INTRODUCTION

Most of the aerospace and automobile components work under severe dynamic conditions. They require good service properties in order to provide efficient and long performance under service [1]. Their performance is primarily affected by failure mechanisms caused by service loads. Fatigue is one of the major reasons for failure of the engineering components. It is of great concern in case of components particularly where safety is paramount. It is also the failure mode in components such as moulds/dies, gears, bearings and shafts, and therefore has a detrimental effect on life cycle/operating costs [2]. Fatigue is failure under a repeated or varying load, never reaching a high enough level to cause failure in a single application. Fatigue failures occur due to the application of fluctuating stresses that are much lower than the stress required to cause failure during a single application of stress. It has been estimated that fatigue contributes to approximately 90% of all mechanical service failures [3]. Fatigue is a problem that can affect any part or component that moves. Automobiles on roads, aircraft wings and fuselages, ships at sea, nuclear reactors, jet engines, and land-based turbines are all subject to fatigue failures. Fatigue was initially recognized as a problem in the early 1800s when investigators in Europe observed that bridge and railroad components were cracking when subjected to repeated loading. As the century progressed and the use of metals expanded with the increasing use of machines, more and more failures of components subjected to repeated loads were recorded.

Fatigue failures are caused by loads varying with time. Fatigue life of a component is generally measured as number of stress reversals before failure and depends on the magnitude of stress. Typical stress versus life behaviour of a component subjected
to fatigue is given by an S-N curve as shown in Figure 1.1. The "S-N" means stress versus cycles to failure, which when plotted uses the stress amplitude, $\sigma_a$ plotted on the vertical axis and the logarithm of the number of cycles to failure on X-axis [3]. As can be observed in this plot, lower the magnitude of stress higher is the fatigue life.

![Typical S-N curve for a part under fatigue loading](image)

**Figure 1.1:** Typical S-N curve for a part under fatigue loading [3]

There are many factors affecting the relationship between stress and the life, like shape and size, surface finish of component. The fatigue failure depends solely on the ultimate strength or yield stress of the material. Unfortunately, in reality, many other factors may shorten the service life of a component. The following are some of these factors:

- **Cyclic stress state:** Depending on the complexity of the geometry and the loading, one or more properties of the stress state need to be considered, such as stress amplitude, mean stress, shear stress, and load sequence.
- **Geometry:** Notches and variation in cross section throughout a part lead to stress concentrations where fatigue cracks initiate.
- **Surface quality:** Surface roughness causes microscopic stress concentrations that lower the fatigue strength.
Material type: Fatigue life as well as the behavior during cyclic loading varies widely for different materials: e.g. composites and polymers differ markedly from metals.

Residual stresses: Welding, cutting, casting, and other manufacturing processes involving heat or deformation can produce high levels of tensile residual stress, which decreases the fatigue strength.

Size and distribution of internal defects: Casting defects such as gas porosity and shrinkage voids can significantly reduce fatigue strength.

Direction of loading: For non-isotropic materials, fatigue strength depends on the direction of the principal stress.

Grain size: For most metals, smaller grains yield longer fatigue lives, however, the presence of surface defects or scratches will have a greater influence than in a coarse grained alloy.

Environment: Environmental conditions can cause erosion, corrosion, or gas-phase embrittlement, which all affect fatigue life.

Temperature: Higher temperatures generally decrease fatigue strength.

It has long been recognized that fatigue occurs due to cracks generally initiated on free surfaces and therefore fatigue life is reliant on the surface finish [2]. It is also known that fatigue life is heavily influenced by residual stresses and microstructure of the part surface [2, 4]. Residual stress in a surface is defined as “the stress resident inside a component or structure after all applied forces have been removed”. It can be either beneficial or detrimental to the fatigue life of a component. Figure 1.2 illustrates that the surface tensile residual stress opens the crack and accelerates crack propagation whereas residual compressive stress will close the crack and slows the crack propagation thus making itself beneficial [5].
Component surfaces are usually in some state of residual stress due to machining process adopted [6]. Nature of residual stress, either compressive or tensile by the process has a direct impact on fatigue life. Residual stresses are generated due to plastic deformation caused by applied mechanical, thermal loads or phase changes during machining. Mechanical and thermal processes applied to a component during service/fabrication also alter its residual stress state. Therefore, knowledge of the residual stress state is important to determine the actual loads experienced by a component. In general, a “compressive stress” in the surface of a component is beneficial. It tends to enhance the fatigue life, by slow the crack propagation, increased resistance to environmentally assisted cracking due to stress corrosion and hydrogen induction. Tensile residual stress on the surface of the component is generally undesirable as it decreases the fatigue life by speeding crack propagation and lowered resistance to environmentally assisted cracking.

