Chapter 2
Lubrication Principles

2-1. Friction

a. Definition of friction.

(1) Friction is a force that resists relative motion between two surfaces in contact. Depending on the application, friction may be desirable or undesirable. Certain applications, such as tire traction on pavement and braking, or when feet are firmly planted to move a heavy object, rely on the beneficial effects of friction for their effectiveness. In other applications, such as operation of engines or equipment with bearings and gears, friction is undesirable because it causes wear and generates heat, which frequently lead to premature failure.

(2) For purposes of this manual, the energy expended in overcoming friction is dispersed as heat and is considered to be wasted because useful work is not accomplished. This waste heat is a major cause of excessive wear and premature failure of equipment. Two general cases of friction occur: sliding friction and rolling friction.

b. Sliding friction.

(1) To visualize sliding friction, imagine a steel block lying on a steel table. Initially a force F (action) is applied horizontally in an attempt to move the block. If the applied force F is not high enough, the block will not move because the friction between the block and table resists movement. If the applied force is increased, eventually it will be sufficient to overcome the frictional resistance force f and the block will begin to move. At this precise instant, the applied force F is equal to the resisting friction force f and is referred to as the force of friction.

(2) In mathematical terms, the relation between the normal load L (weight of the block) and the friction force f is given by the coefficient of friction denoted by the Greek symbol μ. Note that in the present context, "normal" has a different connotation than commonly used. When discussing friction problems, the normal load refers to a load that is perpendicular to the contacting surfaces. For the example used here, the normal load is equal to the weight of the block because the block is resting on a horizontal table. However, if the block were resting on an inclined plane or ramp, the normal load would not equal the weight of the block, but would depend on the angle of the ramp. Since the intent here is to provide a means of visualizing friction, the example has been simplified to avoid confusing readers not familiar with statics.

c. Laws of sliding friction. The following friction laws are extracted from the Machinery Handbook, 23rd Revised Edition.

(1) Dry or unlubricated surfaces. Three laws govern the relationship between the frictional force f and the load or weight L of the sliding object for unlubricated or dry surfaces:

(a) "For low pressures (normal force per unit area) the friction force is directly proportional to the normal load between the two surfaces. As the pressure increases, the friction does not rise proportionally; but when the pressure become abnormally high, the friction increases at a rapid rate until seizing takes place."

(b) The value of f/L is defined as the coefficient of friction μ. "The friction both in its total amount and its coefficient is independent of the area of contact, so long as the normal force remains the same. This is true for moderate pressures only. For high pressures, this law is modified in the same way as the first case."

(c) "At very low velocities, the friction force is independent of the velocity of rubbing. As the velocities increase, the friction decreases."

The third law (c) implies that the force required to set a body in motion is the same as the force required to keep it in motion, but this is not true. Once a body is in motion, the force required to maintain motion is less than the force required to initiate motion and there is some dependency on velocity. These facts reveal two categories of friction: static and kinetic. Static friction is the force required to initiate motion (F_s). Kinetic or dynamic friction is the force required to maintain motion (F_k).

(2) Lubricated surfaces. The friction laws for well lubricated surfaces are considerably different than those for dry surfaces, as follows:

(a) "The frictional resistance is almost independent of the pressure (normal force per unit area) if the surfaces are flooded with oil."

(b) "The friction varies directly as the speed, at low pressures; but for high pressures the friction is very great at low velocities, approaching a minimum at about 2 ft/sec linear velocity, and afterwards increasing approximately as the square root of the speed."

(c) "For well lubricated surfaces the frictional resistance depends, to a very great extent, on the temperature, partly because of the change in viscosity of the oil and partly because, for journal bearings, the diameter of the bearing increases with the rise in temperature more rapidly than the diameter of the shaft, thus relieving the bearing of side pressure."

(d) "If the bearing surfaces are flooded with oil, the friction is almost independent of the nature of the material of the surfaces in contact. As the lubrication becomes less ample, the coefficient of friction becomes more dependent upon the material of the surfaces."

(3) The coefficient of friction. The coefficient of friction depends on the type of material. Tables showing the coefficient of friction of various materials and combinations of materials are available. Common sources for these tables are Marks Mechanical Engineering Handbooks and Machinery's Handbook. The tables show the coefficient of friction for clean dry surfaces and lubricated surfaces. It is important to note that the coefficients shown in these tables can vary.

