Kreith, F.; Berger, S.A.; et. al. "Fluid Mechanics"
Mechanical Engineering Handbook
Ed. Frank Kreith
Boca Raton: CRC Press LLC, 1999

Fluid Mechanics*

Frank Kreith
University of Colorado
Stanley A. Berger
University of California, Berkeley
Stuart W. Churchill
University of Pennsylvania
J. Paul Tullis
Utah State University
Frank M. White
University of Rhode Island
Alan T. McDonald
Purdue University
Ajay Kumar
NASA Langley Research Center
John C. Chen
Lehigh University
Thomas F. Irvine, Jr.
State University of New York, Stony Brook
Massimo Capobianchi
State University of New York, Stony Brook
Francis E. Kennedy
Dartmouth College
E. Richard Booser
Consultant, Scotia, NY
Donald F. Wilcock
Tribolock, Inc.
Robert F. Boehm
University of Nevada-Las Vegas
Rolf D. Reitz
University of Wisconsin
Sherif A. Sherif
University of Florida
Bharat Bhushan
The Ohio State University

* Nomenclature for Section 3 appears at end of chapter.

3.1 Fluid Statics

Stanley A. Berger

Equilibrium of a Fluid Element

If the sum of the external forces acting on a fluid element is zero, the fluid will be either at rest or moving as a solid body — in either case, we say the fluid element is in equilibrium. In this section we consider fluids in such an equilibrium state. For fluids in equilibrium the only internal stresses acting will be normal forces, since the shear stresses depend on velocity gradients, and all such gradients, by the definition of equilibrium, are zero. If one then carries out a balance between the normal surface stresses and the body forces, assumed proportional to volume or mass, such as gravity, acting on an elementary prismatic fluid volume, the resulting equilibrium equations, after shrinking the volume to zero, show that the normal stresses at a point are the same in all directions, and since they are known to be negative, this common value is denoted by $-p$, p being the pressure.

Hydrostatic Pressure

If we carry out an equilibrium of forces on an elementary volume element $dxdydz$, the forces being pressures acting on the faces of the element and gravity acting in the $-z$ direction, we obtain

$$\frac{\partial p}{\partial x} = \frac{\partial p}{\partial y} = 0, \quad \text{and} \quad \frac{\partial p}{\partial z} = -\rho g = -\gamma \tag{3.1.1}$$

The first two of these imply that the pressure is the same in all directions at the same vertical height in a gravitational field. The third, where γ is the specific weight, shows that the pressure increases with depth in a gravitational field, the variation depending on $\rho(z)$. For homogeneous fluids, for which ρ = constant, this last equation can be integrated immediately, yielding

$$p_2 - p_1 = -\rho g\left(z_2 - z_1\right) = -\rho g\left(h_2 - h_1\right) \tag{3.1.2}$$

or

$$p_2 + \rho g h_2 = p_1 + \rho g h_1 = \text{constant} \tag{3.1.3}$$

where h denotes the elevation. These are the equations for the hydrostatic pressure distribution.

When applied to problems where a liquid, such as the ocean, lies below the atmosphere, with a constant pressure p_{atm}, h is usually measured from the ocean/atmosphere interface and p at any distance h below this interface differs from p_{atm} by an amount

$$p - p_{atm} = \rho g h \tag{3.1.4}$$

Pressures may be given either as *absolute pressure,* pressure measured relative to absolute vacuum, or *gauge pressure,* pressure measured relative to atmospheric pressure.

Manometry

The hydrostatic pressure variation may be employed to measure pressure differences in terms of heights of liquid columns — such devices are called manometers and are commonly used in wind tunnels and a host of other applications and devices. Consider, for example the U-tube manometer shown in Figure 3.1.1 filled with liquid of specific weight γ, the left leg open to the atmosphere and the right to the region whose pressure p is to be determined. In terms of the quantities shown in the figure, in the left leg

$$p_0 - \rho g h_2 = p_{atm} \tag{3.1.5a}$$

and in the right leg

$$p_0 - \rho g h_1 = p \tag{3.1.5b}$$

the difference being

$$p - p_{atm} = -\rho g\left(h_1 - h_2\right) = -\rho g d = -\gamma d \tag{3.1.6}$$

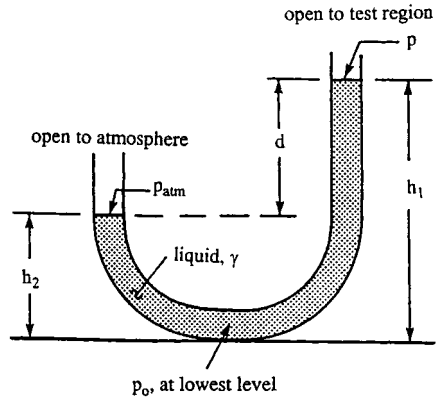


FIGURE 3.1.1 U-tube manometer.

and determining p in terms of the height difference $d = h_1 - h_2$ between the levels of the fluid in the two legs of the manometer.

Hydrostatic Forces on Submerged Objects

We now consider the force acting on a submerged object due to the hydrostatic pressure. This is given by

$$\boldsymbol{F} = \iint p\, d\boldsymbol{A} = \iint p \cdot \boldsymbol{n}\, dA = \iint \rho g h\, d\boldsymbol{A} = p_0 \iint d\boldsymbol{A} \tag{3.1.7}$$

where h is the variable vertical depth of the element $d\boldsymbol{A}$ and p_0 is the pressure at the surface. In turn we consider plane and nonplanar surfaces.

Forces on Plane Surfaces

Consider the planar surface A at an angle θ to a free surface shown in Figure 3.1.2. The force on one side of the planar surface, from Equation (3.1.7), is

$$\boldsymbol{F} = \rho g \boldsymbol{n} \iint_A h\, dA + p_0 A \boldsymbol{n} \tag{3.1.8}$$

but $h = y \sin\theta$, so

$$\iint_A h\, dA = \sin\theta \iint_A y\, dA = y_c A \sin\theta = h_c A \tag{3.1.9}$$

where the subscript c indicates the distance measured to the centroid of the area A. Thus, the total force (on one side) is

$$\boldsymbol{F} = \gamma h_c \boldsymbol{A} + p_0 \boldsymbol{A} \tag{3.1.10}$$

Thus, the magnitude of the force is independent of the angle θ, and is equal to the pressure at the centroid, $\gamma h_c + p_0$, times the area. If we use gauge pressure, the term $p_0 A$ in Equation (3.1.10) can be dropped.

Since p is not evenly distributed over A, but varies with depth, $\boldsymbol{F}$ does not act through the centroid. The point action of $\boldsymbol{F}$, called the *center of pressure*, can be determined by considering moments in Figure 3.1.2. The moment of the hydrostatic force acting on the elementary area dA about the axis perpendicular to the page passing through the point 0 on the free surface is

$$y\, dF = y\left(\gamma\, y \sin\theta\, dA\right) = \gamma\, y^2 \sin\theta\, dA \tag{3.1.11}$$

so if y_{cp} denotes the distance to the center of pressure,

$$y_{cp} F = \gamma \sin\theta \iint y^2\, dA = \gamma \sin\theta\, I_x \tag{3.1.12}$$

where I_x is the moment of inertia of the plane area with respect to the axis formed by the intersection of the plane containing the planar surface and the free surface (say $0x$). Dividing by $F = \gamma h_c A = \gamma\, y_c \sin\theta A$ gives

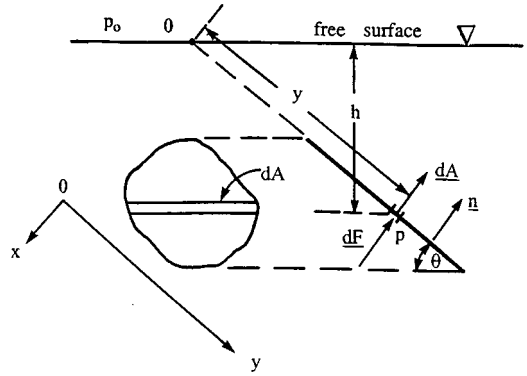


FIGURE 3.1.2 Hydrostatic force on a plane surface.

$$y_{cp} = \frac{I_x}{y_c A} \tag{3.1.13}$$

By using the parallel axis theorem $I_x = I_{xc} + Ay_c^2$, where I_{xc} is the moment of inertia with respect to an axis parallel to $0x$ passing through the centroid, Equation (3.1.13) becomes

$$y_{cp} = y_c + \frac{I_{xc}}{y_c A} \tag{3.1.14}$$

which shows that, in general, the center of pressure lies below the centroid.

Similarly, we find x_{cp} by taking moments about the y axis, specifically

$$x_{cp} F = \gamma \sin\theta \iint xy\, dA = \gamma \sin\theta\, I_{xy} \tag{3.1.15}$$

or

$$x_{cp} = \frac{I_{xy}}{y_c A} \tag{3.1.16}$$

where I_{xy} is the product of inertia with respect to the x and y axes. Again, the parallel axis theorem $I_{xy} = I_{xyc} + Ax_c y_c$, where the subscript c denotes the value at the centroid, allows Equation (3.1.16) to be written

$$x_{cp} = x_c + \frac{I_{xyc}}{y_c A} \tag{3.1.17}$$

This completes the determination of the center of pressure (x_{cp}, y_{cp}). Note that if the submerged area is symmetrical with respect to an axis passing through the centroid and parallel to either the x or y axes that $I_{xyc} = 0$ and $x_{cp} = x_c$; also that as y_c increases, $y_{cp} \to y_c$.

Centroidal moments of inertia and centroidal coordinates for some common areas are shown in Figure 3.1.3.

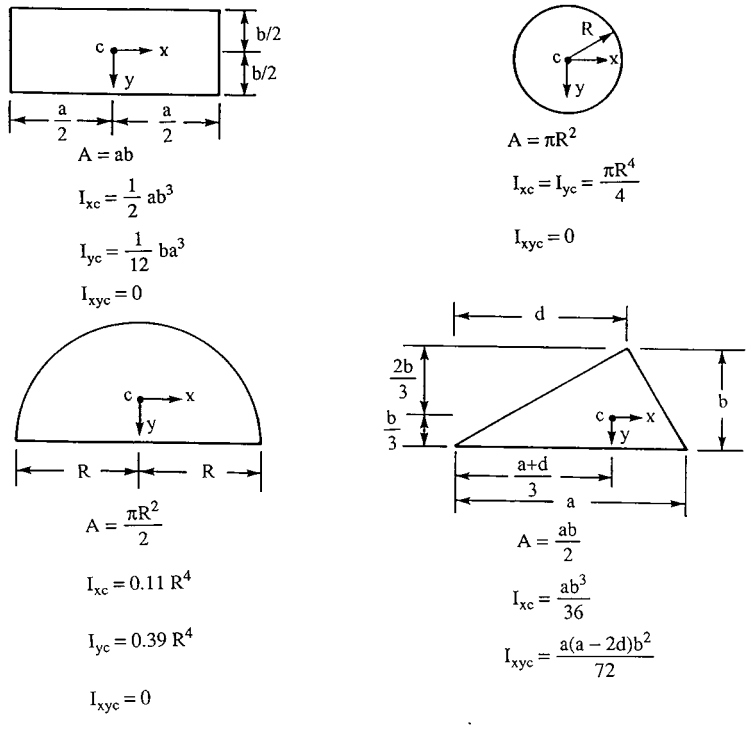


FIGURE 3.1.3 Centroidal moments of inertia and coordinates for some common areas.

Forces on Curved Surfaces

On a curved surface the forces on individual elements of area differ in direction so a simple summation of them is not generally possible, and the most convenient approach to calculating the pressure force on the surface is by separating it into its horizontal and vertical components.

A free-body diagram of the forces acting on the volume of fluid lying above a curved surface together with the conditions of static equilibrium of such a column leads to the results that:

1. The horizontal components of force on a curved submerged surface are equal to the forces exerted on the planar areas formed by the projections of the curved surface onto vertical planes normal to these components, the lines of action of these forces calculated as described earlier for planar surfaces; and
2. The vertical component of force on a curved submerged surface is equal in magnitude to the weight of the entire column of fluid lying above the curved surface, and acts through the center of mass of this volume of fluid.

Since the three components of force, two horizontal and one vertical, calculated as above, need not meet at a single point, there is, in general, no single resultant force. They can, however, be combined into a single force at any arbitrary point of application together with a moment about that point.

Hydrostatic Forces in Layered Fluids

All of the above results which employ the linear hydrostatic variation of pressure are valid only for homogeneous fluids. If the fluid is heterogeneous, consisting of individual layers each of constant density, then the pressure varies linearly with a different slope in each layer and the preceding analyses must be remedied by computing and summing the separate contributions to the forces and moments.

Buoyancy

The same principles used above to compute hydrostatic forces can be used to calculate the net pressure force acting on completely submerged or floating bodies. These laws of buoyancy, the principles of Archimedes, are that:

1. A completely submerged body experiences a vertical upward force equal to the weight of the displaced fluid; and
2. A floating or partially submerged body displaces its own weight in the fluid in which it floats (i.e., the vertical upward force is equal to the body weight).

The line of action of the buoyancy force in both (1) and (2) passes through the centroid of the displaced volume of fluid; this point is called the *center of buoyancy.* (This point need not correspond to the center of mass of the body, which could have nonuniform density. In the above it has been assumed that the displaced fluid has a constant γ. If this is not the case, such as in a layered fluid, the magnitude of the buoyant force is still equal to the weight of the displaced fluid, but the line of action of this force passes through the center of gravity of the displaced volume, not the centroid.)

If a body has a weight exactly equal to that of the volume of fluid it displaces, it is said to be *neutrally buoyant* and will remain at rest at any point where it is immersed in a (homogeneous) fluid.

Stability of Submerged and Floating Bodies

Submerged Body

A body is said to be in stable equilibrium if, when given a slight displacement from the equilibrium position, the forces thereby created tend to restore it back to its original position. The forces acting on a submerged body are the buoyancy force, F_B, acting through the center of buoyancy, denoted by CB, and the weight of the body, W, acting through the center of gravity denoted by CG (see Figure 3.1.4). We see from Figure 3.1.4 that if the CB lies above the CG a rotation from the equilibrium position creates a restoring couple which will rotate the body back to its original position — thus, this is a *stable* equilibrium situation. The reader will readily verify that when the CB lies below the CG, the couple that results from a rotation from the vertical increases the displacement from the equilibrium position — thus, this is an *unstable* equilibrium situation.

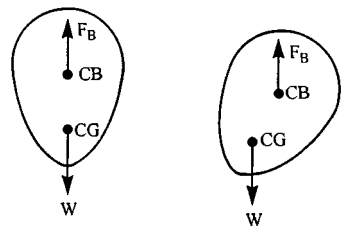


FIGURE 3.1.4 Stability for a submerged body.

Partially Submerged Body

The stability problem is more complicated for floating bodies because as the body rotates the location of the center of buoyancy may change. To determine stability in these problems requires that we determine the location of the *metacenter.* This is done for a symmetric body by tilting the body through a small angle $\Delta\theta$ from its equilibrium position and calculating the new location of the center of buoyancy CB′; the point of intersection of a vertical line drawn upward from CB′ with the line of symmetry of the floating body is the metacenter, denoted by M in Figure 3.1.5, and it is independent of $\Delta\theta$ for small angles. If M lies above the CG of the body, we see from Figure 3.1.5 that rotation of the body leads to

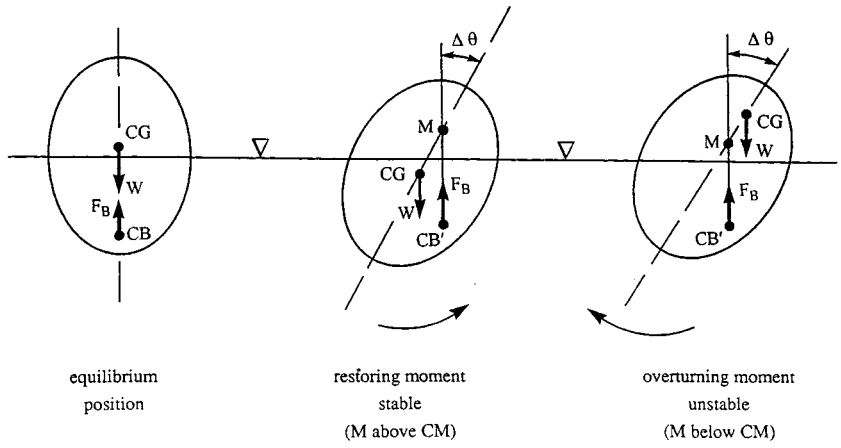


FIGURE 3.1.5 Stability for a partially submerged body.

a restoring couple, whereas *M* lying below the CG leads to a couple which will increase the displacement. Thus, the stability of the equilibrium depends on whether *M* lies above or below the CG. The directed distance from CG to *M* is called the *metacentric height*, so equivalently the equilibrium is stable if this vector is positive and unstable if it is negative; stability increases as the metacentric height increases. For geometrically complex bodies, such as ships, the computation of the metacenter can be quite complicated.

Pressure Variation in Rigid-Body Motion of a Fluid

In rigid-body motion of a fluid all the particles translate and rotate as a whole, there is no relative motion between particles, and hence no viscous stresses since these are proportional to velocity gradients. The equation of motion is then a balance among pressure, gravity, and the fluid acceleration, specifically.

$$\nabla p = \rho(\boldsymbol{g} - \boldsymbol{a}) \tag{3.1.18}$$

where $\boldsymbol{a}$ is the uniform acceleration of the body. Equation (3.1.18) shows that the lines of constant pressure, including a free surface if any, are perpendicular to the direction $\boldsymbol{g} - \boldsymbol{a}$. Two important applications of this are to a fluid in uniform linear translation and rigid-body rotation. While such problems are not, strictly speaking, fluid statics problems, their analysis and the resulting pressure variation results are similar to those for static fluids.

Uniform Linear Acceleration

For a fluid partially filling a large container moving to the right with constant acceleration $\boldsymbol{a} = (a_x, a_y)$ the geometry of Figure 3.1.6 shows that the magnitude of the pressure gradient in the direction $\boldsymbol{n}$ normal to the accelerating free surface, in the direction $\boldsymbol{g} - \boldsymbol{a}$, is

$$\frac{dp}{dn} = \rho\left[a_x^2 + \left(g + a_y\right)^2\right]^{1/2} \tag{3.1.19}$$

and the free surface is oriented at an angle to the horizontal

$$\theta = \tan^{-1}\left(\frac{a_x}{g + a_y}\right) \tag{3.1.20}$$

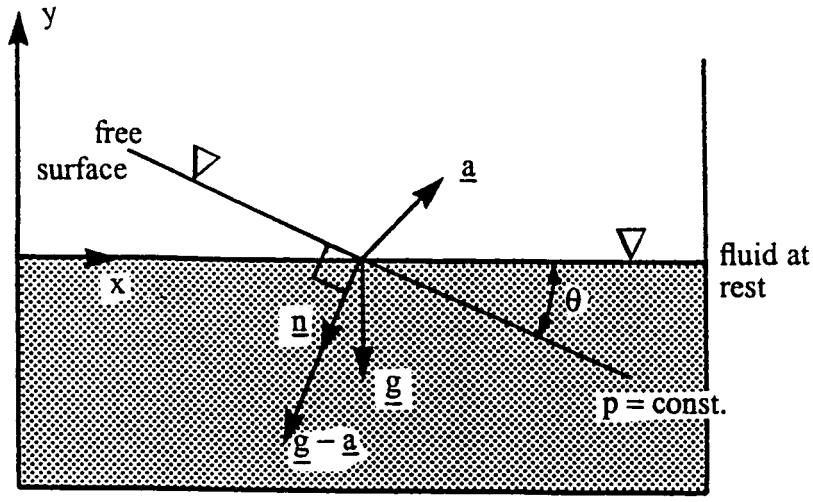


FIGURE 3.1.6 A fluid with a free surface in uniform linear acceleration.

Rigid-Body Rotation

Consider the fluid-filled circular cylinder rotating uniformly with angular velocity $\Omega = \Omega \boldsymbol{e}_r$ (Figure 3.1.7). The only acceleration is the centripetal acceleration $\Omega \times \Omega \times \boldsymbol{r}) = -r\Omega^2 \boldsymbol{e}_r$, so Equation 3.1.18 becomes

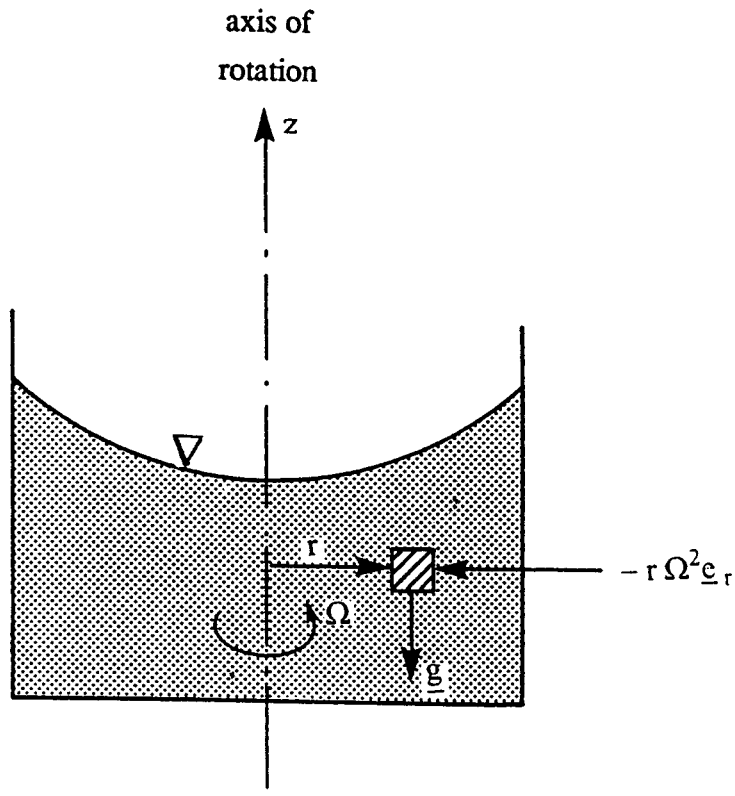


FIGURE 3.1.7 A fluid with a free surface in rigid-body rotation.

$$\nabla p = \frac{\partial p}{\partial r}\boldsymbol{e}_r + \frac{\partial p}{\partial z}\boldsymbol{e}_z = \rho(\boldsymbol{g} - \boldsymbol{a}) = \rho\left(r\Omega^2\boldsymbol{e}_r - g\boldsymbol{e}_z\right) \tag{3.1.21}$$

or

$$\frac{\partial p}{\partial r} = \rho r\Omega^2, \qquad \frac{\partial p}{\partial z} = -\rho g = -\gamma \tag{3.1.22}$$

Integration of these equations leads to

$$p = p_o - \gamma z + \frac{1}{2}\rho r^2\Omega^2 \tag{3.1.23}$$

where p_o is some reference pressure. This result shows that at any fixed r the pressure varies hydrostatically in the vertical direction, while the constant pressure surfaces, including the free surface, are paraboloids of revolution.

Further Information

The reader may find more detail and additional information on the topics in this section in any one of the many excellent introductory texts on fluid mechanics, such as

White, F.M. 1994. *Fluid Mechanics,* 3rd ed., McGraw-Hill, New York.
Munson, B.R., Young, D.F., and Okiishi, T.H. 1994. *Fundamentals of Fluid Mechanics,* 2nd ed., John Wiley & Sons, New York.

3.2 Equations of Motion and Potential Flow

Stanley A. Berger

Integral Relations for a Control Volume

Like most physical conservation laws those governing motion of a fluid apply to material particles or systems of such particles. This so-called Lagrangian viewpoint is generally not as useful in practical fluid flows as an analysis through fixed (or deformable) control volumes — the Eulerian viewpoint. The relationship between these two viewpoints can be deduced from the Reynolds transport theorem, from which we also most readily derive the governing integral and differential equations of motion.

Reynolds Transport Theorem

The *extensive* quantity B, a scalar, vector, or tensor, is defined as any property of the fluid (e.g., momentum, energy) and b as the corresponding value per unit mass (the *intensive* value). The Reynolds transport theorem for a moving and arbitrarily deforming control volume CV, with boundary CS (see Figure 3.2.1), states that

$$\frac{d}{dt}\left(B_{\text{system}}\right) = \frac{d}{dt}\left(\iiint_{\text{CV}} \rho b \, d\upsilon\right) + \iint_{\text{CS}} \rho b\left(\boldsymbol{V}_r \cdot \boldsymbol{n}\right) dA \tag{3.2.1}$$

where B_{system} is the total quantity of B in the system (any mass of fixed identity), $\boldsymbol{n}$ is the outward normal to the CS, $\boldsymbol{V}_r = \boldsymbol{V}(\boldsymbol{r}, t) - \boldsymbol{V}_{\text{CS}}(\boldsymbol{r}, t)$, the velocity of the fluid particle, $\boldsymbol{V}(\boldsymbol{r}, t)$, relative to that of the CS, $\boldsymbol{V}_{\text{CS}}(\boldsymbol{r}, t)$, and d/dt on the left-hand side is the derivative following the fluid particles, i.e., the fluid mass comprising the system. The theorem states that the time rate of change of the total B in the system is equal to the rate of change within the CV plus the net flux of B through the CS. To distinguish between the d/dt which appears on the two sides of Equation (3.2.1) but which have different interpretations, the derivative on the left-hand side, following the system, is denoted by D/Dt and is called the material derivative. This notation is used in what follows. For any function $f(x, y, z, t)$,

$$\frac{Df}{Dt} = \frac{\partial f}{\partial t} + \boldsymbol{V} \cdot \nabla f$$

For a CV fixed with respect to the reference frame, Equation (3.2.1) reduces to

$$\frac{D}{Dt}\left(B_{\text{system}}\right) = \frac{d}{dt}\iiint_{\substack{\text{CV}\\(\text{fixed})}} \left(\rho b\right) d\upsilon + \iint_{\text{CS}} \rho b\left(\boldsymbol{V} \cdot \boldsymbol{n}\right) dA \tag{3.2.2}$$

(The time derivative operator in the first term on the right-hand side may be moved inside the integral, in which case it is then to be interpreted as the partial derivative $\partial/\partial t$.)

