, an increase or a decrease of the effective stress can be experienced by the material, depending on the pre-existing stress state.

Compressive residual stress in the surface layers are generally helpful because the built-in compressive pattern can reduce the effects of imposed tensile stresses that may produce cracking or failure. Compressive stresses therefore contribute to the improvement of fatigue strength and resistance to stress-corrosion cracking in a part and an increase in the bending strength of brittle ceramics and glass.

Tensile residual stress at the surface of a part are usually undesirable because they can effectively increase the stress levels, thus causing unpredicted stress-corrosion cracking (due to the combined effect of stress and environment), fatigue failure, quench cracking, and grinding checks even at low external stresses. Tensile stresses tend to reduce fatigue life and strength of a part, the extent thereof being closer or even larger than the strength of the material.

4.4.3 Development of residual stresses

Variations in stresses, temperature, and chemical species within the body during processing cause the production of macro-residual stresses. Various manufacturing processes such as forming, machining, heat treatment, shot peening, casting, welding, flame cutting, and plating render their characteristic residual stress pattern to processed parts. In heat-treated parts, residual stresses may be classified as those caused by a thermal gradient alone, and a thermal gradient in combination with a structural change (phase transformation). When a steel part is quenched from the austenitizing temperature to room temperature, a residual stress pattern is established due to a combination of thermal gradient and local transformation-induced volume expansion. Thermal contraction develops non-uniform thermal (or quenching) stress due to different rates of cooling experienced by the surface and interior of the steel part. Transformational volume expansion induces transformation stress arising from the transformation of austenite into martensite or other transformation products [64].

Thermal contraction

The relation between the thermal stress σ_{th} during cooling and the corresponding temperature gradient in the component is given by the following equation:

$$\sigma_{th} = E \cdot \varepsilon_{th} = E \cdot \Delta T \cdot \alpha \tag{4.21}$$

where E is the modulus of elasticity or Young's modulus (GPa) and α is the thermal expansion coefficient ($^{\circ}C^{-1}$) of the material. It is thus apparent that thermal stresses are greatest for materials with high elastic modulus and coefficient of thermal expansion. Concerning the temperature gradient, this is depending on several aspects, related both to the material and process side. The so-called thermal diffusivity λ (m^2/s), which is defined as the ratio between the thermal conductivity k (W/m·°C) and the volumetric heat capacity γC_p (J/m^3·°C), provides a measure of the thermal inertia of the material: low values of λ promote large temperature gradient and thermal contraction. On the other hand, the heating and cooling rates imposed to the part , may not be negligible, due to the severity of the austenitizing and/or quenching phases, as well as the heating and cooling uniformity, which become critical issues to be considered when dealing with large part sizes. In general, the higher is the quenching intensity of the coolant, the greater are the maximum residual stress attained upon quenching.

Thermal and Transformational Volume Changes

During quench hardening of a steel part, hard martensite forms at the surface layers, associated with volume expansion, whereas the remainder of the part is still hot and ductile austenite. Later, the remainder austenite transforms into martensite, but its volumetric expansion is restricted by the hardened surface layer. This restraint causes the central portion to be under compression with the outer surface under tension. In some particular conditions, these volumetric changes can produce sufficiently large residual stresses that can cause plastic deformation on cooling, leading to warping or distortion of the steel part. While plastic deformation appears to reduce the severity of quenching stresses, in most severe quenching, such stresses are so high that they do not get sufficiently released by plastic deformation. Consequently, the large residual stress remaining may reach or even exceed fracture stress of steel. This localized rupture or fracture is called quench cracking [65]. It should be emphasized again that for a given grade of steel, both large size of the part and higher quenching speed contribute to the larger value of thermal contraction, as compared to the volumetric expansion, of martensite. In contrast, when the parts are thin and the quenching rate is not high, thermal contraction of the part subsequent to the hardening of the surface will be smaller than the volumetric expansion of martensite. Similarly, for a given quenching rate, the temperature gradients decrease with decreasing section thickness, and consequently the thermal component of the residual stress is also decreased [66].

Although the shallower hardening steels exhibit higher surface compressive stresses, deep hardening steels may develop moderately high surface compressive stresses with severe water quenching. When these deep hardening steels are through-hardened in a less efficient quenchant, they may exhibit surface tensile stresses [67]. However, the residual stress pattern in the hardened steels can be modified either with different transformation characteristics or during the tempering and finish-machining (after hardening) operations.

4.4.4 Measurement of residual stresses

There are two methods of measuring residual stresses: the destructive method, also called the dissection method, comprising the hole-drilling method and the non-destructive methods, comprising mainly x-ray diffraction. A classification of the dif-

ferent techniques with their respective ranges of application is depicted in Figure 4.18 [68].

Destructive methods. Despite they're quite old, time-consuming, and expensive, these methods are reasonably accurate, well-established, and can be employed in confined situations at site. The main drawbacks are related to their destructive, or at best semi-destructive, nature and their ability to measure only the macro-residual stresses.

Non-destructive methods. The main difficulty with the non-destructive methods such as the measurement of crystallographic lattice parameters, ultrasonic velocities, or magnetization changes is that all these quantities are indirectly related to the residual stress, being usually dependent on the material parameters (e.g., metallurgical textures), which are difficult to quantify.

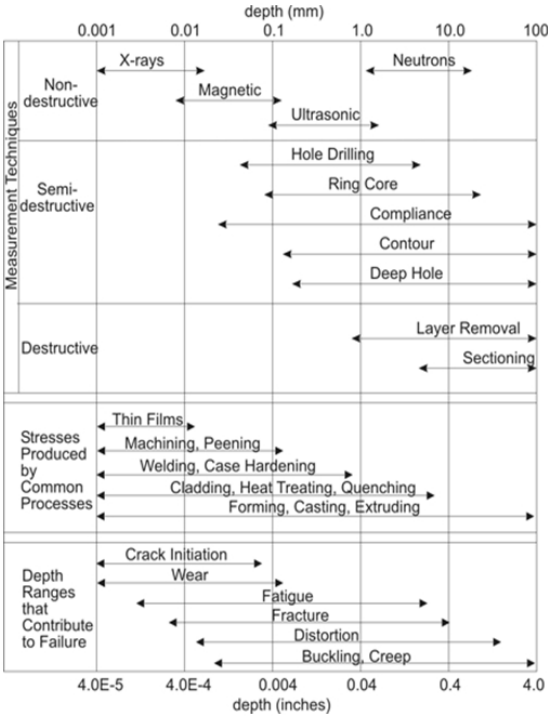


Figure 4.18: Residual stress measurement techniques - Overview and classification of different techniques for the measurement of residual stresses.

The hole-drilling method

The *hole-drilling strain-gauge method* (ASTM standard E837) is used extensively for measuring residual stresses and is based on the dissection approach: it relies on stress relaxation when a hole is drilled into the center of a rosette strain gauge mounted on the surface of the sample (see Figure 4.19). When the material is removed by drilling, the extent of the strain relief is monitored and both the direction and magnitude of the principal stresses can be calculated. A special high speed air turbine drill (see

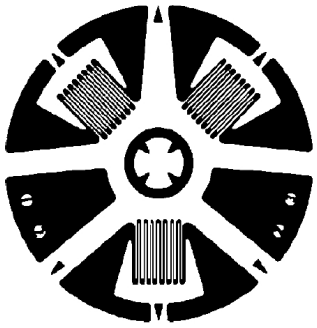


Figure 4.19: Rosette strain gauge - Representation of the strain gauge used for residual stress measurement by the hole-drilling method (according to the ASTM standard E837).

Figure 4.20) is used to first locate the drill in the rosette center and then to remove material to a controlled depth. At each depth increment, the strain relief on each of the gauges is measured and converted into stress. As subsequent material removals occur, the stress distribution as a function of depth can be estimated. The hole-drilling method is used in those situations where the residual stress is relatively uniform over the drilling depth. Thus, it is not intended for situations where the residual stress is superficial. The accuracy of the method is directly related to the ability of locating the hole accurately in the center of the rosette and to drill the relief hole so as not to introduce new stresses. This is best achieved in hard materials by use of a high-speed turbine drill with a very low rate of metal removal, which avoids excessive rubbing of the cutting surface against the hole wall and minimizes the introduction of spurious strains. To ensure that any heat generated has been dissipated, readings are recorded after 15 min of the end of the grinding process.

Figure 4.20: Air turbine drill - High speed air turbine drill used in the hole-drilling strain-gauge method.

The x-ray diffraction method

The *x-ray diffraction method* (XRD) is a well-established technique for measuring both macro- and micro-residual stress non-destructively. It has been employed to provide quantitative values for residual stress profiles in surface or fully hardened components. This technique relies on the determination of lattice strains and the stress-induced differences in the lattice spacing: stress magnitudes are determined through measurement of changes in the materials lattice spacing, due to the presence of a stress; from knowledge of the non-stressed lattice spacing, any stresses present can then be calculated using established equations. Macro-residual strain is measured from the shift of diffraction lines in the peak position using the so-called non-linear $\sin^2\Psi$ method, from which residual stress is calculated. For the measurement of micro-strain, statistical methods based on the broadening analysis of diffraction lines are applied [69].
The $\sin^2\Psi$ method is a sensitive and accurate technique to measure residual stresses in a fine grained, polycrystalline material. As shown in Figure 4.21, the position of a diffraction peak will shift as the sample is rotated by an angle Ψ. The magnitude of the shift will be related to the magnitude of the residual stress. Thus, if there is no residual stress, the shift will be zero. The relationship between the peak shift and the residual stress σ is given by the following equation:

$$\sigma = \frac{E}{(1+\nu)\sin^2\Psi}\frac{(d_n - d_0)}{d_0} \tag{4.22}$$

where E is the Young's modulus (GPa), ν is the Poisson's ratio, Ψ is the tilt angle (deg), d_n and d_0 are the atomic lattice spacings measured at each tilt angle (m) for stressed and unstressed material conditions, respectively. If there are no shear strains present in the sample, such spacings would change linearly with $\sin^2 \Psi$ and a least squares fit to the curve (for multiple values of Ψ) would give σ. However, if shear strains are present, a splitting of the plot will occur and the analysis is more complicated. Finally, if the sample is rotated in-plane, it is possible to determine the principal stresses and their directions. The main drawback of XRD is related to the fact that it

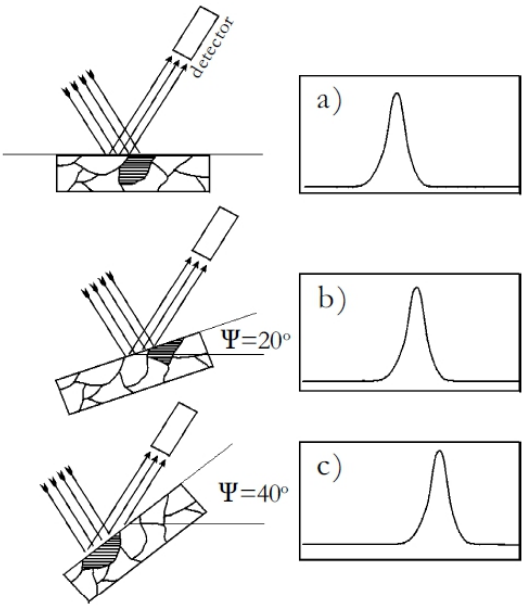


Figure 4.21: XRD rationale - Shift of diffraction peak with increasing change in Ψ value.

is only capable of measuring residual stresses of materials very close to the surface under examination, the measurement being purely surface related (a depth of 10 μm is commonly quoted). However, by layer removal, depth profiles can be determined, thus increasing the spectrum of applications to which it can be applied, while forfeiting its non-destructive nature.

Other methods

The *ultrasonic method* for evaluating residual stress involves ultrasonic stress birefringence or sono-elasticity; this depends upon the linear variation of the velocities of sound in a body (i.e., ultrasonic waves) with the stress. This method has the potential for greater capability, versatility, and usefulness in the future. However, this has the disadvantage, in common with the magnetic methods, that it requires transducers shaped to match the surface being inspected.

The *magnetic method* is based on the stress dependence of the Barkhausen noise amplitude. Each time an alternating magnetic field induced in a ferromagnetic material is reversed, it generates a burst of Barkhausen noise. The peak amplitude of the burst, as determined with an inductive coil near the surface of the component material, varies with the surface stress level. Since Barkhausen noise depends on composition, texture, and work hardening, it is necessary in each application to use calibrated standard (reference) samples with the same processing history and composition as the component being analysed. This method is used to measure residual stresses well below the yield strength of the ferromagnetic materials. This method is rapid, and the measurements are made with the commercially available portable equipment. However, this method is limited to only ferromagnetic materials.

Thermal evaluation for residual stress analysis (TERSA) is a new, non-destructive method that is in an experimental stage. It has the advantage of being completely independent, remote, and non-contacting. It consists of merely directing a controlled amount of energy from a laser energy source into the volume of the material being inspected and then making a precise determination of changes in the resulting temperature rise by infrared radiometry. However, the working instrument will also require some form of display to enable visual examination to be made of any high-stressed regions [70].

5

Simulation models

5.1 Partial differential equations and the finite element method

Partial differential equations (PDEs) arise naturally in science and engineering from complex balance equations. Commonplace PDEs are derived from conservation laws for transport of mass, momentum, species and energy. Because these conservation laws are integral equations over the domain, the PDEs that arise from the continuum hypothesis have a structure that is readily represented by the finite element method as an approximation [71]. PDEs are classified according to their order, boundary condition type, and degree of linearity. Amazingly, most PDEs encountered in science and engineering are second order, that is the highest derivative term is a second partial derivative. Thus, classification and solution of second order spatial temporal systems in 2-D and 3-D are of wide applicability and importance in the sciences and engineering.