Residual stress in a workpiece is a function of its processing method and can enhance the life of a machined part, and thus economics [6]. Generally widely used machining methods such as, turning results in tensile residual stress which is not

**Figure 1.2:** Effect of tensile & residual compressive stress on the crack propagation [5]
beneficial for fatigue life. Hence some secondary processes are adopted to improve the fatigue life. A brief account of these is provided in the next section.

1. **Introduction to fatigue life enhancement and the techniques**

Advanced industrial applications require materials with special surface properties such as high corrosion and wear resistance, and hardness. Alloys possessing these properties are usually very expensive and their utilization drastically increases the cost of the parts. On the other hand, failure or degradation of engineering components due to mechanical and chemical/electrochemical interaction with the surrounding environment is most likely to initiate at the surface because the intensity of external stress and environmental attack are often highest at the surface. Therefore, several surface treatment techniques are employed to improve fatigue life of the component. The different fatigue life enhancement techniques are presented in Figure 1.3.

**Figure 1.3:** Summary of different fatigue life enhancement techniques
Fatigue life enhancement techniques are categorized into two types: thermal and mechanical surface treatment methods. Some of the thermal and mechanical surface treatment methods are discussed below.

1.1. Thermal surface treatment methods

Thermal techniques involve heating of the material to achieve a microstructure beneficial towards improving fatigue life. Some of the thermal surface treatments are furnished below.

- **Nitriding**

Nitridization, also known as nitriding, is a process which introduces nitrogen in the surface of a material. It is used in metallurgy, for example, for surface-hardening treatment of the steel surface. In nitriding, nitrogen is introduced into the surface of a steel component by heating it in a fused salt bath containing nitrogen-bearing salts (typically, sodium cyanide, NaCN) or in a gas stream containing cracked ammonia (NH$_3$). Steels suitable for nitriding contain aluminium, vanadium, tungsten or molybdenum; these form stable nitride precipitates that harden the surface to a depth of about 500 micrometers. The temperature 495 to 565°C is lower than that for carburizing, giving less distortion, and the surface does not require later heat treatment.

- **Case hardening**

Case hardening or surface hardening is the process of hardening the surface of metal, often low carbon steel, by infusing elements into the material's surface and forming a thin layer of a harder alloy. Case hardening is usually done after the part in question has been formed into its final shape, but can also be done to increase the hardening element content of bars to be used in a pattern welding or similar process. Both carbon and alloy
steels are suitable for case-hardening; typically mild steels are used, with low carbon content, usually less than 0.3%. These mild steels are not normally hardenable due to the low quantity of carbon, so the surface of the steel is chemically altered to increase the hardenability. Case hardened steel is usually formed by diffusing carbon nitrogen and/or boron into the outer layer of the steel at high temperature, and then heat treating the surface layer to the desired hardness.

- **Induction hardening**

Induction hardening is a form of heat treatment in which a metal part is heated by induction heating and then quenched. The quenched metal undergoes a martensitic transformation, increasing the hardness and brittleness of the part. Induction hardening is used to selectively harden areas of a part or assembly without affecting the properties of the part as a whole. It is a widely used process for the surface hardening of steel. The components are heated by means of an alternating magnetic field to a temperature within or above the transformation range followed by immediate quenching. This operation is most commonly used in steel alloys. Many mechanical parts, such as shafts, gears, springs, etc. are subjected to surface treatments before the delivering in order to improve wear behavior. The effectiveness of these treatments depends both on surface materials properties modification and on the introduction of residual stresses. Among these treatments, induction hardening is the one that is most widely employed to improve component durability. It determines in the work-piece a tough core with tensile residual stresses and a hard surface layer with compressive stresses, which have proved to be very effective in extending the component fatigue life and wear resistance.