Conservation of Mass

If we apply Equation (3.2.2) for a fixed control volume, with B_{system} the total mass in the system, then since conservation of mass requires that $DB_{\text{system}}/Dt = 0$ there follows, since $b = B_{\text{system}}/m = 1$,

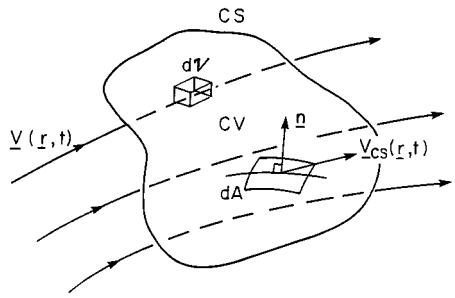


FIGURE 3.2.1 Control volume.

$$\iiint_{\substack{CV\\(fixed)}} \frac{\partial \rho}{\partial t}\, d\upsilon + \iint_{CS} \rho(\boldsymbol{V}\cdot\boldsymbol{n})\, dA = 0 \tag{3.2.3}$$

This is the integral form of the conservation of mass law for a fixed control volume. For a steady flow, Equation (3.2.3) reduces to

$$\iint_{CS} \rho(\boldsymbol{V}\cdot\boldsymbol{n})\, dA = 0 \tag{3.2.4}$$

whether compressible or incompressible. For an incompressible flow, ρ = constant, so

$$\iint_{CS} (\boldsymbol{V}\cdot\boldsymbol{n})\, dA = 0 \tag{3.2.5}$$

whether the flow is steady or unsteady.

Conservation of Momentum

The conservation of (linear) momentum states that

$$F_{\text{total}} \equiv \sum (\text{external forces acting on the fluid system}) = \frac{D\boldsymbol{M}}{Dt} \equiv \frac{D}{Dt}\left(\iiint_{\text{system}} \rho \boldsymbol{V}\, d\upsilon\right) \tag{3.2.6}$$

where $\boldsymbol{M}$ is the total system momentum. For an arbitrarily moving, deformable control volume it then follows from Equation (3.2.1) with b set to $\boldsymbol{V}$,

$$\boldsymbol{F}_{\text{total}} = \frac{d}{dt}\left(\iiint_{CV} \rho \boldsymbol{V}\, d\upsilon\right) + \iint_{CS} \rho \boldsymbol{V}(\boldsymbol{V}_r\cdot\boldsymbol{n})\, dA \tag{3.2.7}$$

This expression is only valid in an inertial coordinate frame. To write the equivalent expression for a noninertial frame we must use the relationship between the acceleration $\boldsymbol{a}_I$ in an inertial frame and the acceleration $\boldsymbol{a}_R$ in a noninertial frame,

$$\boldsymbol{a}_I = \boldsymbol{a}_R + \frac{d^2\boldsymbol{R}}{dt^2} + 2\Omega \times \boldsymbol{V} + \Omega \times (\Omega \times \boldsymbol{r}) + \frac{d\Omega}{dt} \times \boldsymbol{r} \tag{3.2.8}$$

where $\boldsymbol{R}$ is the position vector of the origin of the noninertial frame with respect to that of the inertial frame, Ω is the angular velocity of the noninertial frame, and $\boldsymbol{r}$ and $\boldsymbol{V}$ the position and velocity vectors in the noninertial frame. The third term on the right-hand side of Equation (3.2.8) is the Coriolis acceleration, and the fourth term is the centrifugal acceleration. For a noninertial frame Equation (3.2.7) is then

$$\begin{aligned} \boldsymbol{F}_{\text{total}} - \iiint_{\text{system}} \left[\frac{d^2\boldsymbol{R}}{dt^2} + 2\Omega \times \boldsymbol{V} + \Omega \times (\Omega \times \boldsymbol{r}) + \frac{d\Omega}{dt} \times \boldsymbol{r} \right] \rho \, d\upsilon &= \frac{D}{Dt}\left(\iiint_{\text{system}} \rho \boldsymbol{V} \, d\upsilon \right) \\ = \frac{d}{dt}\left(\iiint_{\text{CV}} \rho \boldsymbol{V} \, d\upsilon \right) + \iint_{\text{CS}} \rho \boldsymbol{V} \cdot (\boldsymbol{V}_r \cdot \boldsymbol{n}) \, dA & \end{aligned} \tag{3.2.9}$$

where the frame acceleration terms of Equation (3.2.8) have been brought to the left-hand side because to an observer in the noninertial frame they act as "apparent" body forces.

For a fixed control volume in an inertial frame for steady flow it follows from the above that

$$\boldsymbol{F}_{\text{total}} = \iint_{\text{CS}} \rho V (\boldsymbol{V} \cdot \boldsymbol{n}) \, dA \tag{3.2.10}$$

This expression is the basis of many control volume analyses for fluid flow problems.

The cross product of $\boldsymbol{r}$, the position vector with respect to a convenient origin, with the momentum Equation (3.2.6) written for an elementary particle of mass dm, noting that $(d\boldsymbol{r}/dt) \times \boldsymbol{V} = 0$, leads to the integral moment of momentum equation

$$\sum \boldsymbol{M} - \boldsymbol{M}_I = \frac{D}{Dt} \iiint_{\text{system}} \rho (\boldsymbol{r} \times \boldsymbol{V}) \, d\upsilon \tag{3.2.11}$$

where $\Sigma \boldsymbol{M}$ is the sum of the moments of all the external forces acting on the system about the origin of $\boldsymbol{r}$, and $\boldsymbol{M}_I$ is the moment of the apparent body forces (see Equation (3.2.9)). The right-hand side can be written for a control volume using the appropriate form of the Reynolds transport theorem.

Conservation of Energy

The conservation of energy law follows from the first law of thermodynamics for a moving system

$$\dot{Q} - \dot{W} = \frac{D}{Dt}\left(\iiint_{\text{system}} \rho e \, d\upsilon \right) \tag{3.2.12}$$

where $\dot{Q}$ is the rate at which heat is added to the system, $\dot{W}$ the rate at which the system works on its surroundings, and e is the total energy per unit mass. For a particle of mass dm the contributions to the specific energy e are the internal energy u, the kinetic energy $V^2/2$, and the potential energy, which in the case of gravity, the only body force we shall consider, is gz, where z is the vertical displacement opposite to the direction of gravity. (We assume no energy transfer owing to chemical reaction as well

as no magnetic or electric fields.) For a fixed control volume it then follows from Equation (3.2.2) [with $b = e = u + (V^2/2) + gz$] that

$$\dot{Q} - \dot{W} = \frac{d}{dt}\left(\iiint_{CV} \rho\left(u + \frac{1}{2}V^2 + gz\right)\, d\upsilon\right) + \iint_{CS} \rho\left(u + \frac{1}{2}V^2 + gz\right)(\boldsymbol{V}\cdot\boldsymbol{n})\, dA \tag{3.2.13}$$

Problem

An incompressible fluid flows through a pump at a volumetric flow rate $\dot{Q}$. The (head) loss between sections 1 and 2 (see Figure 3.2.2) is equal to $\beta\rho V_1^2/2$ (V is the average velocity at the section). Calculate the power that must be delivered by the pump to the fluid to produce a given increase in pressure, $\Delta p = p_2 - p_1$.

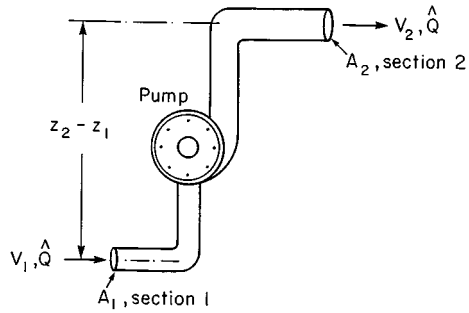


FIGURE 3.2.2 Pump producing pressure increase.

Solution: The principal equation needed is the energy Equation (3.2.13). The term $\dot{W}$, the rate at which the system does work on its surroundings, for such problems has the form

$$\dot{W} = -\dot{W}_{shaft} + \iint_{CS} p\boldsymbol{V}\cdot\boldsymbol{n}\, dA \tag{P.3.2.1}$$

where $\dot{W}_{shaft}$ represents the work done on the fluid by a moving shaft, such as by turbines, propellers, fans, etc., and the second term on the right side represents the rate of working by the normal stress, the pressure, at the boundary. For a steady flow in a control volume coincident with the physical system boundaries and bounded at its ends by sections 1 and 2, Equation (3.2.13) reduces to ($u \equiv 0$),

$$\dot{Q} + \dot{W}_{shaft} - \iint_{CS} p\boldsymbol{V}\cdot\boldsymbol{n}\, dA = \iint_{CS}\left(\frac{1}{2}\rho V^2 + \gamma z\right)(\boldsymbol{V}\cdot\boldsymbol{n})\, dA \tag{P.3.2.2}$$

Using average quantities at sections 1 and 2, and the continuity Equation (3.2.5), which reduces in this case to

$$V_1 A_1 = V_2 A_2 = \dot{Q} \tag{P.3.2.3}$$

We can write Equation (P.3.2.2) as

$$\dot{Q} + \dot{W}_{shaft} - (p_2 - p_1)\dot{Q} = \left[\frac{1}{2}\rho\left(V_2^2 - V_1^2\right) + \gamma(z_2 - z_1)\right]\dot{Q} \tag{P.3.2.4}$$

$\dot{Q}$, the rate at which heat is added to the system, is here equal to $-\beta\rho V_1^2/2$, the head loss between sections 1 and 2. Equation (P.3.2.4) then can be rewritten

$$\dot{W}_{\text{shaft}} = \beta\rho\frac{V_1^2}{2} + (\Delta p)\dot{Q} + \frac{1}{2}\rho\left(V_2^2 - V_1^2\right)\dot{Q} + \gamma\left(z_2 - z_1\right)\dot{Q}$$

or, in terms of the given quantities,

$$\dot{W}_{\text{shaft}} = \frac{\beta\rho\dot{Q}^2}{A_1^2} + (\Delta p)\dot{Q} + \frac{1}{2}\rho\frac{\dot{Q}^3}{A_2^2}\left(1 - \frac{A_2^2}{A_1^2}\right) + \gamma\left(z_2 - z_1\right)\dot{Q} \qquad \text{(P.3.2.5)}$$

Thus, for example, if the fluid is water ($\rho \approx 1000$ kg/m^3, $\gamma = 9.8$ kN/m^3), $\dot{Q} = 0.5$ m^3/sec, the heat loss is $0.2\rho V_1^2/2$, and $\Delta p = p_2 - p_1 = 2 \times 10^5$N/m^2 = 200 kPa, $A_1 = 0.1$ m^2 = $A_2/2$, $(z_2 - z_1) = 2$ m, we find, using Equation (P.3.2.5)

$$\dot{W}_{\text{shaft}} = \frac{0.2(1000)(0.5)^2}{(0.1)^2} + \left(2 \times 10^5\right)(0.5) + \frac{1}{2}(1000)\frac{(0.5)^3}{(0.2)^2}(1 - 4) + \left(9.8 \times 10^3\right)(2)(0.5)$$

$$= 5{,}000 + 10{,}000 - 4{,}688 + 9{,}800 = 20{,}112 \text{ Nm/sec}$$

$$= 20{,}112 \text{ W} = \frac{20{,}112}{745.7}\text{hp} = 27 \text{ hp}$$

Differential Relations for Fluid Motion

In the previous section the conservation laws were derived in integral form. These forms are useful in calculating, generally using a control volume analysis, gross features of a flow. Such analyses usually require some *a priori* knowledge or assumptions about the flow. In any case, an approach based on integral conservation laws cannot be used to determine the point-by-point variation of the dependent variables, such as velocity, pressure, temperature, etc. To do this requires the use of the differential forms of the conservation laws, which are presented below.

Mass Conservation–Continuity Equation

Applying Gauss's theorem (the divergence theorem) to Equation (3.2.3) we obtain

$$\iiint_{\substack{\text{CV}\\(\text{fixed})}}\left[\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho V)\right]d\upsilon = 0 \qquad (3.2.14)$$

which, because the control volume is arbitrary, immediately yields

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{V}) = 0 \qquad (3.2.15)$$

This can also be written as

$$\frac{D\rho}{Dt} + \rho\nabla\cdot\boldsymbol{V} = 0 \qquad (3.2.16)$$

using the fact that

$$\frac{D\rho}{Dt} = \frac{\partial \rho}{\partial t} + \boldsymbol{V} \cdot \nabla \rho \tag{3.2.17}$$

Special cases:

1. Steady flow $[(\partial/\partial t)(\ ) \equiv 0]$

$$\nabla \cdot (\rho \boldsymbol{V}) = 0 \tag{3.2.18}$$

2. Incompressible flow ($D\rho/Dt \equiv 0$)

$$\nabla \cdot \boldsymbol{V} = 0 \tag{3.2.19}$$

Momentum Conservation

We note first, as a consequence of mass conservation for a system, that the right-hand side of Equation (3.2.6) can be written as

$$\frac{D}{Dt}\left(\iiint_{\text{system}} \rho \boldsymbol{V}\, d\upsilon \right) \equiv \iiint_{\text{system}} \rho \frac{D\boldsymbol{V}}{Dt}\, d\upsilon \tag{3.2.20}$$

The total force acting on the system which appears on the left-hand side of Equation (3.2.6) is the sum of body forces $\boldsymbol{F}_b$ and surface forces $\boldsymbol{F}_s$. The body forces are often given as forces per unit mass (e.g., gravity), and so can be written

$$\boldsymbol{F}_b = \iiint_{\text{system}} \rho \boldsymbol{f}\, d\upsilon \tag{3.2.21}$$

The surface forces are represented in terms of the second-order stress tensor* $\underline{\underline{\sigma}} = \{\sigma_{ij}\}$, where σ_{ij} is defined as the force per unit area in the i direction on a planar element whose normal lies in the j direction. From elementary angular momentum considerations for an infinitesimal volume it can be shown that σ_{ij} is a symmetric tensor, and therefore has only six independent components. The total surface force exerted on the system by its surroundings is then

$$\boldsymbol{F}_s = \iint_{\substack{\text{system} \\ \text{surface}}} \underline{\underline{\sigma}} \cdot \boldsymbol{n}\, dA, \quad \text{with } i-\text{component } F_{s_i} = \iint \sigma_{ij} n_j\, dA \tag{3.2.22}$$

The integral momentum conservation law Equation (3.2.6) can then be written

$$\iiint_{\text{system}} \rho \frac{D\boldsymbol{V}}{Dt}\, d\upsilon = \iiint_{\text{system}} \rho \boldsymbol{f}\, d\upsilon + \iint_{\substack{\text{system} \\ \text{surface}}} \underline{\underline{\sigma}} \cdot \boldsymbol{n}\, dA \tag{3.2.23}$$

* We shall assume the reader is familiar with elementary Cartesian tensor analysis and the associated subscript notation and conventions. The reader for whom this is not true should skip the details and concentrate on the final principal results and equations given at the ends of the next few subsections.

The application of the divergence theorem to the last term on the right-side of Equation (3.2.23) leads to

$$\iiint_{\text{system}} \rho \frac{DV}{Dt}\, d\upsilon = \iiint_{\text{system}} \rho \boldsymbol{f}\, d\upsilon + \iiint_{\text{system}} \nabla \cdot \underline{\underline{\sigma}}\, d\upsilon \tag{3.2.24}$$

where $\nabla \cdot \underline{\underline{\sigma}} \equiv \{\partial \sigma_{ij}/x_j\}$. Since Equation (3.2.24) holds for any material volume, it follows that

$$\rho \frac{DV}{Dt} = \rho \boldsymbol{f} + \nabla \cdot \underline{\underline{\sigma}} \tag{3.2.25}$$

(With the decomposition of $\boldsymbol{F}_{\text{total}}$ above, Equation (3.2.10) can be written

$$\iiint_{CV} \rho \boldsymbol{f}\, d\upsilon + \iint_{CS} \underline{\underline{\sigma}} \cdot \boldsymbol{n}\, dA = \iint_{CS} \rho \boldsymbol{V}(\boldsymbol{V} \cdot \boldsymbol{n})\, dA \tag{3.2.26}$$

If ρ is uniform and $\boldsymbol{f}$ is a conservative body force, i.e., $\boldsymbol{f} = -\nabla\Psi$, where Ψ is the force potential, then Equation (3.2.26), after application of the divergence theorem to the body force term, can be written

$$\iint_{CS} \left(-\rho\Psi \boldsymbol{n} + \underline{\underline{\sigma}} \cdot \boldsymbol{n}\right) dA = \iint_{CS} \rho \boldsymbol{V}(\boldsymbol{V} \cdot \boldsymbol{n})\, dA \tag{3.2.27}$$

It is in this form, involving only integrals over the surface of the control volume, that the integral form of the momentum equation is used in control volume analyses, particularly in the case when the body force term is absent.)

Analysis of Rate of Deformation

The principal aim of the following two subsections is to derive a relationship between the stress and the rate of strain to be used in the momentum Equation (3.2.25). The reader less familiar with tensor notation may skip these sections, apart from noting some of the terms and quantities defined therein, and proceed directly to Equations (3.2.38) or (3.2.39).

The relative motion of two neighboring points P and Q, separated by a distance η, can be written (using $\boldsymbol{u}$ for the local velocity)

$$\boldsymbol{u}(Q) = \boldsymbol{u}(P) + (\nabla \boldsymbol{u})\eta$$

or, equivalently, writing $\nabla\boldsymbol{u}$ as the sum of antisymmetric and symmetric tensors,

$$\boldsymbol{u}(Q) = \boldsymbol{u}(P) + \frac{1}{2}\left((\nabla \boldsymbol{u}) - (\nabla \boldsymbol{u})^*\right)\eta + \frac{1}{2}\left((\nabla \boldsymbol{u}) + (\nabla \boldsymbol{u})^*\right)\eta \tag{3.2.28}$$

where $\nabla\boldsymbol{u} = \{\partial u_i/\partial x_j\}$, and the superscript * denotes transpose, so $(\nabla\boldsymbol{u})^* = \{\partial u_j/\partial x_i\}$. The second term on the right-hand side of Equation (3.2.28) can be rewritten in terms of the *vorticity*, $\nabla \times \boldsymbol{u}$, so Equation (3.2.28) becomes

$$\boldsymbol{u}(Q) = \boldsymbol{u}(P) + \frac{1}{2}(\nabla \times \boldsymbol{u}) \times \eta + \frac{1}{2}\left((\nabla \boldsymbol{u}) + (\nabla \boldsymbol{u})^*\right)\eta \tag{3.2.29}$$

which shows that the local rate of deformation consists of a rigid-body translation, a rigid-body rotation with angular velocity $^1/_2\,(\nabla \times \boldsymbol{u})$, and a velocity or rate of deformation. The coefficient of η in the last term in Equation (3.2.29) is defined as the rate-of-strain tensor and is denoted by $\underline{\underline{e}}$, in subscript form

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{3.2.30}$$

From $\underline{\underline{e}}$ we can define a rate-of-strain central quadric, along the principal axes of which the deforming motion consists of a pure extension or compression.

Relationship Between Forces and Rate of Deformation

We are now in a position to determine the required relationship between the stress tensor $\underline{\underline{\sigma}}$ and the rate of deformation. Assuming that in a static fluid the stress reduces to a (negative) hydrostatic or thermodynamic pressure, equal in all directions, we can write

$$\underline{\underline{\sigma}} = -p\underline{\underline{I}} + \underline{\underline{\tau}} \quad \text{or} \quad \underline{\underline{\sigma}}_{ij} = -p\delta_{ij} + \tau_{ij} \tag{3.2.31}$$

where $\underline{\underline{\tau}}$ is the viscous part of the total stress and is called the deviatoric stress tensor, $\underline{\underline{I}}$ is the identity tensor, and δ_{ij} is the corresponding Kronecker delta ($\delta_{ij} = 0$ if $i \neq j$; $\delta_{ij} = 1$ if $i = j$). We make further assumptions that (1) the fluid exhibits no preferred directions; (2) the stress is independent of any previous history of distortion; and (3) that the stress depends only on the local thermodynamic state and the kinematic state of the immediate neighborhood. Precisely, we assume that $\underline{\underline{\tau}}$ is linearly proportional to the first spatial derivatives of $\boldsymbol{u}$, the coefficient of proportionality depending only on the local thermodynamic state. These assumptions and the relations below which follow from them are appropriate for a Newtonian fluid. Most common fluids, such as air and water under most conditions, are Newtonian, but there are many other fluids, including many which arise in industrial applications, which exhibit so-called non-Newtonian properties. The study of such non-Newtonian fluids, such as viscoelastic fluids, is the subject of the field of rheology.

With the Newtonian fluid assumptions above, and the symmetry of $\underline{\underline{\tau}}$ which follows from the symmetry of $\underline{\underline{\sigma}}$, one can show that the viscous part $\underline{\underline{\tau}}$ of the total stress can be written as

$$\underline{\underline{\tau}} = \lambda(\nabla \cdot \boldsymbol{u})\underline{\underline{I}} + 2\mu\underline{\underline{e}} \tag{3.2.32}$$

so the total stress for a Newtonian fluid is

$$\underline{\underline{\sigma}} = -p\underline{\underline{I}} + \lambda(\nabla \cdot \boldsymbol{u})\underline{\underline{I}} + 2\mu\underline{\underline{e}} \tag{3.2.33}$$

or, in subscript notation

$$\sigma_{ij} = -p\delta_{ij} + \lambda\left(\frac{\partial u_k}{\partial x_k}\right)\delta_{ij} + \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{3.2.34}$$

(the Einstein summation convention is assumed here, namely, that a repeated subscript, such as in the second term on the right-hand side above, is summed over; note also that $\nabla \cdot \boldsymbol{u} = \partial u_k/\partial x_k = e_{kk}$.) The coefficient λ is called the "second viscosity" and μ the "absolute viscosity," or more commonly the "dynamic viscosity," or simply the "viscosity." For a Newtonian fluid λ and μ depend only on local thermodynamic state, primarily on the temperature.

We note, from Equation (3.2.34), that whereas in a fluid at rest the pressure is an isotropic normal stress, this is not the case for a moving fluid, since in general $\sigma_{11} \neq \sigma_{22} \neq \sigma_{33}$. To have an analogous quantity to p for a moving fluid we define the pressure in a moving fluid as the negative mean normal stress, denoted, say, by $\bar{p}$

$$\bar{p} = -\frac{1}{3}\sigma_{ii} \tag{3.2.35}$$

(σ_{ii} is the trace of $\underline{\underline{\sigma}}$ and an invariant of $\underline{\underline{\sigma}}$, independent of the orientation of the axes). From Equation (3.2.34)

$$\bar{p} = -\frac{1}{3}\sigma_{ii} = p - \left(\lambda + \frac{2}{3}\mu\right)\nabla \cdot \boldsymbol{u} \tag{3.2.36}$$

For an incompressible fluid $\nabla \cdot \boldsymbol{u} = 0$ and hence $\bar{p} \equiv p$. The quantity $(\lambda + {}^2/_3\mu)$ is called the bulk viscosity. If one assumes that the deviatoric stress tensor τ_{ij} makes no contribution to the mean normal stress, it follows that $\lambda + {}^2/_3\mu = 0$, so again $\bar{p} = p$. This condition, $\lambda = -{}^2/_3\mu$, is called the Stokes assumption or hypothesis. If neither the incompressibility nor the Stokes assumptions are made, the difference between $\bar{p}$ and p is usually still negligibly small because $(\lambda + {}^2/_3\mu)\,\nabla \cdot \boldsymbol{u} << p$ in most fluid flow problems. If the Stokes hypothesis is made, as is often the case in fluid mechanics, Equation (3.2.34) becomes

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\left(e_{ij} - \frac{1}{3}e_{kk}\delta_{ij}\right) \tag{3.2.37}$$

The Navier–Stokes Equations

Substitution of Equation (3.2.33) into (3.2.25), since $\nabla \cdot (\phi \underline{\underline{I}}) = \nabla\phi$, for any scalar function ϕ, yields (replacing $\boldsymbol{u}$ in Equation (3.2.33) by $\boldsymbol{V}$)

$$\rho\frac{D\boldsymbol{V}}{Dt} = \rho\boldsymbol{f} - \nabla p + \nabla\left(\lambda\nabla \cdot \boldsymbol{V}\right) + \nabla \cdot \left(2\mu\underline{\underline{e}}\right) \tag{3.2.38}$$

These equations are the Navier–Stokes equations (although the name is as often given to the full set of governing conservation equations). With the Stokes assumption ($\lambda = -{}^2/_3\mu$), Equation (3.2.38) becomes

$$\rho\frac{D\boldsymbol{V}}{Dt} = \rho\boldsymbol{f} - \nabla p + \nabla \cdot \left[2\mu\left(\underline{\underline{e}} - \frac{1}{3}e_{kk}\underline{\underline{I}}\right)\right] \tag{3.2.39}$$

If the Eulerian frame is not an inertial frame, then one must use the transformation to an inertial frame either using Equation (3.2.8) or the "apparent" body force formulation, Equation (3.2.9).

Energy Conservation — The Mechanical and Thermal Energy Equations

In deriving the differential form of the energy equation we begin by assuming that heat enters or leaves the material or control volume by heat conduction across the boundaries, the heat flux per unit area being $\boldsymbol{q}$. It then follows that

$$\dot{Q} = -\iint \boldsymbol{q} \cdot \boldsymbol{n}\, dA = -\iiint \nabla \cdot \boldsymbol{q}\, d\upsilon \tag{3.2.40}$$

The work-rate term $\dot{W}$ can be decomposed into the rate of work done against body forces, given by

$$-\iiint \rho \boldsymbol{f} \cdot \boldsymbol{V}\, d\upsilon \tag{3.2.41}$$

and the rate of work done against surface stresses, given by

$$-\iint_{\substack{\text{system}\\ \text{surface}}} \boldsymbol{V} \cdot \left(\underline{\underline{\sigma}}\boldsymbol{n}\right) dA \tag{3.2.42}$$

Substitution of these expressions for $\dot{Q}$ and $\dot{W}$ into Equation (3.2.12), use of the divergence theorem, and conservation of mass lead to

$$\rho \frac{D}{Dt}\left(u + \frac{1}{2}V^2\right) = -\nabla \cdot \boldsymbol{q} + \rho \boldsymbol{f} \cdot \boldsymbol{V} + \nabla \cdot \left(\boldsymbol{V}\underline{\underline{\sigma}}\right) \tag{3.2.43}$$

(note that a potential energy term is no longer included in e, the total specific energy, as it is accounted for by the body force rate-of-working term $\rho \boldsymbol{f} \cdot \boldsymbol{V}$).