There are three canonical classed of differential equations that arise in spatial-temporal systems:

Laplace's equation (elliptic):

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{5.1}$$

Diffusion equation (parabolic):

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{5.2}$$

Wave equation (hyperbolic):

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \tag{5.3}$$

The terms elliptic, parabolic and hyperbolic are traditional guides to the features of a PDE system from characterization of the linear terms by reference to the general linear, second-order partial differential equations in one dependent and two independent variables:

$$a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + d\frac{\partial u}{\partial x} + e\frac{\partial u}{\partial y} + fu + g = 0 \tag{5.4}$$

where the coefficients are either functions of the independent variables x and y only, or constant. The three canonical forms are determined by the following criterion:

$$b^2 - ac \begin{cases} < 0 & \text{elliptic;} \\ = 0 & \text{parabolic;} \\ > 0 & \text{hyperbolic.} \end{cases} \tag{5.5}$$

These classifications serve as a rough guide to the information flow in the domain. For instance, in elliptic equations, information from the boundaries is propagated instantaneously to all interior points. Thus, elliptic equations are termed "non-local" meaning that information from far away influences the given position, versus "local", where only information from nearby influences the field variable. In parabolic systems, information "diffuses", spreading out in all directions. In hyperbolic systems, information "propagates" from regions that have already received such information to regions that are going to receive it, and possibly regions that will never receive it. If the system is linear or quasi-linear (i.e., some coefficient depends on the dependent variable or a lower order partial derivative), this classification system and the intuition about how information is transported serves as a robust guide to second order systems. For non-linear systems, however, non-linearity can destroy the information transport structure. In non-linear systems, information may be "bound", thus never be transferred, beyond given attractors, or it may be created from noise or lost by forgetting initial conditions in a given window in time.

Probably one of the most popular numerical methods for solving differential equations sets is the finite element method (FEM). Because of its powerful unstructured meshing capability, FEM is capable of handling problems involving complex geometries

with inhomogeneous media. Although traditionally used for structural mechanics problems, for which it was originally developed, FEM is now widely used in many branches of physics and engineering, from electromagnetism to thermodynamics, from acoustics to optics, from chemistry to hydrodynamics.

Fundamental to the FEM is the concept that any domain can be implemented as a collection of smaller sub-domains of preferred shape. These sub-domains are called finite elements. Corners of an element are called nodes at which the solutions to field variables are computed. There can be nodes in between corner points that are commonly called edge nodes. The discretization process defines the type and the number of elements. The number of elements is directly connected with the accuracy of the solution: the higher the number of elements used, the lesser will be the solution approximation error over the considered region. However, having a large number of elements would be computationally expensive, demanding a large chunk of physical memory (RAM, random access memory) and an extended runtime.

Defining an unnecessary number of elements is a very common practice. There is no precise formula that allows one to choose optimally exact number of elements. Typically, it is only by experience that the right amount of elements to pack in a domain can be decided. Though, some general rules of thumb exist and can be used to achieve a preliminary mesh. As the accuracy of the solution increases with the number of elements N_{el}, there will be a certain critical number N^*_{el}, beyond which the sensitivity of accuracy becomes negligible. It is very simple to verify by performing a bunch of sequential simulations with increasing number of elements that, beyond a certain value N^*_{el} there isn't any considerable improvement on the accuracy. In this way, the appropriate number of elements to be used can be found out.

There are instances where one is interested in a certain region of the domain rather than the whole domain. This is the case in induction heating, where the eddy current distribution is confined to a very narrow, superficial boundary layer of the part. In such cases, very dense meshes should be created in correspondence of the high field density region, whereas coarser discretizations may be adopted in the far field region. Such smart meshing techniques permit to attain the required accuracy of results without almost affecting the computation time. Another point worth mentioning is the skewness of the elements in stretched grids. Skewed elements make the formation of the Jacobian

(i.e., the field space partial derivatives matrix) impossible. Therefore great care should be taken in using stretched meshes.

The type of the elements to be used depends on the problem that has to be solved. The dimensionality of the domain defines the dimensionality of the elements. The most simple is the 1-D element that represents a line segment between two nodes at each end. The most fundamental element in 2-D is a triangle where as in 3-D it is a tetrahedron. Figure 5.1 shows some of the basic elements in use with respect to the dimensionality. Although straight lines are represented, curvilinear segments between nodes would provide a more general form of the elements. The element types are categorized according to the order of their associated polynomial interpolation function (see Figure 5.2).

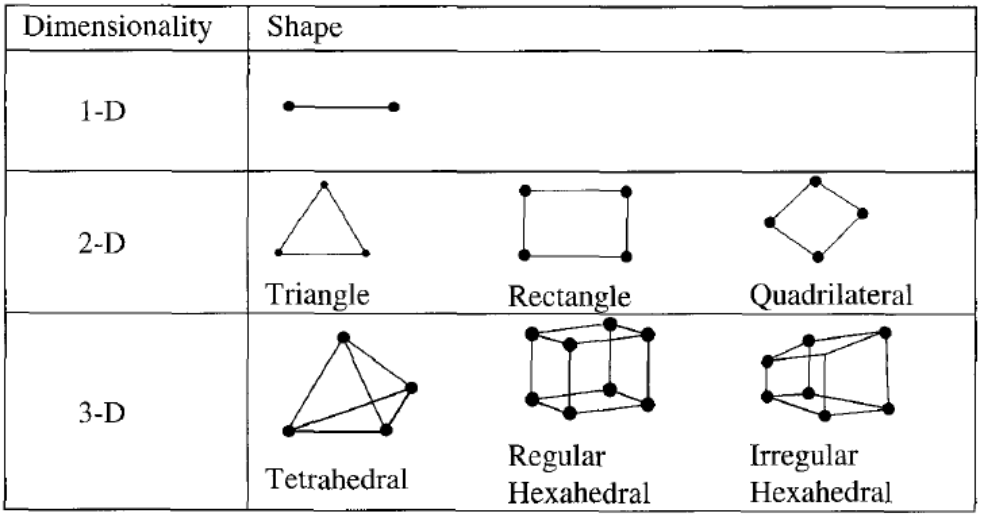


Figure 5.1: FEM element types - Basic types of element for finite element analysis [71].

According to this categorization there are three types of elements: simplex, complex and multiplex. The simplex elements are those for which the approximating polynomials consists of constant and linear terms. The complex elements may have the same shapes as the simplex elements but will have additional boundary nodes and, sometimes, internal nodes. The multiplex elements are those whose boundaries are parallel to the coordinate axes to achieve inter-element continuity, and whose approximating polynomials contain higher order terms [72].

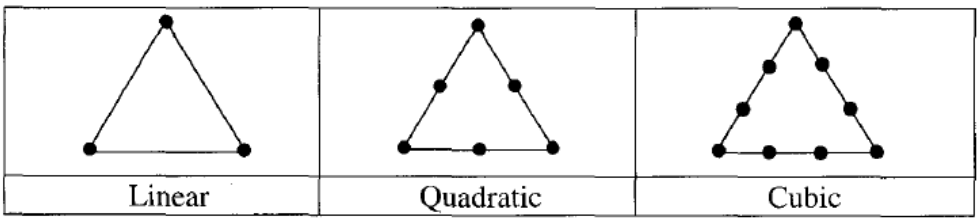


Figure 5.2: FEM element orders - Example of triangular element with increasing order of the associated interpolation function [71].

5.2 Numerical simulation

With the rapid development of computing speed and memory capacity in recent computers, numerical simulation is available for various engineering problems, providing valuable information needed for process design and development. In fact, the cost and lead time for optimising processes are largely reduced by employing numerical tools instead of experimental trials, thus largely addressing industrial needs. Due to the fast development of transistor technology and microelectronics, computing performances are exponentially raising and the applicable range of numerical analysis is increasingly expanding [73]. Moreover, the recent advancements in parallel (e.g., graphics processing units, GPUs) and distributed (e.g., clusters) computing have made it now possible to break the wall of one trillion floating-point operations per second (teraflop) on a single node, thus enabling users to perform very demanding 3-D multi-physics simulations.

Highly effective numerical methods, such as finite difference, finite element, boundary element, and others, are now widely and successfully used in the computation of coupled electromagnetic, heat transfer, and metallurgical phenomena, providing the modern induction heating specialist with a very detailed simulation of the heat treat process in all its phases. Though, any of these computational analyses heavily rely on the correct determination of all the physical properties of the treated materials. In fact, besides the need for a proper definition of the governing equations describing the physics to be simulated, great importance should be given to the theoretical models implementing material behaviours and characteristics, which often present non-linearities and cross-couplings among each other.

Over the years, FEM simulation of gear surface hardening processes has continuously evolved both in terms of complexity of the implemented models and number of

considered physical phenomena. The use of 2-D planar and semi-3-D models allow a rapid definition of the geometrical model and a direct understanding of the influence of process parameters (frequency in particular) on the heating distribution; by restricting the analysis to a thin axial cross-section of the component, such models provide a quite limited description of the real process, as they are not taking into account extremity effects both on the gear and the coil [35, 74, 75, 76].

Such limitations may be overcome by using full 3-D models, which can be reduced by exploiting symmetries and periodicities of the toothed profile. In order to optimise the number of necessary mesh elements, regions of induced currents can be usually described with specific conditions at boundaries, of the so-called "surface impedance" type, which approximate the distribution of heat sources as if they were confined on a very thin, superficial layer [77, 78]. The main drawback of this approach, which requires the penetration depth to be much smaller than the tooth dimension, is evident when dealing with processes which heat the part above the Curie temperature; in such cases, the above requirement on the penetration depth doesn't apply any more, due to the abrupt drop of the magnetic permeability around 750 °C.

As a consequence of the above considerations, the implementation of an accurate 3-D simulation model, taking into account the complete geometrical description of the process, together with the non-linear physical behaviour of the materials, significantly advance the current state-of-the-art in finite element analysis of gear surface hardening processes.

5.2.1 Finite element formulations

Over the years, many formulations have been proposed for the computation of three dimensional problems with eddy currents induced in non-linear, ferromagnetic media [79, 80]. In such problems, both the electric and the magnetic field must be described in conductors; these fields can be derived from the relative potentials in various ways, the main difference being that of choosing either a vector or scalar representation of the magnetic potential [81].

The so-called vector potential formulations solve the following differential equation for the magnetic vector potential $\mathbf{A}$ and the electric scalar potential V:

$$\nabla \times \nu \nabla \times \mathbf{A} + \sigma \frac{\partial \mathbf{A}}{\partial t} + \sigma \nabla V = 0 \tag{5.6}$$

where the reciprocal of the magnetic permeability ν appears. The derivation of the static fields can then be made by using the relations:

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{5.7a}$$

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} - \nabla V \tag{5.7b}$$

On the other hand, the so-called magnetic potential formulations solve the following differential equation for the electric vector potential $\mathbf{T}$ and the magnetic scalar potential Ψ:

$$\nabla \times \rho \nabla \times \mathbf{T} + \mu \frac{\partial \mathbf{T}}{\partial t} - \mu \nabla \frac{\partial \Psi}{\partial t} = 0 \tag{5.8}$$

where the reciprocal of the electrical conductivity ρ appears. The derivation of the static fields can then be made by using the relations:

$$\mathbf{J} = \nabla \times \mathbf{T} \tag{5.9a}$$

$$\mathbf{H} = \mathbf{T} - \nabla \Psi \tag{5.9b}$$

The duality between the two presented formulations is evident and can be reflected on the results of fluxes and currents obtained upon simulation. In fact, the distribution of currents obtained by the scalar formulation is generally more regular, whereas concerning the magnetic field distribution, it is the vector formulation which gives best results [82].

5.3 Gear case-study

In this section, the test-article chosen for the surface hardening simulation study will be presented. The reference work-piece (see Figure 5.3) is a 57-teeth cylindrical spur gear, characterised by an outer diameter of 150 mm (M=2.54 mm). Thanks to their relatively easy design and symmetric geometry, spur gears are relatively simple to manufacture and heat treat, thus representing an ideal candidate for the purposes of this study. The employed material is a low-alloy, medium carbon steel with 0.40% carbon content (AISI 4340), exhibiting good hardenability and high superficial hardness, both beneficial factors for the selected application.

Figure 5.3: Gear test-article - Representation of the test-case used for surface hardening simulation and testing.

5.3.1 Modelling of material properties

The physical properties of steel are primarily a function of its chemical composition: although carbon content and alloying elements are mainly responsible for the mechanical response of the given material, in some cased they may also affect its electromagnetic and/or thermal behaviour. Moreover, prior metallurgical structure is found to have big influence on the material's response to heat treatment.

As it's previously been widely discussed, material physical properties of the part undergoing surface induction hardening dramatically change in the course of the treatment, mainly due to the effect of temperature. In fact, both the electrical and thermal properties show appreciable variations of their value as the work-piece is heated from room to the hardening temperature. Moreover, high values of the magnetic field strength (i.e., high power densities) tend to "saturate" the material, thus reducing the heating efficiency in the superficial layers, where the peak value of the magnetic field is located.

Chemical composition and prior micro-structure

The AISI 4340 is a high strength steel grade commonly used for high quality mechanical components such as gears. Thanks to the medium carbon content and presence of chromium, nickel, and molybdenum, it is particularly suitable for induction surface hardening applications requiring good hardenability and mechanical properties of the

Table 5.1: Chemical composition of the AISI 4340 steel.

Element	Weight %
Carbon	0.38-0.43
Manganese	0.60-0.80
Phosphor	0.035 (max)
Sulphur	0.04 (max)
Silicon	0.15-0.30
Chromium	0.70-0.90
Nickel	1.75-1.90
Molybdenum	0.20-0.30

core. Its chemical composition is reported in Table 5.1.

The considered steel grade is used in the quenched & tempered (Q&T) condition, which is obtained by means of a prior heat treatment producing a homogeneous martensitic micro-structure, conferring relatively high hardness to the core (i.e., 35-40 HRC). By surface hardening such a kind of structure it is possible to obtain a very hard case (i.e., 55-60 HRC) at considerable depths (i.e., 10-15 mm), without excessively increasing quenching intensity, which may result in cracking of the part upon severe cooling. Figure 5.4 illustrates the hardenability diagram for the chose steel.

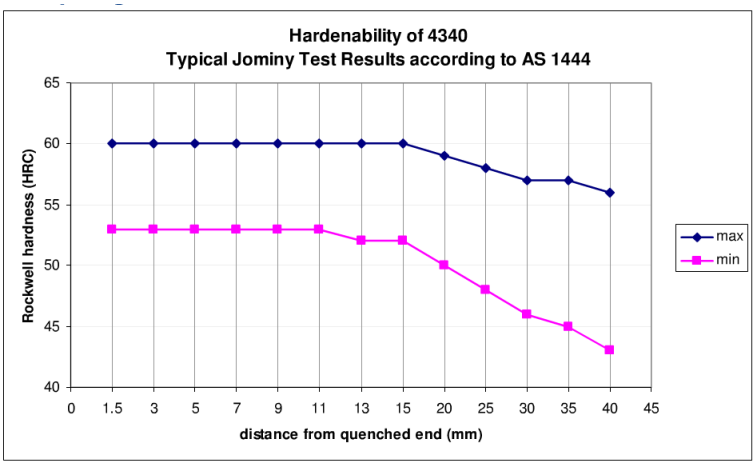


Figure 5.4: Hardenability diagram for AISI 4340 - Jominy test results for the AISI 4340 steel.

