Tribology in manufacturing has seen tremendous development in the past 100 years and essentially was transformed from a known and unavoidable nuisance to a sophisticated discipline. This is demonstrated by the reliability and quality of products that result in modern metal forming and cutting. This paper provides a brief summary of the state-of-the-art in manufacturing tribology, discussing some of the important developments in the field.

[DOI: 10.1115/1.4047723]

Keywords: bulk deformation processes (e.g., extrusion, forging, wire drawing, etc.), machining processes, modeling and simulation, sheet and tube metal forming, tribology in manufacturing

Introduction

This paper summarizes the major developments in manufacturing tribology in the past 100 years, corresponding to the lifetime of the ASME Manufacturing Engineering Division. It is worth noting that research in tribology predates the term; tribology was first coined in 1966 at Batelle laboratories when Peter Jost was interviewing researchers while preparing the now-famous Jost Report [1]. The term literally means “the art and science of rubbing,” but generally refers to friction, wear, lubrication, and surface engineering.

Most of the research in tribology in the post-World War II era has focused on machinery elements, but manufacturing undoubtedly presents a far more demanding set of problems. Temperatures are often above one-half the material melting temperature; surfaces evolve (flattened or roughened); emulsions are common; surface finishes set by customers can be extremely demanding; reliability of processes is a premium; and environmental considerations are extremely important. Indeed, tribology was promoted by the Jost report as being a technology able to greatly improve energy efficiency.

No brief introduction can reasonably cover all manufacturing operations. For the reader who requires additional information, they are directed to the classic manufacturing texts by Schey [2] and Kalpakjian and Schmid [3,4], the general tribology history text by Dowson [5], and especially the remarkable text on tribology in manufacturing by Schey [6], which is still the most comprehensive and most often-cited general reference on this important topic.

Overview of Manufacturing Tribology

In cutting and forming operations, the pressure in the tool-workpiece contact usually greatly exceeds the workpiece yield strength. The use of a coefficient of friction is thus of limited use; instead, unique friction and wear models and mechanisms may become apparent in forming. Thus, while some basic principles used in analysis of machinery are still valid, the extension to manufacturing presents unique challenges.

Friction Theory. The classic model of friction, often taught in high school physics classes, is Coulomb friction, where relative motion between two objects in contact is opposed by a friction force proportional to the normal force between the objects; the constant of proportionality is the Coulomb factor or friction coefficient, $\mu$. Models of asperities show that Coulomb friction is useful when the asperities in contact are remote from each other; however, if their stress fields interact (generally the case in forming operations), then the friction force no longer is linearly related to normal force. Coulomb friction also usually infers that friction is also independent of area of contact and that steady-state friction force does not depend on sliding velocity as long as the sliding velocity exists. These don’t hold in metal forming. For example, in extrusion, the die-workpiece friction force is not proportional to normal force; a coefficient of friction is not a constant but depends on the extrusion pressure; and friction depends on sliding speed due to lubrication, often by melting glass, and viscoplastic effects in the hot extruded material. Figure 1 shows friction behavior as a function of normal force, showing that real contact areas quickly approach the apparent contact area and saturate, which is a very different situation than Coulomb friction. For cases when the fractional contact area saturates as shown, a Tresca friction model [7], where friction force no longer is linearly related to normal force. Coulomb friction does not work in manufacturing except for clamps, chucks, and other machine elements.

The role of asperities in contacting surfaces was classically evaluated by Greenwood and Williamson [8] and Greenwood and Tripp [9], but in manufacturing the surfaces have much higher fractional contact areas and evolving surface profiles, limiting the utility of
such approaches. Further, plastic deformation in the bulk of the workpiece effectively softens the workpiece and leads to dramatic increase in real area of contact.

Sheu and Wilson [10] considered a plastically deforming workpiece and developed an analytical expression for asperity hardness as

\[ H = \frac{2}{f_1(\alpha E) + f_2(\alpha)} \]  

where \( H \) is the effective hardness of the workpiece:

\[ H = \frac{p_u - p_b}{k} \]

\( E \) is the dimensionless strain rate,

\[ E = \frac{\dot{\epsilon}l}{v} \]

\( \alpha \) is the indenter area ratio, \( \alpha = a/l \), \( v_f \) is the asperity flattening rate, and \( f_1 \) and \( f_2 \) are functions defined by

\[ f_1(\alpha) = 0.515 + 0.345\alpha - 0.860\alpha^2 \]

\[ f_2(\alpha) = \frac{1}{2.571 - \alpha - \alpha \ln(1 - \alpha)} \]

The effective hardness given by Eq. (2) is shown in Fig. 2, and it can be used to determine fractional contact areas. An extremely important discovery can be seen: a substrate strain rate (that is, in the bulk of the workpiece) causes the asperity hardness to fall dramatically. This means that in metal forming, asperities will flatten appreciably; this effect is not as significant in conventional tribology applications. This observation is very important for and explaining the use of Tresca friction in metal forming. This is well understood by manufacturers: because plastically deforming workpieces act soft, the fractional contact is quickly saturated, and the surface of the tooling can be impressed onto the workpiece. Of course, anyone who has seen a shiny coin understands that the smooth surface in the die is impressed onto the workpiece, when such an outcome is not predicted by conventional tribology.