(4) Asperities. Regardless of how smooth a surface may appear, it has many small irregularities called asperities. In cases where a surface is extremely rough, the contacting points are significant, but when the surface is fairly smooth, the contacting points have a very modest effect. The real or true surface area refers to the area of the points in direct contact. This area is considerably less than the apparent geometric area.

(5) Adhesion. Adhesion occurs at the points of contact and refers to the welding effect that occurs when two bodies are compressed against each other. This effect is more commonly referred to as "cold welding" and is attributed to pressure rather than heat, which is associated with welding in the more familiar sense. A shearing force is required to separate cold-welded surfaces.

(6) Shear strength and pressure. As previously noted, the primary objective of lubrication is to reduce friction and wear of sliding surfaces. This objective is achieved by introducing a material with a low shear strength or coefficient of friction between the wearing surfaces. Although nature provides such materials in the form of oxides and other contaminants, the reduction in friction due to their presence is insufficient for machinery operation. For these conditions, a second relationship is used to define the coefficient of friction: $\mu = S/P$, where S is the shear strength of the material and P is pressure (or force) contributing to compression. This relationship shows that the coefficient of friction is a function of the force required to shear a material.

(7) Stick-slip. To the unaided eye the motion of sliding objects appears steady. In reality this motion is jerky or intermittent because the objects slow during shear periods and accelerate following the shear. This process is continuously repeated while the objects are sliding. During shear periods, the static friction force F_s controls the speed. Once shearing is completed, the kinetic friction force F_k controls the speed and the object accelerates. This effect is known as stick-slip. In well lubricated machinery operated at the proper speed, stick-slip is insignificant, but it is responsible for the squeaking or chatter sometimes heard in machine operation. Machines that operate over long sliding surfaces, such as the ways of a lathe, are subject to stick-slip. To prevent stick-slip, lubricants are provided with additives to make F_s less than F_k.

d. Rolling friction.

(1) When a body rolls on a surface, the force resisting the motion is termed rolling friction or rolling resistance. Experience shows that much less force is required to roll an object than to slide or drag it. Because force is required to initiate and maintain rolling motion, there must be a definite but small amount of friction involved. Unlike the coefficient of sliding friction, the coefficient of rolling friction varies with conditions and has a dimension expressed in units of length.

(2) Ideally, a rolling sphere or cylinder will make contact with a flat surface at a single point or along a line (in the case of a cylinder). In reality, the area of contact is slightly larger than a point or line due to elastic deformation of either the rolling object or the flat surface, or both. Much of the friction is attributed to elastic hysteresis. A perfectly elastic object will spring back immediately after relaxation of the deformation. In reality, a small but definite amount of time is required to restore the object to original shape. As a result, energy is not entirely returned to the object or surface but is retained and converted to heat. The source of this energy is, in part, the rolling frictional force.

(3) A certain amount of slippage (which is the equivalent of sliding friction) occurs in rolling friction. If the friction of an unhoused rolling object is measured, slippage effects are minimal. However, in practical applications such as a housed ball or roller bearing, slippage occurs and contributes to rolling friction. Neglecting slippage, rolling friction is very small compared to sliding friction.

e. Laws of rolling friction. The laws for sliding friction cannot be applied to rolling bodies in equally quantitative terms, but the following generalities can be given:

(1) The rolling friction force F is proportional to the load L and inversely proportional to the radius of curvature r, or $F = \mu_r L/r$, where μ_r is the coefficient of rolling resistance, in meters (inches). As the radius increases, the frictional force decreases.

(2) The rolling friction force F can be expressed as a fractional power of the load L times a constant k, or $F = kL^n$ where the constant k and the power n must be determined experimentally.

(3) The friction force F decreases as the smoothness of the rolling element improves.