Though the thermal surface treatment techniques are used to improve the fatigue life of component, hardening, plating etc. leave a brittle surface after processing.
Sometimes it can damage or weaken surface grain boundaries and also result in fatigue debits. Therefore, mechanical surface treatment methods are employed to overcome the limitations of thermal treatments. The following section discusses the different mechanical surface enhancement methods employed to improve the life of components along with their advantages and limitations.

1.2. Mechanical surface enhancement (MSE) Techniques

Mechanical surface enhancement (MSE) techniques such as shot peening, laser shock peening, low plasticity burnishing and deep cold rolling are used for improving the resistance to wear and stress corrosion, also which affect the fatigue life of highly-stressed components used in variety of engineering applications [1, 8, 9]. They induce plastic deformation in near-surface area of the components resulting in formation of residual compressive stresses, work hardening and changes in surface topography. Therefore, MSE techniques provide the components with improved surface characteristics after their manufacture. Some of the MSE techniques including deep cold rolling are deliberated below.

- Shot peening (SP)

It is a cold working process in which a stream of metal, glass or silica particles at high velocity is imparted against the surface of the metallic components in a defined and controlled manner. Each piece of shot striking the material acts as a tiny peening hammer imparting a small indentation or dimple to the surface. The principle of shot peening process is shown in Figure 1.4. In order to create the dimple, the surface layer of the material must be yielded in tension. Below the surface, the layers try to restore the surface to its original shape, thereby producing below the dimple a hemisphere of cold-worked material highly stressed in compression. These characteristics are called the “peening
effect”. It is very useful for the improvement of the surface properties. Large residual compressive stresses in the surface layer, as mentioned above, are very efficient in improving the fatigue strength of component.

![Schematic view of shot peening process](image)

**Figure 1.4**: Schematic view of shot peening process

The equipment is simple and inexpensive. However, some of the limitations of shot peening process are summarized below:

- Regarding the quality of peening process, regular monitoring of the shot state is essential, as the shape and the size spread of the shot change during application due to fragmentation and wear.

- It is usually difficult to achieve the full coverage of the surface to be treated, especially for complex geometries. (some areas are inaccessible for impact of shots)

- When using more complex surfaces it is difficult to obtain sufficient levels of residual compressive stress on the internal surfaces.

- Sometimes excessive surface damage is caused by the high intensity of balls.
- Laser shock peening (LSP)

An alternative mechanical surface enhancement technique, namely laser shock processing (also known as laser peening), can induce greater depths of residual stress into metal surfaces using high-power, Q-switched laser pulses [11]. This process drives a high-amplitude shock wave into material surface using a high energy pulsed laser. Figure 1.5 shows the principle of LSP process. Before processing, an opaque overlay (typically black paint or tape) and a transparent overlay (typically flowing water) are applied to the surface to be processed. The laser pulse passes through the transparent overlay and strikes the opaque overlay causing it to vaporize. The vapour absorbs the remaining laser light and produces a rapidly expanding plasma plume. Since the expanding plasma is confined between the part surface and the transparent overlay, a rapidly rising high-pressure shock wave propagates into the material. When the peak stress created by the shockwave is above the yield stress of the metal, metal is “cold worked” or plastically deformed at the surface.

![Figure 1.5: Principle of laser shock peening process [12]](image-url)
The plastic deformation caused by the shock wave results in residual compressive stresses in the surface of the part. Residual compressive stresses typically extend as deep as 1 to 1.5 mm below the surface [13]. These residual compressive stresses increase the resistance of materials to surface-related fatigue failure. The main disadvantage of LSP is that it requires an expensive equipment and specially protected environment for treatment. Skilled operators are needed before/during/after processing.

- **Burnishing**

Burnishing is a process by which a hard roller/ball (using sufficient pressure) is rubbed on the metal surface. This process flattens the high spots by causing plastic flow of the metal. Burnishing improves the finish and size of surfaces of revolution such as cylinders and conical surfaces. Both internal and external surfaces can be burnished using appropriate tools. High-speed steel or carbide rollers/balls are put in contact with surface through a slight indentation in material and then rolled over the surface. The schematic diagram of ball burnishing process is shown in Figure 1.6.