The situation depicted in Fig. 1 is suitable for unlubricated metal forming, but the use of Coulomb friction or Tresca friction requires insight and more elaborate models. In mixed or boundary lubrication, the friction stress depends on both the lubricant and the asperities. Wilson et al. [11] suggest a friction stress expression of

\[ \tau_f = c_1 kA \theta_{fi} + \theta_{ki} H + \tau(1 - A) \]

where \( c \) is an adhesion coefficient, \( k \) is the material shear strength, \( A \) is the fractional contact area, \( \theta_i \) is a plowing coefficient proportional to the surface slope, \( \tau_i \) is a shear stress developed in the lubricant, and \( H \) is the non-dimensional surface hardness, which is the ratio of hardness to shear strength. Christensen [12] and Wilson and Marsault [13] calculate real contact areas in metal forming using large fractional contact areas, but these need the lubricant film thickness to be known, as described below.

Friction in metal fabrication processes cause strain localization and associated microstructural rearrangement near the interface [14-17]. This rearrangement results in the formation of fine grained and hardened layers. This narrow hard layer not merely has significant effect on the performance of machine components, it provides a natural functionally graded material, the type that has been of pursuit for demanding machinery performance [18,19].

**Lubrication**

**Conversion Coatings.** Lubrication in metalworking is complicated by the fact that entrainment velocities are generally too low to generate thick lubricant films. Phosphate and other conversion coatings are used, where the workpiece is placed in a phosphoric acid solution and a porous phosphate salt coating is applied. These coatings help entrain more lubricant and bring them into the die-workpiece contact region, helping to reduce friction and wear. A complete discussion of conversion coating processes and chemistry are given by Bay [20] and Schmid and Wilson [21]. The main effect of conversion coatings is that the volume of lubricant entrained is larger than with smooth surfaces. The theoretical modeling of such surfaces requires a modified Reynolds equation approach which is beyond the scope of this review. The use of flow factor approaches for incorporating surface transport in the Reynolds equation was developed by Patir and Cheng [22] and for metal forming situations by Wilson and Marsault [13].

**The Reynolds Equation.** It is well-known that the presence of a liquid lubricant, when between the workpiece and tool in sufficient thickness, will significantly reduce friction and wear. If the tool and workpiece are completely separated by a lubricant film, and the normal force between tool and workpiece is transmitted through the liquid, then a full film (also called thick film or hydrodynamic lubrication) of lubricant has been generated. The Reynolds equation is used to determine the film thickness. This equation is derived (in multiple methods) in Hamrock et al. [23].

**Regimes of Lubrication.** Wilson [24] defined the Regimes of Lubrication in a manufacturing context (Fig. 3). These are intended to show the mechanism by which a load is transferred between two surfaces; either by direct action from contacting asperities, through a pressurized lubricant film, or a combination.

In **boundary lubrication**, there is significant asperity contact. The lubricant film is insufficient to separate the surfaces, and the lubricant transmits negligible load. The frictional characteristics depend on the chemistry of the lubricant at the interface. Surface films are
sens. Therefore, cold working takes place below around 250 °C.

Owing to high contact stress and friction, solid lubricants, including soaps and waxes, are used for low-speed operations. In high-speed continuous operations, solid lubricants may be solid, but are usually liquid. Liquids are generally 1–10 nm, depending on the specific lubricant or additive molecular size.

At the other extreme, hydrodynamic films are generally thick enough to prevent contact by opposing asperities, so that all load is transmitted through the lubricant. Hydrodynamic films result in low friction and wear. However, this condition is undesirable in manufacturing because of surface roughening, called orange peel because the appearance of the surface is similar to an orange peel.

If load is transferred between asperities and a pressurized liquid lubricant, the condition is known as partial lubrication or mixed lubrication. This can be thought of as a mixture of hydrodynamic and boundary effects. A typical lubricant film thickness in a partial lubrication is typically between 0.001 and 1 μm.

**Lubricants.** Metal forming lubricants are unique; liquids can be used for room temperature applications, but solid lubricants or multi-phase lubricants are common. Lubricants are expensive, challenging to apply uniformly, difficult to remove and often an environmental impossibility to discard. They create problems in subsequent operations such as painting or adhesive bonding. However, lubricants are critical to control surface quality, reduce tool wear and control power consumption.

Metal forming operations are usually classified as cold or hot based on the homologous temperature, which is the ratio of the operating temperature to the material’s melting temperature on an absolute scale [4,25]. This distinction is grounded in metallurgy, for approximate temperature ranges where recrystallization and annealing occur in metals. The temperatures can be extreme: hot working is defined as an operation above half the melting temperature, but can often take place at 70–80% of the melting temperature.

**Emulsions.** An emulsion is a mixture of at least one immiscible liquid dispersed in the other in the form of droplets whose diameters exceed 0.1 μm. Emulsions combine excellent cooling ability with surprisingly good lubrication capability considering that they are often more than 90% water. They are commonly used as lubricants and coolants in metal working and cutting applications because of this unique combination of advantages. An emulsion is formulated and provided by a lubricant supplier and contains all necessary additives such as emulsifiers, brighteners, anti-foaming agents, and biocides. Deionized water is added by the user, and the mixture is agitated to form the emulsion when necessary. Some emulsifiers are effective in producing an emulsion without agitation.

Figure 3 Regimes of lubrication for metalworking applications [24]: (a) thick film, (b) thin film, (c) mixed, and (d) boundary film.