2-2. Wear

Wear is defined as the progressive damage resulting in material loss due to relative contact between adjacent working parts. Although some wear is to be expected during normal operation of equipment, excessive friction causes premature wear, and this creates significant economic costs due to equipment failure, cost for replacement parts, and downtime. Friction and wear also generate heat, which represents wasted energy that is not recoverable. In other words, wear is also responsible for overall loss in system efficiency.

   a. Wear and surface damage. The wear rate of a sliding or rolling contact is defined as the volume of material lost from the wearing surface per unit of sliding length, and is expressed in units of $[length]^2$. For any specific sliding application, the wear rate depends on the normal load, the relative sliding speed, the initial temperature, and the mechanical, thermal, and chemical properties of the materials in contact.

   (1) The effects of wear are commonly detected by visual inspection of surfaces. Surface damage can be classified as follows:

   (a) Surface damage without exchange of material:

   !   Structural changes: aging, tempering, phase transformations, and recrystallization.

   !   Plastic deformation: residual deformation of the surface layer.

   !   Surface cracking: fractures caused by excessive contact strains or cyclic variations of thermally or mechanically induced strains.

   (b) Surface damage with loss of material (wear):

   !   Characterized by wear scars of various shapes and sizes.

   !   Can be shear fracture, extrusion, chip formation, tearing, brittle fracture, fatigue fracture, chemical dissolution, and diffusion.

   (c) Surface damage with gain of material:

   !   Can include pickup of loose particles and transfer of material from the opposing surface.

   !   Corrosion: Material degradation by chemical reactions with ambient elements or elements from the opposing surface.

   (2) Wear may also be classified as mild or severe. The distinguishing characteristics between mild and severe wear are as follows (Williams 1994):

(a) Mild

! Produces extremely smooth surfaces - sometimes smoother than the original.

! Debris is extremely small, typically in the range of 100 nanometers (nm) (3.28×10^{13} ft) in diameter.

! High electrical contact resistance, but little true metallic contact.

(b) Severe

! Rough, deeply torn surfaces - much rougher than the original.

! Large metallic wear debris, typically up to 0.01 mm (3.28×10^5 ft) in diameter.

! Low contact resistance, but true metallic junctions are formed.

b. Types of wear. Ordinarily, wear is thought of only in terms of abrasive wear occurring in connection with sliding motion and friction. However, wear also can result from adhesion, fatigue, or corrosion.

(1) Abrasive wear. Abrasive wear occurs when a hard surface slides against and cuts grooves from a softer surface. This condition is frequently referred to as two-body abrasion. Particles cut from the softer surface or dust and dirt introduced between wearing surfaces also contribute to abrasive wear. This condition is referred to as three-body abrasion.

(2) Adhesive wear. Adhesive wear frequently occurs because of shearing at points of contact or asperities that undergo adhesion or cold welding, as previously described. Shearing occurs through the weakest section, which is not necessarily at the adhesion plane. In many cases, shearing occurs in the softer material, but such a comparison is based on shear tests of relatively large pure samples. The adhesion junctions, on the other hand, are very small spots of weakness or impurity that would be insignificant in a large specimen but in practice may be sufficient to permit shearing through the harder material. In some instances the wearing surfaces of materials with different hardness can contain traces of material from the other face. Theoretically, this type of wear does not remove material but merely transfers it between wearing surfaces. However, the transferred material is often loosely deposited and eventually flakes away in microscopic particles; these, in turn, cause wear.

(3) Pitting wear.

(a) Pitting wear is due to surface failure of a material as a result of stresses that exceed the endurance (fatigue) limit of the material. Metal fatigue is demonstrated by bending a piece of metal wire, such as a paper clip, back and forth until it breaks. Whenever a metal shape is deformed repeatedly, it eventually fails. A different type of deformation occurs when a ball bearing under a load rolls along its race. The bearing is flattened somewhat and the edges of contact are extended outward. This repeated flexing eventually results in microscopic flakes being removed from the bearing. Fatigue wear also occurs during sliding motion. Gear teeth frequently fail due to pitting.

(b) While pitting is generally viewed as a mode of failure, some pitting wear is not detrimental. During the break-in period of new machinery, friction wears down working surface irregularities. This

condition is considered to be nonprogressive and usually improves after the break-in period. However, parts that are continuously subjected to repeated stress will experience destructive pitting as the material's endurance limit is reached.

(4) Corrosive wear.