One side-effect connected to heat treating Q&T steels is related to the so-called

over-tempering region. This is defined as a localized softening of the material in the heat affected zone (HAZ), caused by an intermediate heating of the untransformed, sub-superficial layers of the gear, which thus tend to a stable state [83, 84]. In fact, when pre-existing tempered martensite is heated to temperatures ranging between 200 and 600 °C, it tends to transform into more stable (and softer) ferrite, thus giving origin to a hardness value drop in the transition zone between the hardened layer and core material (see Figure 5.5).

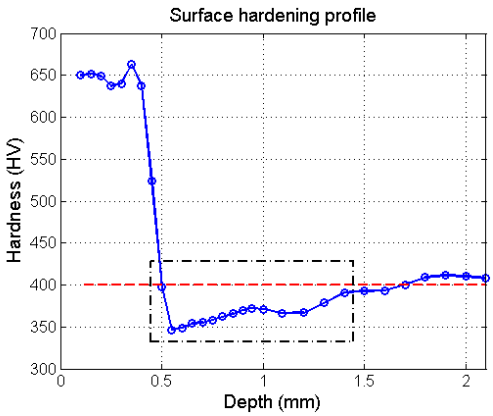


Figure 5.5: Over-tempering region - Illustration of the hardness drop region due to over-tempering.

Prior metallurgical structure is also particularly important for the determination of the upper critical temperature for complete austenitization of the material. In general, the more disperse the carbides are, the lower the Ac_3 point will be. This can be readily explained by the fact that the diffusion distance to redistribute carbon is shorter in finer micro-structures than in coarser ones. Figure 5.6 shows that the increase in critical temperature depends on the initial micro-structure. Moreover, continuous-heating critical temperatures are generally higher than equilibrium ones and this difference increases with the heating rate.

Electrical resistivity

The electrical resistivity is a measure of how strongly a material opposes the current flowing. In other words, it stands for the capacity of the material that contains movable electric charges to make them flow under the action of the electric field: the lower the

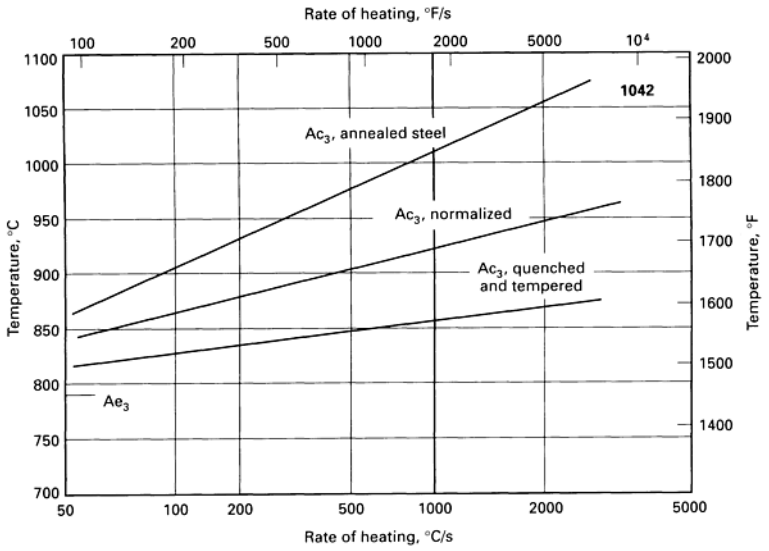


Figure 5.6: CHT diagram for several micro-structures - Effect of prior structure and rate of heating on Ac_3 transformation temperature of 1042 steel [4].

resistivity value is, the higher the flow of charge will be. Resistivity is commonly expressed as a linear function of temperature, according to the following relation:

$$\rho(T) = \rho_0 \left(1 + \alpha_\rho (T - T_0)\right) \tag{5.10}$$

where T_0 is the reference temperature (°C), ρ_0 is the resistivity value at the reference temperature (Ω·m), and α_ρ is the temperature coefficient ($^\circ C^{-1}$). In the simulation model however, available experimentally-determined data has been used, as illustrated in Figure 5.7.

Thermal conductivity

Thermal conductivity is a thermo-physical material property, which characterizes the quantity of heat transferred through a unit surface area in a unit time when the heat transfer is generated by a temperature gradient. In other words, it represents the ability of a material to transfer heat. Thermal conductivity is usually modelled with the following temperature-dependent, exponential relation:

$$k(T) = k_a + k_b \exp\left(-\frac{T - T_0}{\tau}\right) \tag{5.11}$$

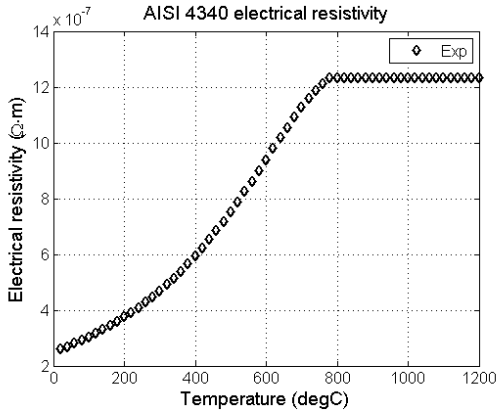


Figure 5.7: Electrical resistivity data for AISI 4340 - Representation of the experimentally-determined data used for modelling the electrical resistivity behaviour at different temperatures.

where T_0 is the reference temperature (°C), k_a is the asymptotic "steady-state" conductivity value (W/m·°C), k_b is the "residual" conductivity value at the reference temperature (W/m·°C), and τ is the temperature constant (°C). Again, available experimentally-determined data has been used in the simulation model, as illustrated in Figure 5.8.

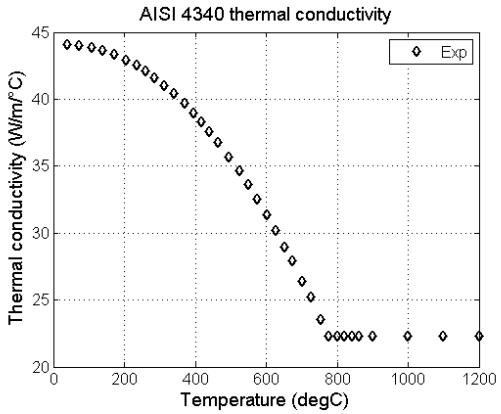


Figure 5.8: Thermal conductivity data for AISI 4340 - Representation of the experimentally-determined data used for modelling the thermal conductivity behaviour at different temperatures.

Table 5.2: Parameters used for the volumetric heat capacity of the AISI 4340.

γC_{P0} **(J/m^3·°C)**	γC_{Pi} **(J/m^3·°C)**	**E (J/m^3)**	σ_{dev} **(°C)**	τ **(°C)**	$\mathbf{T}_\gamma$ **(°C)**
$3{\cdot}10^6$	$6{\cdot}10^6$	$850{\cdot}10^6$	50	300	870

Heat capacity

The heat capacity of a material is the heat quantity necessary to raise its temperature by one Celsius degree. In other words, it is a quantity that offers the possibility to evaluate the amount of heat exchanged by a body during a process characterized by a temperature variation. The so-called volumetric heat capacity (i.e., specific heat capacity per unit volume) is particularly important as it allows to take the energy of phase transition into account. The volumetric heat capacity can be modelled as a combination of an exponential function and a Gaussian function of temperature, as follows:

$$\gamma C_P(T) = \gamma C_{Pi} + (\gamma C_{P0} - \gamma C_{Pi}) \exp\left(-\frac{T}{\tau}\right) + E \cdot G(T) \tag{5.12a}$$

$$G(T) = \frac{1}{\sigma_{dev}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{T - T_\gamma}{\sigma_{dev}}\right)^2\right] \tag{5.12b}$$

where γC_{P0} is the heat capacity (J/m^3·°C) for T=0 °C, γC_{Pi} is the heat capacity (J/m^3·°C) for T$\rightarrow \infty$, E is the energy of phase transition (J/m^3), σ_{dev} is the typical Gaussian standard deviation (°C), T_γ is the phase transition temperature (°C), and τ is the temperature constant of the exponential function (°C). The values used in the simulations are reported in Table 5.2 and the resulting curve is depicted in Figure 5.9.

Magnetic permeability

Generally speaking, hysteresis is a complex phenomenon related to physically irreversible processes. It consists of the fact that, at a given moment, the value of a material property depends not only on the intrinsic properties of the material, but also on its history. Magnetic materials are generally characterized by a hysteresis cycle, which is represented by a closed surface in the (H, B) space. As it is difficult to correctly model hysteresis, a univocal $B(H)$ relationship is usually adopted for magnetic materials; soft magnetic materials, which are characterised by very low values of the

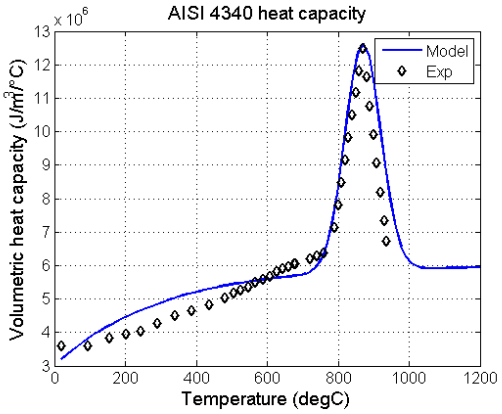


Figure 5.9: Volumetric heat capacity model for AISI 4340 - Representation of the exponential Gaussian model for the volumetric heat capacity: fitting of the experimental data.

Table 5.3: Parameters used for the magnetic model of the AISI 4340.

μ_{r0}	J_s (T)	a	T_C (°C)	C (°C)
500	1.8	0.05	785	80

coercive field, can be modelled by their curve of first magnetization, thus neglecting hysteresis losses.

The $B(H)$ dependency used in the simulations is defined by the following relations:

$$B(H) = \mu_{r0} H + J_s \frac{H_a + 1 - \sqrt{(H_a+1)^2 - 4H_a(1-a)}}{2(1-a)} \tag{5.13a}$$

$$H_a = \mu_{r0} H \frac{\mu_r - 1}{J_s} \tag{5.13b}$$

where μ_{r0} is the initial value of the relative magnetic permeability, J_s is the magnetic flux density saturation value (T), and a is an adjustment coefficient for the sharpness of the transition knee from unsaturated to saturated conditions ($0<a\leq0.5$; the smaller the coefficient, the sharper the curve transition knee). These coefficients have been tuned using a fitting technique based on experimental data taken from [85]. The obtained values are reported in Table 5.3 and the resulting curve is depicted in Figure 5.10.

The presented model can also be extended by introducing the effect of material demagnetization when heated above the Curie temperature. This aspect can be taken into account by means of a temperature coefficient based on exponential functions, as

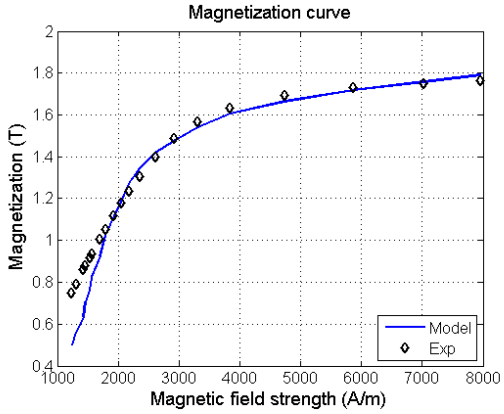


Figure 5.10: $B(H)$ model for AISI 4340 - Results of fitting experimental data with the described 3-parameter model.

follows:

$$COEF(T) = \begin{cases} 1 - \exp\left(\frac{T-T_C}{C}\right), & \text{for } T < T_1 \quad \Rightarrow (T_1 - T_C) = C \ln(0.9) \\ \exp\left(\frac{10(T_2-T)}{C}\right), & \text{for } T > T_1 \quad \Rightarrow (T_2 - T_1) = 0.1 \cdot C \ln(0.1) \end{cases} \tag{5.14}$$

where T_C is the Curie temperature (°C) adn C is the temperature constant (°C). The exponential decay (5.14) is more or less pronounced depending on the value of the temperature constant: higher values of C (e.g., C=100) correspond to faster decays, lower values (e.g., C=10) to slower ones. The used values are reported in Table 5.3 and the resulting curves are depicted in Figure 5.11.

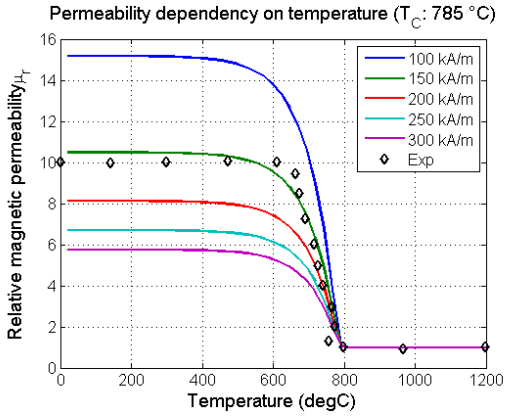


Figure 5.11: Permeability model for AISI 4340 - Demagnetization effect around the Curie temperature value (T_C=785 °C): typical fitting for different applied field strengths.

Hardness and over-tempering prediction

In order to evaluate the resulting hardness at the end of the treatment, an automated routine has been developed, processing thermal history data and estimating the hardened profile along any given path of the geometry. This procedure is mainly based on the characteristic ferrite-to-austenite transformation temperatures and hardenability diagram of the considered steel. Basically, the hardness value is assigned to each point based on the temperature reached at the end of the heating stage: temperatures above Ac_3 correspond to a fully-hardened region; temperatures below Ac_1 correspond to a non-hardened region; temperatures in between Ac_1 and Ac_3 are related to partly-hardened regions.

This theoretical approach has been extended in order to take into account the incomplete carbide dissolution into the matrix and the over-tempering transition from the hardened layer to the softer core [86]:

$$\gamma_r = \exp\left\{-b(T)\cdot(M_s-20)\left[1-f\left(\Delta t_{300}^{700}\right)\right]\right\} \tag{5.15}$$

where b is the same diffusion-less coefficient found in (4.18) and M_s is martnesite start temperature. The function f is defined as follows:

$$f\left(\Delta t_{300}^{700}\right) = 0.41\cdot\left[1-\exp\left(0.03\left(\Delta t_{300}^{700}\right)^{0.6}\right)\right] \tag{5.16}$$

with Δt_{300}^{700} being the time taken to cool each point from 700 °C to 300 °C. In practice, by knowing the martensite formation start temperature M_s (around 300 °C for the given steel), an average cooling rate is computed for each point in the range 700-300 °C and used to estimate the percentage of residual austenite retained in the final micro-structure. An experimentally-tuned, analytic relation is then used to predict over-tempering region, according to the following equation [84]:

$$HV(t,T) = -9.348\cdot\ln(t) - 0.4493\cdot T + 706.13 \tag{5.17}$$

which empirically estimates the hardness drop value in Vickers units.

