Overview of the Mechanisms of Failure in Heat Treated Steel Components


“Primum non nocere” — “First do no harm,” attributed to the ancient Roman physician Galen. “Declare the past, diagnose the present, foretell the future; practice these acts . . . make a habit of two things — to help, or at least to do no harm” (Ref 1).

 Failures in steel components, like any other material, may have various consequences, such as:

- Making the device or component completely inoperable
- Preventing an operable device from functioning satisfactorily
- Making the device or component unsafe or unreliable, with immediate removal from service required

Many aspects may also be involved in tracing back to the possible sources of failure of a component. Some of these sources include:

- Design
- Material issues, such as improper materials selection or material imperfections (laps, seams, inclusions, porosity, etc.)
- Fabrication and processing
- Rework
- Assembly
- Inspection
- Storage and shipment
- Service conditions
- Maintenance
- Unanticipated service conditions

Many times, more than one factor contributes to a part failure. Rarely is it only one factor.

General Sources of Failure

Design deficiencies are a common source of component failure. Examples include the presence of a sharp notch in regions of high stress or a fillet radii that is too sharp. Using a component design for a new application can also lead to unanticipated failures. Higher stresses or unanticipated service conditions can cause unforeseen failure because of complex or increased stress fields. Stress concentrations may become more critical because of the increase in loading for the new application.

Insufficient design criteria can also be the cause of unforeseen failures. Inadequate knowledge of the stress state in the component or inadequate stress calculations can contribute to failures. Much higher stress states than initially assumed or improper stress assumptions can result in premature service failures. Lack of consideration of severe environmental, fatigue, or impact conditions may contribute to failure.

Material issues can usually be attributed to either selection of material or material imperfections rendering it unsuitable for service. Inadequate material data can also result in conditions that may contribute to failure. For example, adequate fatigue data, elevated-temperature tensile data, or creep or corrosion data may not be available, and the designer may have to extrapolate or estimate the effects or these properties.

Other sources of failure can be attributed to material imperfections. For wrought products, this could be related to segregation, inclusions, porosity, laps, and seams. For castings, these imperfections could be cold shuts, inclusions, shrinkage, voids, and porosity. Forgings can have laps, seams, segregation, and anisotropy in properties from forging flow lines.

In one example (Fig. 1), a large roll was heat treated, and several large cracks were observed after inspection. This was originally attributed to quench cracking. On further examination, it was determined that a lap was present in the forging,
indicated by the presence of high-temperature oxides in the crack along the crack faces.

**Manufacture and Processing.** Processing can have a large influence on properties and the resulting residual stresses. Typically, this is related to wrong procedures or improperly specified procedures. Ambiguous processes or specifications can also contribute to failures due to interpretation or application. Simple things like improper selection of processing sequences or procedures or specifications that were not followed can also contribute to failure.

Cold forming, such as stretching or deep drawing, can develop highly localized residual stresses. Local changes in microstructure can occur. Because of the changes in reduction, a large anisotropy in material properties also results. Due to the drawing operation, cracks or microcracking can occur. This could be due to improper lubrication or improper die design. The localized changes in ductility can also contribute to failure.

Machining and grinding can create high residual stresses from either machining practice (feeds and speeds) or improper cutting tool selection, material, or geometry. Grinding, if abusive, can cause large temperature gradients and localized overheating. This overheating can cause changes in microstructure—either localized softening of the material or localized transformation to martensite and other transformation products—resulting in hard spots.

In Fig. 2, a large gear was ground after heat treatment. Because of abusive grinding, local temperatures exceeded the austenitization temperature, and transformation to martensite occurred upon cooling. This transformation and the resulting residual stresses caused cracking of the gear. Temper etch examination of the gear using dilute nitric acid in water in the regions of cracking showed evidence of localized abusive grinding.

Identification of parts can also cause failure to initiate. This is from localized impact or electroetching. Localized mechanical stress concentrations or changes in microstructure can occur. This creates either a mechanical or microstructural notch or stress concentration.

Heat treatment can cause a variety of different root causes for failures. Overheating, decarburization, quenching, tempering, annealing, and other heat treatments can cause failure to occur.