(a) Corrosive wear occurs as a result of a chemical reaction on a wearing surface. The most common form of corrosion is due to a reaction between the metal and oxygen (oxidation); however, other chemicals may also contribute. Corrosion products, usually oxides, have shear strengths different from those of the wearing surface metals from which they were formed. The oxides tend to flake away, resulting in the pitting of' wearing surfaces. Ball and roller bearings depend on extremely smooth surfaces to reduce frictional effects. Corrosive pitting is especially detrimental to these bearings.

(b) American National Standards Institute (ANSI) Standard ANSI/AGMA 1010-E95 provides numerous illustrations of wear in gears and includes detailed discussions of the types of wear mentioned above and more. Electric Power Research Institute (EPRI) Report EPRI GS-7352 provides illustrations of bearing failures.

(c) Normal wear is inevitable whenever there is relative motion between surfaces. However, wear can be reduced by appropriate machinery design, precision machining, material selection, and proper maintenance, including lubrication. The remainder of this manual is devoted to discussions on the fundamental principles of lubrication that are necessary to reduce wear.

2-3. Lubrication and Lubricants

a. Purpose of lubrication. The primary purpose of lubrication is to reduce wear and heat between contacting surfaces in relative motion. While wear and heat cannot be completely eliminated, they can be reduced to negligible or acceptable levels. Because heat and wear are associated with friction, both effects can be minimized by reducing the coefficient of friction between the contacting surfaces. Lubrication is also used to reduce oxidation and prevent rust; to provide insulation in transformer applications; to transmit mechanical power in hydraulic fluid power applications; and to seal against dust, dirt, and water.

b. Lubricants. Reduced wear and heat are achieved by inserting a lower-viscosity (shear strength) material between wearing surfaces that have a relatively high coefficient of friction. In effect, the wearing surfaces are replaced by a material with a more desirable coefficient of friction. Any material used to reduce friction in this way is a lubricant. Lubricants are available in liquid, solid, and gaseous forms. Industrial machinery ordinarily uses oil or grease. Solid lubricants such as molybdenum disulfide or graphite are used when the loading at contact points is heavy. In some applications the wearing surfaces of a material are plated with a different metal to reduce friction.

2-4. Hydrodynamic or Fluid Film Lubrication

a. General. In heavily loaded bearings such as thrust bearings and horizontal journal bearings, the fluid's viscosity alone is not sufficient to maintain a film between the moving surfaces. In these bearings higher fluid pressures are required to support the load until the fluid film is established. If this pressure is supplied by an outside source, it is called hydrostatic lubrication. If the pressure is generated internally, that is, within the bearing by dynamic action, it is referred to as hydrodynamic lubrication. In hydrodynamic lubrication, a fluid wedge is formed by the relative surface motion of the journals or the

thrust runners over their respective bearing surfaces. The guide bearings of a vertical hydroelectric generator, if properly aligned, have little or no loading and will tend to operate in the center of the bearing because of the viscosity of the oil.

    b. Thrust bearings.

    (1) In hydrodynamic lubrication, sometimes referred to as fluid film lubrication, the wearing surfaces are completely separated by a film of oil. This type of lubricating action is similar to a speedboat operating on water. When the boat is not moving, it rests on the supporting water surface. As the boat begins to move, it meets a certain amount of resistance or opposing force due to viscosity of the water. This causes the leading edge of the boat to lift slightly and allows a small amount of water to come between it and supporting water surface. As the boat's velocity increases, the wedge-shaped water film increases in thickness until a constant velocity is attained. When the velocity is constant, water entering under the leading edge equals the amount passing outward from the trailing edge. For the boat to remain above the supporting surface there must be an upward pressure that equals the load.

    (2) The same principle can be applied to a sliding surface. Fluid film lubrication reduces friction between moving surfaces by substituting fluid friction for mechanical friction. To visualize the shearing effect taking place in the fluid film, imagine the film is composed of many layers similar to a deck of cards. The fluid layer in contact with the moving surface clings to that surface and both move at the same velocity. Similarly, the fluid layer in contact with the other surface is stationary. The layers in between move at velocities directly proportional to their distance from the moving surface. For example, at a distance of ½ h from Surface 1, the velocity would be ½ V. The force F required to move Surface 1 across Surface 2 is simply the force required to overcome the friction between the layers of fluid. This internal friction, or resistance to flow, is defined as the viscosity of the fluid. Viscosity will be discussed in more detail later.