5.3.2 Description of simulation models

The implementation of the numerical simulations has followed an increasing complexity strategy, which starting from very simplified sliced models, characterised by crude physical assumptions but fast solving time, moved then to more refined approaches, taking full geometrical details and modelling physics into account.

3-D sliced models

A sliced model is obtained by representing only a very thin, central section of the whole gear and placing consistent boundary conditions in order to constrain the magnetic field to be exactly parallel to the tooth surface. In this way it is possible to simulate a 2-D physics into a 3-D model, thus overcoming the impossibility of dealing with magnetic non-linearities in a 2-D H-normal configuration. The original circular geometry is first reduced exploiting the tooth periodicity and its symmetry along the middle radial plane. Second, it is further sliced into a 0.1 mm layer in order to contain the total number of elements and reduce the computation time (see Figure 5.12). Since it is assumed that the achieved solution is constant along the thickness of the model (as in the case of pure 2-D simulations), meshing can be structured accordingly, placing one single element in the axial direction. This simplification does not allow to take into account the real field behaviour at the end extremities of the gear (the so-called edge effects), but still provides a good approximation of the real power distribution over a mid-plane cross section, located far from edges.

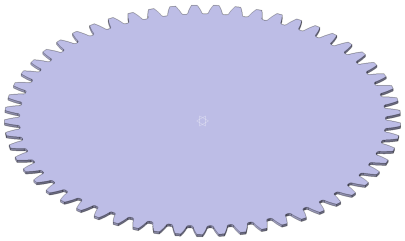


Figure 5.12: Sliced model geometry - Representation of the sliced model simplification.

A single-turn induction coil encircling the gear is added in order to impose the given current source. The inductor is modelled as a coil conductor, which enables the representation of non-magnetic conducting regions with sources. This solution does not account for the real current density distribution within the coil region, which is higher on the inner profile due to the electromagnetic ring effect. The above assumptions allow faster computations without affecting the resulting power density profile, which remains qualitatively unaltered.

The described geometry is meshed using both tetrahedral and brick elements of the first order, obtaining the discretization shown in Figure 5.13. The resulting mesh consists of 6766 volume elements and 9166 nodes. On the tooth profile region, a progressive layered mesh is used, accounting for at least three elements per penetration depth. The eddy current problem is solved by means of a scalar T-Ψ formulation, resulting in lighter computations thanks to the reduced number of required degrees of freedom. In fact, the main advantage of this type of formulation is its reduced memory occupancy and faster transient resolution time, compared to the A-V: a typical simulation can thus be performed in about 40 minutes on a standard single-core workstation with 4 GB of RAM.

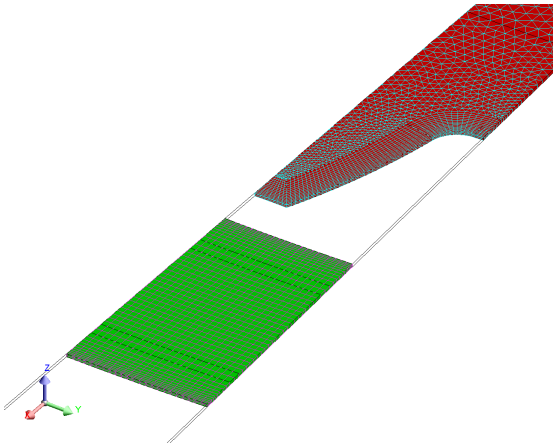


Figure 5.13: Sliced model mesh - Example of discretization for 3-D sliced models.

3-D solid models

A solid model is obtained by extrusion of the sliced one in the axial direction (see Figure 5.14). For this reason, the thickness of the gear-coil assembly can be fully considered and represented, allowing to take extremity effects into account. These effects appear at the edges of the teeth and are caused by local inhomogeneities of the magnetic field distribution at the air-to-metal interface, which produce high current density spots and, thus, overheating areas. In order to take the real field distribution into proper account, the conductive regions must be surrounded by an air box sufficiently large not to distort the flux density lines. Moreover, "infinite" elements can be used to exponentially approximate the field decay at boundaries with a reduced number of elements,

thus optimizing the size of the problem.

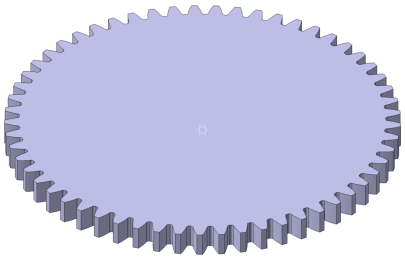


Figure 5.14: Solid model geometry - Representation of the sliced model simplification.

The inductor is modelled as an active solid conductor, in which the skin effect is evident and the current density is non uniform on the cross-section of the coil. This choice allows to take both ring and proximity effects into consideration. These effects lead to a concentration of the current density on the inner layers of the coil cross-section, thus increasing the coupling between the gear and the coil and enhancing the overall efficiency of the process. The implementation of real current distributions within the inductor comes at the cost of a finer meshing requirement, which has to reflect the actual penetration depth at the considered frequency.

The meshing approach is based on the one adopted for the sliced models (see Figure 5.15). However, since the real thickness of the system is considered, a finer discretization must be used in the proximity of sharp edges at the extremities. As a consequence, interfaces between structured (i.e., brick elements) and free (i.e., tetragonal elements) meshing might be critical, in particular in the transition areas within the same region (e.g., gear case-to-core transition) or between different regions (e.g., gear-to-air box transition). The resulting mesh, consisting of 55,731 nodes and 158,317 volume elements, is associated with a vector A-V formulation and allow the problem to be solved in approximately 200 minutes on a professional multi-core[1] workstation with 16 GB of RAM.

3-D complete models

When dealing with complex geometries like the one here considered, which is characterized by a grooved profile in correspondence of the gearing and by a series of lightening

[1]In order to fully exploit multiple core architectures, a suitable parallel solver must be used.

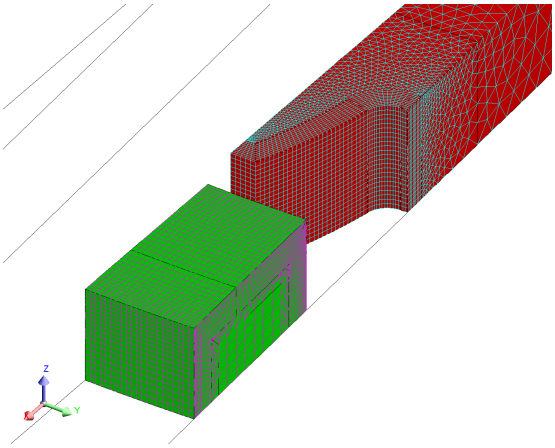
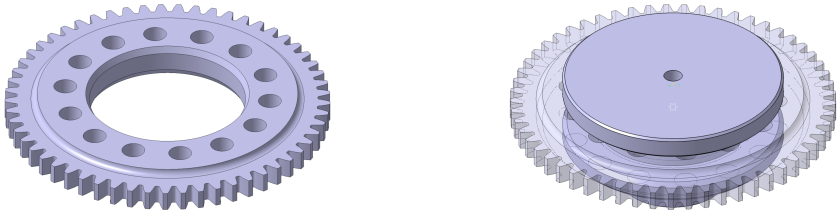


Figure 5.15: Solid model mesh - Example of discretization for 3-D solid models.

holes around the central bore (see Figure 5.16), including such details in the simulation model allows to increase the accuracy of the obtained results. Full models can also consider the effect of the "machine environment", for example the type of support and fixation used to put the gear in rotation during the treatment. This aspect is particularly important as the simulated thermal fluxes might be modified by the presence of highly conductive materials placed in contact with the gear during the process.

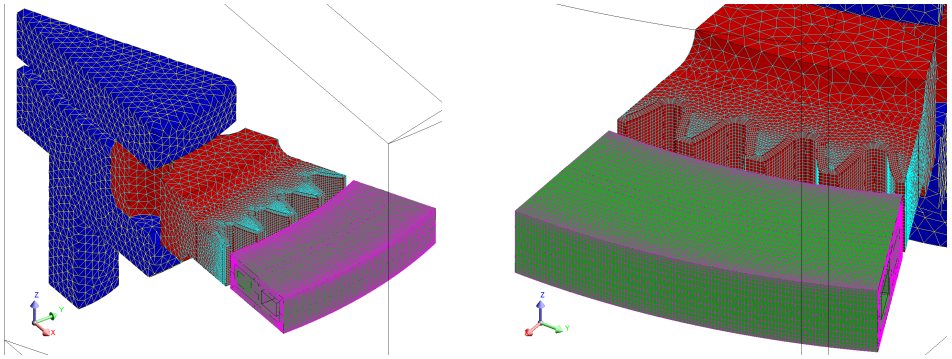


(a) Fully detailed geometry.

(b) Particular of the gear support and fixation cap.

Figure 5.16: Complete model geometry - Representation of the complete model.

In the considered model, the mesh consists of 101,874 nodes and 477,709 volume elements (see Figure 5.17). A single computation can be performed in approximately 500 minutes on a professional multi-core workstation with 24 GB of RAM, by using the scalar A-V formulation.



(a) Geometry and mesh overview. **(b)** Close-up on mesh.

Figure 5.17: Complete model mesh - Example of discretization for 3-D complete models.

Process parameters

Practically, a variety of heating recipes can be used to successfully contour heat treat the described gear. Here, a standard procedure is performed using the double frequency approach, consisting of:

1. a pre-heating stage using middle frequency (10 kHz) and low power densities to bring the teeth to a temperature of about 500 °C;
2. a soaking stage to let the temperature partially uniform on the contour of teeth;
3. a heating stage using high frequency (180 kHz) and high power densities to austenitize the surface of the teeth.

After a careful calibration of the different simulation models previously presented, the 3-D solid model has been chosen and used to compare and validate three different recipes, which are reported in Table 5.4. First, the optimal HF power setting is chosen, which allows to heat the tooth from the pre-heating temperature to the austenitizing temperature with acceptable uniformity and without overheating the edges. Then, different power-over-time ratios are compared for the MF stage. The quenching phase is simulated by switching the power off and adjusting the heat transfer coefficient on the surface of the tooth in order to fit the behaviour of the quenching liquid. In this case, a high-cooling-rate fluid with 10% polymer concentration and a temperature of 25 °C has been considered (see Figure 5.18).

Table 5.4: Heating recipes tested and compared by means of 3-D solid models.

Case	MF (10 kHz)	Soaking	HF (180 kHz)
A	2.0 s - 33 kW (0.78 kW/cm^2)	0.3 s	0.5 s - 111 kW (2.64 kW/cm^2)
B	4.4 s - 18 kW (0.43 kW/cm^2)	0.1 s	0.5 s - 110 kW (2.62 kW/cm^2)
C	6.5 s - 16 kW (0.38 kW/cm^2)	0.1 s	0.5 s - 110 kW (2.61 kW/cm^2)

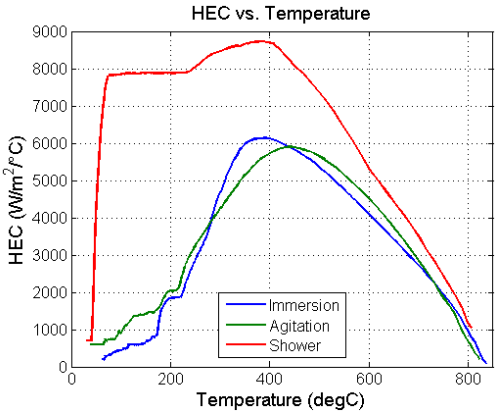


Figure 5.18: HEC as a function of temperature - Example of cooling curves for a PAG quenching fluid with 10% polymer concentration at 25 °C.

5.4 The multi-physics simulation procedure

The presented problem can be analysed in two main steps: the heating phase, consisting of a fast electromagnetic heating caused by the current induced inside the workpiece, and the cooling phase, during which a quenching fluid is sprayed directly over the heated profile in order to rapidly cool it down and obtain the hardened martensitic layer.

While the heating phase is simulated by solving the coupled 3-D eddy current distribution problem coupled with the time-dependent heat transfer equation, the cooling phase deals with thermal aspects only. In fact, the quenching intensity of the cooling fluid can be represented by means of a sudden variation in the heat flux losses of the quenched surface. Such variations are usually modelled with a temperature dependent HEC, which is tuned on the specific cooling curves of the employed quenching medium. The cooling rate of polymer-based quenching liquids, which are the most commonly used for gear-hardening applications, mainly depends on the concentration of the solution, its temperature, and the degree of agitation. Thus, HEC results to be greatly

dependent on the superficial temperature of the metal part and is modelled accordingly: at each time-step, the average temperature over the tooth profile is evaluated and used to update the current value of the HEC. Particular care must be taken in handling the sudden change in the HEC value resulting from the beginning of the quenching phase. In fact, an abrupt increase of such coefficient requires a refinement of the time-stepping in order to accurately describe the thermal dynamics at the end of the heating phase.

In order to translate the above considerations into the simulation software, its multi-physics capabilities must be fully exploited. A standalone model is built for simulating each of the main concurrent stages involved in the hardening process (i.e., eddy current losses, heating and cooling phases), and a scripting procedure is written in order to couple the different phenomena (i.e., magnetic and thermal). In practice, by means of a continuous data exchange and synchronization between the applications, the thermal variations of the main electric and magnetic parameters of the gear are taken into consideration (see Figure 5.19). This process is carried out in two steps, which are iterated until a specific precision of the residual is reached. First, the coordinates of the nodes of the mesh are transferred through a data-exchange text file. The physical quantities of interest are then synchronized and updated.