---

**Fig. 1** A large roll was found to have cracks on the outer and inner surfaces of the forging. These cracks were found during final inspection. During examination of metallographic sections taken from the roll, high-temperature oxides were found on the crack faces, strongly suggesting forging laps.
This could also include improper austenitization temperatures and times. Decarburization is the result of a low-carbon surface from improper atmosphere control. Typically, there is a depleted carbon layer at the surface that, when quenched, is softer than the core material. This soft layer can be completely devoid of carbon (complete decarburization) or only partially depleted in carbon (partial decarburization). This decarburized layer can contribute to premature fatigue failures, because the surface material is different than the designer expected, or failure can result from high residual stresses created at the surface from the quenching operation. The low-carbon surface area can also result in distortion—again, high residual tensile stresses at the surface with low surface hardness. Carburization is similar to the effects of decarburization. In this case, there is a higher surface carbon than expected. High residual tensile stresses can result as well as increased distortion.

Quenching can also contribute to high residual stresses or the formation of cracks or microcracking. Transformation stresses from quenching cause the high residual stresses. These high residual tensile stresses can drastically reduce the fatigue strength or have other ramifications in service.

Overheating can cause excessive grain growth, with resulting increases in hardenability and increased embrittlement. Underheating can cause poor mechanical properties, because there was an incomplete transformation to austenite and therefore an incomplete transformation to martensite. Poor mechanical properties, such as low tensile and yield stress, and poor impact properties may occur.

There are also several embrittlement mechanisms caused by the use of improper tempering temperatures. Temper embrittlement and blue brittleness are just two of the common mechanisms that can occur from improper heat treatment and tempering operations.

Cleaning, pickling, and electroplating operations can also cause potential failures or contribute to them. Hydrogen charging of high-strength steels from the dissociation of hydrogen on the surface of high-strength steel can occur from cleaning operations in acids. Charging of hydrogen from high current densities in electroplating can cause hydrogen embrittlement unless proper baking procedures are used to allow the hydrogen to diffuse out. Electroplating can also cause high residual tensile stresses, which can contribute to crack initiation.

Welding can cause many different problems. These problems can be cracks that are initiated from improper welding procedures, high residual stresses, porosity from inadequately dried weld rods, or dirty workpieces. Microstructural notches or stress concentrations from the heat-affected zone and the transition to the base material can be the result of improper preheat and postheat. Improper weld penetration, weld geometry, and excessive weld current (undercutting) can also cause mechanical stress concentrations (Fig. 3).

The mast arm failure shown in Fig. 3 (Ref 2) was the result of weld bead undercutting and poor weldment design. Fatigue cracking initiated at the site of the weld toe undercut. This location was a highly stressed area and the location of a large mechanical stress-concentration factor because of the weld toe undercut. Typical
causes of undercutting include excessive weld current.

The assembly of a group of components can also cause eventual failure. Force-fitting a component creates high residual stresses or damage and causes premature failure to occur. Incorrect placement of a component or incorrect assembly order can also cause high residual stresses or failure to occur. Improper specifications or torque requirements can also cause premature failure. Misalignment of components within the assembly could also result in inadequate service life, because the stresses are not what the designer had anticipated.

Service conditions obviously can have a large role in the failure of a component. The service conditions could be normal operations but unanticipated by the designer. It could also be abnormal operations, such as speed, temperature (high or low), or a chemical environment, that were also unanticipated. The lack of proper scheduled maintenance can be a major contributor to premature failure. Maintenance procedures are often reduced as a cost-savings measure. Inadequate lubrication or improper lubrication can also play a role in failure (Fig. 4). In the case of Fig. 4 (Ref 3), the lubrication schedule was extended to reduce aircraft
downtime. This, and other contributing factors, resulted in the loss of 88 lives.

Stresses from startup can also contribute, along with rapid temperature gradients and rapid localized changes in the environment. Start-up procedures and maintenance are critical for intermittent operations. Shut-down procedures and resulting stresses are just as critical as proper startup. Inspection procedures to prevent failure are also important. Failure to properly inspect for problems or cracking can be catastrophic (Fig. 5), (Ref 4). In this case, maintenance and inspection personnel failed to detect a fatigue crack in the compressor stage of an aircraft engine. Upon application of power, the compressor stage ruptured, with shrapnel severing fuel lines and igniting the fuel, ultimately leading to the loss of the aircraft.