    (3) The principle of hydrodynamic lubrication can also be applied to a more practical example related to thrust bearings used in the hydropower industry. Thrust bearing assembly is also known as tilting pad bearings. These bearings are designed to allow the pads to lift and tilt properly and provide sufficient area to lift the load of the generator. As the thrust runner moves over the thrust shoe, fluid adhering to the runner is drawn between the runner and the shoe causing the shoe to pivot, and forming a wedge of oil. As the speed of the runner increases, the pressure of the oil wedge increases and the runner is lifted as full fluid film lubrication takes place. In applications where the loads are very high, some thrust bearings have high pressure-pumps to provide the initial oil film. Once the unit reaches 100 percent speed, the pump is switched off.

    c. Journal bearings. Although not as obvious as the plate or thrust bearing examples above, the operation of journal or sleeve bearings is also an example of hydrodynamic lubrication. When the journal is at rest, the weight of the journal squeezes out the oil film so that the journal rests on the bearing surface. As rotation starts, the journal has a tendency to roll up the side of the bearing. At the same time fluid adhering to the journal is drawn into the contact area. As the journal speed increases an oil wedge is formed. The pressure of the oil wedge increases until the journal is lifted off the bearing. The journal is not only lifted vertically, but is also pushed to the side by the pressure of the oil wedge. The minimum fluid film thickness at full speed will occur at a point just to the left of center and not at the bottom of the bearing. In both the pivoting shoe thrust bearing and the horizontal journal bearing, the minimum thickness of the fluid film increases with an increase in fluid viscosity and surface speed and decreases with an increase in load.

d. Film thickness. The preceding discussion is a very simplified attempt to provide a basic description of the principles involved in hydrodynamic lubrication. For a more precise, rigorous interpretation refer to American Society for Metals Handbook Volume 18, listed in the Appendix A. Simplified equations have been developed to provide approximations of film thickness with a considerable degree of precision. Regardless of how film thickness is calculated, it is a function of viscosity, velocity, and load. As viscosity or velocity increases, the film thickness increases. When these two variables decrease, the film thickness also decreases. Film thickness varies inversely with the load; as the load increases, film thickness decreases. Viscosity, velocity, and operating temperature are also interrelated. If the oil viscosity is increased the operating temperature will increase, and this in turn has a tendency to reduce viscosity. Thus, an increase in viscosity tends to neutralize itself somewhat. Velocity increases also cause temperature increases that subsequently result in viscosity reduction.

e. Factors influencing film formation. The following factors are essential to achieve and maintain the fluid film required for hydrodynamic lubrication:

! The contact surfaces must meet at a slight angle to allow formation of the lubricant wedge.

! The fluid viscosity must be high enough to support the load and maintain adequate film thickness to separate the contacting surfaces at operating speeds.

! The fluid must adhere to the contact surfaces for conveyance into the pressure area to support the load.

! The fluid must distribute itself completely within the bearing clearance area.

! The operating speed must be sufficient to allow formation and maintenance of the fluid film.

! The contact surfaces of bearings and journals must be smooth and free of sharp surfaces that will disrupt the fluid film.

Theoretically, hydrodynamic lubrication reduces wear to zero. In reality, the journal tends to move vertically and horizontally due to load changes or other disturbances and some wear does occur. However, hydrodynamic lubrication reduces sliding friction and wear to acceptable levels.

2-5. Boundary Lubrication

a. Definition of boundary lubrication. When a complete fluid film does not develop between potentially rubbing surfaces, the film thickness may be reduced to permit momentary dry contact between wear surface high points or asperities. This condition is characteristic of boundary lubrication. Boundary lubrication occurs whenever any of the essential factors that influence formation of a full fluid film are missing. The most common example of boundary lubrication includes bearings, which normally operate with fluid film lubrication but experience boundary lubricating conditions during routine starting and stopping of equipment. Other examples include gear tooth contacts and reciprocating equipment.

b. Oiliness.