5.4.1 Results of the heating simulations

Recipe A

The first analysed recipe is the one characterized by the shorter global treatment time. By inspection of Figure 5.20 it is possible to appreciate the power densities distributions at the beginning of the MF and HF stage, respectively. It can be observed that the HF "deposition pattern" is quite different from that one might expect, since it presents a concentration of currents in the root area of the tooth. Also, the general distribution looks similar, with only slight changes in the relative intensity. A couple of considerations arise:

- the considered HF value is probably still too low for letting eddy current concentrate on the tip area of the tooth: the small module of the considered gear would require much higher frequency values to confine the power density on the contour;

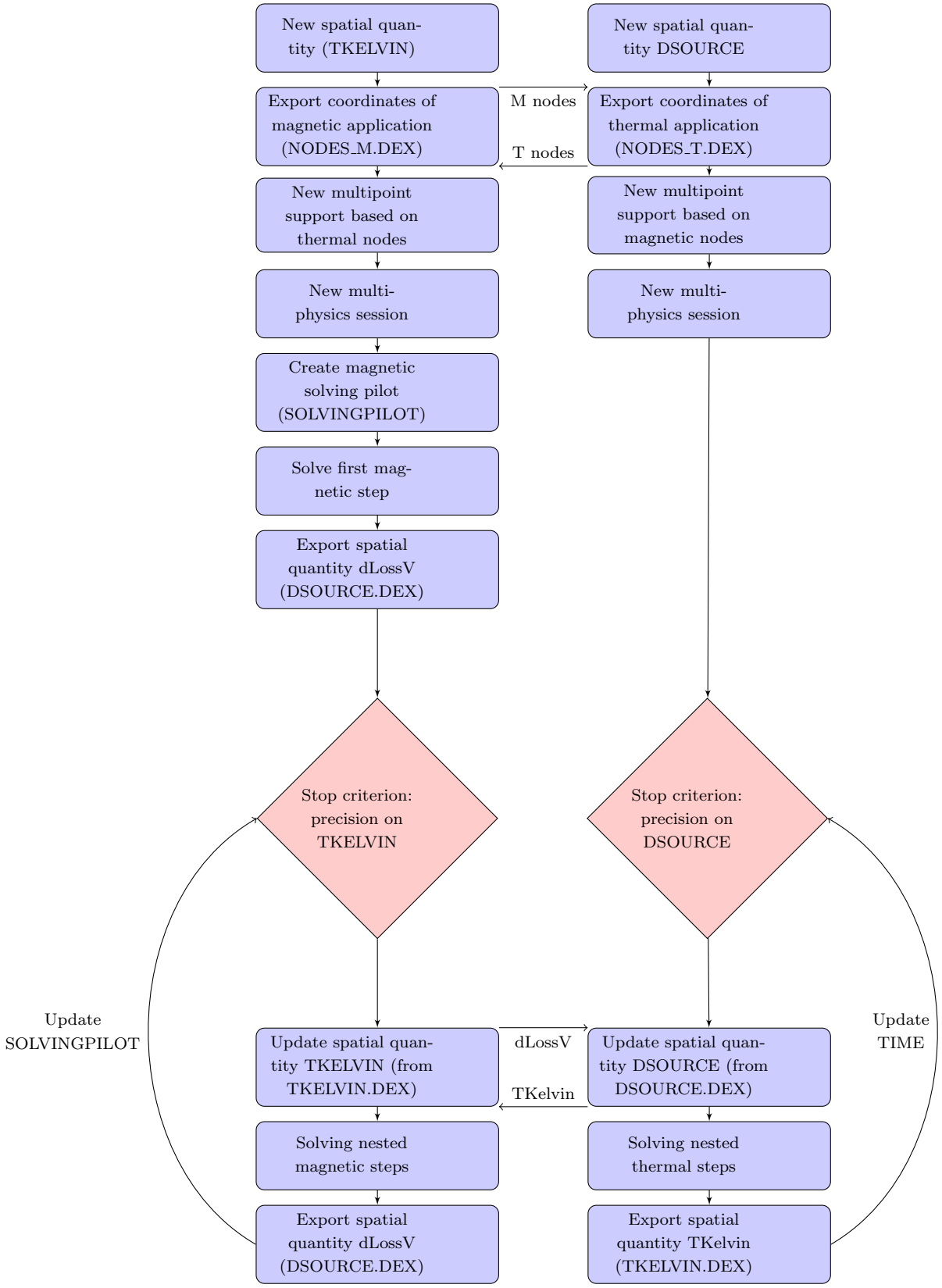
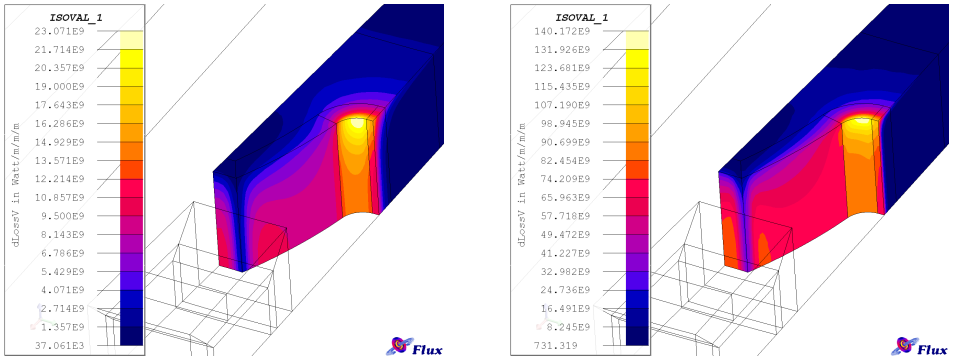


Figure 5.19: Multi-physics simulation - Flow chart of the multi-physics solving procedure used.

- the MF pattern refers to the first heating stage of the treatment, in which the gear is still below the Curie point, whilst the HF one corresponds to the second heating stage, where the superficial layers are well above the Curie point.



(a) MF pre-heating stage. **(b)** HF austenitizing stage.

Figure 5.20: Power density distributions - Illustration of the simulated heat source deposition patterns in the course of a double frequency surface hardening treatment.

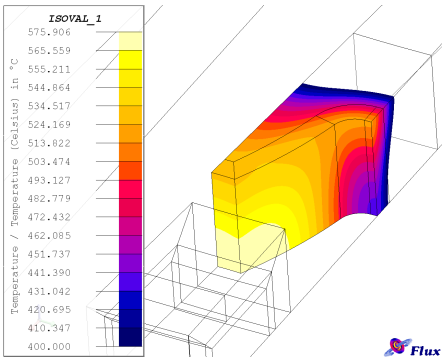
These evidences explain why heat sources tend to concentrate on the free extremities of the gear, in particular in the root area. In fact, when looking at the temperature distributions of each stage (see Figure 5.21), it can be seen that after the pre-heating phase, the tip of the tooth is hotter than the root, thus causing the magnetic permeability value to be higher. As a consequence, the magnetic flux density will remain concentrated on the root area, even during the HF stage.

Recipe B

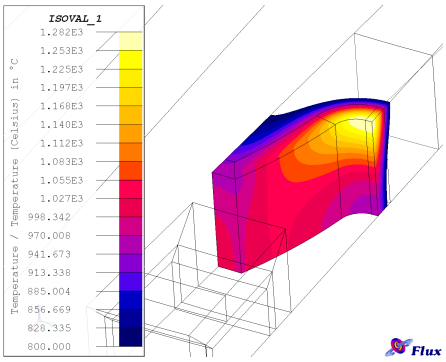
Compared to the first one, the second recipe is characterized by a longer MF pre-heating time and a shorter soaking stage before HF heating. The resulting temperature distributions at the end of each heating stage are depicted in Figure 5.22.

Recipe C

The third recipe is the one characterized by the longer global treatment time. The resulting temperature distributions at the end of each heating stage are depicted in Figure 5.23.

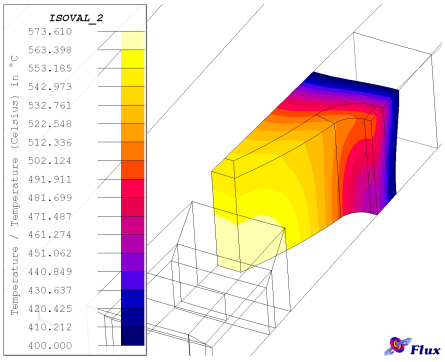


(a) MF pre-heating stage (400 °C isothermal).

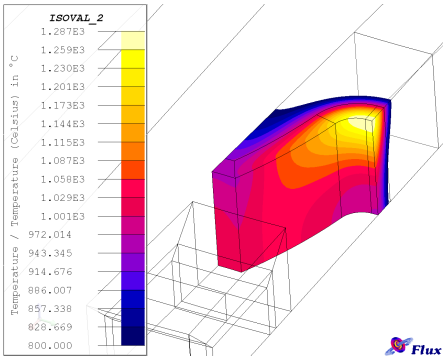


(b) HF austenitizing stage (800 °C isothermal).

Figure 5.21: Recipe A: temperature distribution during heating - Illustration of the simulated temperature distributions after each stage of a double frequency surface hardening treatment (Recipe A).

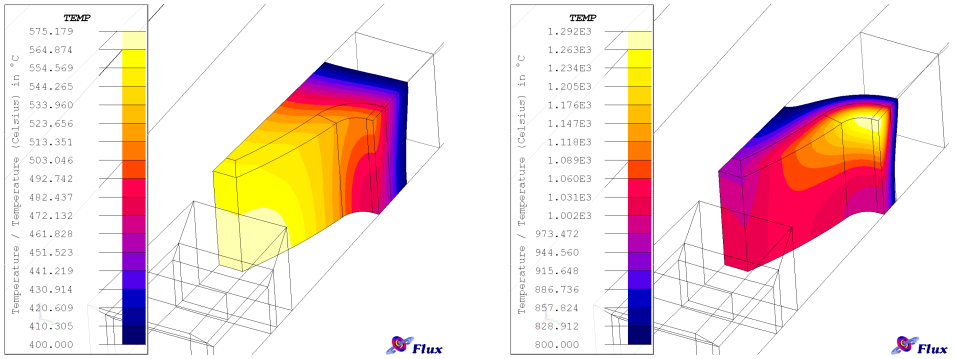


(a) MF pre-heating stage (400 °C isothermal).



(b) HF austenitizing stage (800 °C isothermal).

Figure 5.22: Recipe B: temperature distribution during heating - Illustration of the simulated temperature distributions after each stage of a double frequency surface hardening treatment (Recipe B).



(a) MF pre-heating stage (400 °C isothermal).

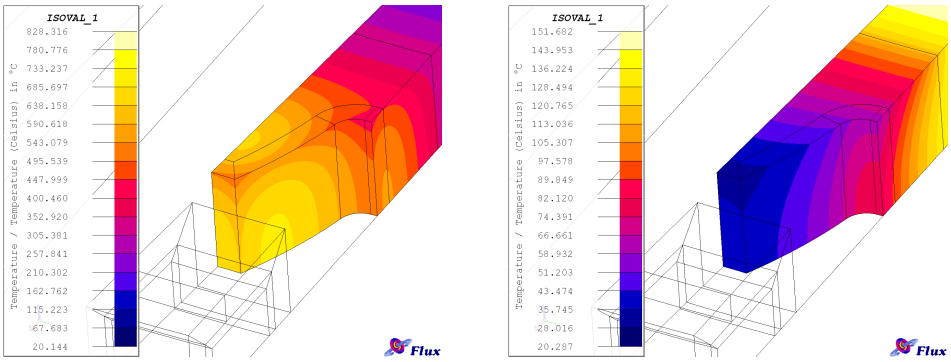
(b) HF austenitizing stage (800 °C isothermal).

Figure 5.23: Recipe C: temperature distribution during heating - Illustration of the simulated temperature distributions after each stage of a double frequency surface hardening treatment.

5.4.2 Results of the cooling simulations

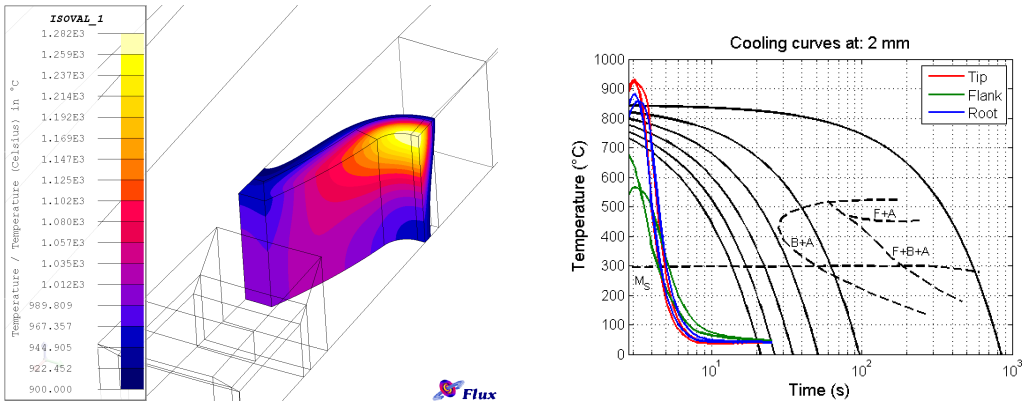
After the simulation of the heating process, the quenching phase is considered, consisting of a superficial, forced cooling of the gear by means of a sprayed polymer-based fluid. In order to correctly take into account such phenomenon, a specific multi-physics procedure is used, which turns the power source off, updates the thermal exchange coefficients of the gear surface, and runs the transient simulation based on the provided input parameters. Usually, a dwell time is considered before starting the actual quenching phase, corresponding to a natural heat diffusion stage which can range from some fractions of a second to a few seconds, depending on the metallurgical (i.e., heating uniformity) and technological (i.e., work-piece handling) constraints.

In the post-processing stage, the thermal history of the gear is processed along several virtual paths located in critical areas of the tooth, namely the tip, the flank and the root. These data provide the input for the subsequent hardness prediction routine, whose results will be described in the Chapter 6. The following results consider a fixed dwell time of 0.3 seconds and an overall process time (i.e., heating + cooling phases) of 25 seconds.



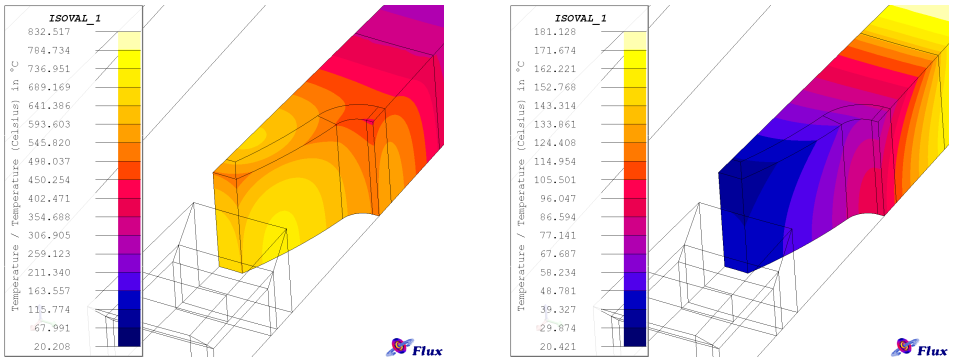
(a) After 1 s of quenching. **(b)** After 5 s of quenching.