Equation (3.2.43) is the total energy equation showing how the energy changes as a result of working by the body and surface forces and heat transfer. It is often useful to have a purely thermal energy equation. This is obtained by subtracting from Equation (3.2.43) the dot product of $\boldsymbol{V}$ with the momentum Equation (3.2.25), after expanding the last term in Equation (3.2.43), resulting in

$$\rho \frac{Du}{Dt} = \frac{\partial V_i}{\partial x_j}\sigma_{ij} - \nabla \cdot \boldsymbol{q} \tag{3.2.44}$$

With $\sigma_{ij} = -p\delta_{ij} + \tau_{ij}$, and the use of the continuity equation in the form of Equation (3.2.16), the first term on the right-hand side of Equation (3.2.44) may be written

$$\frac{\partial V_i}{\partial x_j}\sigma_{ij} = -\rho \frac{D\left(\dfrac{p}{\rho}\right)}{Dt} + \frac{Dp}{Dt} + \Phi \tag{3.2.45}$$

where Φ is the rate of dissipation of mechanical energy per unit mass due to viscosity, and is given by

$$\Phi \equiv \frac{\partial V_i}{\partial x_j}\tau_{ij} = 2\mu\left(e_{ij}e_{ij} - \frac{1}{3}e_{kk}^2\right) = 2\mu\left(e_{ij} - \frac{1}{3}e_{kk}\delta_{ij}\right)^2 \tag{3.2.46}$$

With the introduction of Equation (3.2.45), Equation (3.2.44) becomes

$$\rho \frac{De}{Dt} = -p\nabla \cdot \boldsymbol{V} + \Phi - \nabla \cdot \boldsymbol{q} \tag{3.2.47}$$

or

$$\rho \frac{Dh}{Dt} = \frac{Dp}{Dt} + \Phi - \nabla \cdot \boldsymbol{q} \tag{3.2.48}$$

where $h = e + (p/\rho)$ is the specific enthalpy. Unlike the other terms on the right-hand side of Equation (3.2.47), which can be negative or positive, Φ is always nonnegative and represents the increase in internal energy (or enthalpy) owing to irreversible degradation of mechanical energy. Finally, from elementary thermodynamic considerations

$$\frac{Dh}{Dt} = T\frac{DS}{Dt} + \frac{1}{\rho}\frac{Dp}{Dt}$$

where S is the entropy, so Equation (3.2.48) can be written

$$\rho T\frac{DS}{Dt} = \Phi - \nabla\cdot\boldsymbol{q} \tag{3.2.49}$$

If the heat conduction is assumed to obey the Fourier heat conduction law, so $\boldsymbol{q} = -k\nabla$T, where k is the thermal conductivity, then in all of the above equations

$$-\nabla\cdot\boldsymbol{q} = \nabla\cdot\left(k\nabla T\right) = k\nabla^2 T \tag{3.2.50}$$

the last of these equalities holding only if k = constant.

In the event the thermodynamic quantities vary little, the coefficients of the constitutive relations for $\underline{\underline{\sigma}}$ and $\boldsymbol{q}$ may be taken to be constant and the above equations simplified accordingly.

We note also that if the flow is incompressible, then the mass conservation, or continuity, equation simplifies to

$$\nabla\cdot\boldsymbol{V} = 0 \tag{3.2.51}$$

and the momentum Equation (3.2.38) to

$$\rho\frac{D\boldsymbol{V}}{Dt} = \rho\boldsymbol{f} - \nabla p + \mu\nabla^2\boldsymbol{V} \tag{3.2.52}$$

where ∇^2 is the Laplacian operator. The small temperature changes, compatible with the incompressibility assumption, are then determined, for a perfect gas with constant k and specific heats, by the energy equation rewritten for the temperature, in the form

$$\rho c_v\frac{DT}{Dt} = k\nabla^2 T + \Phi \tag{3.2.53}$$

Boundary Conditions

The appropriate boundary conditions to be applied at the boundary of a fluid in contact with another medium depends on the nature of this other medium — solid, liquid, or gas. We discuss a few of the more important cases here in turn:

1. *At a solid surface:* $\boldsymbol{V}$ and T are continuous. Contained in this boundary condition is the "no-slip" condition, namely, that the tangential velocity of the fluid in contact with the boundary of the solid is equal to that of the boundary. For an inviscid fluid the no-slip condition does not apply, and only the normal component of velocity is continuous. If the wall is permeable, the tangential velocity is continuous and the normal velocity is arbitrary; the temperature boundary condition for this case depends on the nature of the injection or suction at the wall.

2. *At a liquid/gas interface:* For such cases the appropriate boundary conditions depend on what can be assumed about the gas the liquid is in contact with. In the classical liquid free-surface problem, the gas, generally atmospheric air, can be ignored and the necessary boundary conditions are that (a) the normal velocity in the liquid at the interface is equal to the normal velocity of the interface and (b) the pressure in the liquid at the interface exceeds the atmospheric pressure by an amount equal to

$$\Delta p = p_{\text{liquid}} - p_{\text{atm}} = \sigma\left(\frac{1}{R_1} + \frac{1}{R_2}\right) \tag{3.2.54}$$

where R_1 and R_2 are the radii of curvature of the intercepts of the interface by two orthogonal planes containing the vertical axis. If the gas is a vapor which undergoes nonnegligible interaction and exchanges with the liquid in contact with it, the boundary conditions are more complex. Then, in addition to the above conditions on normal velocity and pressure, the shear stress (momentum flux) and heat flux must be continuous as well.

For interfaces in general the boundary conditions are derived from continuity conditions for each "transportable" quantity, namely continuity of the appropriate intensity across the interface and continuity of the normal component of the flux vector. Fluid momentum and heat are two such transportable quantities, the associated intensities are velocity and temperature, and the associated flux vectors are stress and heat flux. (The reader should be aware of circumstances where these simple criteria do not apply, for example, the velocity slip and temperature jump for a rarefied gas in contact with a solid surface.)

Vorticity in Incompressible Flow

With μ = constant, ρ = constant, and $\boldsymbol{f} = -\boldsymbol{g} = -g\boldsymbol{k}$ the momentum equation reduces to the form (see Equation (3.2.52))

$$\rho\frac{D\boldsymbol{V}}{Dt} = -\nabla p - \rho g\boldsymbol{k} + \mu\nabla^2\boldsymbol{V} \tag{3.2.55}$$

With the vector identities

$$(\boldsymbol{V}\cdot\nabla)\boldsymbol{V} = \nabla\left(\frac{V^2}{2}\right) - \boldsymbol{V}\times(\nabla\times\boldsymbol{V}) \tag{3.2.56}$$

and

$$\nabla^2\boldsymbol{V} = \nabla(\nabla\cdot\boldsymbol{V}) - \nabla\times(\nabla\times\boldsymbol{V}) \tag{3.2.57}$$

and defining the *vorticity*

$$\boldsymbol{\zeta} \equiv \nabla\times\boldsymbol{V} \tag{3.2.58}$$

Equation (3.2.55) can be written, noting that for incompressible flow $\nabla\cdot\boldsymbol{V} = 0$,

$$\rho\frac{\partial\boldsymbol{V}}{\partial t} + \nabla\left(p + \frac{1}{2}\rho V^2 + \rho gz\right) = \rho\boldsymbol{V}\times\boldsymbol{\zeta} - \mu\nabla\times\boldsymbol{\zeta} \tag{3.2.59}$$

The flow is said to be *irrotational* if

$$\zeta \equiv \nabla \times \boldsymbol{V} = 0 \tag{3.2.60}$$

from which it follows that a *velocity potential* Φ can be defined

$$\boldsymbol{V} = \nabla\Phi \tag{3.2.61}$$

Setting $\zeta = 0$ in Equation (3.2.59), using Equation (3.2.61), and then integrating with respect to all the spatial variables, leads to

$$\rho\frac{\partial\Phi}{\partial t} + \left(p + \frac{1}{2}\rho V^2 + \rho g z\right) = F(t) \tag{3.2.62}$$

(the arbitrary function $F(t)$ introduced by the integration can either be absorbed in Φ, or is determined by the boundary conditions). Equation (3.2.62) is the unsteady *Bernoulli equation* for irrotational, incompressible flow. (Irrotational flows are always potential flows, even if the flow is compressible. Because the viscous term in Equation (3.2.59) vanishes identically for $\zeta = 0$, it would appear that the above Bernoulli equation is valid even for viscous flow. Potential solutions of hydrodynamics are in fact exact solutions of the full Navier–Stokes equations. Such solutions, however, are not valid near solid boundaries or bodies because the no-slip condition generates vorticity and causes nonzero ζ; the potential flow solution is invalid in all those parts of the flow field that have been "contaminated" by the spread of the vorticity by convection and diffusion. See below.)

The curl of Equation (3.2.59), noting that the curl of any gradient is zero, leads to

$$\rho\frac{\partial\zeta}{\partial t} = \rho\nabla \times \left(\boldsymbol{V} \times \zeta\right) - \mu\nabla \times \nabla \times \zeta \tag{3.2.63}$$

but

$$\begin{aligned}\nabla^2\zeta &= \nabla\left(\nabla \cdot \zeta\right) - \nabla \times \nabla \times \zeta \\ &= -\nabla \times \nabla \times \zeta\end{aligned} \tag{3.2.64}$$

since div curl ( ) $\equiv 0$, and therefore also

$$\nabla \times \left(\boldsymbol{V} \times \zeta\right) \equiv \zeta\left(\nabla\boldsymbol{V}\right) + \boldsymbol{V}\nabla \cdot \zeta - \boldsymbol{V}\nabla\zeta - \zeta\nabla \cdot \boldsymbol{V} \tag{3.2.65}$$

$$= \zeta\left(\nabla\boldsymbol{V}\right) - \boldsymbol{V}\nabla\zeta \tag{3.2.66}$$

Equation (3.2.63) can then be written

$$\frac{D\zeta}{Dt} = \left(\zeta \cdot \nabla\right)\boldsymbol{V} + \nu\nabla^2\zeta \tag{3.2.67}$$

where $\nu = \mu/\rho$ is the kinematic viscosity. Equation (3.2.67) is the vorticity equation for incompressible flow. The first term on the right, an inviscid term, increases the vorticity by vortex stretching. In inviscid, two-dimensional flow both terms on the right-hand side of Equation (3.2.67) vanish, and the equation reduces to $D\zeta/Dt = 0$, from which it follows that the vorticity of a fluid particle remains constant as it moves. This is Helmholtz's theorem. As a consequence it also follows that if $\zeta = 0$ initially, $\zeta \equiv 0$ always;

i.e., initially irrotational flows remain irrotational (for inviscid flow). Similarly, it can be proved that $D\Gamma/Dt = 0$; i.e., the circulation around a material closed circuit remains constant, which is Kelvin's theorem.

If $\nu \neq 0$, Equation (3.2.67) shows that the vorticity generated, say, at solid boundaries, diffuses and stretches as it is convected.

We also note that for steady flow the Bernoulli equation reduces to

$$p + \frac{1}{2}\rho V^2 + \rho g z = \text{constant} \tag{3.2.68}$$

valid for steady, irrotational, incompressible flow.

Stream Function

For two-dimensional flows the continuity equation, e.g., for plane, incompressible flows ($\boldsymbol{V} = (u, \nu)$)

$$\frac{\partial u}{\partial x} + \frac{\partial \nu}{\partial y} = 0 \tag{3.2.69}$$

can be identically satisfied by introducing a stream function ψ, defined by

$$u = \frac{\partial \psi}{\partial y}, \qquad \nu = -\frac{\partial \psi}{\partial x} \tag{3.2.70}$$

Physically ψ is a measure of the flow between streamlines. (Stream functions can be similarly defined to satisfy identically the continuity equations for incompressible cylindrical and spherical axisymmetric flows; and for these flows, as well as the above planar flow, also when they are compressible, but only then if they are steady.) Continuing with the planar case, we note that in such flows there is only a single nonzero component of vorticity, given by

$$\boldsymbol{\zeta} = \left(0,0,\zeta_z\right) = \left(0,0,\frac{\partial \nu}{\partial x} - \frac{\partial u}{\partial y}\right) \tag{3.2.71}$$

With Equation (3.2.70)

$$\zeta_z = -\frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial y^2} = -\nabla^2 \psi \tag{3.2.72}$$

For this two-dimensional flow Equation (3.2.67) reduces to

$$\frac{\partial \zeta_z}{\partial t} + u\frac{\partial \zeta_z}{\partial x} + \nu\frac{\partial \zeta_z}{\partial y} = \nu\left(\frac{\partial^2 \zeta_z}{\partial x^2} + \frac{\partial^2 \zeta_z}{\partial y^2}\right) \tag{3.2.73}$$

Substitution of Equation (3.2.72) into Equation (3.2.73) yields an equation for the stream function

$$\frac{\partial\left(\nabla^2 \psi\right)}{\partial t} + \frac{\partial \psi}{\partial y}\frac{\partial\left(\nabla^2 \psi\right)}{\partial x} - \frac{\partial \psi}{\partial x}\frac{\partial}{\partial y}\left(\nabla^2 \psi\right) = \nu \nabla^4 \psi \tag{3.2.74}$$

where $\nabla^4 = \nabla^2(\nabla^2)$. For uniform flow past a solid body, for example, this equation for Ψ would be solved subject to the boundary conditions:

$$\begin{aligned} &\frac{\partial \psi}{\partial x} = 0, \quad \frac{\partial \psi}{\partial y} = V_\infty \quad \text{at infinity} \\ &\frac{\partial \psi}{\partial x} = 0, \quad \frac{\partial \psi}{\partial y} = 0 \quad \text{at the body (no-slip)} \end{aligned} \tag{3.2.75}$$

For the special case of irrotational flow it follows immediately from Equations (3.2.70) and (3.2.71) with $\zeta_z = 0$, that ψ satisfies the Laplace equation

$$\nabla^2 \psi = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} = 0 \tag{3.2.76}$$

Inviscid Irrotational Flow: Potential Flow

For irrotational flows we have already noted that a velocity potential Φ can be defined such that $\boldsymbol{V} = \nabla\Phi$. If the flow is also incompressible, so $\nabla \cdot \boldsymbol{V} = 0$, it then follows that

$$\nabla \cdot (\nabla \Phi) = \nabla^2 \Phi = 0 \tag{3.2.77}$$

so Φ satisfies Laplace's equation. (Note that unlike the stream function ψ, which can only be defined for two-dimensional flows, the above considerations for Φ apply to flow in two and three dimensions. On the other hand, the existence of ψ does not require the flow to be irrotational, whereas the existence of Φ does.)

Since Equation (3.2.77) with appropriate conditions on $\boldsymbol{V}$ at boundaries of the flow completely determines the velocity field, and the momentum equation has played no role in this determination, we see that inviscid irrotational flow — *potential theory* — is a purely kinematic theory. The momentum equation only enters after Φ is known in order to calculate the pressure field consistent with the velocity field $\boldsymbol{V} = \nabla\Phi$.

For both two- and three-dimensional flows the determination of Φ makes use of the powerful techniques of potential theory, well developed in the mathematical literature. For two-dimensional planar flows the techniques of complex variable theory are available, since Φ may be considered as either the real or imaginary part of an analytic function (the same being true for ψ, since for such two-dimensional flows Φ and ψ are conjugate variables.)

Because the Laplace equation, obeyed by both Φ and ψ, is linear, complex flows may be built up from the superposition of simple flows; this property of inviscid irrotational flows underlies nearly all solution techniques in this area of fluid mechanics.

Problem

A two-dimensional inviscid irrotational flow has the velocity potential

$$\Phi = x^2 - y^2 \tag{P.3.2.6}$$

What two-dimensional potential flow does this represent?

Solution. It follows from Equations (3.2.61) and (3.2.70) that for two-dimensional flows, in general

$$u = \frac{\partial \Phi}{\partial x} = \frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \Phi}{\partial y} = -\frac{\partial \psi}{\partial x} \tag{P.3.2.7}$$

It follows from using Equation (P.3.2.6) that

$$u = \frac{\partial \psi}{\partial y} = 2x, \quad v = -\frac{\partial \psi}{\partial x} = -2y \tag{P.3.2.8}$$

Integration of Equation (P.3.2.8) yields

$$\psi = 2xy \tag{P.3.2.9}$$

The streamlines, ψ = constant, and equipotential lines, Φ = constant, both families of hyperbolas and each family the orthogonal trajectory of the other, are shown in Figure 3.2.3. Because the x and y axes are streamlines, Equations (P.3.2.6) and (P.3.2.9) represent the inviscid irrotational flow in a right-angle corner. By symmetry, they also represent the planar flow in the upper half-plane directed toward a stagnation point at $x = y = 0$ (see Figure 3.2.4). In polar coordinates (r, θ), with corresponding velocity components (u_r, u_θ), this flow is represented by

$$\Phi = r^2 \cos 2\theta, \quad \psi = r^2 \sin 2\theta \tag{P.3.2.10}$$

with

$$\begin{aligned} u_r &= \frac{\partial \Phi}{\partial r} = \frac{1}{r}\frac{\partial \psi}{\partial \theta} = 2r\cos 2\theta \\ u_\theta &= \frac{1}{r}\frac{\partial \Phi}{\partial \theta} = -\frac{\partial \psi}{\partial r} = -2r\sin 2\theta \end{aligned} \tag{P.3.2.11}$$

For two-dimensional planar potential flows we may also use complex variables, writing the complex potential $f(z) = \Phi + i\psi$ as a function of the complex variable $z = x + iy$, where the complex velocity is given by $f'(z) = w(z) = u - iv$. For the flow above

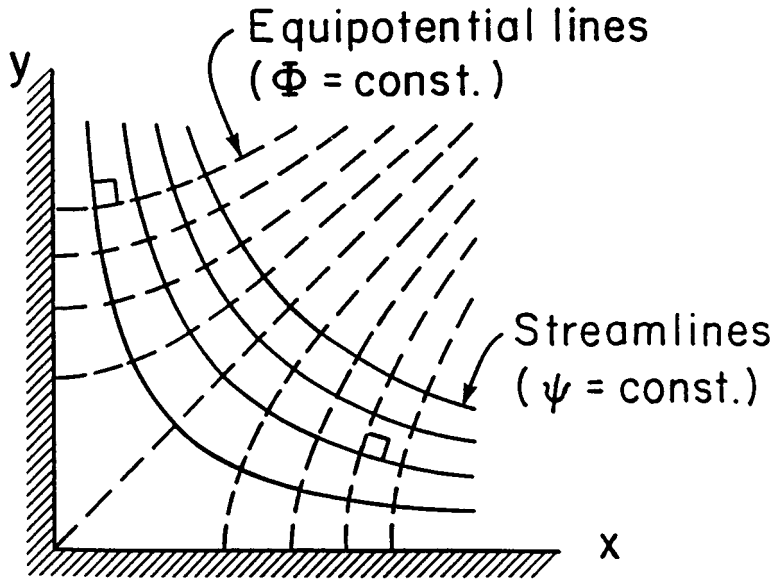


FIGURE 3.2.3 Potential flow in a 90° corner.

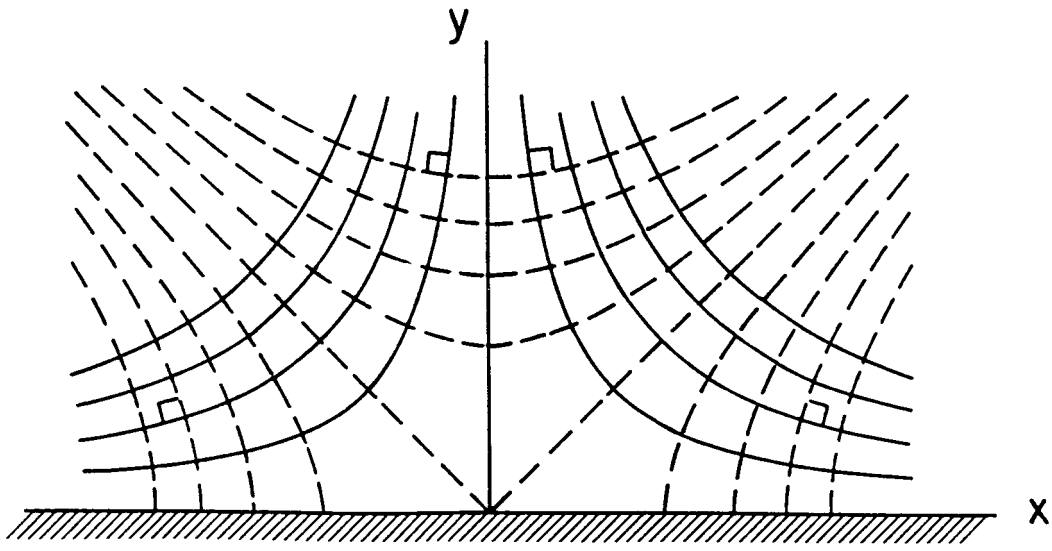


FIGURE 3.2.4 Potential flow impinging against a flat (180°) wall (plane stagnation-point flow).

$$f(z) = z^2 \tag{P.3.2.12}$$

Expressions such as Equation (P.3.2.12), where the right-hand side is an analytic function of z, may also be regarded as a conformal mapping, which makes available as an aid in solving two-dimensional potential problems all the tools of this branch of mathematics.

Further Information

More detail and additional information on the topics in this section may be found in more advanced books on fluid dynamics, such as

Batchelor, G.K. 1967. *An Introduction to Fluid Dynamics,* Cambridge University Press, Cambridge, England.

Warsi, Z.U.A. 1993. *Fluid Dynamics. Theoretical and Computational Approaches,* CRC Press, Boca Raton, FL.

Sherman, F.S. 1990. *Viscous Flow*, McGraw-Hill, New York.

Panton, R.L. 1984. *Incompressible Flow,* John Wiley & Sons, New York.

3.3 Similitude: Dimensional Analysis and Data Correlation

Stuart W. Churchill

Dimensional Analysis

Similitude refers to the formulation of a description for physical behavior that is general and independent of the individual dimensions, physical properties, forces, etc. In this subsection the treatment of similitude is restricted to *dimensional analysis*; for a more general treatment see Zlokarnik (1991). The full power and utility of dimensional analysis is often underestimated and underutilized by engineers. This technique may be applied to a complete mathematical model or to a simple listing of the variables that define the behavior. Only the latter application is described here. For a description of the application of dimensional analysis to a mathematical model see Hellums and Churchill (1964).

General Principles

Dimensional analysis is based on the principle that all additive or equated terms of a complete relationship between the variables must have the same net dimensions. The analysis starts with the preparation of a list of the individual dimensional variables (dependent, independent, and parametric) that are presumed to define the behavior of interest. The performance of dimensional analysis in this context is reasonably simple and straightforward; the principal difficulty and uncertainty arise from the identification of the variables to be included or excluded. If one or more important variables are inadvertently omitted, the reduced description achieved by dimensional analysis will be incomplete and inadequate as a guide for the correlation of a full range of experimental data or theoretical values. The familiar band of plotted values in many graphical correlations is more often a consequence of the omission of one or more variables than of inaccurate measurements. If, on the other hand, one or more irrelevant or unimportant variables are included in the listing, the consequently reduced description achieved by dimensional analysis will result in one or more unessential dimensionless groupings. Such excessive dimensionless groupings are generally less troublesome than missing ones because the redundancy will ordinarily be revealed by the process of correlation. Excessive groups may, however, suggest unnecessary experimental work or computations, or result in misleading correlations. For example, real experimental scatter may inadvertently and incorrectly be correlated in all or in part with the variance of the excessive grouping.

In consideration of the inherent uncertainty in selecting the appropriate variables for dimensional analysis, it is recommended that this process be interpreted as a *speculative* and subject to correction of the basis of experimental data or other information. Speculation may also be utilized as a formal technique to identify the effect of eliminating a variable or of combining two or more. The latter aspect of speculation, which may be applied either to the original listing of dimensional variables or to the resulting set of dimensionless groups, is often of great utility in identifying possible limiting behavior or dimensionless groups of marginal significance. The systematic speculative elimination of all but the most certain variables, one at a time, followed by regrouping, is recommended as a general practice. The additional effort as compared with the original dimensional analysis is minimal, but the possible return is very high. A general discussion of this process may be found in Churchill (1981).

The minimum number of independent dimensionless groups i required to describe the fundamental and parametric behavior is (Buckingham, 1914)

$$i = n - m \tag{3.3.1}$$

where n is the number of variables and m is the number of fundamental dimensions such as mass M, length L, time θ, and temperature T that are introduced by the variables. The inclusion of redundant dimensions such as force F and energy E that may be expressed in terms of mass, length, time, and

temperature is at the expense of added complexity and is to be avoided. (Of course, mass could be replaced by force or temperature by energy as alternative fundamental dimensions.) In some rare cases i is actually greater than $n - m$. Then

$$i = n - k \tag{3.3.2}$$

where k is the maximum number of the chosen variables that cannot be combined to form a dimensionless group. Determination of the minimum number of dimensionless groups is helpful if the groups are to be chosen by inspection, but is unessential if the algebraic procedure described below is utilized to determine the groups themselves since the number is then obvious from the final result.

The *particular* minimal set of dimensionless groups is arbitrary in the sense that two or more of the groups may be multiplied together to any positive, negative, or fractional power as long as the number of independent groups is unchanged. For example, if the result of a dimensional analysis is

$$\phi\left\{XY^{1/2}, Z/Y^2, Z\right\} = 0 \tag{3.3.3}$$

where X, Y, and Z are independent dimensionless groups, an equally valid expression is

$$\phi\left\{X, Y, Z\right\} = 0 \tag{3.3.4}$$

Dimensional analysis itself does not provide any insight as to the best choice of equivalent dimensionless groupings, such as between those of Equations (3.3.3) and (3.3.4). However, isolation of each of the variables that are presumed to be the most important in a separate group may be convenient in terms of interpretation and correlation. Another possible criterion in choosing between alternative groupings may be the relative invariance of a particular one. The functional relationship provided by Equation (3.3.3) may equally well be expressed as

$$X = \phi\left\{Y, Z\right\} \tag{3.3.5}$$

where X is implied to be the dependent grouping and Y and Z to be independent or parametric groupings.