**General Practice Conducting a Failure Analysis**

The primary objective of any failure analysis is to determine the primary root cause of failure and to establish the appropriate corrective action. There are several stages of an analysis, which can proceed one after the other or occur at the same time. There is no set “fixed-in-stone” procedure, because it is highly dependent on the part and procedures/capabilities of the specific laboratory.

These stages of analysis are:

- Collection of background data
- Preliminary visual examination
- Nondestructive testing
- Selection and preservation of specimens
- Mechanical testing
- Macroexamination
- Microexamination
- Metallographic examination
- Determination of the fracture mechanism
- Chemical analysis (bulk and microanalysis)
- Exemplar testing
- Analysis and writing the report

These stages are described as follows, and additional information on failure analysis procedures is given in the chapter “General Aspects of Failure Analysis” in this book.

**Collection of Background Information**

During the collection of background data, the engineer is trying to gather an understanding of the purpose of the part. The engineer is attempting to discern the design criteria, service conditions, and failure conditions. In the background information, the operating details and manufacturing history should be examined and collected. This manufacturing history should include all the mechanical processing, thermal history or processing, and any chemical process performed on the part.

The service history should include all the maintenance records of the part. It should also include the expected environment and loading at the time of failure, as well as the normal environment and loading. Any quality records should be examined for discrepancies. Unfortunately, these records are not always available, and it is often up to the experience of the engineer to determine the quality of the part.

**Preliminary Visual Examination**

Documenting the failure or fracture is extremely important. There can never be too many drawings or photographs. The cost of photographs (especially digital) is cheap compared to analysis. A high-quality camera with macro-capability is very important and is one of the best tools that a failure analysis laboratory can have. The use of gray cards to ensure proper color rendition is also very important, because the color of scale or oxides can often give an indication of the temperatures that the part has experienced.

Sample selection is also very important. All associated debris should be collected and identified. Similar parts should also be collected for
comparison. In the case of a fastener failure, it is important that the nut and washer be collected, too. All mating pieces should be gathered for subsequent analysis.

Any abnormal conditions should be observed and compared with new and used components. Any discoloration or debris should be noted and collected. Any distortion of the part should be noted, along with dimensions of the part. Weather conditions at the time of failure should be collected, as well as all bearing and lubrication conditions and records.

During the initial wreckage analysis, the determination of all wreckage should be identified and located on a map or grid before any is touched or moved. Photograph each piece of wreckage and its surroundings. Inventory the parts present or missing. Determine the operating conditions at time of failure. This should include the position of control surfaces, power settings, position of throttles, and any lights or annunciations that occurred.

As best as possible during the initial examination of the wreckage, the sequence of failure should be determined. This can be accomplished by examining chevron markings and crack order. The parts should then be closely examined and reassembled. Do NOT allow the fracture surfaces to touch each other, because this can cause potential damage to the delicate surfaces. This analysis can also help determine the sequence of events leading up to failure. Preliminary examination of the part should note any paint, debris, or deposits present. Always remember to “do no harm.”

The visual examination should be detailed. Fracture surface crack directions should be noted, identified, and documented. Any abuse or discoloration should be identified, and a general assessment of the workmanship should be determined. Document all findings with photography, with multiple photographs taken from different directions. The incorporation of rulers or scales is important to determine the size and direction of fracture.

**Nondestructive Testing**

Nondestructive testing is very useful for determining the extent of cracking. Magnetic particle inspection is useful for ferrous alloys, with dye-penetrant and ultrasonic inspection as additional methods available for initial inspection.

Magnetic particle inspection uses discontinuities in the magnetic field to identify cracks or discontinuities. Fluorescent dyes with small magnetic particles are used. These magnetic particles gather at the discontinuities in the magnetic field, indicating flaws or indications. It is a common, sensitive, and reliable method that is simple to learn and use. This method has no limitation in part size but is limited to magnetic materials. No elaborate precleaning of the surfaces is necessary. Detection is limited to the surface of the part or section examined. Care must be exercised to prevent local arcing.