Figure 5.24: Recipe A: temperature distributions during quenching - Illustration of the simulated temperature distributions in the course of the quenching phase.



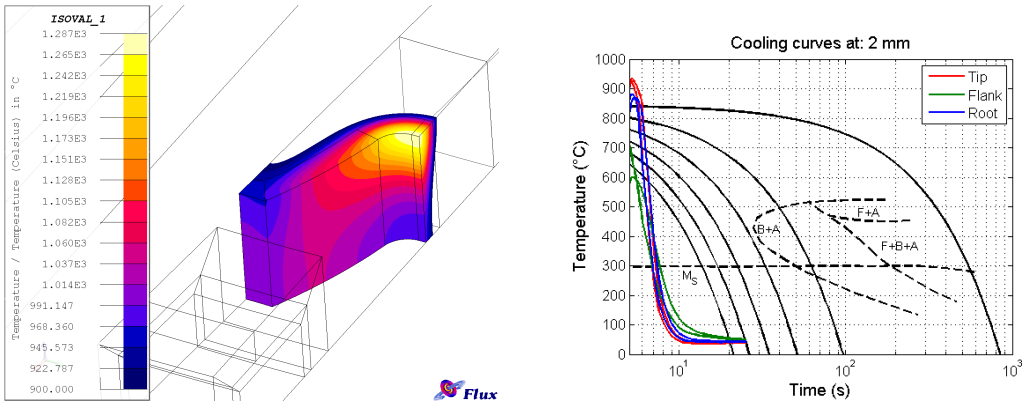
(a) HF austenitizing stage (900 °C isothermal). **(b)** Cooling curves over the CCT diagram.

Figure 5.25: Recipe A: example of cooling history - Illustration of the simulated temperature distributions at the end of the heating phase and cooling curves on critical locations at a depth of 2 mm under the surface.



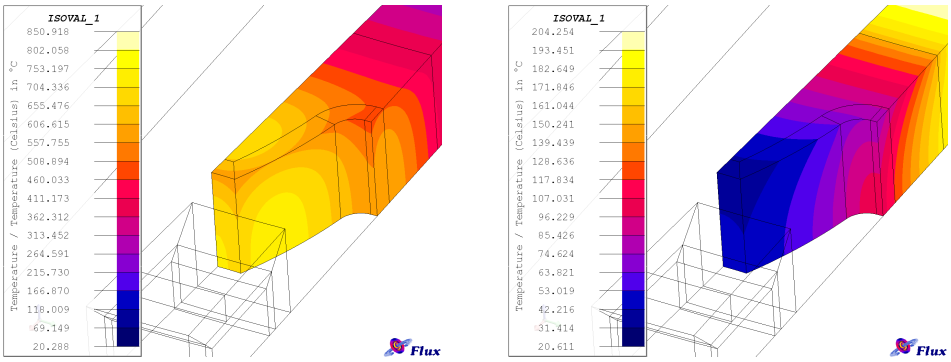
(a) After 1 s of quenching. **(b)** After 5 s of quenching.

Figure 5.26: Recipe B: temperature distributions during quenching - Illustration of the simulated temperature distributions in the course of the quenching phase.



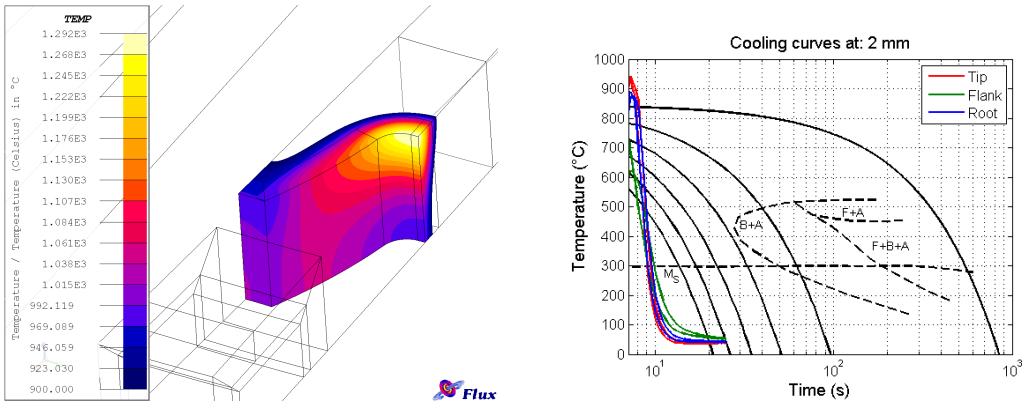
(a) HF austenitizing stage (900 °C isothermal). **(b)** Cooling curves over the CCT diagram.

Figure 5.27: Recipe B: example of cooling history - Illustration of the simulated temperature distributions at the end of the heating phase and cooling curves on critical locations at a depth of 2 mm under the surface.



(a) After 1 s of quenching. **(b)** After 5 s of quenching.

Figure 5.28: Recipe C: temperature distributions during quenching - Illustration of the simulated temperature distributions in the course of the quenching phase.



(a) HF austenitizing stage (900 °C isothermal). **(b)** Cooling curves over the CCT diagram.

Figure 5.29: Recipe C: example of cooling history - Illustration of the simulated temperature distributions at the end of the heating phase and cooling curves on critical locations at a depth of 2 mm under the surface.

6

Experimental results

The validation of the results obtained by simulation is performed by testing the three recipes presented in Chapter 5 on a dedicated simultaneous dual frequency induction hardening machine (see Figure 6.1). Such piece of equipment is capable of providing up to 800 kW MF power at a fixed frequency of about 10 kHz and up to 600 kW HF power at a frequency ranging between 150 and 250 kHz, depending on the global impedance of the system (i.e., coil used and treated part) which is defined by the chosen application.

Figure 6.1: Induction equipment - Picture of the equipment used for the experimental tests.

Since the power requirement determined by simulation represents the losses by Joule

effect evaluated on the gear and inductor regions only, the regulation of the experimental machine must be performed by taking into considerations all the additional losses related to the power conversion chain from the network to the legs of the inductor. For the described machine, such losses are known to be around 20% of the nominal setting. Moreover, as the power level is set as a percentage of the maximum value, a calibration procedure is needed in order to provide the prescribed power to the gear. The results of such calibration are illustrated in Figure 6.2.

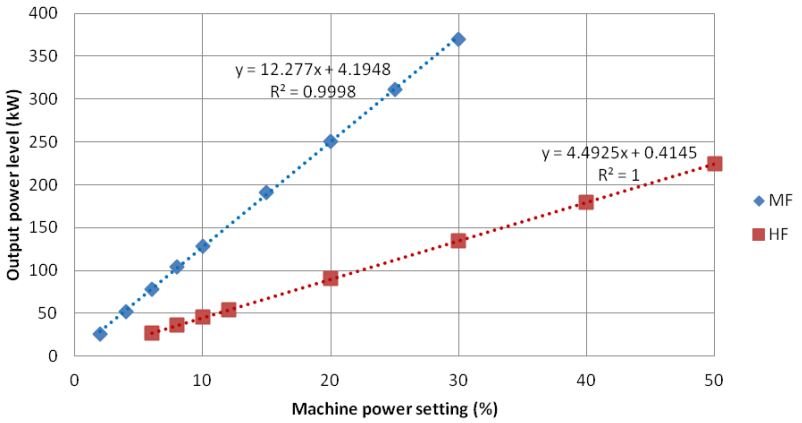


Figure 6.2: Machine calibration - Correspondence between machine power setting and effective output power.

In the next sections, a description of the experimental set-up used to perform temperature and hardness values measurements will be given.

6.1 Temperature measurement

Measuring the temperature of a surface induction hardened component is a challenging task, mainly due to the very high heating rate imposed to the part. In the case of contour hardening, the physical rotation of the gear represents a further complication, as it prevents any contact-based method from being used. An overview of common temperature measurement techniques is presented in Table 6.1.

It is quite clear that the ideal technique should be characterized by a wide temperature range, in order to follow the treatment from the beginning to the end, a large measuring surface, for providing point-wise information over an entire region, a fast response time, for having a good temporal resolution, and a fine accuracy, for having a

Table 6.1: Classification of common temperature measurement techniques.

Method	Range	Area	Response	Accuracy
Thermocouple	0-1250 °C	Surface spot	Variable	±2%
IR pyrometer	0-1650 °C	Surface spot	≈10-100 ms	±1%
IR camera	0-1200 °C	Full field	≈10-100 ms	±1%
Temperature lacquers	25-1050 °C (discrete)	Full field	≈1 ms	±1%

good spatial resolution. On one side, the main limitation of *thermocouples* is related to the fact that they must be placed in contact with the part to be measured; furthermore, they might be susceptible to the electromagnetic field and have a non-linear response for temperatures above the Curie point. On the other side, contact-less, infra-red based devices, like *pyrometers* or thermal cameras, respectively suffer from a limited field of view and a quite low spatial resolution. Finally, *thermal lacquers* alone are available in discrete temperature values and don't provide any data acquisition capability.

A very smart approach to overcome the main limitations of standard techniques is to combine the use of thermal lacquers with high speed video recording [87]. In this way, a very accurate method is obtained, which is able to capture temperature variation with high spatial resolution and high acquisition rate. The core of such new method is based on the proper analysis of huge imaging datasets (up to 3000 frames per second), in order to detect the border between the evaporated and non-evaporated portion of the paint. In practice, after having covered several teeth sectors of the gear with different thermal lacquers (see Figure 6.3), sequential pictures of the part are taken in the course of the treatment and then processed to extract the needed information. A complex procedure is necessary at this stage, mainly consisting of image registration, thresholding, and segmentation operations.

6.2 Hardness measurements

Following to heat treatment, hardness measurements have been performed on each tested recipe, using a Vickers micro-indenter and a test load of 500 gf. Radial hardness profiles have been taken on the tip (point spacing: 200 μm) and the root (point spacing: 100μm) of the tooth over three different cross-sections of the part: the upper surface, the central section, and the lower surface. The experimental points have then

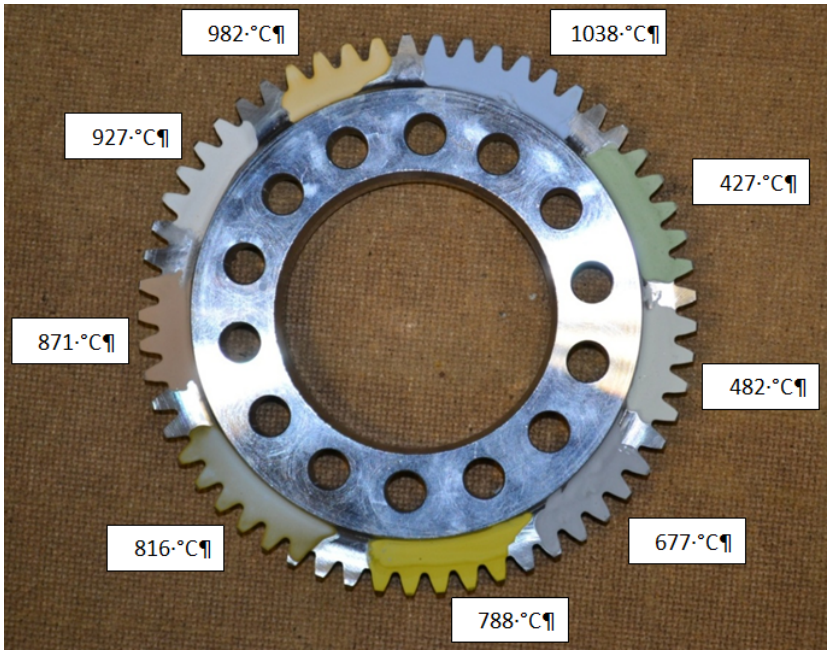
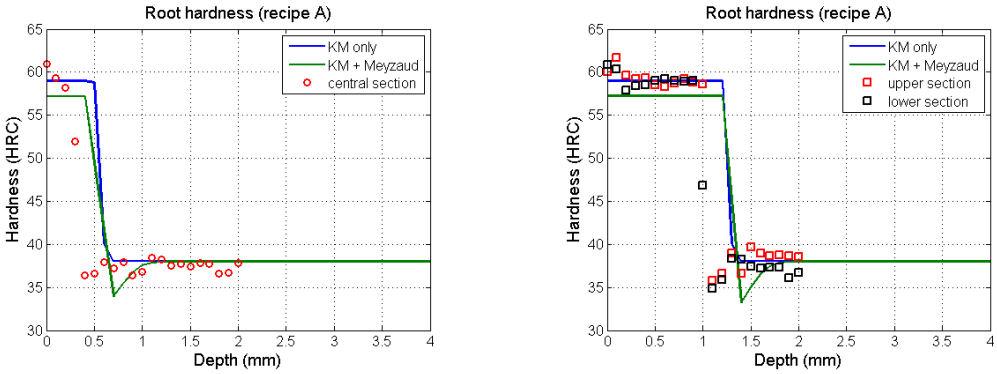


Figure 6.3: Use of thermal lacquers - Example of using several thermal lacquers for precise temperature measurement.

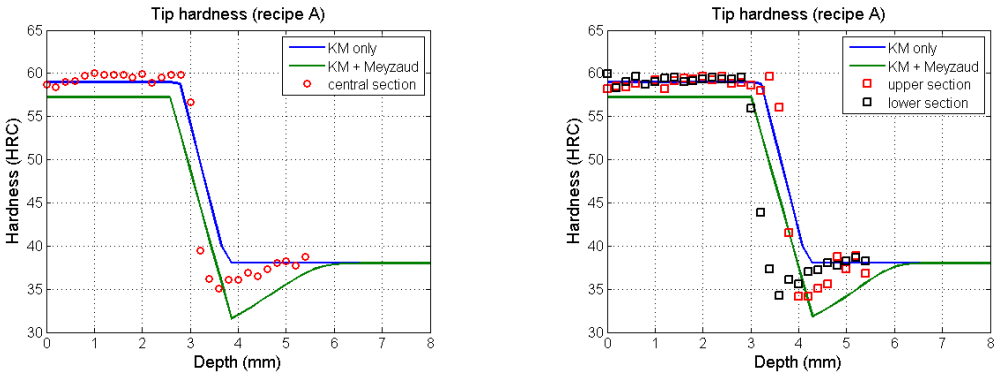
been compared with the calculated profile by applying the hardness prediction routine described in Chapter 4 to the thermal distributions obtained from the simulations. The obtained results are presented in Figures 6.4-6.9, where the blue line correspond to the prediction based on the Koistinen-Marburger model only and the green line is the same prediction corrected with the Meyzaud's equation for the retained austenite and the Ducassy's equation for the over-tempering region.