Three primary methods of determining a minimal set of dimensionless variables are (1) by inspection; (2) by combination of the residual variables, one at a time, with the set of chosen variables that cannot be combined to obtain a dimensionless group; and (3) by an algebraic procedure. These methods are illustrated in the examples that follow.

Example 3.3.1: Fully Developed Flow of Water Through a Smooth Round Pipe

Choice of Variables. The shear stress τ_w on the wall of the pipe may be postulated to be a function of the density ρ and the dynamic viscosity μ of the water, the inside diameter D of the pipe, and the space-mean of the time-mean velocity u_m. The limitation to fully developed flow is equivalent to a postulate of independence from distance x in the direction of flow, and the specification of a smooth pipe is equivalent to the postulate of independence from the roughness e of the wall. The choice of τ_w rather than the pressure drop per unit length $-dP/dx$ avoids the need to include the acceleration due to gravity g and the elevation z as variables. The choice of u_m rather than the volumetric rate of flow V, the mass rate of flow w, or the mass rate of flow per unit area G is arbitrary but has some important consequences as noted below. The postulated dependence may be expressed functionally as $\phi\{\tau_w, \rho, \mu, D, u_m\} = 0$ or $\tau_w = \phi\{\rho, \mu, D, u_m\}$.

Tabulation. Next prepare a tabular listing of the variables and their dimensions:

	τ_w	ρ	μ	D	u_m
M	1	1	1	0	0
L	–1	–3	–1	1	1
θ	–2	0	–1	0	–1
T	0	0	0	0	0

Minimal Number of Groups. The number of postulated variables is 5. Since the temperature does not occur as a dimension for any of the variables, the number of fundamental dimensions is 3. From Equation (3.3.1), the minimal number of dimensionless groups is 5 – 3 = 2. From inspection of the above tabulation, a dimensionless group cannot be formed from as many as three variables such as D, μ, and ρ. Hence, Equation (3.3.2) also indicates that $i = 5 - 3 = 2$.

Method of Inspection. By inspection of the tabulation or by trial and error it is evident that only two independent dimensionless groups may be formed. One such set is

$$\phi\left\{\frac{\tau_w}{\rho u_m^2}, \frac{Du_m\rho}{\mu}\right\} = 0$$

Method of Combination. The residual variables τ_w and μ may be combined in turn with the noncombining variables ρ, D, and u_m to obtain two groups such as those above.

Algebraic Method. The algebraic method makes formal use of the postulate that the functional relationship between the variables may in general be represented by a power series. In this example such a power series may be expressed as

$$\tau_w = \sum_{i=1}^{N} A_i \rho^{a_i} \mu^{b_i} D^{c_i} u_m^{d_i}$$

where the coefficients A_i are dimensionless. Each additive term on the right-hand side of this expression must have the same net dimensions as τ_w. Hence, for the purposes of dimensional analysis, only the first term need be considered and the indexes may be dropped. The resulting highly restricted expression is $\tau_w = A\rho^a\mu^b D^c u_m^d$. Substituting the dimensions for the variables gives

$$\frac{M}{L\theta^2} = A\left(\frac{M}{L^3}\right)^a \left(\frac{M}{L\theta}\right)^b L^c \left(\frac{L}{\theta}\right)^d$$

Equating the sum of the exponents of M, L, and θ on the right-hand side of the above expression with those of the left-hand side produces the following three simultaneous linear algebraic equations: $1 = a + b$; $-1 = -3a - b + c + d$; and $-2 = -b - d$, which may be solved for a, c, and d in terms of b to obtain $a = 1 - b$, $c = -b$, and $d = 2 - b$. Substitution then gives $\tau_w = A\rho^{1-b}\mu^b D^{-b} u_m^{2-b}$ which may be regrouped as

$$\frac{\tau_w}{\rho u_m^2} = A\left(\frac{\mu}{Du_m\rho}\right)^b$$

Since this expression is only the first term of a power series, it should *not* be interpreted to imply that $\tau_w/\rho u_m^2$ is necessarily proportional to some power at $\mu/Du_m\rho$ but instead only the equivalent of the expression derived by the method of inspection. The inference of a power dependence between the

dimensionless groups is the most common and serious error in the use of the algebraic method of dimensional analysis.

Speculative Reductions. Eliminating ρ as a variable on speculative grounds to

$$\phi\left\{\frac{\tau_w D}{\mu u_m}\right\} = 0$$

or its exact equivalent:

$$\frac{\tau_w D}{\mu u_m} = A$$

The latter expression with $A = 8$ is actually the exact solution for the laminar regime ($Du_m\rho/\mu < 1800$). A relationship that does not include ρ may alternatively be derived directly from the solution by the method of inspection as follows. First, ρ is eliminated from one group, say $\tau_w/\rho u_m^2$, by multiplying it with $Du_m\rho/\mu$ to obtain

$$\phi\left\{\frac{\tau_w D}{\mu u_m}, \frac{Du_m\rho}{\mu}\right\} = 0$$

The remaining group containing ρ is now simply dropped. Had the original expression been composed of three independent groups each containing ρ, that variable would have to be eliminated from two of them before dropping the third one.

The relationships that are obtained by the speculative elimination of μ, D, and u_m, one at a time, do not appear to have any range of physical validity. Furthermore, if w or G had been chosen as the independent variable rather than u_m, the limited relationship for the laminar regime would not have been obtained by the elimination of ρ.

Alternative Forms. The solution may also be expressed in an infinity of other forms such as

$$\phi\left\{\frac{\tau_w D^2\rho}{\mu^2}, \frac{Du_m\rho}{\mu}\right\} = 0$$

If τ_w is considered to be the principal dependent variable and u_m the principal independent variable, this latter form is preferable in that these two quantities do not then appear in the same grouping. On the other hand, if D is considered to be the principal independent variable, the original formulation is preferable. The variance of $\tau_w/\rho u_m^2$ is less than that of $\tau_w D/\mu u_m$ and $\tau_w D^2\rho/\mu^2$ in the turbulent regime while that of $\tau_w D/\mu u_m$ is zero in the laminar regime. Such considerations may be important in devising convenient graphical correlations.

Alternative Notations. The several solutions above are more commonly expressed as

$$\phi\left\{\frac{f}{2}, \mathrm{Re}\right\} = 0$$

$$\phi\left\{\frac{f\,\mathrm{Re}}{2}, \mathrm{Re}\right\} = 0$$

or

$$\phi\left\{\frac{f\mathrm{Re}^2}{2}, \mathrm{Re}\right\} = 0$$

where $f = 2\,\tau_w/\rho u_m^2$ is the *Fanning friction factor* and $Re = Du_m\rho/\mu$ is the *Reynolds number.*

The more detailed forms, however, are to be preferred for purposes of interpretation or correlation because of the explicit appearance of the individual, physically measurable variables.

Addition of a Variable. The above results may readily be extended to incorporate the roughness e of the pipe as a variable. If two variables have the same dimensions, they will always appear as a dimensionless group in the form of a ratio, in this case e appears most simply as e/D. Thus, the solution becomes

$$\phi\left\{\frac{\tau_w}{\rho u_m^2}, \frac{Du_m\rho}{\mu}, \frac{e}{D}\right\} = 0$$

Surprisingly, as contrasted with the solution for a smooth pipe, the speculative elimination of μ and hence of the group $Du_m\rho/\mu$ now results in a valid asymptote for $Du_m\rho/\mu \to \infty$ and all finite values of e/D, namely,

$$\phi\left\{\frac{\tau_w}{\rho u_m^2}, \frac{e}{D}\right\} = 0$$

Example 3.3.2: Fully Developed Forced Convection in Fully Developed Flow in a Round Tube

It may be postulated for this process that $h = \phi\{D, u_m, \rho, \mu, k, c_p\}$, where here h is the local heat transfer coefficient, and c_p and k are the specific heat capacity and thermal conductivity, respectively, of the fluid. The corresponding tabulation is

	h	D	u_m	ρ	μ	k	c_p
M	1	0	0	1	1	1	0
L	0	1	1	–3	–1	1	2
θ	–3	0	–1	0	–1	–3	–2
T	–1	0	0	0	0	–1	–1

The number of variables is 7 and the number of independent dimensions is 4, as is the number of variables such as D, u_m, ρ, and k that cannot be combined to obtain a dimensionless group. Hence, the minimal number of dimensionless groups is 7 – 4 = 3. The following acceptable set of dimensionless groups may be derived by any of the procedures illustrated in Example 1:

$$\frac{hD}{k} = \phi\left\{\frac{Du_m\rho}{\mu}, \frac{c_p\mu}{k}\right\}$$

Speculative elimination of μ results in

$$\frac{hD}{k} = \phi\left\{\frac{Du_m\rho c_p}{k}\right\}$$

which has often erroneously been inferred to be a valid asymptote for $c_p\mu/k \to 0$. Speculative elimination of D, u_m, ρ, k, and c_p individually also does not appear to result in expressions with any physical validity. However, eliminating c_p and ρ or u_m gives a valid result for the laminar regime, namely,

$$\frac{hD}{k} = A$$

The general solutions for flow and convection in a smooth pipe may be combined to obtain

$$\frac{hD}{k} = \phi\left\{\frac{\tau_w D^2 \rho}{\mu^2}, \frac{c_p \mu}{k}\right\}$$

which would have been obtained directly had u_m been replaced by τ_w in the original tabulation. This latter expression proves to be superior in terms of speculative reductions. Eliminating D results in

$$\frac{h\mu}{k(\tau_w \rho)^{1/2}} = \phi\left\{\frac{c_p \mu}{k}\right\}$$

which may be expressed in the more conventional form of

$$\mathrm{Nu} = \mathrm{Re}\left(\frac{f}{2}\right)^{1/2} \phi\{\mathrm{Pr}\}$$

where Nu = hD/k is the *Nusselt number* and Pr = $c_p\mu/k$ is the *Prandtl number*. This result appears to be a valid asymptote for Re $\to \infty$ and a good approximation for even moderate values (>5000) for large values of Pr. Elimination of μ as well as D results in

$$\frac{h}{c_p(\tau_w \rho)^{1/2}} = A$$

or

$$\mathrm{Nu} = A\,\mathrm{Re}\,\mathrm{Pr}\left(\frac{f}{2}\right)^{1/2}$$

which appears to be an approximate asymptote for Re $\to \infty$ and Pr $\to 0$. Elimination of both c_p and ρ again yields the appropriate result for laminar flow, indicating that ρ rather than u_m is the meaningful variable to eliminate in this respect.

The numerical value of the coefficient A in the several expressions above depends on the mode of heating, a true variable, but one from which the purely functional expressions are independent. If j_w the heat flux density at the wall, and $T_w - T_m$, the temperature difference between the wall and the bulk of the fluid, were introduced as variables in place of $h \equiv j_w/(T_w - T_m)$, another group such as $c_p(T_w - T_m)(D\rho/\mu)^2$ or $\rho c_p(T_w - T_m)/\tau_w$ or $c_p(T_w - T_m)/u_m^2$, which represents the effect of viscous dissipation, would be obtained. This effect is usually but not always negligible. (See Chapter 4.)

Example 3.3.3: Free Convection from a Vertical Isothermal Plate

The behavior for this process may be postulated to be represented by

$$h = \phi\{g, \beta, T_w - T_\infty, x, \mu, \rho, c_p, k\}$$

where g is the acceleration due to gravity, β is the volumetric coefficient of expansion with temperature, T_∞ is the unperturbed temperature of the fluid, and x is the vertical distance along the plate. The corresponding tabulation is

	h	g	β	$T_w - T_\infty$	x	μ	ρ	c_p	k
M	1	0	0	0	0	1	1	0	1
L	0	1	0	0	1	–1	–3	2	1
θ	–3	–2	0	0	0	–1	0	–2	–3
T	–1	0	–1	1	0	0	0	–1	1

The minimal number of dimensionless groups indicated by both methods is 9 – 4 = 5. A satisfactory set of dimensionless groups, as found by any of the methods illustrated in Example 1 is

$$\frac{hx}{k} = \phi\left\{\frac{\rho^2 g x^3}{\mu^2}, \frac{c_p\mu}{k}, \beta\left(T_w - T_\infty\right), c_p\left(T_w - T_\infty\right)\left(\frac{\rho x}{\mu}\right)^2\right\}$$

It may be reasoned that the buoyant force which generates the convective motion must be proportional to $\rho g\beta(T_w - T_\infty)$, thus, g in the first term on the right-hand side must be multiplied by $\beta(T_w - T_\infty)$, resulting in

$$\frac{hx}{k} = \phi\left\{\frac{\rho^2 g\beta\left(T_w - T_\infty\right)x^3}{\mu^2}, \frac{c_p\mu}{k}, \beta\left(T_w - T_\infty\right), c_p\left(T_w - T_\infty\right)\left(\frac{\rho x}{\mu}\right)^2\right\}$$

The effect of expansion other than on the buoyancy is now represented by $\beta(T_w - T_\infty)$, and the effect of viscous dissipation by $c_p(T_w - T_\infty)(\rho x/\mu)^2$. Both effects are negligible for all practical circumstances. Hence, this expression may be reduced to

$$\frac{hx}{k} = \phi\left\{\frac{\rho^2 g\beta\left(T_w - T_\infty\right)x^3}{\mu^2}, \frac{c_p\mu}{k}\right\}$$

or

$$\mathrm{Nu}_x = \phi\left\{\mathrm{Gr}_x, \mathrm{Pr}\right\}$$

where $\mathrm{Nu}_x = hx/k$ and $\mathrm{Gr}_x = \rho^2 g\beta(T_w - T_\infty)x^3/\mu^2$ is the *Grashof number.*

Elimination of x speculatively now results in

$$\frac{hx}{k} = \left(\frac{\rho^2 g\beta\left(T_w - T_\infty\right)x^3}{\mu^2}\right)^{1/3} \phi\left\{\mathrm{Pr}\right\}$$

or

$$\mathrm{Nu}_x = \mathrm{Gr}_x^{1/3}\phi\left\{\mathrm{Pr}\right\}$$

This expression appears to be a valid asymptote for $\mathrm{Gr}_x \to \infty$ and a good approximation for the entire turbulent regime. Eliminating μ speculatively rather than x results in

$$\frac{hx}{k} = \phi\left\{\frac{\rho^2 c_p^2 g\beta\left(T_w - T_\infty\right)x^3}{k^2}\right\}$$

or

$$\mathrm{Nu}_x = \phi\left\{\mathrm{Gr}_x \mathrm{Pr}^2\right\}$$

The latter expression appears to be a valid asymptote for $\mathrm{Pr} \to 0$ for all Gr_x, that is, for both the laminar and the turbulent regimes. The development of a valid asymptote for large values of Pr requires more subtle reasoning. First $c_p\mu/k$ is rewritten as $\mu/\rho\alpha$ where $\alpha = k/\rho c_p$. Then ρ is eliminated speculatively except as it occurs in $\rho g\beta(T_w - T_\infty)$ and $k/\rho c_p$. The result is

$$\frac{hx}{k} = \phi\left\{\frac{c_p \rho^2 g\beta\left(T_w - T_\infty\right)x^3}{\mu k}\right\}$$

or

$$\mathrm{Nu}_x = \phi\left\{\mathrm{Ra}_x\right\}$$

where

$$\mathrm{Ra}_x = \frac{c_p \rho^2 g\beta\left(T_w - T_\infty\right)x^3}{\mu k} = \mathrm{Gr}_x \mathrm{Pr}$$

is the *Rayleigh number*. The expression appears to be a valid asymptote for $\mathrm{Pr} \to \infty$ and a reasonable approximation for even moderate values of Pr for all Gr_x, that is, for both the laminar and the turbulent regimes.

Eliminating x speculatively from the above expressions for small and large values of Pr results in

$$\mathrm{Nu}_x = A\left(\mathrm{Gr}_x \mathrm{Pr}^2\right)^{1/3} = A\left(\mathrm{Ra}_x \mathrm{Pr}\right)^{1/3}$$

and

$$\mathrm{Nu}_x = B\left(\mathrm{Gr}_x \mathrm{Pr}\right)^{1/3} = B\left(\mathrm{Ra}_x\right)^{1/3}$$

The former appears to be a valid asymptote for $\mathrm{Pr} \to 0$ and $\mathrm{Gr}_x \to \infty$ and a reasonable approximation for very small values of Pr in the turbulent regime, while the latter is well confirmed as a valid asymptote for $\mathrm{Pr} \to \infty$ and $\mathrm{Gr}_x \to \infty$ and as a good approximation for moderate and large values of Pr over the entire turbulent regime. The expressions in terms of Gr_x are somewhat more complicated than those in terms of Ra_x, but are to be preferred since Gr_x is known to characterize the transition from laminar to turbulent motion in natural convection just as Re_D does in forced flow in a channel. The power of speculation combined with dimensional analysis is well demonstrated by this example in which valid asymptotes are thereby attained for several regimes.

Correlation of Experimental Data and Theoretical Values

Correlations of experimental data are generally developed in terms of dimensionless groups rather than in terms of the separate dimensional variables in the interests of compactness and in the hope of greater generality. For example, a complete set of graphical correlations for the heat transfer coefficient h of Example 3.3.2 above in terms of each of the six individual independent variables and physical properties might approach book length, whereas the dimensionless groupings both imply that a single plot with one parameter should be sufficient. Furthermore, the reduced expression for the turbulent regime implies that a plot of $\mathrm{Nu}/\mathrm{Re}\,f^{1/2}$ vs. Pr should demonstrate only a slight parametric dependence on Re or $\mathrm{Re}\,f^{1/2}$. Of course, the availability of a separate correlation for f as a function of Re is implied.

Theoretical values, that is, ones obtained by numerical solution of a mathematical model in terms of either dimensional variables or dimensionless groups, are presumably free from imprecision. Even so, because of their discrete form, the construction of a correlation or correlations for such values may be essential for the same reasons as for experimental data.

Graphical correlations have the merit of revealing general trends, of providing a basis for evaluation of the choice of coordinates, and most of all of displaying visually the scatter of the individual experimental values about a curve representing a correlation or their behavior on the mean. (As mentioned in the previous subsection, the omission of a variable may give the false impression of experimental error in such a plot.) On the other hand, correlating equations are far more convenient as an input to a computer than is a graphical correlation. These two formats thus have distinct and complementary roles; both should generally be utilized. The merits and demerits of various graphical forms of correlations are discussed in detail by Churchill (1979), while the use of logarithmic and arithmetic coordinates, the effects of the appearance of a variable in both coordinates, and the effects of the distribution of error between the dependent and independent variable are further illustrated by Wilkie (1985).

Churchill and Usagi (1972; 1974) proposed general usage of the following expression for the formulation of correlating equations:

$$y^n\{x\} = y_0^n\{x\} + y_\infty^n\{x\} \tag{3.3.6}$$

where $y_o\{x\}$ and $y_\infty\{x\}$ denote asymptotes for small and large values of x, respectively, and n is an arbitrary exponent. For convenience and simplicity, Equation (3.3.6) may be rearranged in either of the following two forms:

$$\left(Y(x)\right)^n = 1 + Z^n\{x\} \tag{3.3.7}$$

or

$$\left(\frac{Y\{x\}}{Z\{x\}}\right)^n = 1 + \frac{1}{Z^n\{x\}} \tag{3.3.8}$$

where $Y\{x\} \equiv y\{x\}/y_o\{x\}$ and $Z\{x\} \equiv y_\infty\{x\}/y_o\{x\}$. Equations (3.3.6), (3.3.7), and (3.3.9) are hereafter denoted collectively as the CUE (Churchill–Usagi equation). The principle merits of the CUE as a canonical expression for correlation are its simple form, generality, and minimal degree of explicit empiricism, namely, only that of the exponent n, since the asymptotes $y_o\{x\}$ and $y_\infty\{x\}$ are ordinarily known in advance from theoretical considerations or well-established correlations. Furthermore, as will be shown, the CUE is quite insensitive to the numerical value of n. Although the CUE is itself very simple in form, it is remarkably successful in representing closely very complex behavior, even including the dependence on secondary variables and parameters, by virtue of the introduction of such dependencies through $y_o\{x\}$ and $y_\infty\{x\}$. In the rare instances in which such dependencies are not represented in the asymptotes, n may be correlated as a function of the secondary variables and/or parameters. Although

the CUE usually produces very close representations, it is empirical and not exact. In a few instances, numerical values of n have been derived or rationalized on theoretical grounds, but even then some degree of approximation is involved. Furthermore, the construction of a correlating expression in terms of the CUE is subject to the following severe limitations:

1. The asymptotes $y_o\{x\}$ and $y_\infty\{x\}$ must intersect once and only once;
2. The asymptotes $y_o\{x\}$ and $y_\infty\{x\}$ must be free of singularities. Even though a singularity occurs beyond the asserted range of the asymptote, it will persist and disrupt the prediction of the CUE, which is intended to encompass all values of the independent variable x; and
3. The asymptotes must both be upper or lower bounds.

In order to avoid or counter these limitations it may be necessary to modify or replace the asymptotes with others. Examples of this process are provided below. A different choice for the dependent variable may be an option in this respect. The suitable asymptotes for use in Equation (3.3.6) may not exist in the literature and therefore may need to be devised or constructed. See, for example, Churchill (1988b) for guidance in this respect. Integrals and derivatives of the CUE are generally awkward and inaccurate, and may include singularities not present or troublesome in the CUE itself. It is almost always preferable to develop a separate correlating equation for such quantities using derivatives or integrals of $y_o\{x\}$ and $y_\infty\{x\}$, simplified or modified as appropriate.

The Evaluation of *n*

Equation (3.3.6) may be rearranged as

$$n = \frac{\ln\left\{1 + \left(\dfrac{y_\infty\{x\}}{y_0\{x\}}\right)^n\right\}}{\ln\left\{\dfrac{y\{x\}}{y_0\{x\}}\right\}} \tag{3.3.9}$$

and solved for n by iteration for any known value of $y\{x\}$, presuming that $y_o\{x\}$ and $y_\infty\{x\}$ are known. If $y\{x^*\}$ is known, where x^* represents the value of x at the point of intersection of the asymptotes, that is, for $y_o\{x\} = y_\infty\{x\}$, Equation (3.3.9) reduces to

$$n = \frac{\ln\{2\}}{\ln\left\{\dfrac{y\{x^*\}}{y_0\{x^*\}}\right\}} \tag{3.3.10}$$

and iterative determination of n is unnecessary.

A graphical and visual method of evaluation of n is illustrated in Figure 3.3.1 in which $Y\{Z\}$ is plotted vs. Z for $0 \le Z \le 1$ and $Y\{Z\}/Z$ vs. $1/Z$ for $0 \le 1/Z \le 1$ in arithmetic coordinates with n as a parameter. Values of $y\{x\}$ may be plotted in this form and the best overall value of n selected visually (as illustrated in Figure 3.3.2). A logarithmic plot of $Y\{Z\}$ vs. Z would have less sensitivity relative to the dependence on n. (See, for example, Figure 1 of Churchill and Usagi, 1972.) Figure 3.3.1 explains in part the success of the CUE. Although y and x may both vary from 0 to ∞, the composite variables plotted in Figure 3.3.1 are highly constrained in that the compound independent variables Z and $1/Z$ vary only between 0 and 1, while for $n \ge 1$, the compound dependent variables $Y\{Z\}$ and $Y\{Z\}/Z$ vary only from 1 to 2.

Because of the relative insensitivity of the CUE to the numerical value of n, an integer or a ratio of two small integers may be chosen in the interest of simplicity and without significant loss of accuracy. For example, the maximum variance in Y (for $0 \le Z \le 1$) occurs at $Z = 1$ and increases only $100(2^{1/20} - 1) = 3.5\%$ if n is decreased from 5 to 4. If $y_o\{x\}$ and $y_\infty\{x\}$ are both lower bounds, n will be positive,

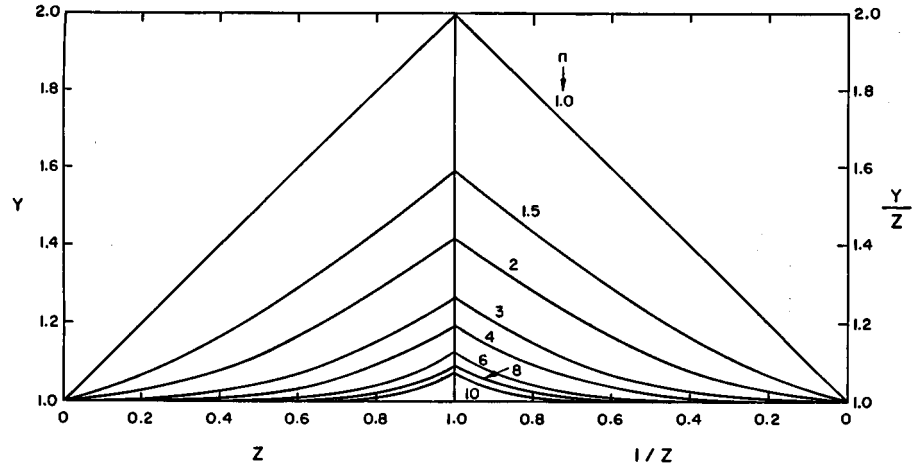


FIGURE 3.3.1 Arithmetic, split-coordinate plot of Equation 3.3.10. (From Churchill, S.W. and Usagi, R. *AIChE J.* 18(6), 1123, 1972. With permission from the American Institute of Chemical Engineers.)