Graphite and molybdenum are useful at elevated temperature. Conversion coatings are used with these solids to ensure entrainment and proper performance.

Environmentally friendly lubrication systems have been pursued as long as there have been lubricants, but this has become a special focus recently. This is challenging mainly as applied to boundary additives and their unique chemistries [27]. Alternative lubricant systems for sheet forming have been investigated by Altan and co-workers [28–30]. The common alternative tribological systems of interest are as follows:

1. Because of sludge and associated heavy metal content that can contaminate soils, traditional zinc phosphate additives have been gradually replaced with more environmentally friendly chemistries, a topic that continues to receive considerable attention [31,32].
2. Graphite-based, or so-called “black” lubricants, are being replaced by “white” lubricants because of occupational safety issues and potential environmental effects. Many researchers are actively pursuing this area [31].
3. In sheet forming, chlorinated paraffin oils are often applied to prevent galling for stainless steel, titanium, and advanced high-strength steels. These are also a serious environmental concern, and much research has been directed toward finding alternative lubricants with anti-seizure properties [32–34].
4. Emulsions have become the preferred lubricant for metal rolling. Emulsions are mostly water, so that their cooling ability and environmental impact are inherently superior to the base oil alone.
5. Bio-based lubricants have become much more popular and continue to be actively researched, as their disposal is greatly simplified [27,35].
6. Solid films [36–39], powder coatings, or polymer films placed between die and workpiece [40,41] have been applied in sheet working.
7. Hot sheet stamping operations developed for new generation high-strength steels use AlCrNi coatings to reduce friction [42].
Wilson et al. [43] divided the inlet zone into three areas, namely:

1. The Deposition or Supply Region. In this region, particles are captured by the tooling or rejected by the backflowing fluid. Inlet zone numerical modeling far from the edge of contact (where the droplets are smaller than the gap) suggests that particles segregate to equilibrium locations; this location depends on the particle size and explains larger droplets having a greater propensity for entrainment.

2. The Concentration Region. In the concentration region, oil droplets are already captured and is preferentially further entrained because of their higher viscosity compared to water. Water backflows out as a result. When the concentration is high enough, the emulsion inverts (i.e., it becomes a water-in-oil emulsion) and then has superior film-generating ability.

3. A Pressurization Region. After inversion, the piezoviscous effect, whereby the lubricant viscosity increases with pressure, causes a rapid pressure increase, leading to good lubricant films. This observation has allowed the application of starvation analogies in metal rolling with very good results [46].

Figure 4 also shows the contact region, which is not part of the inlet zone, but is the location where plastic deformation occurs.

Montmitonnet et al. [48] built upon the work of Cassarini et al. [49] using the DCT framework to produce a complete model of cold strip rolling emulsion lubrication in mixed lubrication. The model predicts a maximum oil pool thickness or "plated out" film thickness. Comparison with experiments confirmed film thickness predictions. In particular, film thickness decreases as speed increases with a 4% emulsion, seen as a strong starvation effect, and a departure from the behavior of the neat oil case. Therefore, friction and roll force increase for the emulsion, but they decrease for the pure oil.

Although there have been considerable successes in understanding emulsions in recent years, a quantitative agreement requires an empirical estimate of likelihood of droplet capture. Schmid et al. [50–53] evaluated the capture mechanisms of droplets in converging channels and demonstrated that capture is more efficient at higher speeds. This remains an area of active research; notably Liang et al. [54] conducted microscopic, high-speed investigations of film formation ability in rolling at the nanoscale.

**Metal Rolling**

Rolling is an extremely demanding process. It is also arguably the most important metal forming process, since most metal parts are rolled at some point. Rolling can be a hot or cold forming process; in hot rolling, surface finish requirements are not stringent, and large-scale hot rolling is usually performed in reversing mills and produce larger thickness products. Cold rolling has much more strict surface finish requirements, often requiring mirror-like surfaces, and are available in thinner gages. Cold rolling is usually performed on tandem or cluster (Sendzimir) mills [4].

An early model of metal rolling is due to Orowan [55], who used a slab method to evaluate the roll stresses, forces, and torques. Rolling mechanics cannot be understood without recognizing the existence of a neutral point (Fig. 5). Since strip rolling is essentially plane strain, the workpiece velocity increases as the workpiece thickness decreases. However, the work rolls have a constant surface velocity, which is initially higher than the workpiece velocity, but is slower than the workpiece near the outlet. The position where the roll and workpiece have the same velocity is called the no-slip or neutral point. The friction forces acting on the workpiece change direction at this location.

The location of the neutral point can be controlled and adjusted through the use of front and back strip tension. The neutral point can even occur outside the roll bite with certain combinations of front and back tensions. It is then useful to express the forward slip [4]:

\[ S_f = \frac{v_1 - v_r}{v_r} \]  \hspace{1cm} (8)

where \( S_f \) is the forward slip, \( v_1 \) is the final workpiece velocity, and \( v_r \) is the roll velocity. Forward slip can be directly related to the neutral point location and can be readily measured.