The dye-penetrant method is useful for examining surface flaws or cracks. It is used primarily for nonferrous alloys but is used for examining ferrous weldments for cracks and porosity. In this method, a high-wetting liquid is spread on the surface of the part. Excess liquid is wiped off. A developer is applied to the part surface. Any cracks, flaws, or other indications will appear. Limitations of this method are the necessity of cleaning the surface prior to and after application of the indication fluid and developer solution. Surface features may also mask indications. It is simple to use, but an understanding of the limitations must be understood prior to application to a part.

Eddy-current methods depend on the principle that all metals conduct electricity. An alternating current is applied, and eddy currents occur by electromagnetic induction. Cracks or other flaws cause distortions in the electromagnetic fields, with a result of changing the field impedance. The advantage of this method is that subsurface discontinuities can be detected. No special skill is required to use this method, and the method can be automated. Probe contact with the part is not needed. Limitations of this method are that the depth penetration is limited, and the part must be capable of conducting electricity. Reference standards are needed for specific flaw sizes and materials. Many things can influence readings, including segregation, carburized layers, and changes in profile.

Ultrasonic testing uses high-frequency sound waves transmitted through a conducting medium. Any discontinuous boundary can cause a deflection. This method is very sensitive and has high penetration. It is possible to get accurate measurements of flaw position and size, but reference standards must be used. Shape and size can cause errors in interpretation. Experienced operators are required to properly interpret the
results of testing. Effects of grain size, porosity, and inclusions can also hinder interpretation.

Radiography, using x-rays, neutrons, or gamma rays, is also often used to examine structures. Film or sensors (charge-coupled devices) pick up the emitting radiation, with the intensity proportional to the density of the sample. Light areas indicate a dense region, and dark areas indicate a greater exposure or less dense region. Advantages of radiography are the detection of subsurface and internal features at various depths and the documentation of these features by film or other imaging techniques. The primary disadvantage is that reference standards must be used, and the area for testing must be enclosed to prevent radiation from leaking out.

**Mechanical Testing**

Mechanical testing is useful to determine the properties of the part and to verify that it meets expected properties and specifications. There are many types of mechanical testing available, including hardness, tensile testing, and impact fracture testing.

Hardness testing is probably the most versatile and widely used. It is often used to evaluate heat treatment and can be used as an approximation for tensile strength. It can be used to detect the presence of work hardening or softening and hardening or softening from localized thermal events such as grinding. For the most part, it is a nondestructive test. For microhardness testing, it is necessary to use a metallographic specimen.

Tensile testing is used more to establish conformance to specification. It is not necessary to show adequate ductility because of service loads. Because of the size of the tensile specimen, it may not be possible to excise an appropriately sized sample from the part. Anisotropy of properties can be expected to lower measured tensile and yield strength properties.

Impact and fracture toughness testing is typically used to determine conformance to specifications. Charpy impact testing has a high variability in results and may be temperature related. Results must be taken with temperature in mind and may not correlate with real results because of size limitations. Fracture toughness testing and the results from \( K_{IC} \) testing can be used in design, and the results are useful for calculating critical flaw sizes. It can also be used to examine estimated crack growth rates; however, samples are difficult to prepare and test. These methods also do not incorporate the effects of residual stresses.

**Selection and Preservation of Specimens**

The selection and preservation of fracture surfaces is vital to prevent the destruction of evidence. Unprotected, the fracture surfaces or parts can become mechanically or chemically damaged. This damage can obliterate evidence and make the determination of fracture difficult or impossible. Both sides of the fracture must be protected. This is in the event that if one surface is damaged, the other side can be examined. Protection of the specimens during shipment is also very important, because evidence could be destroyed. Avoid touching surfaces with the hands, because the chemicals and acids present can cause artifacts or destroy data. NEVER fit surfaces together, because the delicate fracture features can be destroyed. Since both surfaces would be damaged, it could destroy the chances for determining the fracture mechanism.

Cleaning of specimens is to be done only when absolutely necessary. For the most part, it is required to prepare the sample for the scanning electron microscope (SEM). Dry air blasts or soft artist brushes are typically all that is needed. Rinsing in organic solvents then evaporating the solvent with dry air is useful for preparing specimens for the SEM.