![Figure 1.6: Schematic diagram of ball burnishing process [14]](image-url)
The resulting plastic deformation leaves residual compressive stresses with an added benefit of smoothing out the surface finish. A mirror finish can be achieved in some cases. Multiple passes of the rollers also cold work the surface. Cold working increases the fatigue strength of the material. This can improve wear resistance of the surface [15, 16]. Burnishing can be very cost effective and, in some cases, can be used to finish a component instead of a secondary operation like grinding or honing. The biggest drawback for roller burnishing is process control. These tools rely on previous machining operations for consistency. Surface finish and residual stress can vary widely depending on the pre-finish of the surface. However, if the pre-finish parameters are well controlled, excellent results can be achieved with tools that can last for hundreds of components.

- **Low plasticity burnishing (LPB)**

Low plasticity burnishing, is a patented method of controlled burnishing. It is usually performed using a single pass of a smooth free rolling ball under a normal force sufficient to plastically deform the surface of the material. The schematic representation of LPB process is shown in Figure 1.7.

![Schematic representation of LPB process](image)  

**Figure 1.7:** Schematic representation of LPB process [13]
Hertzian loading creates a layer of residual compressive stress to a maximum depth of 1mm [17-19]. LPB is performed in a machine shop environment using conventional or CNC machine tools at speeds comparable to machining operations. The machine tool’s coolant is used to pressurize the bearing and “float the ball. The ball is loaded normal to the surface of a component with a hydraulic cylinder that is in the body of the tool. The ball rolls across the surface of a component in a linear stepping pattern. Since there is no shear being applied to the ball, it is free to roll in any direction. As the ball rolls over the component, the pressure from the ball causes plastic deformation to occur in the surface of the material just under the ball. This causes the peaks of the metallic surface to spread out permanently, when the applied burnishing pressure exceeds the yield strength of the steel to fill the valleys [20].

The surface of the metallic material will be smoothed out and because of the plastic deformation the surface becomes work hardened, the material being left with a residual stress distribution that is comprehensive in the surface [20]. No material is removed during the process, and if is only displaced inward by a few ten-thousandths of an inch. The burnishing ball is the only wear prone component of the LPB tooling. High chromium steel, beta-silicon nitride, and sintered tungsten carbide balls can be used successfully in the apparatus. The subsurface residual stress depends upon the normal force, feed and mechanical properties of both the ball and work piece [22, 23]. Lateral plastic deformation of the surface is necessary to achieve surface compression.

The major benefit of LPB is the improved high cycle fatigue life. An LPB treated surface is resistant to foreign object damage and stress corrosion cracking. In contrast, shot peening typically produces 20% to 70% cold work and much shallower compression [8]. The LPB manufacturing process is highly controllable and operates at cycle times
that greatly lower cost of production and provide processed parts in minutes, thus preventing production bottleneck constraints typical of LSP processing.

1.3. Deep Cold Rolling (DCR)

DCR is a surface treatment technique which is performed using a roller or ball type instrument to produce a surface residual compressive stress that enhances the fatigue life of engineering components [9]. It is a non-cutting production method which next to finish rolling and size rolling is counted among the fine surface rolling methods according to VDI guideline [24]. Finish rolling aims at creating surfaces with a plateau-like profile and especially low roughness, without significantly altering the geometric shape, in order to achieve good sliding and anti-frictional qualities and to reduce wear. Size rolling is used to generate precise dimensions and thus to create work pieces with an accurate fit. By contrast, the objective of actual deep cold rolling is to introduce work hardening and residual compressive stresses into near surface regions in order to increase the fatigue strength [25, 26].

![Figure 1.8: Principle of deep cold rolling process [26]](image)

Figure 1.8: Principle of deep cold rolling process [26]
DCR process can be performed with a hydrostatic ballpoint tool or mechanical tools. The load is generated by pressure of specific fluid, which is supplied from a high-pressure hydraulic system. The ball, which is hydraulically loaded, forms a seal with the tool in the idle condition. The principle of DCR process is as shown in Figure 1.8. During the process, the ball is pushed back against the hydrostatic pressure and a small pressure fluid escape path is opened. The constant gap between the ball and the retainer is independent of the distance between the tool and the work piece. The ball then rotates freely as the tool is moved relative to the surface being treated while a constant pressure is maintained. During the treatment, a film of oil exists on the part surface. The rolling force of mechanical tools is determined indirectly by measuring and monitoring spring deflection with mechanical dial indicators or inductive sensors. The rolling force is slowly increased at the beginning of the process and slowly decreased at the end by slowly in feeding the tool or pulling it out. This procedure prevents stress transitions.