In the absence of friction, work rolls will slip on the workpiece surface and no workpiece can be drawn into the gap. Therefore, hydrodynamic films and the low associated friction that results are rarely encountered. However, the mechanisms of hydrodynamic lubrication entrainment are still important. The Wilson–Walowit equation [56] provided the first lubricant film thickness prediction for isothermal metal rolling:

\[ h = \frac{6\eta \gamma U}{\tan \theta (1 - e^{-\eta \gamma})} \]  \hspace{1cm} (9)

In Eq. (9), \( h \) is the film thickness, \( \eta \) is the lubricant viscosity, \( \gamma \) is the viscosity pressure coefficient, \( U \) is the rolling speed, and \( \theta \) is the bite angle (the contact angle between roll and workpiece). Usually, \( (\gamma \eta) \) is a large number; the denominator term in parentheses is usually very close to unity. The Wilson–Walowit equation represents the first successful prediction of the film thickness in lubricated rolling, and as such can predict final surface workpiece roughness or process parameters such as forward slip discussed below.

**Hot Rolling.** When metal is continuously cast into slab, ingot, or bloom, it is fairly thick. Plates, sheets, and foils require a large reduction in thickness, which can only be accomplished with ductile workpieces. This necessitates the use of elevated temperatures. Hot rolling reductions are as high as 50% per pass, but more modest drafts (thickness reductions) per pass are normal. Hot rolling is done on reversing mills; a thick slab can be rolled repeatedly by changing the roll rotation direction and clearance after each pass. Large drafts require high friction in order to draw
the workpiece through the rolls, and hot rolling work rolls are therefore fairly rough to aid in workpiece entainment.

Elevated temperatures in hot rolling, along with the propensity to expose nascent surface from the workpiece suggest that material transfer to the rolls is very likely. The transferred metal can be stable and form a protective coating, or roll coat, or else it can be loosely adhered and then be redeposited on the workpiece surface, reducing surface quality. Billets also may have a very hard, thick, brittle, and abrasive oxide layer that compromises roll surface finish [2]. This explains why surfaces in hot rolling are not as good as those in cold rolling.

Glasses, which are liquid at hot forging temperatures, can develop full hydrodynamic films. Glass lubricants are rare in rolling practice, but are common in hot extrusion [6]. An area that has seen dramatic improvements in recent years and still has significant research interest is the use of textures on hot rolling work rolls [57].

Cold Rolling. Cold rolling is usually conducted on hot rolled workpieces. Smaller material thicknesses require higher rolling speeds for economic considerations. Lubrication is vital to reduce friction, roll bending, and material transfer and retransfer. Although carbide rolls are used in foil rolling on cluster (Szendzimer) mills, most rolling is conducted with cast iron or steel rolls. Chrome plating of work rolls to reduce wear and material transfer is common.

Smooth surfaces can be generated by flattening mechanisms as described above. However, large fractional contact areas result in pickup and work roll roughening; this forces refinishing or wire brushing to remove pickup. This is especially common for aluminum rolling. Smut or smudge is a concern; Reich et al. [58] suggest that effective lubricants result in more smudge, because the lubricant additives reduce surface energy and thereby ease particle removal from workpiece surfaces. Maintenance of the lubricant is difficult. Residual lubricant from hot rolling that is conveyed to a cold rolling mill as a film on the workpiece can contaminate the cold rolling lubricants. Another contaminant is leakage of greases and heavy oils from bearings, gears, and hydraulic pipes of the rolling mill into the lubricant. With multiple hot and cold rolling operations, the residual oil is increasingly complicated and becomes increasingly difficult to control.

Forging

Forging uses compressive stresses to deform metal. Forging is usually classified as “open die” or “closed” or “impression die” forging. Open die forging is always performed hot, and involves large reductions and simple tooling. Closed die forging may occasionally be done on cold workpieces, and involve more intricate final shapes.

Hot Forging. Friction is of major concern and influences workpiece strain. Friction strongly affects required forging force, forging energy and maximum die pressure. The die pressure distribution in upsetting of a cylindrical workpiece is given by

\[ p = S_y \frac{e^{(r - r_o) \tan \theta}}{h} \]  

where \( p \) is the die pressure at radius \( r \), \( r_o \) is the outer radius, \( S_y \) is the material yield strength, \( \mu \) is the friction coefficient, and \( h \) is the workpiece thickness. The pressure at the center of the workpiece is very high, especially for thin workpieces or large diameters. Due to the shape of the pressure distribution, this phenomenon is called the friction hill [4].

Cold Forging. Cold forging results in improved material properties and better surface finishes compared to hot forging. A much larger variety of lubricants is available because of the lower process temperatures; these lubricants can be optimized for a specific workpiece through proper additive selection. Lubricants reduce friction and use additive packages to reduce boundary friction. An additional role of lubricant is to thermally insulate the workpiece. This has two beneficial effects.

1. Since the tooling is usually much cooler than the workpiece, the lubricant slows conduction of heat from the workpiece to the tooling. Cooler workpieces have a higher flow strength; according to the Tresca friction law, maintaining higher temperatures reduces friction.
2. A good thermal barrier slows die wear.

There are a large number of lubricants used in metal forging, including:

1. Metal coatings, such as zinc, tin, or copper. These metallic coating reduce material transfer to the die as well as friction.
2. Solid lubricants, mainly graphite and molybdenum disulfide.
3. Polymer coatings, usually as an intermediate film or laminated or adhered onto the workpiece.
4. For light duty applications, liquid lubricants are applied.
5. Forging usually requires phosphate conversion coatings, as described above.

One of the main drawbacks to a thick lubricating film is associated with surface roughening, a phenomenon known as orange peel. Wilson et al. [59–61] describe the mechanisms associated with surface roughening.