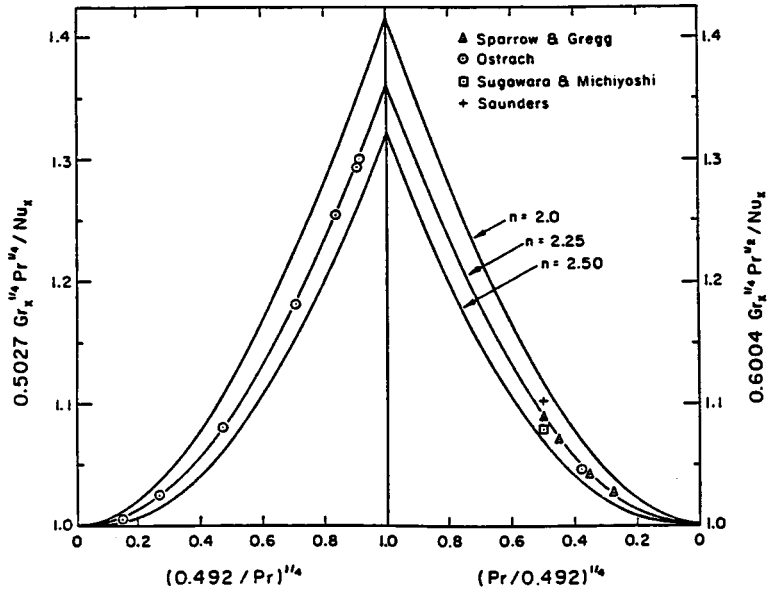


FIGURE 3.3.2 Arithmetic, split-coordinate plot of computed values and experimental data for laminar free convection from an isothermal vertical plate. (From Churchill, S.W. and Usagi, R. *AIChE J.* 18(6), 1124, 1972. With permission from the American Institute of Chemical Engineers.)

and if they are both upper bounds, n will be negative. To avoid extending Figure 3.3.1 for negative values of n, $1/y\{x\}$ may simply be interpreted as the dependent variable.

Intermediate Regimes

Equations (3.3.6), (3.3.7), and (3.3.8) imply a slow, smooth transition between $y_o\{x\}$ and $y_\infty\{x\}$ and, moreover, one that is symmetrical with respect to $x^*(Z = 1)$. Many physical systems demonstrate instead a relatively abrupt transition, as for example from laminar to turbulent flow in a channel or along a flat plate. The CUE may be applied serially as follows to represent such behavior if an expression $y_i\{x\}$ is

postulated for the intermediate regime. First, the transition from the initial to the intermediate regime is represented by

$$y_1^n = y_0^n + y_i^n \tag{3.3.11}$$

Then the transition from this combined regime to the final regime by

$$y^m = y_1^m + y_\infty^m = \left(y_0^n + y_i^n\right)^{m/n} + y_\infty^m \tag{3.3.12}$$

Here, and throughout the balance of this subsection, in the interests of simplicity and clarity, the functional dependence of all the terms on x is implied rather written out explicitly. If y_o is a lower bound and y_i is implied to be one, y_1 and y_∞ must be upper bounds. Hence, n will then be positive and m negative. If y_o and y_i are upper bounds, y_1 and y_∞ must be lower bounds; then n will be negative and m positive. The reverse formulation starting with y_∞ and y_1 leads by the same procedure to

$$y^n = y_0^n + \left(y_i^m + y_\infty^m\right)^{n/m} \tag{3.3.13}$$

If the intersections of y_i with y_o and y_∞ are widely separated with respect to x, essentially the same pair of values for n and m will be determined for Equations (3.3.12) and (3.3.13), and the two representations for y will not differ significantly. On the other hand, if these intersections are close in terms of x, the pair of values of m and n may differ significantly and one representation may be quite superior to the other. In some instances a singularity in y_o or y_∞ may be tolerable in either Equation (3.3.12) or (3.3.13) because it is overwhelmed by the other terms. Equations (3.3.12) and (3.3.13) have one hidden flaw. For $x \to 0$, Equation (3.3.12) reduces to

$$y \to y_0\left[1 + \left(\frac{y_\infty}{y_0}\right)^m\right]^{1/m} \tag{3.3.14}$$

If y_o is a lower bound, m is necessarily negative, and values of y less than y_o are predicted. If y_o/y_∞ is sufficiently small or if m is sufficiently large in magnitude, this discrepancy may be tolerable. If not, the following alternative expression may be formulated, again starting from Equation (3.3.11):

$$\left(y^n - y_0^n\right)^m = y_i^{nm} + \left(y_\infty^n - y_0^n\right)^m \tag{3.3.15}$$

Equation (3.3.15) is free from the flaw identified by means of Equation (3.3.14) and invokes no additional empiricism, but a singularity may occur at $y_\infty = y_o$, depending on the juxtapositions of y_o, y_i, and y_∞. Similar anomalies occur for Equation (3.3.13) and the corresponding analog of Equation (3.3.14), as well as for behavior for which $n < 0$ and $m > 0$. The preferable form among these four is best chosen by trying each of them.

One other problem with the application of the CUE for a separate transitional regime is the formulation of an expression for $y_i\{x\}$, which is ordinarily not known from theoretical considerations. Illustrations of the empirical determination of such expressions for particular cases may be found in Churchill and Usagi (1974), Churchill and Churchill (1975), and Churchill (1976; 1977), as well as in Example 3.3.5 below.

Example 3.3.4: The Pressure Gradient in Flow through a Packed Bed of Spheres

The pressure gradient at asymptotically low rates of flow (the creeping regime) can be represented by the Kozeny–Carman equation, $\Phi = 150\ \mathrm{Re}_p$, and at asymptotically high rates of flow (the inertial regime)

by the Burke–Plummer equation, $\Phi = 1.75\ (\mathrm{Re}_p)^2$, where $\Phi = \rho\varepsilon^2 d_p(-dP_f/dx)\mu^2(1-\varepsilon)$, $\mathrm{Re}_p = d_p u_o \rho/\mu(1-\varepsilon)$, d_p = diameter of spherical particles, m, ε = void fraction of bed of spheres, dP_f/dx = dynamic pressure gradient (due to friction), Pa/m, and u_o = superficial velocity (in absence of the spheres), m/sec. For the origin of these two asymptotic expressions see Churchill (1988a). They both have a theoretical structure, but the numerical coefficients of 150 and 1.75 are basically empirical. These equations are both lower bounds and have one intersection. Experimental data are plotted in Figure 3.3.3, which has the form of Figure 3.3.1 with $Y = \Phi/150\ \mathrm{Re}_p$, $Y/Z = \Phi/(1.75\ \mathrm{Re}_p)^2$ and $Z = 1.75\ \mathrm{Re}_p^2/150\ \mathrm{Re}_p = \mathrm{Re}_p/85.7$. A value of $n = 1$ is seen to represent these data reasonably well on the mean, resulting in

$$\Phi = 150\mathrm{Re}_p + 1.75\left(\mathrm{Re}_p\right)^2$$

which was originally proposed as a correlating equation by Ergun (1952) on the conjecture that the volumetric fraction of the bed in "turbulent" flow is proportional to Re_p. The success of this expression in conventional coordinates is shown in Figure 3.3.4. The scatter, which is quite evident in the arithmetic split coordinates of Figure 3.3.3, is strongly suppressed in a visual sense in the logarithmic coordinates of Figure 3.3.4.

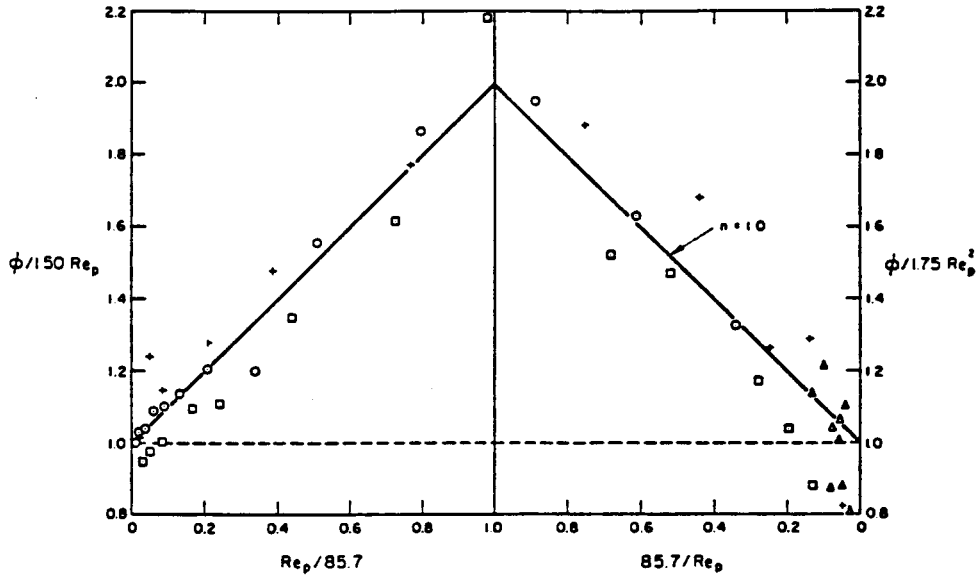


FIGURE 3.3.3 Arithmetic, split-coordinate plot of experimental data for the pressure drop in flow through a packed bed of spheres. (From Churchill, S.W. and Usagi, R. *AIChE J.* 18(6), 1123, 1972. With permission from the American Institute of Chemical Engineers.)

Example 3.3.5: The Friction Factor for Commercial Pipes for All Conditions

The serial application of the CUE is illustrated here by the construction of a correlating equation for both smooth and rough pipes in the turbulent regime followed by combination of that expression with ones for the laminar and transitional regimes.

The Turbulent Regime. The Fanning friction factor, f_F, for turbulent flow in a smooth round pipe for asymptotically large rates of flow (say $\mathrm{Re}_D > 5000$) may be represented closely by the empirical expression:

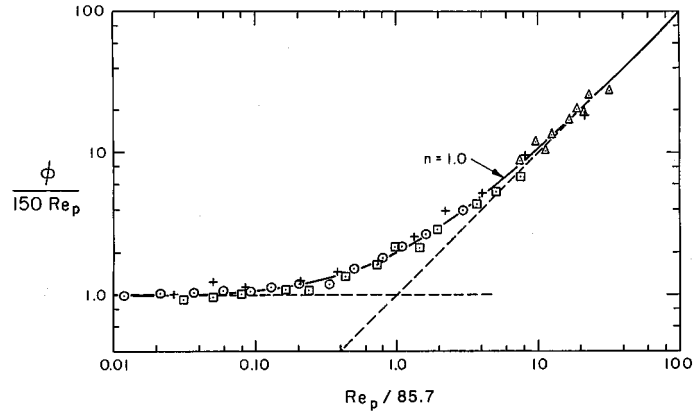


FIGURE 3.3.4 Logarithmic correlation of experimental data for the pressure drop in flow through a packed bed of spheres. (From Churchill, S.W. and Usagi, R. *AIChE J.* 18(6), 1123, 1972. With permission from the American Institute of Chemical Engineers.)

$$\left(\frac{2}{f_F}\right)^{1/2} = 0.256 + 2.5\ln\left\{\left(\frac{f_F}{2}\right)^{1/2}\mathrm{Re}_D\right\}$$

A corresponding empirical representation for naturally rough pipe is

$$\left(\frac{2}{f_F}\right)^{1/2} = 3.26 + 2.5\ln\left\{\frac{D}{e}\right\}$$

Direct combination of these two expressions in the form of the CUE does not produce a satisfactory correlating equation, but their combination in the following rearranged forms:

$$e^{(1/2.5)(2/f_F)^{1/2}} = 1.108\left(\frac{f_F}{2}\right)^{1/2}\mathrm{Re}_D$$

and

$$e^{(1/2.5)(2/f_F)^{1/2}} = 3.68\left(\frac{D}{e}\right)$$

with $n = -1$ results in, after the reverse rearrangement,

$$\left(\frac{2}{f_F}\right)^{1/2} = 0.256 + 2.5\ln\left\{\frac{\left(\frac{f_F}{2}\right)^{1/2}\mathrm{Re}_D}{1 + 0.3012\left(\frac{e}{D}\right)\left(\frac{f_F}{2}\right)^{1/2}\mathrm{Re}_D}\right\}$$

The exact equivalent of this expression in structure but with the slightly modified numerical coefficients of 0.300, 2.46, and 0.304 was postulated by Colebrook (1938–1939) to represent his own experimental data. The coefficients of the expression given here are presumed to be more accurate, but the difference in the predictions of f_F with the two sets of coefficients is within the band of uncertainty of the

experimental data. The turbulent regime of the "friction-factor" plot in most current textbooks and handbooks is simply a graphical representation of the Colebrook equation. Experimental values are not included in such plots since e, the effective roughness of commercial pipes, is simply a correlating factor that forces agreement with the Colebrook equation. Values of e for various types of pipe in various services are usually provided in an accompanying table, that thereby constitutes an integral part of the correlation.

The Laminar Region. The Fanning friction factor in the laminar regime of a round pipe ($\mathrm{Re}_d < 1800$) is represented exactly by the following theoretical expression known as Poiseuille's law: $f_F = 16/\mathrm{Re}_D$. This equation may be rearranged as follows for convenience in combination with that for turbulent flow:

$$\left(\frac{2}{f_F}\right)^{1/2} = \frac{\mathrm{Re}_D\left(f_F/2\right)^{1/2}}{8}$$

The Transitional Regime. Experimental data as well as semitheoretical computed values for the limiting behavior in the transition may be represented closely by $(f_F/2) = (\mathrm{Re}_D/37500)^2$. This expression may be rewritten, in terms of $(2/f_F)^{1/2}$ and $\mathrm{Re}_D(f_F/2)^{1/2}$, as follows:

$$\left(\frac{f_F}{2}\right)^{1/2} = \left(\frac{37500}{\mathrm{Re}_D\left(f_F/2\right)^{1/2}}\right)^{1/2}$$

Overall Correlation. The following correlating equation for all $\mathrm{Re}_D(f_F/2)^{1/2}$ and e/D may now be constructed by the combination of the expressions for the turbulent and transition regimes in the form of the CUE with $n = 8$, and then that expression and that for the laminar regime with $n = -12$, both components being chosen on the basis of experimental data and predicted values for the full regime of transition:

$$\left(\frac{2}{f_F}\right)^{1/2} = \left[\left(\frac{8}{\mathrm{Re}_D\left(f_F/2\right)^{1/2}}\right)^{12} + \left(\left[\frac{37500}{\mathrm{Re}_D\left(f_F/2\right)^{1/2}}\right]^4 + \left|2.5\ln\left\{\frac{1.108\,\mathrm{Re}_D\left(f_F/2\right)^{1/2}}{1+0.3012\left(\frac{e}{a}\right)\mathrm{Re}_D\left(f_F/2\right)^{1/2}}\right\}\right|^8\right)^{-3/2}\right]^{-1/12}$$

The absolute value signs are only included for aesthetic reasons; the negative values of the logarithmic term for very small values of $\mathrm{Re}_D(f_F/2)^{1/2}$ do not affect the numerical value of $(2/f_F)^{1/2}$ in the regime in which they occur. This overall expression appears to have a complicated structure, but it may readily be recognized to reduce to its component parts when the corresponding term is large with respect to the other two. It is insensitive to the numerical values of the two arbitrary exponents. For example, doubling their values would have almost no effect on the predictions of $(f_F/2)^{1/2}$. The principal uncertainty is associated with the expression for the transition regime, but the overall effect of the corresponding term is very small. The uncertainties associated with this correlating equation are common to most graphical correlations and algebraic expressions for the friction factor, and are presumed to be fairly limited in magnitude and to be associated primarily with the postulated value of e. Although the overall expression is explicit in $\mathrm{Re}_D(f_F/2)^{1/2}$ rather than Re_D, the latter quantity may readily be obtained simply by multiplying the postulated value of $\mathrm{Re}_D(f_F/2)^{1/2}$ by the computed values of $(2/f_F)^{1/2}$.

References

Buckingham, E. 1914. On physically similar systems; illustrations of the use of dimensional equations. *Phys. Rev., Ser. 2,* 4(4):345–375.

Churchill, S.W. 1976. A comprehensive correlating equation for forced convection from plates. *AIChE J.* 22(2):264–268.

Churchill, S.W. 1977. Comprehensive correlating equation for heat, mass and momentum transfer in fully developed flow in smooth tubes. *Ind. Eng. Chem. Fundam.* 16(1):109–116.

Churchill, S.W. 1979. *The Interpretation and Use of Rate Data. The Rate Process Concept,* rev. printing, Hemisphere Publishing Corp., Washington, D.C.

Churchill, S.W. 1981. The use of speculation and analysis in the development of correlations. *Chem. Eng. Commun.* 9:19–38.

Churchill, S.W. 1988a. Flow through porous media, Chapter 19 in *Laminar Flows. The Practical Use of Theory,* pp. 501–538, Butterworths, Boston.

Churchill, S.W. 1988b. Derivation, selection, evaluation and use of asymptotes. *Chem. Eng. Technol.* 11:63–72.

Churchill, S.W. and Churchill, R.U. 1975. A general model for the effective viscosity of pseudoplastic and dilatant fluids. *Rheol. Acta.* 14:404–409.

Churchill, S.W. and Usagi, R. 1972. A general expression for the correlation of rates of transfer and other phenomena. *AIChE J.* 18(6):1121–1128.

Churchill, S.W. and Usagi, R. 1974. A standardized procedure for the production of correlations in the form of a common empirical equation. *Ind. Eng. Chem. Fundam.* 13(1):39–44.

Colebrook, C.R. 1938–1939. Turbulent flow in pipes with particular reference to the transition region between the smooth and rough pipe laws. *J. Inst. Civ. Eng.* 11(5024):133–156.

Ergun, S. 1952. Fluid flow through packed beds. *Chem. Eng. Prog.* 48(2):81–96.

Hellums, J.D. and Churchill, S.W. 1964. Simplifications of the mathematical description of boundary and initial value problems. *AIChE J.* 10(1):110–114.

Wilkie, D. 1985. The correlation of engineering data reconsidered. *Int. J. Heat Fluid Flow.* 8(2):99–103.

Zlokarnik, M. 1991. *Dimensional Analysis and Scale-Up in Chemical Engineering*. Springer-Verlag, Berlin.

3.4 Hydraulics of Pipe Systems

J. Paul Tullis

Basic Computations

Equations

Solving fluid flow problems involves the application of one or more of the three basic equations: continuity, momentum, and energy. These three basic tools are developed from the law of conservation of mass, Newton's second law of motion, and the first law of thermodynamics.

The simplest form of the continuity equation is for one-dimensional incompressible steady flow in a conduit. Applying continuity between any two sections gives

$$A_1V_1 = A_2V_2 = Q \tag{3.4.1}$$

For a variable density the equation can be written

$$\rho_1 A_1 V_1 = \rho_2 A_2 V_2 = \dot{m} \tag{3.4.2}$$

in which A is the cross-sectional area of the pipe, V is the mean velocity at that same location, Q is the flow rate, ρ is the fluid density, and $\dot{m}$ is the mass flow rate. The equations are valid for any rigid conduit as long as there is no addition or loss of liquid between the sections.

For steady state pipe flow, the momentum equation relates the net force in a given direction (F_x) acting on a control volume (a section of the fluid inside the pipe), to the net momentum flux through the control volume.

$$F_x = \rho_2 A_2 V_2 V_{2x} - \rho_1 A_1 V_1 V_{1x} \tag{3.4.3}$$

For incompressible flow this equation can be reduced to

$$F_x = \rho Q\left(V_{2x} - V_{1x}\right) \tag{3.4.4.}$$

These equations can easily be applied to a three-dimensional flow problem by adding equations in the y and z directions.

A general form of the energy equation (see Chapter 2) applicable to incompressible pipe or duct flow

$$\frac{P_1}{\gamma} + Z_1 + \frac{V_1^2}{2g} = \frac{P_2}{\gamma} + Z_2 + \frac{V_2^2}{2g} - H_p + H_t + H_f \tag{3.4.5}$$

The units are energy per unit weight of liquid: ft · lb/lb or N · m/N which reduce to ft or m. The first three terms are pressure head (P/γ), elevation head (Z) (above some datum), and velocity head ($V^2/2g$). The last three terms on the right side of the equation are the total dynamic head added by a pump (H_p) or removed by a turbine (H_t) and the friction plus minor head losses (H_f). The sum of the first three terms in Equation 3.4.5 is defined as the total head, and the sum of the pressure and elevation heads is referred to as the piezometric head.

The purpose of this section is to determine the pressure changes resulting from incompressible flow in pipe systems. Since pipes of circular cross sections are most common in engineering application, the analysis in this section will be performed for circular geometry. However, the results can be generalized for a pipe of noncircular geometry by substituting for the diameter D in any of the equations, the hydraulic diameter, D_h, defined as

$$D_h = 4 \times \frac{\text{the cross sectional area}}{\text{the wetted perimeter}}$$

The analysis in this section can also be applied to gases and vapors, provided the Mach number in the duct does not exceed 0.3. For greater values of the Mach number, the compressibility effect becomes significant and the reader is referred to Section 3.7 on compressible flow.

Fluid Friction

The calculation of friction loss in pipes and ducts depends on whether the flow is laminar or turbulent. The Reynolds number is the ratio of inertia forces to viscous forces and is a convenient parameter for predicting if a flow condition will be laminar or turbulent. It is defined as

$$\text{Re}_D = \frac{\rho VD}{\mu} = \frac{VD}{\nu} \tag{3.4.6}$$

in which V is the mean flow velocity, D diameter, ρ fluid density, μ dynamic viscosity, and ν kinematic viscosity.

Friction loss (H_f) depends on pipe diameter (d), length (L), roughness (e), fluid density (ρ) or specific weight (γ), viscosity (ν), and flow velocity (V). Dimensional analysis can be used to provide a functional relationship between the friction loss H_f, pipe dimensions, fluid properties, and flow parameters. The resulting equation is called the Darcy–Weisbach equation:

$$H_f = \frac{fLV^2}{2gd} = \frac{fLQ^2}{1.23gD^5} \tag{3.4.7}$$

The friction factor f is a measure of pipe roughness. It has been evaluated experimentally for numerous pipes. The data wee used to create the Moody friction factor chart shown as Figure 3.4.1. For Re < 2000, the flow in a pipe will be laminar and f is only a function of Re_D. It can be calculated by

$$f = \frac{64}{\text{Re}_D} \tag{3.4.8}$$

At Reynolds numbers between about 2000 and 4000 the flow is unstable as a result of the onset of turbulence (critical zone in Figure 3.4.1). In this range, friction loss calculations are difficult because it is impossible to determine a unique value of f. For Re > 4000 the flow becomes turbulent and f is a function of both Re and relative pipe roughness (e/d). At high Re, f eventually depends only on e/d; defining the region referred to as fully turbulent flow. This is the region in Figure 3.4.1 where the lines for different e/d become horizontal (e is the equivalent roughness height and d pipe diameter). The Re_D at which this occurs depends on the pipe roughness. Laminar flow in pipes is unusual. For example, for water flowing in a 0.3-m-diameter pipe, the velocity would have to be below 0.02 m/sec for laminar flow to exist. Therefore, most practical pipe flow problems are in the turbulent region.

Using the Moody chart in Figure 3.4.1 to get f requires that Re and e/d be known. Calculating Re is direct if the water temperature, velocity, and pipe diameter are known. The problem is obtaining a good value for e. Typical values of e are listed in Figure 3.4.1. These values should be considered as guides only and not used if more–exact values can be obtained from the pipe supplier.

Since roughness may vary with time due to buildup of solid deposits or organic growths, f is also time dependent. Manufacturing tolerances also cause variations in the pipe diameter and surface roughness. Because of these factors, the friction factor for any pipe can only be approximated. A designer is required to use good engineering judgment in selecting a design value for f so that proper allowance is made for these uncertainties.

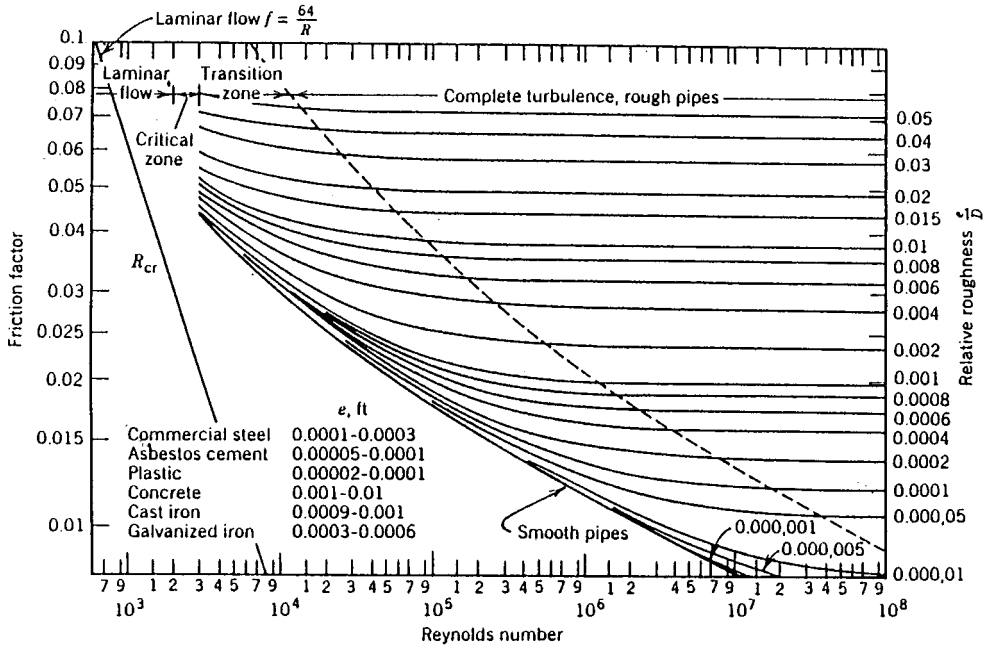


FIGURE 3.4.1 The Moody diagram.

For noncircular pipes, the only change in the friction loss equation is the use of an equivalent diameter — based on the hydraulic radius (R), i.e., $d = 4R$ — in place of the circular pipe diameter d. R is the ratio of the flow area to the wetter perimeter.