Although the tool designs and hydraulic systems differ, the LPB tooling is similar to “deep rolling” tools using a hydrostatically supported burnishing ball. The LPB and deep rolling processes differ in the method of use and the level of cold work generated in developing the compressive layer [26]. Deep cold rolling has significant advantages over other mechanical surface enhancement techniques. e.g.:

- Higher residual stresses extending further below the surface
- Smoother and more hardened work piece surfaces
- Can be utilized on any conventional CNC machine
- Few parameters to be controlled
- Deeper penetration
- Cost effective by completion in one setting after the cutting process
Applications of DCR:

- **Automotive industry**

  There are several mostly rotation-symmetrical components in the automotive industry that are highly loaded and deep rolled such as axles, shafts and especially steering knuckles. Since weight saving is especially attractive for components of the steering wheels or propulsion system, deep rolling is efficiently used here. Applications range from cars to farm tractors. Sophisticated rolling tools enable machining of hardened components leading to further applications such as fretting fatigue loaded roller bearer steels. The combination of thermal and mechanical surface treatments significantly enhances the fatigue strength leading to a pronounced reduction of weight.

- **Aerospace and aircraft industry/power plant industry**

  In the aerospace and aircraft industry weight reduction is a major aim of component designers. The optimization of near-surface properties by shot peening has been well established here. Since deep rolling offers a variety of economic and technological advantages, a lot of applications can also be found here. Typical examples are wheel rims of military aircraft, highly loaded screws and bolts, as well as propulsion components such as turbine discs and compressor fan blades.

- **Medical applications**

  There are several examples in medical applications especially in the field of endoprothesis components that are highly loaded and must exhibit high cyclic strength. Typical examples are implants for artificial hip joints and the spinal chord. In order to guarantee durability, surface optimization is an absolute must. Moreover components of
surgical instruments that are loaded cyclically in acidic and corrosive environment can be improved by deep rolling in terms of their stress corrosion behaviour.

**Factors affecting the compressive stress and fatigue life in deep cold rolling**

Figure 1.9 summarizes the essential variables governing deep rolling process, which may be classified according to work piece, tool, method or device used.

![Figure 1.9: Parameters influencing the results of deep rolling process [24]](image)

The study of deep cold rolling process could be undertaken either analytically or experimentally to assess the effects of these variables and optimize them for better responses. With modern analytical and computational techniques it is often possible to estimate the effects of various process parameters of deep rolling on the residual compressive stress to which the component is subjected in service. However, there are numerous parameters to be considered. Importantly, following the discussion above, the effects of each of these parameters dependent on other variables. This interaction between
variables would make the normal approach of a parametric study followed by simple
inspection of the results prohibitively complex for more than two or possibly three
variables. Therefore, the methodology of statistical experimental design has been
developed to address such problems where multiple, interacting variables with significant
experimental variation make interpretation of the data complex and difficult. So, design
of experiments can be used for planning the experiments so that the data obtained can be
analyzed to yield valid and objective conclusions [27].

1.4. Design of experiments (DOE)

DOE is a very powerful analytical method that can be used for providing a cost-
effective and organized approach to conduct industrial experiments. Multiple product
design and/or process variables can be studied at the same time with these efficient
designs, instead of in a hit-and-miss approach, providing very reproducible results [28].
Due to the statistical balance of the designs, thousands of potential combinations of
numerous variables (at different settings or levels) can be evaluated for the best overall
combination, in a very small number of experiments.