Extrusion and Drawing

When simple sketches of extrusion and drawing operations are made, the only apparent difference is that the workpiece is pushed through dies in extrusion, while it is pulled through the dies in drawing. In actuality, the processes have significant differences, perhaps the most important being that extrusion is a batch process, while drawing can be a continuous process. More subtly, extrusion will use far greater reductions in area per pass than drawing, and will be performed at a higher temperature to obtain greater workpiece ductility. Good general references on extrusion are Saha [62] and Schey [6].

Extrusion. The starting material in extrusion is a cast or previously rolled billet, placed in a container and pushed by a hydraulic press/dummy block assembly through a die. Direct, indirect, and hydrostatic are the most common forms of extrusion, but the discussion here can be limited to direct extrusion since the tribology and mechanics involves are common to these types. In direct extrusion, the material is pushed through the container in the same direction as the ram, generating significant friction between the workpiece and container. In indirect or reverse extrusion, the billet is at rest in the container, and the stem pushes against this stationary billet. Friction is therefore only generated at the die/billet interface. Hydrostatic extrusion uses a pressurized liquid (usually a low-viscosity oil) to force material through a die; while a complex tooling arrangement, hydrostatic extrusion develops considerable ductility in the workpiece.

Extrusion can be a hot or cold process, as described previously. Cold extrusion is often indistinguishable from cold forging operation and has the following characteristics:

1. Oxide films are a much lower concern than in hot extrusion. With proper lubrication, a very good surface finish can be achieved in cold extrusion without secondary finishing.
2. Lubricants used are typically liquid viscous lubricants or emulsions. Lubricants are not always required.
3. Extrusion ratios are smaller than in hot extrusion.
Hot extrusion is characterized by:

1. Carbon steel extrusion involves thick, hard, abrasive oxides that can lead to excessive die wear and compromised surface finish.
2. Greases, graphite, or molten solids such as glass are used as lubricants. The Séjournet process uses glass powder applied as a pillow in front of the billet. Another approach is to coat the hot billet with glass powder before insertion into the cylinder. This causes particles to adhere to and melt on the billet surface, and has the added benefit of insulating the billet and reducing heat transfer to the container.
3. Hot extrusion has large billets and very large extrusion ratios, so that very long workpieces are the norm.

Extrusion lubrication theory follows from the Reynolds equation as discussed above. However, there is a lubrication breakdown situation that is unique to extrusion. At startup, the billet has a layer of lubricant, but the billet plastically deforms and stretches, reducing the film thickness. This lubricant breakdown causes workpiece and tooling asperities to contact, leading to high forces and workpiece and tooling surface damage. Wilson [63] described the lubrication mechanics involved and explained that when this situation occurs, the billet can be periodically backed away from the die. When a steady-state is reached, continuous extrusion can progress.

A number of simplified equations exist for the lubricant thickness in extrusion and are summarized by Wilson [63]. For a viscous lubricant with a piezoviscous effect (pressure–viscosity dependence), the film thickness is given by

$$h = \frac{3\eta_0 U}{\tan \theta (1 - e^{-\gamma \sigma})}$$  \hspace{1cm} (11)

where $h$ is the film thickness, $\eta_0$ is the ambient lubricant viscosity, $U$ is the extrusion speed, $\theta$ is the die angle, and $\sigma$ is the flow stress of the material. The pressure exponent of viscosity, $\gamma$, is used to account for the increase in viscosity with pressure according to

$$\eta = \eta_0 e^{\gamma \sigma}$$  \hspace{1cm} (12)

Equation (12) is known as the Barus law [64]; many other rheological models exist for metalworking fluids, as discussed by Hamrock et al. [23]. The similarity between Eqs. (9) and (11) is not coincidence, as both derived directly from the Reynolds equation.

**Drawing.** Continuous lengths of wire and tube are produced by drawing. Drawing has stringent surface finish requirements and is commonly performed cold. The use of a mandrel allows drawing of tubes.

Liquid lubricants are common, although soap and wax are sometimes used for steel extrusion. In drawing, a number of die features are used to produce uniform and consistent lubricant films. These include the following:

1. A bell entrance section of the die encourages formation of a lubricant meniscus, encouraging consistent flow by preventing starvation.
2. Schey [6] discusses inlet tubes that can develop a hydrostatic pressure in a viscous lubricant just before a drawing die. The use of an inlet tube requires relatively long tubes (50–400 mm). Contaminant particles must also be carefully filtered.
3. Multiple dies can be used. Here the first die acts as a seal to allow lubricant, usually grease or soap, to pass through and become pressurized in the chamber between dies.
4. Hydrostatic pressure can be introduced using a double die arrangement. Schey [6] notes that the pressure needs to be high enough to cause yielding in the chamber.

Most research has been performed on isothermal drawing, but Schey [6] notes that there is a significant die temperature increase over the first meter or so of a drawing operation.

**Sheet Metal Forming**

Sheet metal forming consists of a wide variety of operations, ranging from stamping of automobile bodies from steel and drawing and ironing of beverage containers. Tribology has important implications on tooling design and formability in sheet metal forming operations.

**Friction Modeling in Sheet Forming.** Friction in sheet forming is complex, as discussed above. The use of a Coulomb friction model leads to highly inaccurate strain predictions, requiring a more sophisticated approach toward friction.