Chemical cleaning is generally not recommended under any circumstance. Foreign substances such as scale or debris should be preserved. Do not use rust inhibitors, because of the inevitable damage to the part and fracture surfaces. These rust inhibitors are also extremely difficult to remove. Avoid washing the sample or parts with water unless seawater or other chemical is present. In this case, gently wash with distilled water and follow that with high-quality alcohol or acetone. Allow to dry and place in a dessicator.

Plastic replicas are useful in preserving fracture surfaces and removing debris for further analysis. Softening replica tape (available at transmission electron microscope supply houses) with small amounts of acetone forms plastic replicas. The softened tape is pressed gently onto the fracture surface. Additional layers of tape, softened with acetone, are applied to the fracture surface. After multiple layers have been applied, the entire replica is allowed to dry and then is placed in a dessicator. When the part is ready to be examined, the replica is
carefully removed using tweezers. Any debris on the surface is also preserved for further analysis in the replica. Multiple plastic replicas can be used to clean a surface of a part. This can be repeated as necessary.

Sectioning of Specimens

Sectioning is very important, because it captures the portion of the fracture surface for examination or the appropriate metallographic specimen. The biggest limitation is size. It is important that the portion to be removed is documented by photographs and sketches, showing the location of the specimen to be removed. Preserve any fracture surface by plastic replicas or other method to prevent damage or attack. Regions adjacent to cracks are also to be preserved and protected. Cutting the specimens should be done very carefully so as not to cause any heat damage. Coolants are not recommended, unless the material cannot be cut without heat generation. The use of plastic replicas is useful for protecting surfaces and preserving any debris present.

Opening secondary cracks is useful when the primary fracture surface is damaged. These secondary cracks may provide better information, because they are tightly closed, and the fracture surfaces are not exposed to surface contaminants and corrosion. Care must be taken not to damage the primary fracture surface. Bending to open the crack is preferable, to expose the crack face. Often, the use of a sawcut to the back of the part will reduce the amount of force necessary to open the secondary crack. Another method is to use a tensile machine to open the crack face. The crack opening should be measured prior to opening, and the crack opening displacement can also be measured as the crack is slowly opened and exposed. One technique is to immerse the specimen in liquid nitrogen and impact the part so that the fracture surfaces are rapidly opened. One problem with this method is that it is very easy to damage the fracture surface from a misapplied hammer hit.

Macroscopic Examination

The macroscopic examination is conducted by a detailed examination at 1 to 100× by eye or binocular microscope. High-quality optics with excellent depth of field are required to properly examine the fracture surfaces. This detailed macroscopic examination can reveal a wealth of information on the location of fracture origins, direction of cracking, configuration of the stress state, and the last region to fail (shear lip). The presence of chevron marks can indicate the direction of rapid crack growth, and the different textures of the fracture can differentiate between fast final fracture and the initiating mechanism of fracture. Different textures from the region of fast fracture can indicate a different mechanism, such as fatigue, stress-corrosion cracking, or hydrogen embrittlement.

Microscopic Examination

The microscopic examination is usually conducted with an SEM (Fig. 6). This instrument is probably the most useful of all instruments for determining the mechanism of failure. It is capable of a large depth of field, with magnifications of 10 to 300,000×. It allows for direct examination of specimens, and when coupled with an energy-dispersive spectrometer, very small regions can be examined and analyzed for chemistry. It is very easy to use and requires very little training to take quality images. Interpretation of the images requires experience and understanding of the four basic modes of failure: dimpled rupture, cleavage, brittle intergranular, and fatigue. From these four basic modes, the detailed mode can be examined, and the failure mechanism is fit to the evidence. A greater discussion of the mechanisms of failure is found later in this chapter and elsewhere in this book.

Metallography

Metallography is a vital part of a failure analysis investigation. It can examine crack
morphology and its relationship with the microstructure present. It can help determine the thermal history of a component or region of a part and can show if work hardening was present. There can never be too many photographs and metallographic sections. Metallographic sections should be taken away from the crack and near the determined origins of cracking. Because this method is destructive, it is undertaken last. Typically, the crack face and edges are protected from rounding by applying support. This support can be electroless nickel plate or the use of alumina beads or steel shot in the metallographic specimen, adjacent to the surface.