(1) Lubricants required to operate under boundary lubrication conditions must possess an added quality referred to as "oiliness" or "lubricity" to lower the coefficient of friction of the oil between the rubbing surfaces. Oiliness is an oil enhancement property provided through the use of chemical additives

known as antiwear (AW) agents. AW agents have a polarizing property that enables them to behave in a manner similar to a magnet. Like a magnet, the opposite sides of the oil film have different polarities. When an AW oil adheres to the metal wear surfaces, the sides of the oil film not in contact with the metal surface have identical polarities and tend to repel each other and form a plane of slippage. Most oils intended for use in heavier machine applications contain AW agents.

(2) Examples of equipment that rely exclusively on boundary lubrication include reciprocating equipment such as engine and compressor pistons, and slow-moving equipment such as turbine wicket gates. Gear teeth also rely on boundary lubrication to a great extent.

2-6. Extreme Pressure (EP) Lubrication

a.   Definition.   AW agents are effective only up to a maximum temperature of about 250 °C (480 °F). Unusually heavy loading will cause the oil temperature to increase beyond the effective range of the antiwear protection. When the load limit is exceeded, the pressure becomes too great and asperities make contact with greater force. Instead of sliding, asperities along the wear surfaces experience shearing, removing the lubricant and the oxide coating. Under these conditions the coefficient of friction is greatly increased and the temperature rises to a damaging level.

b.   Extreme pressure additives.   Applications under extreme pressure conditions rely on additives. Lubricants containing additives that protect against extreme pressure are called EP lubricants, and oils containing additives to protect against extreme pressure are classified as EP oils. EP lubrication is provided by a number of chemical compounds. The most common are compounds of boron, phosphorus, sulfur, chlorine, or combinations of these. The compounds are activated by the higher temperature resulting from extreme pressure, not by the pressure itself. As the temperature rises, EP molecules become reactive and release derivatives of phosphorus, chlorine, or sulfur (depending on which compound is used) to react with only the exposed metal surfaces to form a new compound such as iron chloride or iron sulfide. The new compound forms a solid protective coating that fills the asperities on the exposed metal. Thus, the protection is deposited at exactly the sites where it is needed. AW agents in the EP oil continue to provide antiwear protection at sites where wear and temperature are not high enough to activate the EP agents.

2-7. Elastohydrodynamic (EHD) Lubrication

a.   Definition of EHD lubrication.   The lubrication principles applied to rolling bodies, such as ball or roller bearings, is known as elastohydrodynamic (EHD) lubrication.

b.   Rolling body lubrication.   Although lubrication of rolling objects operates on a considerably different principle than sliding objects, the principles of hydrodynamic lubrication can be applied, within limits, to explain lubrication of rolling elements. An oil wedge, similar to that which occurs in hydrodynamic lubrication, exists at the lower leading edge of the bearing. Adhesion of oil to the sliding element and the supporting surface increases pressure and creates a film between the two bodies. Because the area of contact is extremely small in a roller and ball bearing, the force per unit area, or load pressure, is extremely high. Roller bearing load pressures may reach 34,450 kPa (5000 lb/sq in) and ball bearing load pressures may reach 689,000 kPa (1,000,000 lb/sq in). Under these pressures, it would appear that the oil would be entirely squeezed from between the wearing surfaces. However, viscosity increases that occur under extremely high pressure prevent the oil from being entirely squeezed out. Consequently, a thin film of oil is maintained.

    c.   Effect of film thickness and roughness.

    (1) The roughness of the wearing surfaces is an important consideration in EHD lubrication. Roughness is defined as the arithmetic average of the distance between the high and low points of a surface, and is sometimes called the centerline average (CLA).

    (2) As film thickness increases in relation to roughness fewer asperities make contact. Engineers use the ratio of film thickness to surface roughness to estimate the life expectancy of a bearing system. The relation of bearing life to this ratio is very complex and not always predictable. In general, life expectancy is extended as the ratio increases. Full film thickness is considered to exist when the value of this ratio is between 2 and 4. When this condition prevails, fatigue failure is due entirely to subsurface stress. However, in most industrial applications, a ratio between 1 and 2 is achieved. At these values surface stresses occur, and asperities undergo stress and contribute to fatigue as a major source of failure in antifriction bearings.