Any of the lubrication regimes (boundary, mixed, thin film, full film) can occur simultaneously at different locations in the tooling–workpiece interface, or at different times at the same location. Wilson et al. [11] defined the local film thickness in a stretch forming operation using an internal variable approach. They included a lubrication analysis based on the Reynolds’ equation and a friction stress calculation using fractional contact areas. Figure 6 compares the predicted strains with the measurements performed by Hector et al. [65] for forming a brass dome. The approach used can predict strain distributions accurately.

**Lubricants.** Sheet metal forming operations are generally performed with liquid lubricants. Formulations consist of base oil and additive packages (boundary, EP additives, brighteners, corrosion inhibitors, etc.), the chemistries and concentrations of which are carefully protected. In addition, oil-in-water emulsions are commonly used in stamping and can making.

Soft solid coatings (waxes, soaps, polymers, etc) are also important in forming difficult sheet metal parts. Coduti [66] found such coatings to have better lubricating properties than liquids. They used polymer sheets placed between tooling and workpiece, but it is more common to use a polymer coatings. Franks and Plevy [67] observed that teflon coating applied to ceramic substrate fired onto aluminum sheet was effective in reducing friction and also increased the drawing depth without rupture. Jaworski et al. [36–38] presented mathematical and experimental results of ironing with laminated polymer coatings on steel.

Metal coatings can be applied to sheet metal, especially steels. For example, tin coated steels provide low friction and inhibit corrosion in food containers; corrosion inhibition for food containers; zinc applied to steel provides corrosion resistance in automotive bodies.

Another approach is to apply surface engineering principles, either by applying specialty coatings or imparting a desired

![Fig. 6 Prediction of strain in axisymmetric stretch forming of brass over a spherical punch [11]](image-url)
surface. Dohda [68] describes thin hard coatings used in sheet forming, while Sheu et al. [69] give an excellent overview of surface engineering and tooling surface texture development.

Smudge, Orange Peel, and Galling. Smooth surfaces provide a marketing advantage for containers and automotive sheet metal parts; they also can provide a performance improvement. Major obstacles superior surface generation are smudge, orange peel, and galling.

Smudge and galling are related. When tooling asperities interact with workpieces, wear fragments are occasionally produced. Hector et al. [70], suggested that microcutting is the dominant mechanism for wear debris generation. Such particles can be easily removed from a surface with a soft cloth. This has the same appearance as cleaning a dirty surface, hence the term “smudge.” Particles adhering to the tooling can result in scratches or machining marks, or more serious surface damage, known as galling.

Effective boundary additives limit galling, as does cooling of tooling. Most galling mechanisms are material-specific. For example, 3004 aluminum alloy was developed for producing ironed containers and limits material transfer to tooling. This is due to the hard $A_2(Mn|Fe)_3Si$ inclusions that continuously scrape the tooling surface and remove transferred aluminum [71].

Orange peel is plastic deformation-induced roughening. Workpieces may roughen or smoothen in sheet forming; Saha et al. [72] found both behaviors with different aluminum alloys.

Machining

The science of metal cutting has advanced dramatically in the last one hundred years. In the 1940s, Merchant established a research group at the Cincinnati Milacron company that made significant contributions to a fundamental understanding of machining mechanics. Because this material removal occurs at high strains, strain rates, temperatures, pressures, and velocities, understanding the tribology of the cutting edge-workpiece interface is essential for successful part production. It is therefore not surprising that Merchant started as a tribologist, and was awarded a gold medal in 1980 from the Institution of Mechanical Engineers for his work that increased understanding of the friction stress on tool rake faces [73].

A depiction of shear stresses based on Merchant’s initial work is shown in Fig. 7. As can be seen, the shear stress is fairly constant near the tool tip, but drops further up the tool or rake face. This is now understood to represent a transition from Tresca-type friction (labeled “Sticking” in Fig. 7) to Coulomb friction (labeled “Sliding”). The saturation of contact area and effect on friction depicted in Fig. 1 is understandable—temperatures, shear strains, and pressures are very high the deformation zone, which can be taken as a shear plane for most materials.

Machining tribology poses a significant challenge due to the multiple parameters that must be simultaneously considered to arrive at a cost-minimized solution in production. These include, but are not limited to:

1. Machining parameters such as chip width, chip thickness, and cutting speed; these are specified by the depth of cut (both axial and radial depth for milling), feed per revolution for turning (or feed per tooth for milling), and spindle speed
2. Cutting tool material and coating—many tool material options with and without single and multi-layer coatings are available
3. Cutting tool geometry—basic parameters include side rake angle, clearance angle, cutting edge radius, and number of teeth and spacing (for milling), but actual cutting tools, whether solid or inserted designs, include many additional geometric options
4. Workpiece material and microstructure
5. Cutting fluid type and application method
6. CNC part path—the strategy used by the computer numerically controlled (CNC) program can affect the time to machine and the tool wear-rate; for example, up or down (conventional or climb) milling and the path geometry, such as spiral in/out, constant radial engagement, trochoidal, and zig-zag strategies for milling, must be selected by the process planner.

An excellent summary of the mechanical and tribological concerns in cutting is given by Shaw [74].