An important ethos of the experimental design philosophy is that all
experimentation should not be carried out in one stage. A cyclic approach is taken
whereby an initial experiment is planned, executed and analyzed and this data is used to
plan the next phase of experimentation and so on [28]. The initial or ‘screening’ study
serves to identify the important variables and, most importantly, also the interactions
between them. This gives an overview of the behaviour of the system or process and
allows the efficient planning of a next stage of experimentation with the aim of
investigating the effects of the pertinent variables in more detail.
In a characterization experiment, we are usually interested in determining which process variables affect the response. A logical next step is to optimize, that is to determine the region in the important factors that leads to the best possible response. Usually, the characteristics of a product quality are related to the various process parameters and noise factors through a complicated, nonlinear function. It is possible to find many combinations of process parameter values that can give the desired target value of the product’s quality characteristic under nominal noise conditions. However, due to the nonlinearity, these different process parameter combinations can give quite different variations in the quality characteristic even when the noise factor variations are the same. The principal goal of the robust design is to exploit the nonlinearity to find a combination of process parameter values that gives the smallest variation in the value of the quality characteristics around the desired target value [29].

![Figure 1.10: Block diagram of a process: P Diagram](image)

The block diagram representation of a process is shown in Figure 1.10. The response or output of the process is denoted Y which represent the quality characteristic considered for the purpose of optimization. A number of parameters can influence the
quality characteristic or response of a process. These parameters or factors can be classified into the following three classes:

- Signal factors (M): These are the parameters set by the user or operator to express the intended value for the response of the process. The signal factors are selected by the design engineer based on the engineering knowledge of the process being developed. Sometimes two or more signal factors are used in combinations to express the desired response.

- Noise factors (X): Certain parameters cannot be controlled by the designer and are called noise factors. Parameters whose settings are difficult to control in the field or whose levels are expensive to control are also considered noise factors. The levels of the noise factors change from one unit to another, from one environment to another, and from time to time. Therefore, noise factors cause the response to deviate from the target due to variation of process parameters from unit to unit and product deterioration.

- Control factors (Z): These are parameters that can be specified freely by the designer. In fact, it is the designer’s responsibility to determine the best values of these parameters. Each control factor can take multiple values called levels. When the levels of certain control factors are changed, the manufacturing cost does not change; however, when the levels of others are changed, the manufacturing cost also changes.

To use the statistical approach in designing and analyzing an experiment, it is necessary to have a clear idea in advance of exactly what is to be studied, how the data are to be collected, and how these data are to be analyzed. The different steps involved in designing an experiment are given in Figure 1.11.
Recognition and statement of the problem: A clear statement of the problem often contributes substantially to a better understanding of the phenomena and final solution of the problem.

Choice of factors, levels and ranges: The experimenter must choose the factors to be varied in the experiment, the ranges over which these factors will be varied and the specific levels at runs will be made. Selection of factors and their levels require the process knowledge.

Selection of the response variable: It includes the variable to be measured which give useful information about the process.

Choice of experimental design: It involves the consideration of sample size, the selection of suitable run order for the experimental trials, and determination of whether or not blocking or other randomization restrictions are involved.

Performing the experiment: When running the experiment, it is vital to monitor the process carefully to ensure that everything is being done according to the plan.

Statistical analysis of the data: Statistical methods are used to analyze the data so that the results and conclusions are objective. Residual analysis and model adequacy checking are important analysis techniques.

Conclusions and recommendations: Once the data have been analyzed, the experimenter must draw practical conclusions about the result and recommended course of action.

Confirmation Test: Before presenting the results to the others and taking a practical course of action the experimenter needs to carry out confirmation tests to evaluate the conclusions.
Figure 1.11: Procedure and steps involved in designing an experiment [28]

- **Design of Experiments Terminology**

  In this case a specific form of experimental design will be used known as factorial experimental design. Here the experiment consists of a number of experimental runs where the experimental parameters or factors are varied between each run and the experimental response is measured in each case. The experimental design consists of a plan of the specific values, or levels of the factors to be used for each run. Factor levels may be quantitative or qualitative. The interpretation of the latter type often requires extra care. The objective of the experiment is to determine how changes in the values or levels of each factor change the response, and these effects are called factor effects. An Interaction effect quantifies any differences between the effects of one factor at different values of another factor.
The simplest factorial design, but also the one that requires the most experimental resources, is one in which all combinations of the factors levels are used and is termed a ‘Full Factorial’. ‘Fractional Factorial’ designs, in which a carefully chosen, balanced subset of the possible runs is selected, can substantially reduce the number of experimental runs required to obtain the same amount of information [30].