It can be observed by comparing the experimental hardness profiles at the two extremities of the gear, that a slightly different case depth value is obtained. This effect is particularly visible in recipe C, the one characterized by the longer pre-heating time. Though yet no investigation has been carried out to clarify the origins of this evidence, the main cause might be related to the particular stainless steel support used as a fixation for the gear to the rotating shaft of the machine (see Figure 6.10). In fact, the lower support results to be more massive than the upper cap, thus possibly acting as a "passive cooling device" for the lower extremity of the gear, which may not be symmetrically heated up. Furthermore, since heat diffusion is a time-dependent process, the more marked influence on recipe C would also be explained, since more



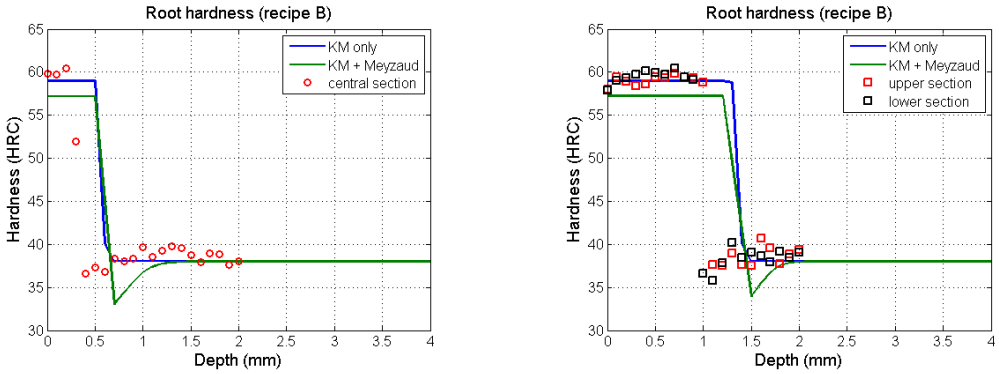
(a) Central section. **(b)** Extremities.

Figure 6.4: Recipe A: root hardness results - Illustration of the simulated hardness profiles on the tooth root and obtained experimental results.



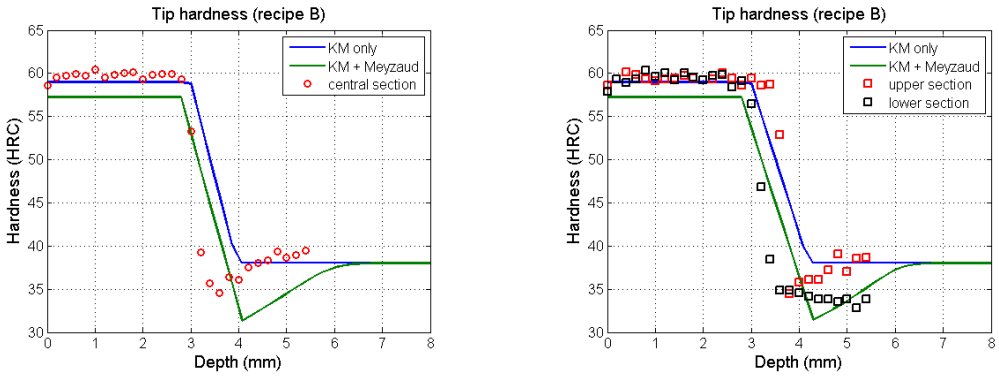
(a) Central section. **(b)** Extremities.

Figure 6.5: Recipe A: hardness results - Illustration of the simulated hardness profiles on the tooth tip and obtained experimental results.



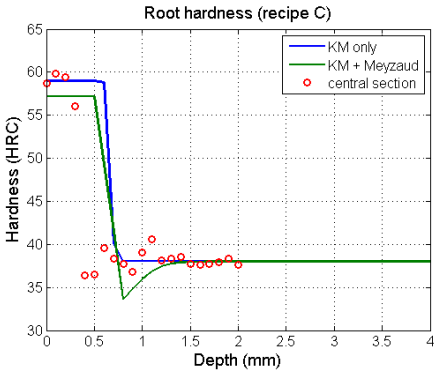
(a) Central section. **(b)** Extremities.

Figure 6.6: Recipe B: root hardness results - Illustration of the simulated hardness profiles on the tooth root and obtained experimental results.

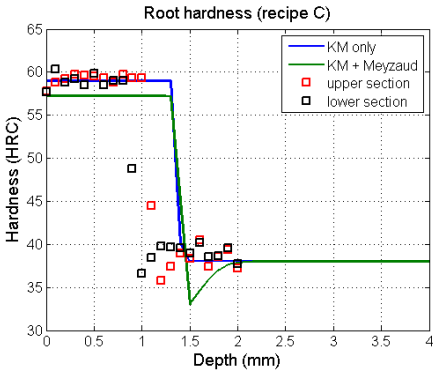


(a) Central section. **(b)** Extremities.

Figure 6.7: Recipe B: hardness results - Illustration of the simulated hardness profiles on the tooth tip and obtained experimental results.

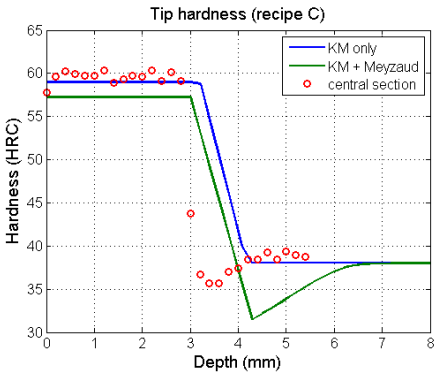


(a) Central section.

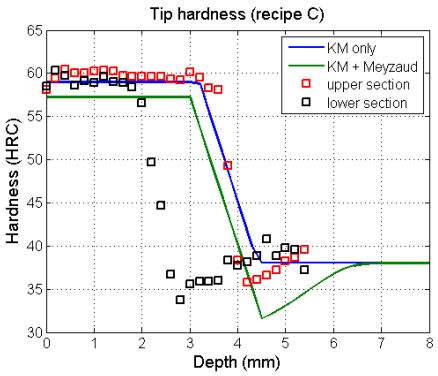


(b) Extremities.

Figure 6.8: Recipe C: root hardness results - Illustration of the simulated hardness profiles on the tooth root and obtained experimental results.



(a) Central section.



(b) Extremities.

Figure 6.9: Recipe C: hardness results - Illustration of the simulated hardness profiles on the tooth tip and obtained experimental results.

time is left for heat to diffuse into the support.

Although a compensation of this effect could be introduced by scanning the gear in the axial direction during the heating phase, in order to compensate for the heat sink effect of the steel support, such a kind of strategy would be quite tedious to tune, so the re-design of supports would be preferable.

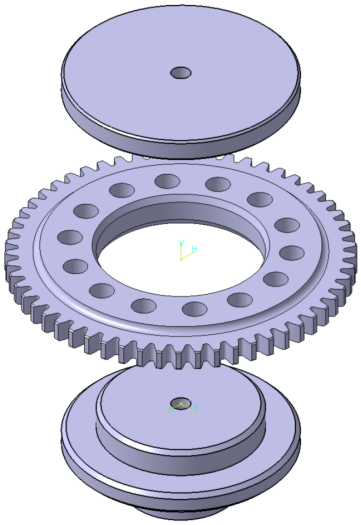


Figure 6.10: Gear support - Exploded view of the gear support and fixation cap used.

6.3 Residual stress measurements

A crucial quantity to be determined, which has a great impact on the mechanical properties and fatigue life of the gear, are residual stresses. The typical stress distribution obtained upon surface hardening, presents a compressive pattern corresponding to the hardened layer which then steeply switches to traction in the core.

A disquisition upon different techniques that can be used for the determination of residual stresses has already been done in a previous chapter. Here, x-ray diffraction has been used to measure axial residual stress values as a function of the depth from the surface of the tooth. XRD exploits the $\sin^2 \Psi$ method, which is based on the detection of diffraction angles for different positions of the sample. By knowing the Young's modulus and Poisson's ratio of the inspected material, the stress value at the given

depth can be calculated from the slope of regression line relating the atomic lattice spacings and the diffraction angle.

Since XRD is a surface analysis technique, electrochemical polishing (see Figure 6.11) has been used to continuously and slowly remove very thin material layers without compromising the existing tensional state of the gear. However, a slight redistribution of such tensions still occurs and the calculated stress profile should be corrected with a specific calibration procedure in order to obtain the real values. Although only the non-calibrated, measured values are here considered, these data provide a qualitative idea of the residual stress state of the gear after heat treatment.

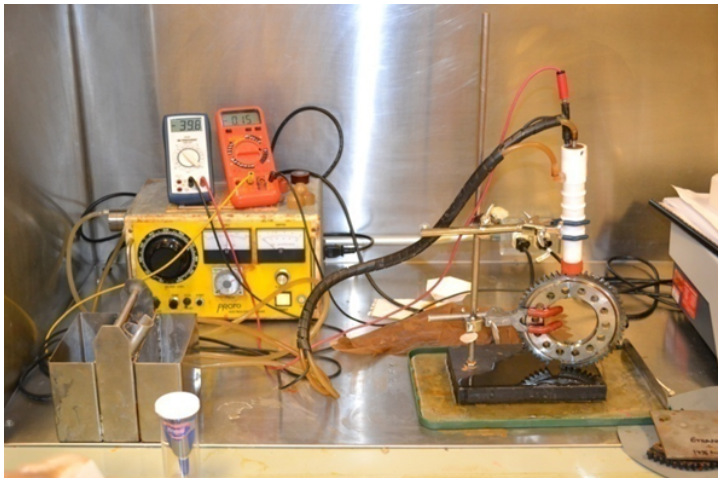


Figure 6.11: Electrochemical polishing test bench - Overview of the electrochemical polishing device.

In Figure 6.12, a comparison of the stress profiles obtained with each one of the three tested recipes is illustrated. Whilst similar stress values are found in the case and in the core of the gear, the main difference between the curves can be observed in the transition zone, where a progressively smoother trend is shown when moving from recipe A to recipe C. This evidence seems to be correlated with the longer pre-heating stage used in the latter, which allows more time for heat to diffuse inside the part before quenching, thus lowering the thermal gradients and more evenly distributing the stress state. The differential start of the quenching phase is illustrated in Figure 6.13, where the HEC behaviour is plot versus treatment time.

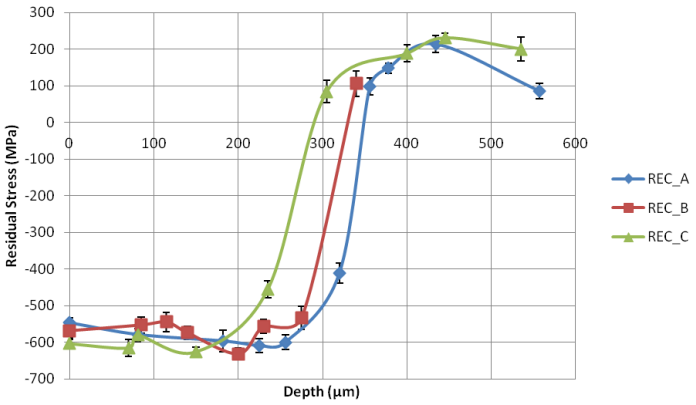


Figure 6.12: Residual stress profiles - Comparison of the measure residual stress profiles on the three tested recipes.

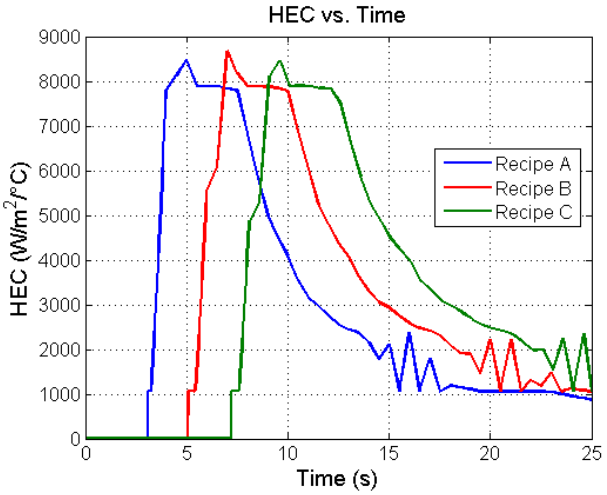


Figure 6.13: HEC behaviour during quenching - Variation of the HEC value in the course of the cooling phase; note the delay in quench phase initiation between the three tested recipes.

7

Conclusion

In the present work, the induction surface hardening process has been discussed and analysed, thus providing a detailed description of the different steps needed to successfully represent its multi-physical dimension by simulation. In particular, the double-frequency gearwheels contour hardening application has been used as paradigm. The conventional approach, considering the heating stage alone, has been extended to properly handle the non-linear heat exchange coefficient during quenching, as well as to provide a built-in hardness evaluation routine based on thermal history analysis. Although satisfactory results have been obtained with this simple approach, the procedure can be further enhanced by taking material phase transitions into account, both in the course of the heating and cooling phases.

The development of the numerical procedures and the implementation of the simulation models have been complemented by the experimental testing activity on a dedicated induction hardening machine, with the main aim of comparing and possibly validating the computed results. Moreover, a big effort has been put in the standardization of the simulation procedure, which can be now carried out in a quite general way, by providing only the necessary process parameters. This point is of particular importance as it could make it easier to couple finite element analysis with numerical optimization techniques, thus providing users with a powerful and comprehensive design tool for gear induction hardening applications.