Wood (1966) developed equations which can be used in place of the Moody diagram to estimate f for $\mathrm{Re} > 10^4$ and $10^{-5} < k < 0.04$ ($k = e/d$).

$$f = a + b\mathrm{Re}^{-c} \tag{3.4.9}$$

$$a = 0.094k^{0.225} + 0.53k, \quad b = 88k^{0.44}, \quad c = 1.62k^{0.134}$$

The practical problem is still obtaining a reliable value for e. It cannot be directly measured but must be determined from friction loss tests of the pipe.

An exact solution using the Darcy–Weisbach equation can require a trial-and-error solution because of the dependency of f on Re if either the flow or pipe diameter are not known. A typical approach to solving this problem is to estimate a reasonable fluid velocity to calculate Re and obtain f from the Moody chart or Equation (3.4.9). Next, calculate a new velocity and repeat until the solution converges. Converging on a solution is greatly simplified with programmable calculators and a variety of software available for computer.

For long gravity flow pipelines, the starting point in selecting the pipe diameter is to determine the smallest pipe that can pass the required flow without friction loss exceeding the available head. For pumped systems, the selection must be based on an economic analysis that compares the pipe cost with the cost of building and operating the pumping plant.

Local Losses

Flow through valves, orifices, elbows, transitions, etc. causes flow separation which results in the generation and dissipation of turbulent eddies. For short systems containing many bends, valves, tees,

etc. local or minor losses can exceed friction losses. The head loss h_l associated with the dissipation caused by a minor loss is proportional to the velocity head and can be accounted for as a minor or local loss using the following equation.

$$h_l = K_l \frac{Q^2}{\left(2gA_m^2\right)} \tag{3.4.10}$$

in which K_l is the minor loss coefficient and A_m is the area of the pipe at the inlet to the local loss. The loss coefficient K_l is analogous to fL/d in Equation 3.4.7.

The summation of all friction and local losses in a pipe system can be expressed as

$$H_f = h_f + h_l \tag{3.4.11}$$

$$H_f = \left[\sum\left(\frac{fL}{2gdA_p^2}\right) + \sum\left(\frac{K_l}{2gA_m^2}\right)\right]Q^2 = CQ^2 \tag{3.4.12}$$

in which

$$C = \sum\left(\frac{fL}{2gdA_p^2}\right) + \sum\left(\frac{K_l}{2gA_m^2}\right) \tag{3.4.13}$$

It is important to use the correct pipe diameter for each pipe section and local loss.

In the past some have expressed the local losses as an equivalent pipe length: $L/d = K_l/f$. It simply represents the length of pipe that produces the same head loss as the local or minor loss. This is a simple, but not a completely accurate method of including local losses. The problem with this approach is that since the friction coefficient varies from pipe to pipe, the equivalent length will not have a unique value. When local losses are truly minor, this problem becomes academic because the error only influences losses which make up a small percentage of the total. For cases where accurate evaluation of all losses is important, it is recommended that the minor loss coefficients K_l be used rather than an equivalent length.

The challenging part of making minor loss calculations is obtaining reliable values of K_l. The final results cannot be any more accurate than the input data. If the pipe is long, the friction losses may be large compared with the minor losses and approximate values of K_l will be sufficient. However, for short systems with many pipe fittings, the local losses can represent a significant portion of the total system losses, and they should be accurately determined. Numerous factors influence K_l. For example, for elbows, K_l is influenced by the shape of the conduit (rectangular vs. circular), by the radius of the bend, the bend angle, the Reynolds number, and the length of the outlet pipe. For dividing or combining tees or Y-branches, the percent division of flow and the change in pipe diameter must also be included when estimating K_l. One factor which is important for systems where local losses are significant is the interaction between components placed close together. Depending on the type, orientation, and spacing of the components, the total loss coefficient may be greater or less than the simple sum of the individual K_l values.

Comparing the magnitude of $\Sigma(fL/2gA^2)$ to $\Sigma(K_l/2gA_m^2)$ will determine how much care should be given to the selection of the K_l values. Typical values of K_l are listed in Table 3.4.1 (Tullis, 1989). When more comprehensive information on loss coefficients is needed, the reader is referred to Miller (1990).

TABLE 3.4.1 Minor Loss Coefficients

Item	K_l Typical Value	Typical Range
Pipe inlets		
Inward projecting pipe	0.78	0.5–0.9
Sharp corner-flush	0.50	—
Slightly rounded	0.20	0.04–0.5
Bell mouth	0.04	0.03–0.1
Expansions[a]	$(1 - A_1/A_2)^2$ (based on V_1)	
Contractions[b]	$(1/C_c - 1)^2$ (based on V_2)	

A_2/A_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
C_c	0.624	0.632	0.643	0.659	0.681	0.712	0.755	0.813	0.892

Bends[c]		
Short radius, $r/d = 1$		
90	—	0.24
45	—	0.1
30	—	0.06
Long radius, $r/d = 1.5$		
90	—	0.19
45	—	0.09
30	—	0.06
Mitered (one miter)		
90	1.1	—
60	0.50	0.40–0.59
45	0.3	0.35–0.44
30	0.15	0.11–0.19
Tees	[c]	—
Diffusers	[c]	—
Valves		
Check valve	0.8	0.5–1.5
Swing check	1.0	0.29–2.2
Tilt disk	1.2	0.27–2.62
Lift	4.6	0.85–9.1
Double door	1.32	1.0–1.8
Full-open gate	0.15	0.1–0.3
Full-open butterfly	0.2	0.2–0.6
Full-open globe	4.0	3–10

[a] See Streeter and Wylie, 1975, p. 304.
[b] See Streeter and Wylie, 1975, p. 305.
[c] See Miller, 1990.
[d] See Kalsi Engineering and Tullis Engineering Consultants, 1993.

Pipe Design

Pipe Materials

Materials commonly used for pressure pipe transporting liquids are ductile iron, concrete, steel, fiberglass, PVC, and polyolefin. Specifications have been developed by national committees for each of these pipe materials. The specifications discuss external loads, internal design pressure, available sizes, quality of materials, installation practices, and information regarding linings. Standards are available from the following organizations:

American Water Works Association (AWWA)
American Society for Testing and Materials (ASTM)
American National Standards Institute (ANSI)
Canadian Standards Association (CSA)

Federal Specifications (FED)
Plastic Pipe Institute (PPI)

In addition, manuals and other standards have been published by various manufacturers and manufacturer's associations. All of these specifications and standards should be used to guide the selection of pipe material. ASCE (1992) contains a description of each of these pipe materials and a list of the specifications for the various organizations which apply to each material. It also discusses the various pipe-lining materials available for corrosion protection.

For air- and low-pressure liquid applications one can use unreinforced concrete, corrugated steel, smooth sheet metal, spiral rib (sheet metal), and HDPE (high-density polyethylene) pipe. The choice of a material for a given application depends on pipe size, pressure requirements, resistance to collapse from internal vacuums, external loads, resistance to internal and external corrosion, ease of handling and installing, useful life, and economics.

Pressure Class Guidelines

Procedures for selecting the pressure class of pipe vary with the type of pipe material. Guidelines for different types of materials are available from AWWA, ASTM, ANSI, CSA, FED, PPI and from the pipe manufacturers. These specifications should be obtained and studied for the pipe materials being considered.

The primary factors governing the selection of a pipe pressure class are (1) the maximum steady state operating pressure, (2) surge and transient pressures, (3) external earth loads and live loads, (4) variation of pipe properties with temperature or long-time loading effects, and (5) damage that could result from handling, shipping, and installing or reduction in strength due to chemical attack or other aging factors. The influence of the first three items can be quantified, but the last two are very subjective and are generally accounted for with a safety factor which is the ratio of the burst pressure to the rated pressure.

There is no standard procedure on how large the safety factor should be or on how the safety factor should be applied. Some may feel that it is large enough to account for all of the uncertainties. Past failures of pipelines designed using this assumption prove that it is not always a reliable approach. The procedure recommended by the author is to select a pipe pressure class based on the internal design pressure (IDP) defined as

$$\text{IDP} = \left(P_{\max} + P_s\right)\text{SF} \tag{3.4.14}$$

in which $P_{\max}$ is the maximum steady state operating pressure, P_s is the surge or water hammer pressure, and SF is the safety factor applied to take care of the unknowns (items 3 to 5) just enumerated. A safety factor between 3 and 4 is typical.

The maximum steady state operating pressure ($P_{\max}$) in a gravity flow system is usually the difference between the maximum reservoir elevation and the lowest elevation of the pipe. For a pumped system it is usually the pump shutoff head calculated based on the lowest elevation of the pipe.

Surge and transient pressures depend on the specific pipe system design and operation. Accurately determining P_s requires analyzing the system using modern computer techniques. The most commonly used method is the "Method of Characteristics" (Tullis, 1989; Wylie and Streeter, 1993). Some of the design standards give general guidelines to predict P_s that can be used if a detailed transient analysis is not made. However, transients are complex enough that simple "rules of thumb" are seldom accurate enough. Transients are discussed again in a later subsection.

Selection of wall thickness for larger pipes is often more dependent on collapse pressure and handling loads than it is on burst pressure. A thin-wall, large-diameter pipe may be adequate for resisting relatively high internal pressures but may collapse under negative internal pressure or, if the pipe is buried, the soil and groundwater pressure plus live loads may be sufficient to cause collapse even if the pressure inside the pipe is positive.

External Loads

There are situations where the external load is the controlling factor determining if the pipe will collapse. The magnitude of the external load depends on the diameter of the pipe, the pipe material, the ovality (out of roundness) of the pipe cross section, the trench width, the depth of cover, the specific weight of the soil, the degree of soil saturation, the type of backfill material, the method used to backfill, the degree of compaction, and live loads. The earth load increases with width and depth of the trench, and the live load reduces with depth of cover. The cumulative effect of all these sources of external loading requires considerable study and analysis.

There are no simple guidelines for evaluating external pipe loads. Because of the complexity of this analysis, the default is to assume that the safety factor is adequate to account for external loads as well as the other factors already mentioned. One should not allow the safety factor to replace engineering judgment and calculations. One option to partially compensate for the lack of a detailed analysis is to use a higher-pressure class of pipe in areas where there will be large live loads or where the earth loading is unusually high. One should consider the cost of a pipe failure caused by external loads compared with the cost of using a thicker pipe or the cost of performing a detailed analysis. Those interested in the details of performing calculations of earth loading should be Spranger and Handy, 1973.

Limiting Velocities

There are concerns about upper and lower velocity limits. If the velocity is too low, problems may develop due to settling of suspended solids and air being trapped at high points and along the crown of the pipe. The safe lower velocity limit to avoid collecting air and sediment depends on the amount and type of sediment and on the pipe diameter and pipe profile. Velocities greater than about 1 m/sec (3 ft/sec) are usually sufficient to move trapped air to air release valves and keep the sediment in suspension.

Problems associated with high velocities are (1) erosion of the pipe wall or liner (especially if coarse suspended sediment is present), (2) cavitation at control valves and other restrictions, (3) increased pumping costs, (4) removal of air at air release valves, (5) increased operator size and concern about valve shaft failures due to excessive flow torques, and (6) an increased risk of hydraulic transients. Each of these should be considered before making the final pipe diameter selection. A typical upper velocity for many applications if 6 m/sec (20 ft/sec). However, with proper pipe design and analysis (of the preceding six conditions), plus proper valve selection, much higher velocities can be tolerated. A typical upper velocity limit for standard pipes and valves is about 6 m/sec (20 ft/sec). However, with proper design and analysis, much higher velocities can be tolerated.

Valve Selection

Valves serve a variety of functions. Some function as isolation or block valves that are either full open or closed. Control valves are used to regulate flow or pressure and must operate over a wide range of valve openings. Check valves prevent reverse flow, and air valves release air during initial filling and air that is collected during operation and admit air when the pipe is drained.

Control Valves

For many flow control applications it is desirable to select a valve that has linear control characteristics. This means that if you close the valve 10%, the flow reduces about 10%. Unfortunately, this is seldom possible since the ability of a valve to control flow depends as much on the system as it does on the design of the valve. The same valve that operates linearly in one system may not in another.

Selecting the proper flow control valve should consider the following criteria:

1. The valve should not produce excessive pressure drop when full open.
2. The valve should control over at least 50% of its movement.
3. At maximum flow, the operating torque must not exceed the capacity of the operator or valve shaft and connections.
4. The valve should not be subjected to excessive cavitation.

5. Pressure transients should not exceed the safe limits of the system.
6. Some valves should not be operated at very small openings. Other valves should be operated near full open.

Controllability. To demonstrate the relationship between a valve and system, consider a butterfly valve that will be used to control the flow between two reservoirs with an elevation difference of ΔZ. System A is a short pipe (0.3 m dia., 100 m long, ΔZ = 10 m) where pipe friction is small $fL/2gdA_p^2 = 46.9$, and System B is a long pipe (0.3 m dia., 10,000 m long, ΔZ = 200 m) with high friction $fL/2gdA_p^2 =$ 4690. Initially, assume that the same butterfly valve will be used in both pipes and it will be the same size as the pipe. The flow can be calculated using the energy equation (Equation 3.4.5) and the system loss equation (Equation (3.4.12):

$$Q = \sqrt{\frac{\Delta Z}{\left[\sum\left(\frac{fL}{2gdA_p^2}\right) + \sum\left(\frac{K_l}{2gA_m^2}\right)\right]}} \tag{3.4.15}$$

For the valve, assume that the K_l full open is 0.2 and at 50% open it is 9.0. Correspondingly, $K_l/2gA_m^2$ = 1.905 and 85.7. For System A, the flow with the valve full open will be 0.453 m^3/sec and at 50% open 0.275 m^3/sec, a reduction of 39%. Repeating these calculations over the full range of valve openings would show that the flow for System A reduces almost linearly as the valve closes.

For System B, the flow with the valve full open will be 0.206 m^3/sec and at 50% open 0.205 m^3/sec, a reduction of less than 1%. For System B the valve does not control until the valve loss, expressed by $K_l/2gA_m^2$ becomes a significant part of the friction term (4690). The same valve in System B will not start to control the flow until it has closed more than 50%. A line-size butterfly valve is obviously not a good choice for a control valve in System B. One solution to this problem is to use a smaller valve. If the butterfly valve installed in System B was half the pipe diameter, it would control the flow over more of the stroke of the valve.

The range of opening over which the valve controls the flow also has a significant effect on the safe closure time for control valves. Transient pressures are created when there is a sudden change in the flow. Most valve operators close the valve at a constant speed. If the valve does not control until it is more than 50% closed, over half of the closing time is wasted and the effective valve closure time is less than half the total closing time. This will increase the magnitude of the transients that will be generated.

Torque. To be sure that the valve shaft, connections, and operator are properly sized, the maximum torque or thrust must be known. If the maximum force exceeds operator capacity, it will not be able to open and close the valve under extreme flow conditions. If the shaft and connectors are underdesigned, the valve may fail and slam shut causing a severe transient.

For quarter-turn valves, the force required to operate a valve consists of seating, bearing, and packing friction, hydrodynamic (flow) forces, and inertial forces. These forces are best determined experimentally. A key step in applying experimental torque information is the determination of the flow condition creating maximum torque. This requires that the system be analyzed for all possible operating conditions and valve openings. For a given size and type of valve, the flow torque depends on the torque coefficient (which is dependent on the specific valve design) and the pressure drop which, in turn, depends on the flow. In short systems where there is little friction loss and high velocities, a quarter-turn valve will see maximum torques at large openings where the flow is high. In long systems with high reservoir heads and smaller velocities, the same valve will see maximum torque at small openings where the pressure drop is high.

One situation where it is easy to overlook the condition causing maximum torque is with parallel pumps. Each pump normally will have a discharge control valve. The maximum system flow occurs with all three pumps operating. However, the flow and the torque for any of the pump discharge valves

is maximum for only one pump operating. One specific example (Tullis, 1989) showed that the torque on a butterfly valve was three times higher when one pump was operating compared with three pumps operating in parallel.

Cavitation. Cavitation is frequently an important consideration in selection and operation of control valves. It is necessary to determine if cavitation will exist, evaluate its intensity, and estimate its effect on the system and environment. Cavitation can cause noise, vibration, and erosion damage and can decrease performance. The analysis should consider the full range of operation. Some valves cavitate worst at small openings and others cavitate heavily near full open. It depends on both the system and the valve design. If cavitation is ignored in the design and selection of the valves, repairs and replacement of the valves may be necessary. Information for making a complete cavitation analysis is beyond the scope of this section. Detailed information on the process to design for cavitation is contained in Tullis (1989; 1993).

The first step in a cavitation analysis is selecting the acceptable level of cavitation. Experimental data are available for four limits: incipient (light, intermittent noise), critical (light, steady noise), incipient damage (pitting damage begins), and choking (very heavy damage and performance drops off). Limited cavitation data are available for each of these limits (Tullis, 1989; 1993). Choosing a cavitation limit depends on several factors related to the operating requirements, expected life, location of the device, details of the design, and economics. For long-term operation of a control valve in a system where noise can be tolerated, the valve should never operate beyond incipient damage. In systems where noise is objectionable, critical cavitation would be a better operating limit.

Using a choking cavitation as a design limit is often misused. It is generally appropriate as a design limit for valves that only operate for short periods of time, such as a pressure relief valve. The intensity of cavitation and the corresponding noise vibration and erosion damage at the valve are at their maximum just before a valve chokes. If the valve operates beyond choking (sometimes referred to as supercavitation), the collapse of the vapor cavities occurs remote from the valve. Little damage is likely to occur at the valve, but farther downstream serious vibration and material erosion problems can occur.

If the cavitation analysis indicates that the valve, orifice, or other device will be operating at a cavitation level greater than can be tolerated, various techniques can be used to limit the level of cavitation. One is to select a different type of valve. Recent developments in valve design have produced a new generation of valves that are more resistant to cavitation. Most of them operate on the principle of dropping the pressure in stages. They usually have multiple paths with numerous sharp turns or orifices in series. Two limitations to these valves are that they often have high pressure drops (even when full open), and they are only usable with clean fluids.

A similar approach is to place multiple conventional valves in series or a valve in series with orifice plates. Proper spacing of valves and orifices placed in series is important. The spacing between valves depends upon the type. For butterfly valves it is necessary to have between five and eight diameters of pipe between valves to prevent flutter of the leaf of the downstream valve and to obtain the normal pressure drop at each valve. For globe, cone, and other types of valves, it is possible to bolt them flange to flange and have satisfactory operation.

For some applications another way to suppress cavitation is to use a free-discharge valve that is vented so cavitation cannot occur. There are valves specifically designed for this application. Some conventional valves can also be used for free discharge, if they can be adequately vented.

Cavitation damage can be suppressed by plating critical areas of the pipe and valve with cavitation-resistant materials. Based on tests using a magnetostriction device, data show that there is a wide variation in the resistance of the various types of material. Limited testing has been done on the erosion resistance of different materials and coating to cavitation in flowing systems. The available data show that there is less variation in the damage resistance of materials in actual flowing systems. However, experience has shown the plating parts of the valve with the right material will extend valve life.

Injecting air to suppress cavitation is a technique which has been used for many years with varying degrees of success. The most common mistake is placing the air injection port in the wrong location so

the air does not get to the cavitation zone. If an adequate amount of air is injected into the proper region, the noise, vibrations, and erosion damage can be significantly reduced. The air provides a cushioning effect reducing the noise, vibration, and erosion damage. If the system can tolerate some air being injected, aeration is usually the cheapest and best remedy for cavitation.

Transients. Transient pressures can occur during filling and flushing air from the line, while operating valves, and when starting or stopping pumps. If adequate design provisions and operational procedures are not established, the transient pressure can easily exceed the safe operating pressure of the pipe. A system should be analyzed to determine the type and magnitudes of possible hydraulic transients. The basic cause is rapid changes in velocity. The larger the incremental velocity change and the faster that change takes place, the greater will be the transient pressure. If the piping system is not designed to withstand the high transient pressures, or if controls are not included to limit the pressure, rupture of the pipe or damage to equipment can result.

All pipelines experience transients. Whether or not the transient creates operational problems or pipe failure depends upon its magnitude and the ability of the pipes and mechanical equipment to tolerate high pressures without damage. For example, an unreinforced concrete pipeline may have a transient pressure head allowance of only a meter above its operating pressure before damage can occur. For such situations even slow closing of control valves or minor interruptions of flow due to any cause may create sufficient transient pressures to rupture the pipeline. In contrast, steel and plastic pipes can take relatively high transient pressures without failure.

Transients caused by slow velocity changes, such as the rise and fall of the water level in a surge tank, are called surges. Surge analysis, or "rigid column theory" involves mathematical or numerical solution of simple ordinary differential equations. The compressibility of the fluid and the elasticity of the conduit are ignored, and the entire column of fluid is assumed to move as a rigid body.

When changes in velocity occur rapidly, both the compressibility of the liquid and the elasticity of the pipe must be included in the analysis. This procedure is often called "elastic" or "waterhammer" analysis and involves tracking acoustic pressure waves through the pipe. The solution requires solving partial differential equations.

An equation predicting the head rise ΔH caused by a sudden change of velocity $\Delta V = V_2 - V_1$ can be derived by applying the unsteady momentum equation to a control volume of a section of the pipe where the change of flow is occurring. Consider a partial valve closure which instantly reduces the velocity by an amount ΔV. Reduction of the velocity can only be accomplished by an increase in the pressure upstream from the valve. This creates a pressure wave of magnitude ΔH which travels up the pipe at the acoustic velocity **a**. The increased pressure compresses the liquid and expands the pipe. The transient head rise due to an incremental change in velocity is

$$\Delta H = -\mathbf{a}\,\Delta V/g, \quad \text{for } \mathbf{a} \gg \Delta V \tag{3.4.16}$$

This equation is easy to use for multiple incremental changes of velocity as long as the first wave has not been reflected back to the point of origin.

The derivation of Equation (3.4.16) is based on an assumption of an instant velocity change. For a valve closing at the end of the pipe, instant closure actually refers to a finite time. It is the longest time that a valve can be closed and still cause a pressure rise equal to that of an instant closure. Normally, it is equal to $2L/a$ sec (which is the time required for the first pressure wave to travel to and from the other end of the pipe of length L); the head rise at the valve will be the same as if the valve were closed instantly. The $2L/a$ time is therefore often the instant closure time.

For a valve at the end of a long pipeline, the instant closure time can be considerably greater than $2L/a$. This is because when the friction loss coefficient fL/d is much greater than the loss coefficient for the valve K_l, the valve can be closed a long way before the flow changes. This dead time must be added to the $2L/a$ time to identify the actual instant closure time. To avoid the maximum potential transient pressure rise, the valve must be closed much slower than the instant closure time.

Computational techniques for estimating transient pressures caused by unsteady flow in pipelines are too complex to be done with simple hand calculations. The solution involves solving partial differential equations based on the equations of motion and continuity. These equations are normally solved by the method of characteristics. This technique transforms the equations into total differential equations. After integration, the equations can be solved numerically by finite differences. This analysis provides equations that can be used to predict the flow and head at any interior pipe section at any time (Tullis, 1989; Wiley and Streeter, 1993).

To complete the analysis, equations describing the boundary conditions are required. Typical boundary conditions are the connection of a pipe to a reservoir, a valve, changes in pipe diameter or material, pipe junctions, etc. Friction loss is included in the development of the basic equations and minor losses are handled as boundary conditions. The analysis properly models friction and the propagation and reflections of the pressure wave. It can also be used for surge calculations.

It is recommended that every pipe system should have at least a cursory transient analysis performed to identify the possibility of serious transients and decide whether or not a detailed analysis is necessary. If an analysis indicates that transients are a problem, the types of solutions available to the engineer include

1. Increasing the closing time of control valves.
2. Using a smaller valve to provide better control.
3. Designing special facilities for filling, flushing, and removing air from pipelines.
4. Increasing the pressure class of the pipeline.
5. Limiting the flow velocity.
6. Using pressure relief valves, surge tanks, air chambers, etc.

Check Valves

Selecting the wrong type or size of check valve can result in poor performance, severe transients, and frequent repairs (Kalsi, 1993). Proper check valve selection requires understanding the characteristics of the various types of check valves and analyzing how they will function as a part of the system in which they will be installed. A check valve that operates satisfactorily in one system may be totally inadequate in another. Each valve type has unique characteristics that give it advantages or disadvantages compared with the others. The characteristics of check valves that describe their hydraulic performance and which should be considered in the selection process include

1. Opening characteristics, i.e., velocity vs. disk position data.
2. Velocity required to fully open and firmly backseat the disk.
3. Pressure drop at maximum flow.
4. Stability of the disk at partial openings.
5. Sensitivity of disk flutter to upstream disturbances.
6. Speed of valve closure compared with the rate of flow reversal of the system.

Disk stability varies with flow rate, disk position, and upstream disturbances and is an important factor in determining the useful life of a check valve. For most applications it is preferable to size the check valve so that the disk is fully open and firmly backseated at normal flow rates. One of the worst design errors is to oversize a check valve that is located just downstream from a disturbance such as a pump, elbow, or control valve. If the disk does not fully open, it will be subjected to severe motion that will accelerate wear. To avoid this problem, it may be necessary to select a check valve that is smaller than the pipe size.

The transient pressure rise generated at check valve closure is another important consideration. The pressure rise is a function of how fast the valve disk closes compared with how fast the flow in the system reverses. The speed that the flow in a system reverses depends on the system. In systems where rapid flow reversals occur, the disk can slam shut causing a pressure transient (Thorley, 1989).

The closing speed of a valve is determined by the mass of the disk, the forces closing the disk, and the distance of travel from full open to closed. Fast closing valves have the following properties: the disk (including all moving parts) is lightweight, closure is assisted by springs, and the full stroke of the disk is short. Swing check valves are the slowest-closing valves because they violate all three of these criteria; i.e., they have heavy disks, no springs, and long disk travel. The nozzle check valve is one of the fastest-closing valves because the closing element is light, is spring loaded, and has a short stroke. The silent, duo, double door, and lift check valves with springs are similar to nozzle valves in their closing times, mainly because of the closing force of the spring.