- **Fractional Factorial Design**

It is a design in which experimenters perform only a selected subset or "fraction" of the runs in the full factorial design. Fractional factorial designs are a good choice when resources are limited or the number of factors in the design is large because they use fewer runs than the full factorial designs [31]. The number of runs necessary for a 2-level full factorial design is 2^k where k is the number of factors. As the number of factors in a 2^k design increases, the numbers of runs necessary to perform a full factorial design increases rapidly. For example, a 2-level full factorial design with 7 factors requires 128 runs. A 1/16 fraction, fractional factorial design would require only 8 of those runs. In fractional factorial designs, some of the effects are confounded and cannot be separated from other effects [28]. Usually we are not concerned with any terms higher than the 2-way interaction, and we can assume the effects of higher-order interactions are negligible, so there is no need to estimate them.

A fractional factorial design uses a subset of a full factorial design to obtain information about main effects and low-order interactions with fewer runs. The diagrams below show (Figure 1.12) a full factorial design versus a ½ fraction factorial design. The full factorial design contains twice as many design points as the ½ fraction design. The response is only measured at four of the possible eight corner points of the factorial design.
portion of the design. However, with this design the main effects will be confounded with the 2-way interactions.

![Figure 1.12: Representation of a (a) Full factorial design (b) ½ Fractional factorial design](image)

- **Central composite design (CCD)**

  A second-order model can be constructed efficiently with central composite designs. CCD is first-order (2N) designs augmented by additional centre and axial points to allow estimation of the tuning parameters of a second-order model. There are two parameters in the design that must be specified: the distance of the axial runs from the design center and the number of center points. It is important for the second order model to provide good predictions throughout the region of interest [28, 32]. One way to define “good” is to ensure that the model has a reasonably consistent and stable variance of the predicted response at points of interest. Box and Hunter suggested that a second order response surface design should be rotatable. Rotatability is a reasonable basis for the selection of a response surface design. Because the purpose of RSM is optimization and the location of the optimum is unknown prior to running the experiment, it makes sense to use a design that provides equal precision of estimation in all directions.

  Rotatability is a spherical property; that is, it makes the most sense as a design criterion when the region of interest is a sphere. However, it is important to have exact
rotatability to have a good design. In many situations, the region of interest is cuboidal rather than spherical. In these cases, a useful variation of the central composite design is the face centered central composite design. This design locates the axial or star points on the centers of the faces of the cube as shown in Figure 1.13 for 3 factors.

![Figure 1.13: Representation of a face centered central composite design [28]](image)

- **Response surface methodology (RSM)**

  Response surface methodology is a collection of mathematical and statistical techniques for empirical model building. By careful design of experiments, the objective is to optimize a response (output variable) which is influenced by several independent variables (input variables). An experiment is a series of tests, called runs, in which changes are made in the input variables in order to identify the reasons for changes in the output response. Originally, RSM was developed to model experimental responses, and then migrated into the modelling of numerical experiments. The application of RSM to design optimization is aimed at reducing the cost of expensive analysis methods (e.g. finite element method or computational fluid dynamic analysis) and their associated numerical noise.
• **Response surface plots for optimizing individual response**

Response surface plots such as contour and surface plots are useful for establishing desirable response values and operating conditions. In a contour plot, the response surface is viewed as a two-dimensional plane where all points that have the same response are connected to produce contour lines of constant responses. A surface plot generally displays a three-dimensional view that may provide a clearer picture of the response. If the regression model (i.e., first-order model) contains only the main effects and no interaction effect, the fitted response surface will be a plane (i.e., contour lines will be straight). If the model contains interaction effects, the contour lines will be curved and not straight. Both contour and surface plots help experimenters to understand the nature of the relationship between the two factors and the response. In the present work an attempt is made for multi-response optimization of the performance parameters based on the process parameters during deep cold rolling.