Tool Life. Tool wear in machining leads to finite tool life, which is often measured in minutes rather than hours. Because tool costs can be significant, selecting a process plan that limits tool wear to an acceptable level is an important activity for part production by machining. Tool wear is driven by as follows:

1. High temperatures at the tool–chip interface, which can lead to softening of the tool material [75–77]; this tendency of metals to soften with heat addition has actually been leveraged in thermally enhanced machining which uses an external heat source (such as a laser) to heat and soften the workpiece locally in front of the cutting tool [78].
2. Intimate contact between the chip and tool rake face, which encourages diffusion between the chip and tool materials, particularly at high temperatures
3. Hard inclusions in the workpiece material, which can abrade the tool material
4. Large cutting forces and pressures, which can cause plastic deformation and fracture of the tool edge
5. Thermal cycling in interrupted cuts, which can cause cracking and catastrophic tool failure due to thermal fatigue.

Several wear mechanisms have been identified in metal cutting. These include abrasion, adhesion, diffusion, and attrition [79–81]. These wear mechanisms combine to form characteristic wear features on the cutting tool surfaces. Common wear features include the following: flank, crater, and notch wear. In many cases, a single wear feature will dominate, such as flank wear, but all may occur simultaneously.

The tool wear-rate naturally depends on the cutting tool material and any coatings applied to the wear surfaces. Fundamental tool material requirements include the following:

1. a higher hardness than the work material at the cutting temperature
2. high strength to resist cutting stresses
3. high toughness to avoid breakage under impact loads
4. low reactivity (i.e., chemically inert)

Tool materials have progressed from high-speed steels in the 1940s to sintered carbides, such as tungsten carbide or WC and titanium carbide or TiC (1930s), to ceramics, such as $A_2O_3$, cubic boron nitride or CBN, and polycrystalline diamond or PCD (1960s), and coated steels and carbides (1970s). Coatings are now routinely applied by chemical vapor deposition (CVD) and physical vapor deposition (PVD). Common examples include TiC, TiN, $A_2O_3$, and combinations of these in alternating layers [82–85].
One important issue for coating life is the cutting edge preparation. Edge honing can be used to increase the coating persistence at the sharp edge [86–88]. Although it has been understood since the early 1900s that the rate of tool wear depends strongly on the cutting speed, as well as other process parameters and tool geometry, its pre-process prediction remains elusive [89]. Empirical modeling efforts are therefore common. The Taylor-type tool life model, named for the pioneering work by F.W. Taylor, can be expressed for turning operations as

\[ C = v^p f_r^q T \]  

(13)

where \( C \), \( p \), and \( q \) are empirical constants that depend on the tool and workpiece materials, tool geometry, coolant use, and process parameters. The parameters \( v \), \( f_r \), and \( T \) are the cutting speed (which is the product of spindle speed and workpiece radius), feed per revolution, and tool life, respectively, and they are set by the machinist. Other parameters may be added as well, including the depth of cut, \( a \), for example. The cutting parameters are identified in Fig. 8, where \( \Omega \) is the spindle speed and \( r \) is the workpiece radius. The tool life may be defined by the cutting time required to reach a predetermined level for a particular wear feature.

One common tool life indicator is flank wear, since increased levels of flank wear lead to higher friction between the insert flank face and the machined surface [79,80]. This higher friction, in turn, accelerates the tool wear due to increased temperatures. To establish the tool life, an allowable flank wear width, FWW (also denoted Verschleissmarken Breite, or VB, in German), is established and tests are performed at a given combination of process parameters (\( v \) and \( f_r \) from Eq. (13)). The time to reach the predetermined flank wear width is then recorded and defined as the tool life for the selected machining conditions. In practice, cutting proceeds for some time and then the FWW is measured. Additional cutting is performed and the FWW is measured again. This sequence is repeated at different process parameters in order to determine the best-fit constants for Eq. (13). Alternately, an artificial neural network (ANN) may be established which defines the machining parameter-tool life relationship [90–92]. Fundamental limitations of this empirical approach are as follows: (1) a new testing sequence is required for each change in tool geometry, tool material, workpiece material, or coolant delivery; (2) the data cannot, in general, be extrapolated outside the testing range; and (3) only the limiting tool life is examined—the process variation (in force or power, for example) which naturally occurs due to the tool degradation is not typically evaluated.

While measurement of a wear feature provides an identifiable metric to define tool life, it can be inconvenient to obtain this data in production environments. An alternative approach is to monitor an accessible process signal, such as spindle power, cutting force, or acoustic emission and compare this signal level to the nominal value for a new tool throughout the selected CNC part program. This approach is often referred to as “indirect” sensing because a signal which is believed to be indicative of tool wear is queried, rather than a direct physical feature such as the machined surface finish [93–95]. The primary challenges associated with indirect sensing are follows: (1) defining the baseline behavior and (2) setting the limits at which the tool is considered to be worn while avoiding false alarms. Once a tool life model is identified, the tool life can be related to the machining cost and the process parameters can be optimized to arrive at a minimum cost solution.

Tool material research continues, with the goal of extending the useful machining life of tools and matching tool materials to specific workpiece materials to minimize adhesion.

**Cutting Fluids.** Cutting fluids are used to:

1. prevent the tool, workpiece, and machine from overheating and distorting
2. increase tool life
3. improve surface finish
4. prevent oxidation
5. clear away chips.