Systems where rapid flow reversals occur include parallel pumps, where one pump is stopped while the others are still operating, and systems that have air chambers or surge tanks close to the check valve. For these systems there is a high-energy source downstream from the check valve to cause the flow to quickly reverse. As the disk nears its seat, it starts to restrict the reverse flow. This builds up the pressure, accelerates the disk, and slams it into the seat. Results of laboratory experiments, field tests, and computer simulations show that dramatic reductions in the transient pressures at disk closure can be achieved by replacing a slow-closing swing check valve with a fast-acting check valve. For example, in a system containing parallel pumps where the transient was generated by stopping one of the pumps, the peak transient pressure was reduced from 745 to 76 kPa when a swing check was replaced with a nozzle check valve. Such a change improved performance and significantly reduced maintenance.

Air Valves

There are three types of automatic air valves: (1) air/vacuum valves, (2) air release valves, and (3) combination valves. The air/vacuum valve is designed for releasing air while the pipe is being filled and for admitting air when the pipe is being drained. The valve must be large enough that it can admit and expel large quantities of air at a low pressure differential. The outlet orifice is generally the same diameter as the inlet pipe.

These valves typically contain a float, which rises and closes the orifice as the valve body fills with water. Once the line is pressurized, this type of valve cannot reopen to remove air that may subsequently accumulate until the pressure becomes negative, allowing the float to drop. If the pressure becomes negative during a transient or while draining, the float drops and admits air into the line. For thin-walled pipes that can collapse under internal vacuums, the air/vacuum valves should be sized for a full pipe break at the lowest pipe elevation. The vacuum valve must supply an air flow equal to the maximum drainage rate of the water from the pipe break and at an internal pipe pressure above the pipe collapse pressure.

The critical factor in sizing air/vacuum valves is usually the air flow rate to protect the pipe from a full pipe break. Since a pipe is filled much slower than it would drain during a full break, the selected valve will be sized so that the air is expelled during filling without pressurizing the pipe. Sizing charts are provided by manufacturers.

Air release valves contain a small orifice and are designed to release small quantities of pressurized air that are trapped during filling and that accumulate after initial filling and pressurization. The small orifice is controlled by a plunger activated by a float at the end of a lever arm. As air accumulates in the valve body, the float drops and opens the orifice. As the air is expelled, the float rises and closes off the orifice. Sizing air release valves requires an estimate of the amount of pressurized air that must be expelled. This is determined by the filling procedure and any source of air that can be admitted into the pipe or be degassed from the liquid during operation.

The combination valve is actually two valves: a large valve that functions as an air/vacuum valve and a small one that functions as an air release valve. The installation can either consist of an air/vacuum valve and an air release valve plumbed in parallel, or the two can be housed in a single valve body. Most air valve installations require combination valves.

One caution is that manual air release valves should be avoided because improper operation of them can be very dangerous. If the system is pressurized with the manual air valves closed, the trapped air

will be pressurized to full system pressure. When the air valve is manually opened, the pressurized air can cause rapid acceleration of the liquid and generate serious transients when the water is decelerated as it hits the air valve. If manual air valves are installed, they should be very small so the air release rate is controlled to a safe rate.

Locating air valves in a piping system depends on the pipe profile, pipe length, and flow rates. Preferably, pipes should be laid to grade with valves placed at the high points or at intervals if there are no high points. One should use engineering judgment when defining a high point. If the pipe has numerous high points that are close together, or if the high points are not pronounced, it will not be necessary to have an air valve at each high point. If the liquid flow velocity is above about 1 m/sec (3 ft/sec), the flowing water can move the entrained air past intermediate high points to a downstream air valve. Releasing the air through an air valve prevents any sizable air pockets under high pressure from forming in the pipe. Trapped air under high pressure is extremely dangerous.

Velocity of the flow during filling is important. A safe way to fill a pipe is to limit the initial fill rate to an average flow velocity of about 0.3 m/sec (1 ft/sec) until most of the air is released and the air/vacuum valves close. The next step is to flush the system at about 1 m/sec (3 ft/sec), at a low system pressure, to flush the remaining air to an air release valve. It is important that the system not be pressurized until the air has been removed. Allowing large quantities of air under high pressure to accumulate and move through the pipe can generate severe transients. This is especially true if the compressed air is allowed to pass through a control valve or manual air release valve. When pressurized air flows through a partially open valve, the sudden acceleration and deceleration of the air and liquid can generate high pressure transients.

Pump Selection

Optimizing the life of a water supply system requires proper selection, operation, and maintenance of the pumps. During the selection process, the designer must be concerned about matching the pump performance to the system requirements and must anticipate problems that will be encountered when the pumps are started or stopped and when the pipe is filled and drained. The design should also consider the effect of variations in flow requirements, and also anticipate problems that will be encountered due to increased future demands and details of the installation of the pumps.

Selecting a pump for a particular service requires matching the system requirements to the capabilities of the pump. The process consists of developing a system equation by applying the energy equation to evaluate the pumping head required to overcome the elevation difference, friction, and minor losses. For a pump supplying water between two reservoirs, the pump head required to produce a given discharge can be expressed as

$$Hp = \Delta Z + H_f \tag{3.4.17}$$

or

$$Hp = \Delta Z + CQ^2 \tag{3.4.18}$$

in which the constant C is defined by Equation (3.4.13).

Figure 3.4.2 shows a system curve for a pipe having an 82-m elevation lift and moderate friction losses. If the elevation of either reservoir is a variable, then there is not a single curve but a family of curves corresponding to differential reservoir elevations.

The three pump curves shown in Figure 3.4.2 represent different impeller diameters. The intersections of the system curve with the pump curves identify the flow that each pump will supply if installed in that system. For this example both A and B pumps would be a good choice because they both operate at or near their best efficiency range. Figure 3.4.2 shows the head and flow that the B pump will produce

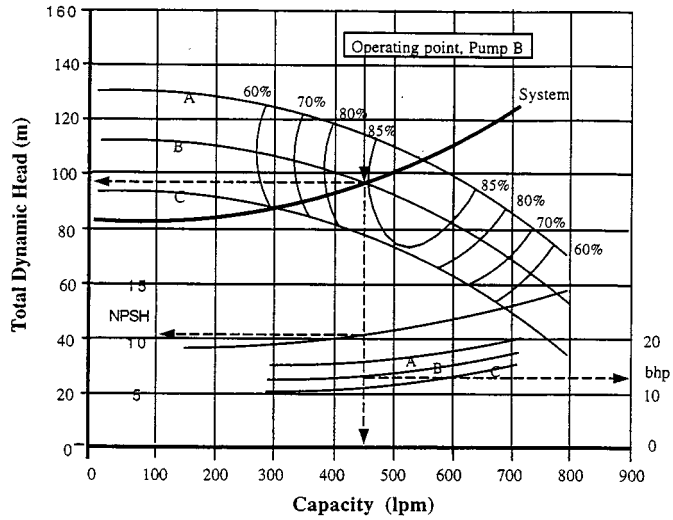


FIGURE 3.4.2 Pump selection for a single pump.

when operating in that system are 97 m and 450 L/m. The net positive suction head (NPSH) and brake horsepower (bhp) are obtained as shown in the figure.

The selection process is more complex when the system demand varies, either due to variations in the water surface elevation or to changing flow requirements. If the system must operate over a range of reservoir elevations, the pump should be selected so that the system curve, based on the mean (or the most frequently encountered) water level, intersects the pump curve near the midpoint of the best efficiency range. If the water level variation is not too great, the pump may not be able to operate efficiently over the complete flow range.

The problem of pump selection also becomes more difficult when planning for future demands or if the pumps are required to supply a varying flow. If the flow range is large, multiple pumps or a variable-speed drive may be needed. Recent developments in variable-frequency drives for pumps make them a viable alternative for systems with varying flows. Selection of multiple pumps and the decision about installing them in parallel or in series depend on the amount of friction in the system. Parallel installations are most effective for low-friction systems. Series pumps work best in high-friction systems.

For parallel pump operation the combined two pump curve is constructed by adding the flow of each pump. Such a curve is shown in Figure 3.4.3 (labeled 2 pumps). The intersection of the two-pump curve with the system curve identifies the combined flow for the two pumps. The pump efficiency for each pump is determined by projecting horizontally to the left to intersect the single-pump curve. For this example, a C pump, when operating by itself, will be have an efficiency of 83%. With two pumps operating, the efficiency of each will be about 72%. For the two pumps to operate in the most efficient way, the selection should be made so the system curve intersects the single-pump curve to the right of its best efficiency point.

Starting a pump with the pipeline empty will result in filling at a very rapid rate because initially there is little friction to build backpressure. As a result, the pump will operate at a flow well above the design flow. This may cause the pump to cavitate, but the more serious problem is the possibility of high pressures generated by the rapid filling of the pipe. Provisions should be made to control the rate of filling to a safe rate. Start-up transients are often controlled by starting the pump against a partially open discharge valve located near the pump and using a bypass line around the pump. This allows the system to be filled slowly and safely. If the pipe remains full and no air is trapped, after the initial filling, subsequent start-up of the pumps generally does not create any serious problem. Adequate air release valves should be installed to release the air under low pressure.

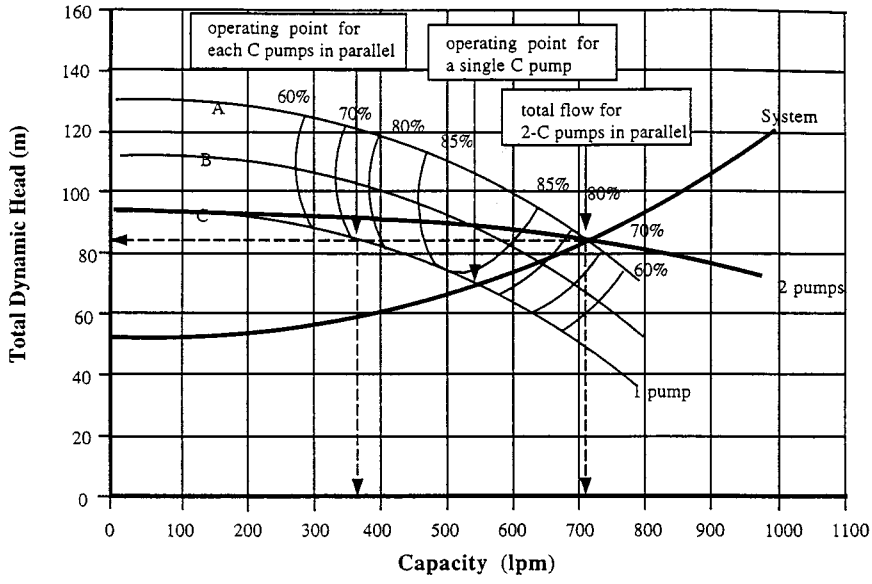


FIGURE 3.4.3 Selection of parallel pumps.

For some systems, stopping the pump, either intentionally or accidentally, can generate high pressures that can damage the pipe and controls. If the design process does not consider these potential problems, the system may not function trouble free. Downtime and maintenance costs may be high. Not all systems will experience start-up and shutdown problems, but the design should at least consider the possibility. The problem is more severe for pipelines that have a large elevation change and multiple high points. The magnitude of the transient is related to the length and profile of the pipeline, the pump characteristics, the magnitude of the elevation change, and the type of check valve used. The downsurge caused by stopping the pump can cause column separation and high pressures due to flow reversals and closure of the check valves. Surge-protection equipment can be added to such systems to prevent damage and excessive maintenance.

Another operational problem occurs with parallel pumps. Each pump must have a check valve to prevent reverse flow. When one of the pumps is turned off, the flow reverses almost immediately in that line because of the high manifold pressure supplied by the operating pumps. This causes the check valve to close. If a slow-closing check valve is installed, the flow can attain a high reverse velocity before the valve closes, generating high pressure transients.

Numerous mechanical devices and techniques have been used to suppress pump shutdown transients. These include increasing the rotational inertia of the pump, use of surge tanks or air chambers near the pump, pressure relief valves, vacuum-breaking valves, and surge-anticipating valves. Selection of the proper transient control device will improve reliability, extend the economic life of the system, and reduce maintenance. Failure to complete a transient analysis and include the required controls will have the opposite effect. A system is only as good as it is designed to be.

Other Considerations

Feasibility Study

Designing pipelines, especially long transmission lines, involves more than just determining the required type and size of pipe. A starting point for major projects is usually a feasibility study which involves

social, environmental, political, and legal issues, as well as an economic evaluation of the engineering alternatives developed during the preliminary design. The preliminary design should identify the scope of the project and all major features that influence the cost or viability. Since local laws, social values, and environmental concerns vary significantly between geographic areas, the engineer must be aware of the problems unique to the area.

Choices that affect the economics of the project include alternative pipe routes, amount of storage and its effect on reliability and controllability of flow, choice of pipe material, diameter and pressure class, provision for future demands, etc. In making decisions one must consider both the engineering and economic advantages of the alternatives. Reliability, safety, maintenance, operating, and replacement costs must all be given their proper value. The analysis should consider (1) expected life of the pipe, which is a function of the type of pipe material and the use of linings or protective coatings; (2) economic life, meaning how long the pipe will supply the demand; (3) planning for future demand; (4) pumping cost vs. pipe cost; and (5) provisions for storage.

During the feasibility study only a general design has been completed so a detailed analysis of all hydraulic problems and their solutions is not available. Even so, it is necessary to anticipate the need for special facilities or equipment and problems such as safe filling, provisions for draining, cavitation at control valves, and transient problems caused by valve or pump operation. Provisions should be made for the cost of the detailed analysis, design, and construction costs required to control special operational problems. Attention should also be given to costs associated with winterizing, stream crossings, highways crossing, special geologic or topographic problems, and any other items that would have a significant influence on the cost, reliability, or safety of the project.

Storage

The purposes of storage tanks and intermediate reservoirs include (1) to supply water when there is a temporary interruption of flow from the supply, (2) to provide supplemental water during peak periods, (3) to sectionalize the pipe to reduce mean and transient pressures, (4) to maintain pressure (elevated storage), and (5) to simplify control. Storage also has a significant impact on the control structures, pumping plants, and general operation of the pipeline. If there is adequate storage, large fluctuations in demand can be tolerated. Any mismatch in supply and demand is made up for by an increase or decrease in storage, and valves in the transmission main will require only infrequent adjustments to maintain storage. Pumps can be activated by level controls at the storage tank and not by fluctuations in demand so they can operate for long periods near their design point.

If there is no storage, the system may have to provide continuous fine adjustment of the flow to provide the required flow within safe pressure limits. For gravity systems this may require automatic pressure- or flow-regulating valves. For pumped systems, the variations in flow can cause constant-speed centrifugal pumps to operate both below and above their design point where power consumption is high, efficiency is low, and where there is more chance of operational problems. Selection of a variable-frequency drive can avoid these problems. The selection of multiple pumps vs. a variable-speed pump is primarily a economic decision.

Thrust Blocks

Any time there is a change of pipe alignment, an unbalanced force is developed. The force required to restrain the pipe can be calculated with the two-dimensional, steady state momentum equation. For buried pipelines, this force can be transmitted to the soil with a thrust block. Determining the size of the block and, consequently, the bearing surface area depends on pipe diameter, fluid pressure, deflection angle of the pipe, and bearing capacity of the soil. A convenient monograph for sizing thrust blocks was published in the *Civil Engineering* in 1969 (Morrison, 1969).

References

ASCE. 1992. *Pressure Pipeline Design for Water and Wastewater.* Prepared by the Committee on Pipeline Planning of the Pipeline Division of the American Society of Civil Engineers, New York.

Kalsi Engineering and Tullis Engineering Consultants. 1993. *Application Guide for Check Valves in Nuclear Power Plants*, Revision 1, NP-5479. Prepared for Nuclear Maintenance Applications Center, Charlotte, NC.

Miller, D.S. 1990. *Internal Flow Systems — Design and Performance Prediction*, 2nd ed. Gulf Publishing Company, Houston.

Morrison, E.B. 1969. Monograph for the design of thrust blocks. *Civil Eng.,* 39, June, 55–51.

Spanger, M.G. and Handy, R.L. 1973. *Soil Engineering,* 3rd ed. Intext Eductational Publishers, New York, Chap. 25 and 26.

Streeter, V.L. and Wylie, E.B. 1975. *Fluid Mechanics,* 6th ed. McGraw-Hill, New York, 752 pp.

Thorley, A.R.D. 1989. Check valve behavior under transient flow conditions: a state-of-the-art review. *ASME*, 111, Vol. 2, June. *J. Fluids Engineering: Transactions of the ASME,* pp. 173–183.

Tullis, J.P. 1989. *Hydraulics of Pipelines — Pumps, Valves, Cavitation, Transients*, John Wiley and Sons, New York.

Tullis, J.P. 1993. Cavitation Guide for Control Valves, NUREG/CR-6031, U.S. Nuclear Regulatory Commission, Washington, D.C.

Wood, D.J. 1966. An explicit friction factor relationship. *Civil Eng.,* 36, December, 60–61.

Wylie, E.B. and Streeter, V.L. 1993. *Fluid Transients in Systems,* Prentice-Hall, Englewood Cliffs, N.J.

Further Information

Bean, H.S., ed. 1971. *Fluid Meters, Their Theory and Application,* 6th ed. The American Society of Mechanical Engineers, New York.

Handbook of PVC-Design and Construction. 1979. Uni-Bell Plastic Pipe Association, Dallas.

King, H.W. 1954. *Handbook of Hydraulics — For the Solution of Hydraulic Problems,* 4th ed. Revised by E.F. Brater. McGraw-Hill, New York.

Stephenson, D. 1981. *Pipeline Design for Water Engineers,* 2nd ed. Elsevier Scientific Publishing Company, Distributed in the U.S. by Gulf Publishing Company, Houston.

Stutsman, R.D. 1993. Steel Penstocks, ASCE Manuals and Reports on Engineering Practice No. 79, Energy Division, American Society of Civil Engineers, New York.

Watkins, R.K. and Spangler, M.G., Some Characteristics of the Modulus of Passive Resistance of Soil: A Study in Similitude. Highway Research Board Proceedings Vol. 37, 1958, pp. 576–583.

Tullis, J.P. 1996. Valves, in Dorf, R.C., Ed., *The Engineering Handbook,* CRC Press, Boca Raton, FL.

3.5 Open Channel Flow*

Frank M. White

Definition

The term *open channel flow* denotes the gravity-driven flow of a liquid with a free surface. Technically, we may study any flowing liquid and any gas interface. In practice, the vast majority of open channel flows concern water flowing beneath atmospheric air in artificial or natural channels.

The geometry of an arbitrary channel is shown in Figure 3.5.1. The area A is for the water cross section only, and b is its top width. The wetted perimeter P covers only the bottom and sides, as shown, not the surface (whose air resistance is neglected). The water depth at any location is y, and the channel slope is θ, often denoted as S_o = sin θ. All of these parameters may vary with distance x along the channel. In unsteady flow (not discussed here) they may also vary with time.

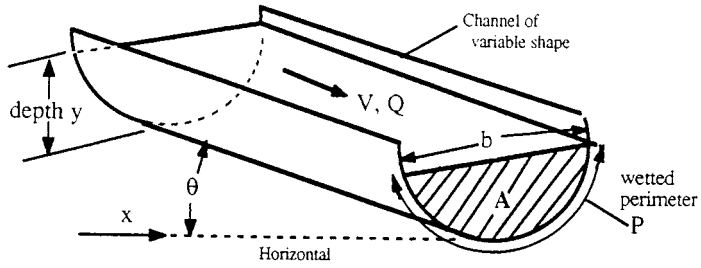


FIGURE 3.5.1 Definition sketch for an open channel.

Uniform Flow

A simple reference condition, called *uniform flow,* occurs in a long straight prismatic channel of constant slope S_o. There is no acceleration, and the water flows at constant depth with fluid weight exactly balancing the wetted wall shear force: $\rho g L A \sin\theta = \tau_w P L$, where L is the channel length. Thus, $\tau_w = \rho g R_h S_o$, where $R_h = A/P$ is called the *hydraulic radius* of the channel. If we relate wall shear stress to the Darcy friction factor f, $\tau_w = (f/8)\rho V^2$, we obtain the basic uniform flow open channel relation:

$$\text{Uniform flow:}\quad V = \sqrt{\frac{8g}{f}}\sqrt{R_h S_o}, \quad \text{where} \quad \sqrt{\frac{8g}{f}} = C = \text{Chézy coefficient} \tag{3.5.1}$$

Antoine Chézy first derived this formula in 1769. It is satisfactory to base f upon the pipe-flow Moody diagram (Figure 3.4.1) using the hydraulic diameter, $D_h = 4R_h$, as a length scale. That is, $f = fcn\,(VD_h/\nu, \varepsilon/D_h)$ from the Moody chart. In ordinary practice, however, engineers assume fully rough, high-Reynolds-number flow and use Robert Manning's century-old correlation:

$$C \approx \frac{\zeta}{n} R_h^{1/6}, \quad \text{or} \quad V_{\text{uniform}} \approx \frac{\zeta}{n} R_h^{2/3} S_o^{1/2} \quad \text{and} \quad Q = VA \tag{3.5.2}$$

where ζ is a conversion factor equal to 1.0 in SI units and 1.486 in English units. The quantity n is Manning's roughness parameter, with typical values, along with the associated roughness heights ε, listed in Table 3.5.1.

* From White, F., *Fluid Mechanics,* 3rd ed., 1994; reproduced with permission of MacGraw-Hill, Inc. Nomenclature appears at end of this section.

TABLE 3.5.1 Average Roughness Parameters for Various Channel Surfaces

		Average Roughness Height ε	
	n	**ft**	**mm**
Artificial lined channels			
Glass	0.010 ± 0.002	0.0011	0.3
Brass	0.011 ± 0.002	0.0019	0.6
Steel; smooth	0.012 ± 0.002	0.0032	1.0
Painted	0.014 ± 0.003	0.0080	2.4
Riveted	0.015 ± 0.002	0.012	3.7
Cast iron	0.013 ± 0.003	0.0051	1.6
Cement; finished	0.012 ± 0.002	0.0032	1.0
Unfinished	0.014 ± 0.002	0.0080	2.4
Planed wood	0.012 ± 0.002	0.0032	1.0
Clay tile	0.014 ± 0.003	0.0080	2.4
Brickwork	0.015 ± 0.002	0.012	3.7
Asphalt	0.016 ± 0.003	0.018	5.4
Corrugated metal	0.022 ± 0.005	0.12	37
Rubble masonry	0.025 ± 0.005	0.26	80
Excavated earth channels			
Clean	0.022 ± 0.004	0.12	37
Gravelly	0.025 ± 0.005	0.26	80
Weedy	0.030 ± 0.005	0.8	240
Stony; cobbles	0.035 ± 0.010	1.5	500
Natural channels			
Clean and straight	0.030 ± 0.005	0.8	240
Sluggish, deep pools	0.040 ± 0.010	3	900
Major rivers	0.035 ± 0.010	1.5	500
Floodplains			
Pasture, farmland	0.035 ± 0.010	1.5	500
Light brush	0.05 ± 0.02	6	2000
Heavy brush	0.075 ± 0.025	15	5000
Trees	0.15 ± 0.05	?	?

Critical Flow

Since the surface is always atmospheric, pressure head is not important in open channel flows. Total energy E relates only to velocity and elevation:

$$\text{Specific energy } E = y + \frac{V^2}{2g} = y + \frac{Q^2}{2gA^2}$$

At a given volume flow rate Q, the energy passes through a minimum at a condition called *critical flow*, where $dE/dy = 0$, or $dA/dy = b = gA^3/Q^2$:

$$A_{crit} = \left(\frac{bQ^2}{g}\right)^{1/3} \qquad V_{crit} = \frac{Q}{A_{crit}} = \left(\frac{gA_{crit}}{b}\right)^{1/2} \tag{3.5.3}$$

where b is the top-surface width as in Figure 3.5.1. The velocity V_{crit} equals the speed of propagation of a surface wave along the channel. Thus, we may define the Froude number Fr of a channel flow, for any cross section, as $Fr = V/V_{crit}$. The three regimes of channel flow are

$$\text{Fr} < 1\text{: subcritical flow;} \quad \text{Fr} = 1\text{: critical flow;} \quad \text{Fr} > 1\text{: supercritical flow}$$

There are many similarities between Froude number in channel flow and Mach number in variable-area duct flow (see Section 3.6).

For a rectangular duct, $A = by$, we obtain the simplified formulas

$$V_{crit} = \sqrt{gy} \qquad \text{Fr} = \frac{V}{\sqrt{gy}} \tag{3.5.4}$$

independent of the width of the channel.

Example 3.5.1

Water (ρ = 998 kg/m^3, μ = 0.001 kg/m · sec) flows uniformly down a half-full brick 1-m-diameter circular channel sloping at 1°. Estimate (a) Q; and (b) the Froude number.

Solution 3.5.1 (a). First compute the geometric properties of a half-full circular channel:

$$A = \frac{\pi}{8}(1\text{ m})^2 = 0.393\text{ m}^2; \quad P = \frac{\pi}{2}(1\text{ m}) = 1.57\text{ m}; \quad R = \frac{A}{P} = \frac{0.393}{1.57} = 0.25\text{ m}$$

From Table 3.5.1, for brickwork, $n \approx 0.015$. Then, Manning's formula, Equation (3.5.2) predicts

$$V = \frac{\zeta}{n} R_h^{1/6} S_o^{1/2} = \frac{1.0}{0.015}(0.25)^{1/6}(\sin 1°)^{1/2} \approx 3.49\frac{\text{m}}{\text{sec}}; \quad Q = 3.49(0.393) \approx \mathbf{1.37}\frac{\mathbf{m^3}}{\mathbf{sec}} \quad \textit{Solution 3.5.1(a)}$$

The uncertainty in this result is about ±10%. The flow rate is quite large (21,800 gal/min) because 1°, although seemingly small, is a substantial slope for a water channel.