Metallographic specimens are prepared using an epoxy or phenolic resin. The sample is placed into a small press, and phenolic resin is poured over the section. The press compacts the resin and forms a small, round sample that is then polished, etched, and examined under a metallographic microscope. When the specimen has cooled, it is taken out of the press and ground through a sequence of sandpapers. Typically, the sequence is 240, 320, 400, and 600 grit. The specimen is ground very flat before polishing. During polishing, the metallographic specimen is polished using a flat platen and 3 μm alumina slurries. Final polish is accomplished using 0.15 μm alumina slurry. Other polishing agents can be used, with diamond being a very common polishing agent. A finished metallographic sample used for the determination of the fracture mechanism in a steel weldment is shown in Fig. 7.

Examination of the metallographic specimen reveals surface imperfections, inclusions, and microstructural details. It can reveal the presence of decarburization and improper heat treatment. It often provides the needed documentation and support for the fracture analysis and determination of the root cause of failure.

Determination of the Fracture Mechanism

Examination of the fracture surface and metallography are used to determine the cause of failure. First, it is necessary to determine the fracture mode. Unfortunately, there is no clear or logical classification of fracture. Generally, classification is based on the crack growth mechanism (see also the chapter “General Aspects of Failure Analysis” in this book).

Ductile Fracture

On a macroscopic scale, a ductile fracture is accompanied by a relatively large amount of plastic deformation before the part fails. After failure, the cross section is reduced or distorted. Shear lips are observed at the latter part of the fracture and indicate the final failure of the part. The fracture surface is dull, with a fibrous appearance. Microscopically, ductile fracture is characterized by several distinct stages (Ref 5–8); an example is shown in Fig. 8. In this case, an ISO 12.9 low-alloy bolt failed by ductile torsional overload. The fracture was smooth, with fracture initiating from the threads. The fracture mode was microvoid coalescence (Ref 9), which occurs by the following process:

- A free surface is created from a small particle. This particle can be a second-phase particle, dispersoid, or inclusion. The separation of the metal matrix from the small particle at the matrix/particle interface can form this free surface, or the fracturing of the small particle can form the free surface.
- The free surface around the small particle creates a void. This void grows by plastic strain and hydrostatic stress.
- Finally, the voids grow to a size that they join or coalesce with adjacent voids.

This process of void formation, growth, and coalescence is shown schematically in Fig. 9. If the particles are well matched to the matrix and form a strong interface between the matrix and the particle, then the initial formation of voids is the critical step. Fracture occurs shortly after
void formation (Ref 10). If the interface between the particles and the matrix is weak, then voids form and grow readily. Substantial plastic deformation occurs. Fracture occurs when the voids reach a critical size. These voids substantially reduce the cross section, with the resulting local plastic instability (Ref 11). These voids coalesce to form a central crack perpendicular to the applied tensile stress. Depending on the applied stresses, the shape and configuration of the dimple shape can be changed (Fig. 10). This fact is important in determining the type of loading during a postfracture investigation. Dimples are small and can only be detected by using electron microscopy (Fig. 11).

The presence of inclusions in steel plays a major role in the ductility of steel. As indicated previously, the inclusions fracture and separate from the matrix during decohesion. Therefore, the deformability of these inclusions is important to determine the ductility of steel.

Nearly all steels have nonmetallic inclusions. The size and frequency of these inclusions is determined by the methods described in ASTM E45 (Ref 12). The cleanliness of the steel is

Fig. 8 Fracture of an ISO 12.9 bolt by ductile torsional overload. (a) Overall view of fracture. (b) Smooth and fibrous fracture as seen through the SEM. (c) Microvoid coalescence (dimples)

Fig. 9 Schematic showing the formation of microvoid coalescence

Fig. 10 Schematic representation of the creation of dimples in a loaded member by (a) simple tension, (b) shear loading, and (c) tearing

Fig. 11 Microvoid coalescence as seen through the SEM