They can serve as a coolant, lubricant, or both. Due to the high pressures between the chip and rake face, this interface is generally inaccessible to externally applied cutting fluids, particularly for continuous cutting. Because the pressure between the tool and chip in the cutting zone is much higher than atmospheric pressure, it is challenging to introduce cutting fluid into this zone. Therefore, new tool designs that include high pressure streams directed at the tool–chip interface have been introduced in an effort to lubricate and cool this location, while encouraging chip breakage for continuous cutting operations [96–98]. The two primary fluid types are neat cutting oils and water-based (oil/synthetic/organic and water emulsion). Emulsions serve as an effective coolant and are often directed at the highest temperature location to flood the cutting zone [6]; it is also possible to target the hot zones using mist delivery at much lower flowrates [99].

High pressure coolant is a more recent approach and is carried out at flowrates that may exceed 100 lpm with pressures up to 200 MPa [99–109]. Tooling with integrated coolant nozzles is available. Selection of the flowrate (lpm) depends on the power (kW) consumed in the cut; a rule of thumb is 2.5 lpm/kW. The required coolant jet orifice diameter is directly related to the velocity, pressure, and flowrate. Therefore, given the anticipated power and a selected pressure, the appropriate orifice diameter can be chosen.

High pressure coolant research and development is motivated by the shortcomings of existing technology. It is proposed that low pressure flood coolant fails because a vapor barrier is created when the coolant is rapidly heated above its boiling point due to contact with the primary shear zone. It is suggested that high pressure coolant penetrates this barrier and is, therefore, capable of removing the heat that is generated by the cutting process. The ability to select the high pressure coolant application point and direction is therefore important.

Tool wear for hard-to-machine materials, such as nickel-based alloys, is a particular challenge. Oil emulsion and water-based cutting fluids are the conventional cooling/lubrication approach in these cases. However, the environmental and health concerns associated with the use of these fluids are driving the demand for alternative machining strategies. These strategies include, for example, dry machining using ceramic cutters, minimum quantity lubrication, chilled air, and cryogenic machining [109].

Among these methods, the use of cryogenic fluids offers an attractive cooling alternative and has received significant attention in the literature [110–112]. Cryogenic fluids cool the cutting zone to very low temperatures using liquefied gases, such as liquid nitrogen (N\(_2\)), helium (He), or carbon dioxide (CO\(_2\)). Investigations involving cryogenic cooling are leading to improved tool performance and overall part quality. The influence of cryogenic nitrogen machining, for example, has improved the tool life, surface finish,
and residual compressive stresses in the turning of magnesium alloys, NiTi-shape memory alloys, and titanium alloys [113–115].

CO₂ cooling has been shown to improve the tool life in polycrys-
talline diamond (PCD) turning of compacted graphite iron (CGI), but the performance is highly dependent on grain size and binder content of the tool material [116]. It has also been shown to also suppress burr formation and limit notch wear in the turning of β-titanium alloys by a factor of two over conventional emulsion flood coolant [117].

Predominantly, cryogenic cooling has positively impacted the cutting performance of hard-to-machine materials during uninterrupted cutting operations. Its effect on interrupted cutting operations, such as milling, is not well understood, although some studies are reported in the literature. For example, Sù et al. observed improved tool life in the high-speed milling of Ti-6Al-4V using compressed nitrogen gas, but also observed instances of dramatically reduced tool life due to thermal fatigue cracking of the tool [118].

Another study examined the cutting performance of milling Haste-
tloy X (a nickel–chromium–iron–molybdenum alloy used for aircraft engine components) using two cooling/lubricating strategies [119]. The effects on tool life, cutting forces, and cutting temperatures were examined using an aqueous-based minimum quantity lubrication (MQL) coolant and external cryogenic CO₂ spray. MQL refers to the use of near-dry conditions as opposed to the traditional flooding of the cutting zone; in theory, the lubricant supply matches the amount needed for lubrication but not heat transfer. Digital micro-
scope images, scanning electron microscope images, cutting force data, and thermal images were used to compare the effectiveness of the two cooling methods. Results showed an approximately 89% longer tool life when using MQL compared to CO₂ for the selected cutting conditions. From microscope images, a clear edge chipping phenomenon occurred when using CO₂. This led to larger cutting forces and, ultimately, sudden catastrophic failure of the tool edge. Possible explanations for the adverse cutting performance using CO₂ include thermal fatigue cycling, material adherence to the cutting edge, coating failure, and changes in workpiece material properties.

Conclusions

The field of tribology has seen enormous development in the past 100 years, and this is especially true for tribology applied to manu-
ufacturing processes. A fundamental understanding of the differ-
ences between manufacturing and machine element tribology has been developed and applied to improve the robustness of manufact-
uring processes. Unique lubricants, such as glass and oil-in-water emulsions, are widespread. Friction rules are complex, and lubricant is necessary for process viability but often a desired level of friction must be preserved. Surfaces are dynamic and produced during manu-
facturing; roughness can decrease due to flattening against a smooth tool or roughen when unconstrained and strained. This paper highlighted the work of two of the most prolific researchers in tribology in manufacturing of the last century: John A. Schey and William R.D. Wilson.

Dedication

Researchers in the field of process tribology cannot help but continually read the papers of two of the most prolific researchers of the past century: William R.D. Wilson and John A. Schey. The authors of this paper acknowledge their significant scientific contributions, leadership to the profession, and also, with great gratitude, their example as mentors and role models.

Data Availability Statement

The authors attest that all data for this study are included in the paper. No data, models, or codes were generated or used for this paper.

References

Hill, New York.