• **Multi response optimization**

Performance characteristics of a process are often characterized by a group of responses which are measurements of one or several characteristics of quality. These responses are generally correlated and can be expressed in different measuring units. In the design stage, the problem is to find the optimal levels of the parameters by combining the different responses. Several criteria were developed by different authors in order to define a global optimization function of the responses (for example: loss function, desirability function, performance index of production).

In a general way, the optimization criteria are based on the following:

- the distance to the target value (to minimize the deviations compared to the target values)
- the variance of the responses (to minimize variability or to maximize the robustness with the noise).

- sensitivity of the response to the weak variations (to maximize the robustness with the fluctuations of the parameters of control).

Development of logical methodologies to optimize the deep cold rolling parameters requires a fundamental knowledge of the prevailing deep cold rolling mechanisms and their influence on the resulting surface properties. Process modeling and optimization are two important concerns in deep cold rolling. The processes are characterized by a multiplicity of dynamically interacting process variables. Residual compressive stress, fatigue life, surface hardness and surface finish are considered to be the important factors in predicting performance of deep cold rolling operation. Hence in the present work, multi objective optimization by desirability function approach is applied.

- **Desirability function approach (DFA)**

The desirability function approach is based on the idea that when one of the quality characteristics of an industrial process or product with many characteristics are not in the desired limits, the overall quality of the industrial process or the product is not desirable. By this approach the process (and/or product) variables which yield the most desirable responses are found. Desirability function is a mathematical method to find the optimum values of input parameters and performance parameters (response) concurrently by using the optimum input parameters levels. It is an effective tool for multi response optimization. The desirability function approach to simultaneously optimize the multiple equations was originally proposed by Harrington (1965) and later improved by Derringer and Suich (1980).
A process performance is a function of responses associated with multiple quality characteristics. Each response has its own effect on the performance where one response may have a stronger influence on process performance than another. Also each response may be measured in different units. Hence, it becomes difficult to combine all the different types of responses into a single entity that would indicate the level of process performance. The desirability function achieves this task by transforming the responses into dimensionless variables called as desirability index $d_i$. The range of a desirability index falls in the closed interval $[0, 1]$. A higher value of the desirability index for a response implies a higher contribution to the product performance by the particular response. The overall assessment of product performance is accomplished by multiplying all desirability indices to yield an aggregate desirability index $D$.

The optimization process is performed using desirability function, the option available in MINITAB 15 software. A prediction profile for the performance parameters consists of a series of graphs, one for each process variable, of the performance parameters at different levels of one process variable, holding the levels of the other process variables constant at specified values, called target values. If appropriate target values for the independent variables have been selected inspecting the prediction profile, it is possible to explain which levels of the process variables produce the most desirable predicted response on the performance variable. In the current study, the objective is to minimize the surface finish, and maximize the fatigue life and surface hardness. It is necessary to specify the lower limit and target value of the response for the maximization problem and target value and minimum value for the minimization problem.

### 1.5. Finite element method (FEM)

Different factors of surface integrity in machining, such as microstructure and residual stresses, can be affected by various machining conditions. In addition to
experimental studies, predictive models are greatly needed to further understand the mechanisms that drive the surface integrity changes and to find the optimum machining conditions that would lead to desirable surface integrity. Residual stress is one of the most commonly studied surface integrity factors in predictive models. Due to the improved performance and power of computers, finite element modeling is becoming one of the widely used approaches to predict residual stresses induced by machining. Major advantages of FEM include the application of complex material behavior models which allow flow stress to change with strain, strain-rate and temperature. Also, user subroutines can be easily implemented to capture complex material changes during the machining process, such as phase transformations, dynamic recrystallization, etc., that influence the surface integrity. Obtaining appropriate experimental data as input to calibrate the model is still an essential part of this implementation.

Thus, a brief introduction to fatigue life improvement by different thermal and mechanical surface enhancement techniques including deep cold rolling are presented in this chapter. From the literature presented in this chapter it could be seen that deep cold rolling process requires expensive tools and set-up. Most of the studies on DCR process are limited to materials like aluminum and titanium alloys. Further, the number of parameters to be controlled during the process is also more. The following chapter focusses on the available literature in the field of mechanical surface enhancement techniques for the improvement in fatigue life, particularly DCR process which is deemed to be beneficial in terms of improvement in fatigue life, surface hardness and finish.