References

[1] S. L. Semiatin and D. E. Stutz, *Induction heat treatment of steel*. American Society for Metals, Jan. 1986. 1, 53

[2] V. Biringuccio, *De la pirotechnia*, 1559. 1

[3] A. Mühlbauer, *History of Induction Heating and Melting*. Vulkan-Verlag GmbH, Jun. 2008. 4, 98

[4] A. S. M. International, *ASM Handbook Volume 4: Heat Treating*, 10th ed. ASM International, Dec. 1991. 6, 19, 60, 61, 113, 149

[5] G. Krauss, *Steels: processing, structure, and performance*. ASM International, Aug. 2005. 10, 12

[6] A. S. f. Metals, *Atlas of Isothermal Transformation and Cooling Transformation Diagrams*. ASM, 1977. 11, 12, 114

[7] A. S. M. International, *ASM Handbook Volume 1: Properties and Selection: Irons, Steels, and High-Performance Alloys*, 10th ed. ASM International, Apr. 1990. 13

[8] J. Kirkaldy and D. Venugopalan, "Prediction of microstructure and hardenability in low-alloy steels," *Phase Transformations in Ferrous Alloys*, pp. 125–148, 1983. 15

[9] G. Krauss, *Steels: heat treatment and processing principles*. ASM International, Jan. 1990. 16

[10] NATEC, "Internet microscope for schools," 2005. [Online]. Available: http://pwatlas.mt.umist.ac.uk/internetmicroscope/micrographs/microstructures/heat-treated-steel.html 17, 22, 27

[11] C. Wick and S. o. M. Engineers, *Tool and Manufacturing Engineers Handbook: Forming*. SME, May 1984. 25

[12] B. S. Corporation and B. S. Company, *Modern steels and their properties: carbon and alloy steel bars*. Bethlehem Steel Corporation, 1966. 29

[13] J. Holloman and L. Jaffe, "Time-Temperature relations in tempering steels," *Trans. AIME*, vol. 162, pp. 223–249, 1945. 28

[14] E. C. Bain and H. W. Paxton, *Alloying elements in steel*. American Society for Metals, 1961. 30

[15] G. Krauss, *Principles of Heat Treatment of Steel*, subsequent ed. American Society for Metals, Nov. 1980. 35

[16] R. Barron and R. Thompson, "Effect of cryogenic treatment on corrosion resistance," *Advances in Cryogenic Engineering: Materials.*, vol. 36, p. 13751379, 1989. 37

[17] K. G. Budinski, *Surface Engineering for Wear Resistance*. Prentice Hall, Sep. 1988. 40

[18] I. F. o. H. T. Engineering and Surface, *Heat treatment and surface engineering: new technology and practical applications : proceedings of the 6th International Congress on Heat Treatment of Materials, held in conjunction with the 1988 World Materials Congress, McCormick Place, Chicago, Illinois, 28-30 September 1988*. ASM International, 1988. 46

[19] G. Krauss, A. I. H. T. Division, C. S. o. M. A. S. P. Center, P. Research, and A. W. u. W. (Germany), *Carburizing: processing and performance : proceedings of an international conference, 12-14 July 1989, Lakewood, Colorado, USA*. ASM International, 1989. 49

[20] J. Crank, *The mathematics of diffusion*. Clarendon Press, 1979. 49

[21] S. Lupi, *Appunti di Elettrotermia*, Padova, 2005. 51, 52

[22] G. E. Totten, *Steel heat treatment: equipment and process design*. CRC Press, Sep. 2006. 56, 57, 59, 61, 62, 63, 64, 65, 66, 68

[23] W. Graupner and H. Conrad, "Anwendung des Dual-Thyristor-Prinzips in HF-Speisequellen für die induktionserwärmung," *Elektrowärme international. Edition B, Industrielle Elektrowärme*, vol. 49, no. 3, pp. B118–B123. 58

[24] V. Esteve, J. Jordan, E. J. Dede, C. Cases, J. M. Magraner, E. Sanchis-Kilders, and E. Maset, "Using pulse density modulation to improve the efficiency of IGBT inverters in induction heating applications," in *IEEE Power Electronics Specialists Conference, 2007. PESC 2007*. IEEE, Jun. 2007, pp. 1370–1373. 59

[25] R. Cook, D. Loveless, and V. Rudnev, "Load matching in modern induction heat treatment," *Industrial heating*, Sep. 1995. 65

[26] ——, "Power supplies for induction heat treating," *Industrial heating*, Jun. 1995. 65

[27] J. R. Davis, *Gear materials, properties, and manufacture*. ASM International, Sep. 2005. 71, 76, 78, 82, 84

[28] P. R. N. Childs, *Mechanical design.* Butterworth-Heinemann, Dec. 2003. 71

[29] Wolff and Mscher, *Einfluss der Legierungselemente auf den Stahlpreis.* [Volkswagenwerk AG], Forschung und Entwicklung, 1979. 79

[30] N. P. Chironis, *Gear design and application.* McGraw-Hill, 1967. 81

[31] F. J. Otto and D. H. Herring, "Gear heat treatment, part i," *Heat Treating Progress*, vol. 2, p. 55, Jun. 2002. 83

[32] A. S. M. International, *ASM Handbook Volume 16: Machining*, 10th ed. ASM International, Jan. 1989. 91

[33] T. Krantz, M. Alanou, H. Evans, and R. Snidle, "Surface fatigue lives of case-carburized gears with an improved surface finish," *Journal of tribology*, vol. 123, p. 709, 2001. 92

[34] T. Bausch, *Innovative Zahnradfertigung: Verfahren, Maschinen und Werkzeuge zur kostengünstigen Herstellung von Stirnrädern mit hoher Qualität*, 4th ed. Expert-Verlag, Dec. 2010. 93

[35] A. Candeo, C. Ducassy, P. Bocher, and F. Dughiero, "Multiphysics modeling of induction hardening of ring gears for the aerospace industry," *IEEE Transactions on Magnetics*, vol. 47, no. 5, pp. 918–921, May 2011. 94, 144

[36] V. Rudnev, D. Loveless, R. Cook, and M. Black, "Induction hardening of gears: a review," *Heat Treatment of Metals*, vol. 30, no. 4, p. 97103, 2003. 95, 98, 99

[37] V. Rudnev, *Handbook of induction heating.* CRC Press, 2003. 105

[38] A. S. M. International, *Asm Handbook Volume 22B: Metals Process Simulation*, 10th ed. ASM International, Sep. 2010. 107, 108, 109

[39] K. W. Andrews, "Empirical formulae for calculation of some transformation temperatures," *Iron and Steel Institute Journal*, vol. 203, no. Part 7, pp. 721–727, 1965. 110

[40] A. S. f. Metals, *Source Book on Heat Treating: Materials and processes.* The Society, 1975. 112

[41] J. Hodge and M. Orehoski, "Relationship between hardenability and percentage of martensite in some low alloy steels'," *Trans. AIME*, vol. 167, p. 62742, 1946. 113

[42] U. S. S. C. R. Laboratory, *Atlas of Isothermal Transformation Diagrams.* United States Steel, 1951. 114

[43] G. F. V. Voort, *Atlas of time-temperature diagrams for irons and steels.* ASM International, 1991. 114

[44] Z. Zhang and R. A. Farrar, *An atlas of continuous cooling transformation (CCT) diagrams applicable to low carbon low alloy weld metals.* Institute of Materials, Nov. 1995. 115

[45] M. Atkins, *Atlas of continuous cooling transformation diagrams for engineering steels.* American Society for Metals, Jun. 1980. 115

[46] W. W. Cias and C. M. Company, *Austenite transformation kinetics of ferrous alloys.* Climax Molybdenum Co., 1978. 115

[47] M. Avrami, "Kinetics of phase change. i general theory," *The Journal of Chemical Physics*, vol. 7, no. 12, pp. 1103–1112, Dec. 1939. 115

[48] ——, "Kinetics of phase change. II TransformationTime relations for random distribution of nuclei," *The Journal of Chemical Physics*, vol. 8, no. 2, pp. 212–224, Feb. 1940. 115

[49] W. A. Johnson and R. F. Mehl, "Reaction kinetics in processes of nucleation and growth," *Trans AIME*, vol. 135, no. 8, p. 416, 1939. 115

[50] E. Scheil, *Anlaufzeit der Austenitumwandlung: (Mitteilg aus d. Kaiser-Wilhelm-Institut f. Metallforschg, Stuttgart).* Verl. Stahleisen, 1935. 115

[51] D. P. Koistinen and R. E. Marburger, "A general equation prescribing the extent of the austenite-martensite transformation in pure iron-carbon alloys and plain carbon steels," *Acta Metallurgica*, vol. 7, no. 1, pp. 59–60, 1959. 115

[52] S. Kang and Y. Im, "Three-dimensional thermo-elasticplastic finite element modeling of quenching process of plain-carbon steel in couple with phase transformation," *International Journal of Mechanical Sciences*, vol. 49, no. 4, pp. 423–439, Apr. 2007. 116

[53] J. Walker, "Boiling and the leidenfrost effect," 2000. 119

[54] S. MacKenzie, "Quenching: mastering the process," 2011. 120, 121

[55] C. E. Bates, "Predicting properties and minimizing residual stress in quenched steel parts," *Journal of Heat Treating*, vol. 6, no. 1, pp. 27–45, Mar. 1988. 125

[56] F. R. Palmer, *Tool Steel Simplified, 4th Edition*, 4th ed. Chilton Book Co, Dec. 1978. 126, 127, 128

[57] G. Parrish and G. S. Harper, *Production Gas Carburizing*, 1st ed. Pergamon Pr, Jan. 1985. 129

[58] M. B. Bever, *Encyclopedia of materials science and engineering.* Pergamon Press, 1986. 129

[59] G. E. Dieter and L. C. Schmidt, *Engineering design.* McGraw-Hill Higher Education, Mar. 2008. 129

[60] S. Heinilä, T. Björk, G. Marquis, M. Bäckström, and R. Ilvonen, "Fatigue crack paths and residual stresses in cold formed rectangular structural tubes," 2007. 130

[61] J. O. Almen and P. H. Black, *Residual stresses and fatigue in metals.* McGraw-Hill, 1963. 130

[62] J. R. Davis, *Metals Handbook Desk Edition 2nd Edition*, 2nd ed. CRC Press, Dec. 1998. 130

[63] H. Thielsch, *Defects and Failures in Pressure Vessels and Piping.* Amer Society for Nondestructive, Jun. 1975. 130

[64] H. K. D. H. Bhadeshia, H. Bhadeshia, and R. W. K. Honeycombe, *Steels: microstructure and properties.* Butterworth-Heinemann, Aug. 2006. 131

[65] R. Abbaschian, L. Abbaschian, and R. E. Reed-Hill, *Physical Metallurgy Principles.* Cengage Learning, Nov. 2008. 132

[66] R. F. Kern and A. S. f. Metals, *Selecting steels and designing parts for heat treatment.* American Society for Metals, 1969. 132

[67] H. P. Kirchner, *Strengthening of ceramics: treatments, tests, and design applications.* M. Dekker, 1979. 132

[68] L. J. V. Walle, A. S. f. M. H. Activity, and O. highway Vehicles, *Residual stress for designers and metallurgists: proceedings of a conference held, 9-10 April 1980, Chicago, Illinois.* The Society, 1981. 133

[69] D. Louër, T. Bataille, T. Roisnel, and J. Rodriguez-Carvajal, "A study of nanocrystalline yttrium oxide from diffraction-line-profile analysis: comparison of methods and crystallite growth," *Powder Diffraction*, vol. 17, no. 4, p. 262, 2002. 135

[70] D. S. Mountain and G. P. Cooper, "TERSAA new technique for assessing residual stress," *Strain*, vol. 25, no. 1, pp. 15–19, Feb. 1989. 137

[71] W. B. J. Zimmerman, *Process Modelling and Simulation with Finite Element Methods*, first edition ed. World Scientific Publishing Company, Aug. 2004. 139, 142, 143

[72] S. S. Rao, *The Finite Element Method in Engineering.* Elsevier, Oct. 2010. 142

[73] K.-i. Mori, *Simulation of materials processing: theory, methods and applications : proceedings of the 7th International Conference on Numerical Methods in Industrial Forming Processes–NUMIFORM 2001, Toyohashi, Japan, 18-20 June 2001.* Taylor & Francis, Jan. 2001. 143

[74] N. Barka, P. Bocher, J. Brousseau, M. Galopin, and S. Sundararajan, "Modeling and sensitivity study of the induction hardening process," *Advanced Materials Research*, vol. 15-17, pp. 525–530, 2007. 144

[75] S. Bonetto, D. Dolega, and M. Forzan, "Metallurgical model for induction pulse hardening of gear wheels," 2007. 144

[76] S. Lupi, F. Dughiero, and M. Forzan, "Modelling single- and double-frequency induction hardening of Gear-Wheels," 2006, p. 473. 144

[77] B. Paya, V. Fireteanu, A. Spahiu, and C. Gurin, "3D magneto-thermal computations of electromagnetic induction phenomena," *COMPEL: Int J for Computation and Maths. in Electrical and Electronic Eng.*, vol. 22, no. 3, p. 744755, 2003. 144

[78] K. Kurek, M. Niklewicz, and A. Smalcerz, "Estimation of chosen parameters influence on induction hardening process of gears," 2008. 144

[79] A. Bossavit and J. Verite, "The TRIFOU code: Solving the 3-D eddy-currents problem by using h as state variable," *IEEE Transactions on Magnetics*, vol. 19, no. 6, pp. 2465–2470, Nov. 1983. 144

[80] O. C. Zienkiewicz and R. L. Taylor, *The finite element method: Basic formulation and linear problems.* McGraw-Hill, Feb. 1989. 144

[81] O. Biro and K. Preis, "Finite element analysis of 3-D eddy currents," *Magnetics, IEEE Transactions on*, vol. 26, no. 2, p. 418423, 1990. 144

[82] G. Meunier, *The finite element method for electromagnetic modeling.* John Wiley & Sons, Nov. 2008. 145

[83] G. Blut, "Effet de la gomtrie sur les paramtres du gnrateur et le profil de duret lors d'une chauffe par induction : exprience et simulation," Ph.D. dissertation, Ecole de technologie superieure, Montreal, QC, Canada, 2010. 148

[84] C. Ducassy, "Prdiction de la duret et de la profondeur de la zone de sur-revenu lors d'une chauffe rapide par induction d'un acier 4340 tremp revenu," Ph.D. dissertation, Ecole de technologie superieure, Montreal, QC, Canada, 2010. 148, 154

[85] D. Colling, "Soft magnetic structural alloys for elevated-temperature applications," *IEEE Transactions on Magnetics*, vol. 7, no. 1, pp. 91– 102, Mar. 1971. 152

[86] Y. Meyzaud and C. Sauzay, "Traitement thermique et austnite rsiduelle," pp. 61–74, 1994. 154

[87] B. Larregain, N. Vanderesse, F. Bridier, P. Bocher, and P. Arkinson, "Method for accurate surface temperature measurements during fast induction heating," Oct. 2011, pp. 60–67. 171