One can also use the Moody chart. with $V \approx 3.49$ m/sec, compute Re = $\rho V D_h/\mu \approx 3.49$ E6 and $\varepsilon/D_h \approx 0.0037$, then compute $f \approx 0.0278$ from the Moody chart. Equation (3.5.1) then predicts

$$V = \sqrt{\frac{8g}{f} R_h S_o} = \sqrt{\frac{8(9.81)}{0.0278}(0.25)(\sin 1°)} \approx 3.51\frac{\text{m}}{\text{sec}}; \quad Q = VA \approx \mathbf{1.38}\frac{\mathbf{m^3}}{\mathbf{sec}}$$

Solution 3.5.1 (b). With Q known from part (a), compute the critical conditions from Equation (3.5.3):

$$A_{crit} = \left(\frac{bQ^2}{g}\right)^{1/3} = \left[\frac{1.0(1.37)^2}{9.81}\right]^{1/3} = 0.576\text{ m}^2, \quad V_{crit} = \frac{Q}{A_{crit}} = \frac{1.37}{0.576} = 2.38\ \frac{\text{m}}{\text{sec}}$$

Hence

$$\text{Fr} = \frac{V}{V_{crit}} = \frac{3.49}{2.38} \approx \mathbf{1.47}\square(\text{supercritical}) \quad \textit{Solution 3.5.1(b)}$$

Again the uncertainty is approximately ±10%, primarily because of the need to estimate the brick roughness.

Hydraulic Jump

In gas dynamics (Section 3.6), a supersonic gas flow may pass through a thin normal shock and exit as a subsonic flow at higher pressure and temperature. By analogy, a supercritical open channel flow may pass through a *hydraulic jump* and exit as a subcritical flow at greater depth, as in Figure 3.5.2. Application of continuity and momentum to a jump in a rectangular channel yields

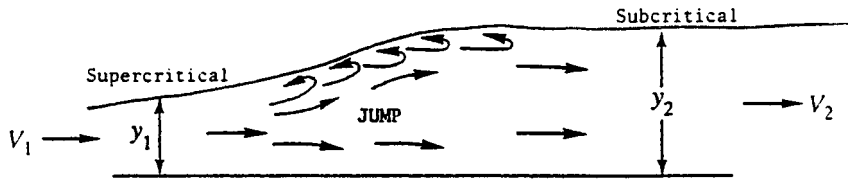


FIGURE 3.5.2 A two-dimensional hydraulic jump.

$$V_2 = V_1 \frac{y_1}{y_2} \qquad y_2 = \frac{y_1}{2}\left[-1 + \sqrt{1 + 8\mathrm{Fr}_1^2}\right] \text{ where } \mathrm{Fr}_1 = \frac{V_1}{\sqrt{g y_1}} > 1 \tag{3.5.5}$$

Both the normal shock and the hydraulic jump are dissipative processes: the entropy increases and the effective energy decreases. For a rectangular jump,

$$\Delta E = E_1 - E_2 = \frac{\left(y_2 - y_1\right)^3}{4 y_1 y_2} > 0 \tag{3.5.6}$$

For strong jumps, this loss in energy can be up to 85% of E_1. The second law of thermodynamics requires $\Delta E > 0$ and $y_2 > y_1$ or, equivalently, $\mathrm{Fr}_1 > 1$,

Note from Figure 3.5.2 that a hydraulic jump is not thin. Its total length is approximately four times the downstream depth. Jumps also occur in nonrectangular channels, and the theory is much more algebraically laborious.

Weirs

If an open channel flow encounters a significant obstruction, it will undergo rapidly varied changes which are difficult to model analytically but can be correlated with experiment. An example is the *weir* in Figure 3.5.3 (colloquially called a *dam*), which forces the flow to deflect over the top. If $L << Y$, the weir is termed *sharp-crested*; if $L = \mathrm{O}(Y)$ it is *broad-crested*. Small details, such as the upper front corner radius or the crest roughness, may be significant. The crest is assumed level and of width b into the paper.

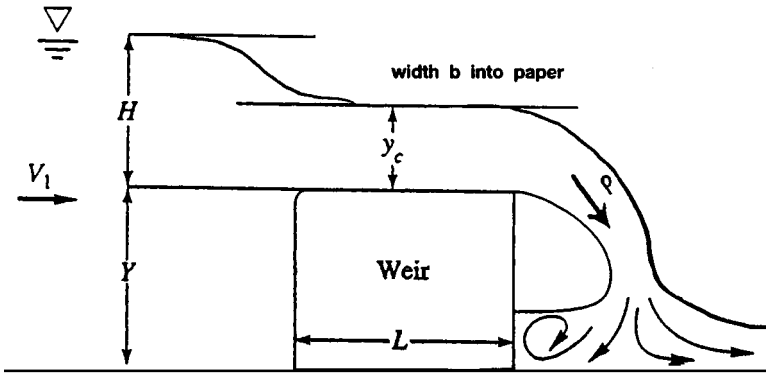


FIGURE 3.5.3 Geometry and notation for flow over a weir.

If there is a free overfall, as in Figure 3.5.3, the flow accelerates from subcritical upstream to critical over the crest to supercritical in the overfall. There is no flow when the excess upstream depth $H = 0$. A simple Bernoulli-type analysis predicts that the flow rate Q over a wide weir is approximately proportional to $bg^{1/2}H^{3/2}$. An appropriate correlation is thus

$$Q_{\text{weir}} = C_d b g^{1/2} H^{3/2}, \quad \text{where } C_d = \text{dimensionless weir coefficient} \tag{3.5.7}$$

If the upstream flow is turbulent, the weir coefficient depends only upon geometry, and Reynolds number effects are negligible. If the weir has sidewalls and is narrow, replace width b by $(b - 0.1H)$.

Two recommended empirical correlations for Equation (3.5.7) are as follows:

$$\begin{aligned} &\text{Sharp-crested:} \quad C_d \approx 0.564 + 0.0846\frac{H}{Y} \quad \text{for} \quad \frac{L}{Y} < 0.07 \\ &\text{Broad-crested:} \quad C_d \approx 0.462 \quad \text{for} \quad 0.08 < \frac{H}{L} < 0.33 \end{aligned} \tag{3.5.8}$$

These data are for wide weirs with a sharp upper corner in front. Many other weir geometries are discussed in the references for this section. Of particular interest is the sharp-edged vee-notch weir, which has no length scale b. If 2θ is the total included angle of the notch, the recommended correlation is

$$\text{Vee-notch, angle } 2\theta: \quad Q \approx 0.44 \tan\theta\, g^{1/2} H^{5/2} \quad \text{for} \quad 10° < \theta \le 50° \tag{3.5.9}$$

The vee-notch is more sensitive at low flow rates (large H for a small Q) and thus is popular in laboratory measurements of channel flow rates.

A weir in the field will tend to spring free and form a natural *nappe*, or air cavity, as in Figure 3.5.3. Narrow weirs, with sidewalls, may need to be aerated artificially to form a nappe and keep the flow from sliding down the face of the weir. The correlations above assume nappe formation.

Gradually Varied Flow

Return to Figure 3.5.1 and suppose that (y, A, b, P, S_o) are all variable functions of horizontal position x. If these parameters are slowly changing, with no hydraulic jumps, the flow is termed *gradually varied* and satisfies a simple one-dimensional first-order differential equation if Q = constant:

$$\frac{dy}{dx} \approx \frac{S_o - S}{1 - \dfrac{V^2 b}{gA}}, \quad \text{where } V = \frac{Q}{A} \quad \text{and} \quad S = \frac{f}{D_h}\frac{V^2}{2g} = \frac{n^2 V^2}{\zeta^2 R_h^{4/3}} \tag{3.5.10}$$

The conversion factor $\zeta^2 = 1.0$ for SI units and 2.208 for English units. If flow rate, bottom slope, channel geometry, and surface roughness are known, we may solve for $y(x)$ for any given initial condition $y = y_o$ at $x = x_o$. The solution is computed by any common numerical method, e.g., Runge–Kutta.

Recall from Equation (3.5.3) that the term $V^2b/(gA) \equiv \text{Fr}^2$, so the sign of the denominator in Equation (3.5.10) depends upon whether the flow is sub- or supercritical. The mathematical behavior of Equation (3.5.10) differs also. If Fr is near unity, the change dy/dx will be very large, which probably violates the basic assumption of "gradual" variation.

For a given flow rate and local bottom slope, two reference depths are useful and may be computed in advance:

(a) The *normal* depth y_n for which Equation (3.5.2) yields the flow rate:
(b) The *critical* depth y_c for which Equation (3.5.3) yields the flow rate.

Comparison of these two, and their relation to the actual local depth y, specifies the type of solution curve being computed. The five bottom-slope regimes (mild M, critical C, steep S, horizontal H, and adverse A) create 12 different solution curves, as illustrated in Figure 3.5.4. All of these may be readily generated by a computer solution of Equation 3.5.10. The following example illustrates a typical solution to a gradually varied flow problem.

Example 3.5.2

Water, flowing at 2.5 m^3/sec in a rectangular gravelly earth channel 2 m wide, encounters a broad-crested weir 1.5 m high. Using gradually varied theory, estimate the water depth profile back to 1 km upstream of the weir. The bottom slope is 0.1^0.

Solution. We are given Q, $Y = 1.5$ m, and $b = 2$ m. We may calculate excess water level H at the weir (see Figure 3.5.3) from Equations (3.5.7) and (3.5.8):

$$Q = 2.5\frac{\text{m}^3}{\text{sec}} = C_d b_{\text{eff}} g^{1/2} H^{3/2} = 0.462(2.0 - 0.1H)(9.81)^{1/2} H^{3/2}, \quad \text{solve for } H \approx \mathbf{0.94}\ \text{m}$$

Since the weir is not too wide, we have subtracted 0.1 H from b as recommended. The weir serves as a "control structure" which sets the water depth just upstream. This is our initial condition for gradually varied theory: $y(0) = Y + H = 1.5 + 0.94 \approx 2.44$ m at $x = 0$. Before solving Equation (3.5.10), we find the normal and critical depths to get a feel for the problem:

$$\text{Normal depth:} \quad Q = 2.5\frac{\text{m}^3}{\text{sec}} = \frac{1.0}{0.025}(2.0y_n)\left(\frac{2.0y_n}{2.0 + 2y_n}\right)^{3/2}\sqrt{\sin 0.1^\circ}, \quad \text{solve } y_n \approx \mathbf{1.14}\ \text{m}$$

$$\text{Critical depth:} \quad A_c = 2.0y_c - \left(\frac{bQ^2}{g}\right)^{1/3} = \left[\frac{2.0(2.5)^2}{9.81}\right]^{1/3}, \quad \text{solve } y_c \approx \mathbf{0.54}\ \text{m}$$

We have taken $n \approx 0.025$ for gravelly earth, from Table 3.5.1. Since $y(0) > y_n > y_c$, we are on a mild slope $M - 1$ "backwater" curve, as in Figure 3.5.4. For our data, Equation (3.5.10) becomes

$$\frac{dy}{dx} \approx \frac{S_o - n^2Q^2/\left(\zeta^2 A^2 R_h^{4/3}\right)}{1 - Q^2 b/\left(gA^3\right)}$$

where $Q = 2.5$, $b = 2$, $\zeta = 1$, $A = 2y$, $S_o = \sin 0.1^\circ$, $R_h = 2y/(2 + 2y)$, $g = 9.81$, $y(0) = 2.44$ at $x = 0$.

Integrate numerically backward, that is, for $\Delta x < 0$, until $x = -1$ km $= -1000$ m. The complete solution curve is shown in Figure 3.5.5. The water depth decreases upstream and is approximately $y \approx 1.31$ m at $x = -1000$ m. If slope and channel width remain constant, the water depth asymptotically approaches the normal depth y_n far upstream.

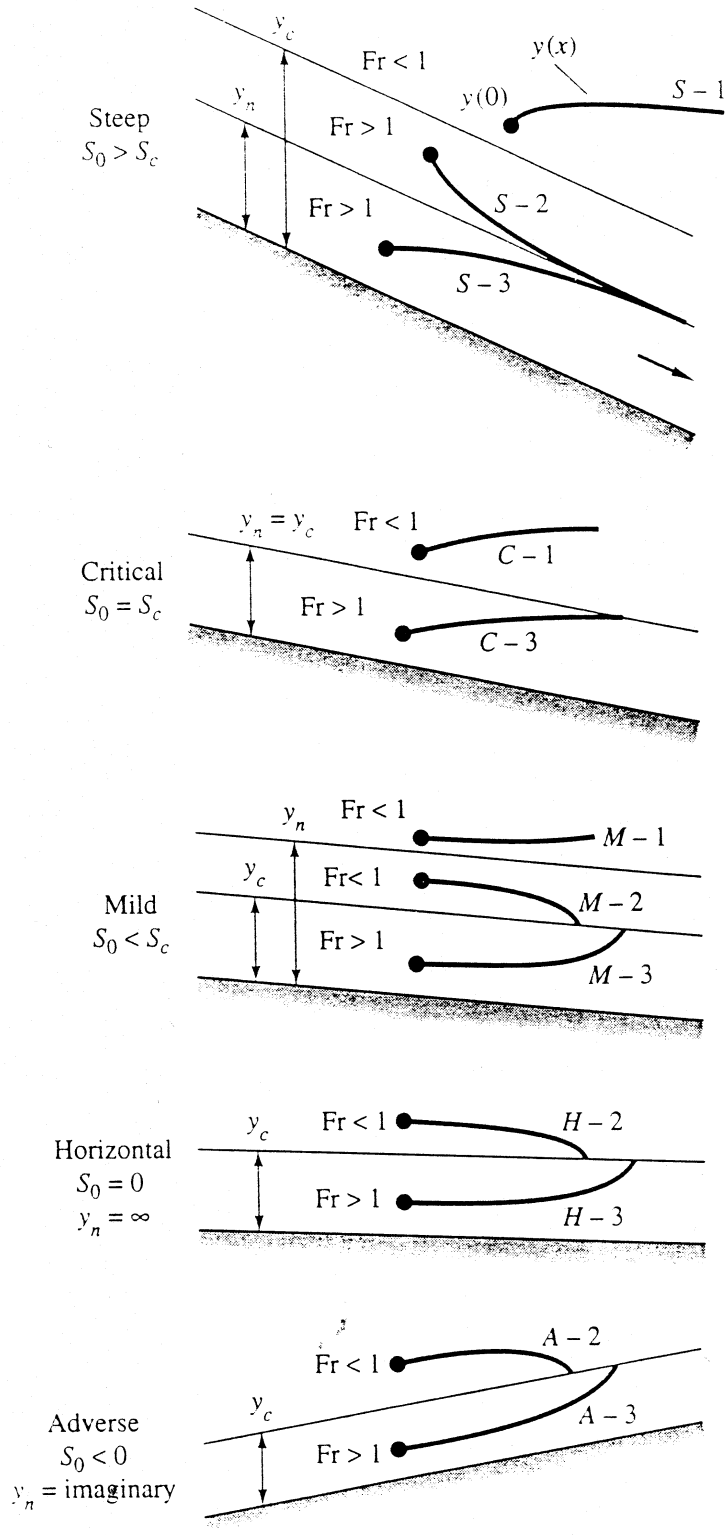


FIGURE 3.5.4 Classification of solution curves for gradually varied flow.

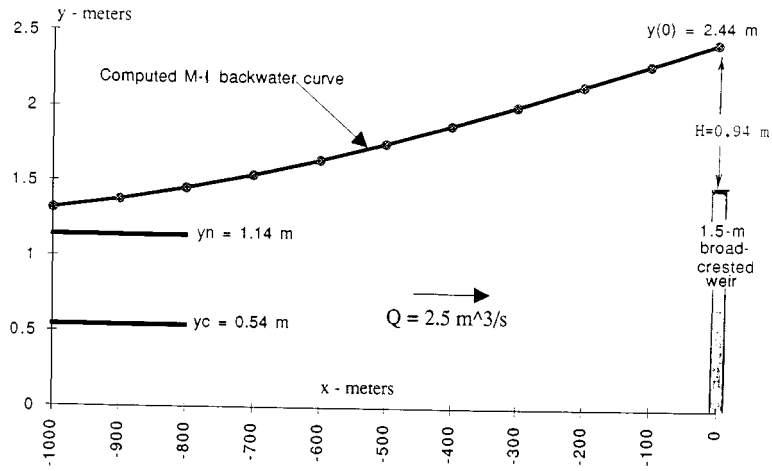


FIGURE 3.5.5 Backwater solution curve for Example 3.5.2.

Nomenclature

English symbols

A = water cross section area
b = channel upper-surface width
C = Chézy coefficient, Equation (3.5.1)
C_d = weir discharge coefficient, Equation (3.5.7)
D_h = hydraulic diameter, = $4R_h$
E = specific energy, = $y + V^2/2g$
f = Moody friction factor
Fr = Froude number, = V/V_{crit}
g = acceleration of gravity
H = excess water level above weir, Figure 3.5.3
L = weir length, Figure 3.5.3
n = Manning roughness factor, Table 3.5.1
P = wetted perimeter
Q = volume flow rate
R_h = hydraulic radius, = A/P
S = frictional slope, Equation (3.5.10)
S_o = bottom slope
V = average velocity
x = horizontal distance along the channel
y = water depth
Y = weir height, Figure (3.5.3

Greek Symbols

ε = wall roughness height, Table 3.5.1
ρ = fluid density
μ = fluid viscosity
ν = fluid kinematic viscosity, = μ/ρ
ζ = conversion factor, = 1.0 (SI) and 1.486 (English)

Subscripts

c,crit = critical, at Fr = 1
n = normal, in uniform flow

References

Ackers, P. et al. 1978. *Weirs and Flumes for Flow Measurement*, John Wiley, New York.

Bos, M.G. 1985. *Long-Throated Flumes and Broad-Crested Weirs,* Martinus Nijhoff (Kluwer), Dordrecht, The Netherlands.

Bos, M.G., Replogle, J.A., and Clemmens, A.J. 1984. *Flow-Measuring Flumes for Open Channel Systems,* John Wiley, New York.

Brater, E.F. 1976. *Handbook of Hydraulics,* 6th ed., McGraw-Hill, New York.

Chow, V.T. 1959. *Open Channel Hydraulics,* McGraw-Hill, New York.

French, R.H. 1985. *Open Channel Hydraulics,* McGraw-Hill, New York.

Henderson, F.M. 1966. *Open Channel Flow,* Macmillan, New York.

Sellin, R.H.J. 1970. *Flow in Channels*, Gordon & Breach, London.

Spitzer, D.W. (Ed.). 1991. *Flow Measurement: Practical Guides for Measurement and Control*, Instrument Society of America, Research Triangle Park, NC.

3.6 External Incompressible Flows

Alan T. McDonald

Introduction and Scope

Potential flow theory (Section 3.2) treats an incompressible *ideal fluid* with zero viscosity. There are no shear stresses; pressure is the only stress acting on a fluid particle. Potential flow theory predicts no drag force when an object moves through a fluid, which obviously is not correct, because all real fluids are viscous and cause drag forces. The objective of this section is to consider the behavior of viscous, incompressible fluids flowing over objects.

A number of phenomena that occur in external flow at high Reynolds number over an object are shown in Figure 3.6.1. The freestream flow divides at the stagnation point and flows around the object. Fluid at the object surface takes on the velocity of the body as a result of the no-slip condition. Boundary layers form on the upper and lower surfaces of the body; flow in the boundary layers is initially laminar, then **transition** to turbulent flow may occur (points "T").

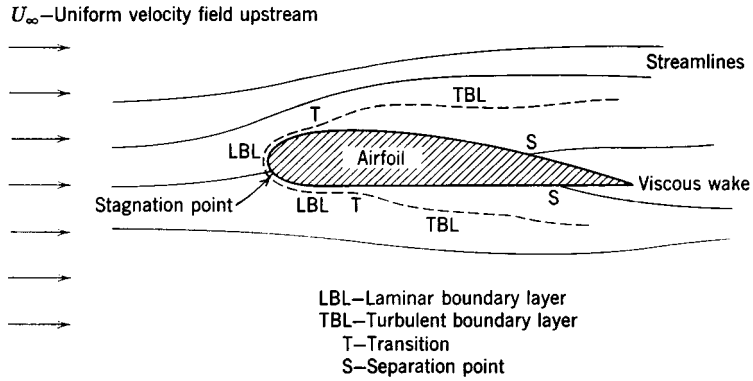


FIGURE 3.6.1 Viscous flow around an airfoil (boundary layer thickness exaggerated for clarity).

Boundary layers thickening on the surfaces cause only a slight displacement of the streamlines of the external flow (their thickness is greatly exaggerated in the figure). **Separation** may occur in the region of increasing pressure on the rear of the body (points "S"); after separation boundary layer fluid no longer remains in contact with the surface. Fluid that was in the boundary layers forms the viscous *wake* behind the object.

The Bernoulli equation is valid for steady, incompressible flow without viscous effects. It may be used to predict pressure variations outside the boundary layer. Stagnation pressure is constant in the uniform inviscid flow far from an object, and the Bernoulli equation reduces to

$$p_\infty + \frac{1}{2}\rho V^2 = \text{constant} \tag{3.6.1}$$

where p_∞ is pressure far upstream, ρ is density, and V is velocity. Therefore, the local pressure can be determined if the local freestream velocity, U, is known.

Boundary Layers

The Boundary Layer Concept

The **boundary layer** is the thin region near the surface of a body in which viscous effects are important. By recognizing that viscous effects are concentrated near the surface of an object, Prandtl showed that only the Euler equations for inviscid flow need be solved in the region outside the boundary layer. Inside the boundary layer, the elliptic Navier-Stokes equations are simplified to boundary layer equations with parabolic form that are easier to solve. The thin boundary layer has negligible pressure variation across it; pressure from the freestream is impressed upon the boundary layer.

Basic characteristics of all laminar and turbulent boundary layers are shown in the developing flow over a flat plate in a semi-infinite fluid. Because the boundary layer is thin, there is negligible disturbance of the inviscid flow outside the boundary layer, and the **pressure gradient** along the surface is close to zero. Transition from laminar to turbulent boundary layer flow on a flat plate occurs when Reynolds number based on x exceeds $\mathrm{Re}_x = 500{,}000$. Transition may occur earlier if the surface is rough, pressure increases in the flow direction, or separation occurs. Following transition, the turbulent boundary layer thickens more rapidly than the laminar boundary layer as a result of increased shear stress at the body surface.

Boundary Layer Thickness Definitions

Boundary layer disturbance thickness, δ, is usually defined as the distance, y, from the surface to the point where the velocity within the boundary layer, u, is within 1% of the local freestream velocity, U. As shown in Figure 3.6.2, the boundary layer velocity profile merges smoothly and asymptotically into the freestream, making δ difficult to measure. For this reason and for their physical significance, we define two integral measures of boundary layer thickness. Displacement thickness, δ^*, is defined as

$$\frac{\delta^*}{\delta} = \int_0^\infty \left(1 - \frac{u}{U}\right) d\left(\frac{y}{\delta}\right) \tag{3.6.2}$$

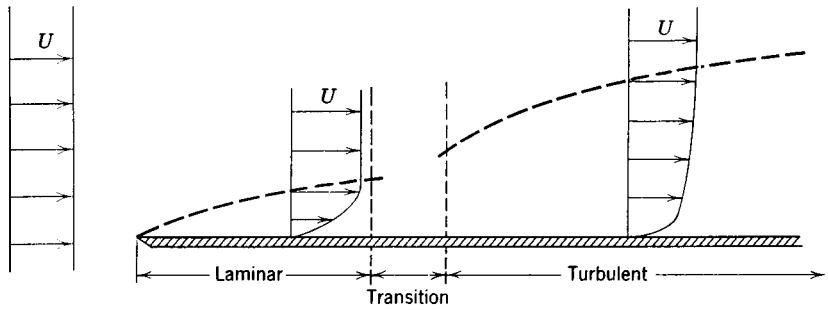


FIGURE 3.6.2 Boundary layer on a flat plate (vertical thickness exaggerated for clarity).

Physically, δ^* is the distance the solid boundary would have to be displaced into the freestream in a frictionless flow to produce the mass flow deficit caused by the viscous boundary layer. Momentum thickness, θ, is defined as

$$\frac{\theta}{\delta} = \int_0^\infty \frac{u}{U}\left(1 - \frac{u}{U}\right) d\left(\frac{y}{\delta}\right) \tag{3.6.3}$$

Physically, θ is the thickness of a fluid layer, having velocity U, for which the momentum flux is the same as the deficit in momentum flux within the boundary layer (momentum flux is momentum per unit time passing a cross section).

Because δ^* and θ are defined in terms of integrals for which the integrand vanishes in the freestream, they are easier to evaluate experimentally than disturbance thickness δ.

Exact Solution of the Laminar Flat-Plate Boundary Layer

Blasius obtained an exact solution for laminar boundary layer flow on a flat plate. He assumed a thin boundary layer to simplify the streamwise momentum equation. He also assumed *similar* velocity profiles in the boundary layer, so that when written as $u/U = f(y/\delta)$, velocity profiles do not vary with x. He used a similarity variable to reduce the partial differential equations of motion and continuity to a single third-order ordinary differential equation.

Blasius used numerical methods to solve the ordinary differential equation. Unfortunately, the velocity profile must be expressed in tabular form. The principal results of the Blasius solution may be expressed as

$$\frac{\delta}{x} = \frac{5}{\sqrt{\mathrm{Re}_x}} \tag{3.6.4}$$

and

$$C_f = \frac{\tau_w}{\frac{1}{2}\rho U^2} = \frac{0.664}{\sqrt{\mathrm{Re}_x}} \tag{3.6.5}$$

These results characterize the laminar boundary layer on a flat plate; they show that laminar boundary layer thickness varies as $x^{1/2}$ and wall shear stress varies as $1/x^{1/2}$.

Approximate Solutions

The Blasius solution cannot be expressed in closed form and is limited to laminar flow. Therefore, approximate methods that give solutions for both laminar and turbulent flow in closed form are desirable. One such method is the *momentum integral equation* (MIE), which may be developed by integrating the boundary layer equation across the boundary layer or by applying the streamwise momentum equation to a differential control volume (Fox and McDonald, 1992). The result is the ordinary differential equation

$$\frac{d\theta}{dx} = \frac{\tau_w}{\rho U^2} - \left(\frac{\delta^*}{\theta} + 2\right)\frac{\theta}{U}\frac{dU}{dx} \tag{3.6.6}$$

The first term on the right side of Equation (3.6.6) contains the influence of wall shear stress. Since τ_w is always positive, it always causes θ to increase. The second term on the right side contains the pressure gradient, which can have either sign. Therefore, the effect of the pressure gradient can be to either increase or decrease the rate of growth of boundary layer thickness.

Equation (3.6.6) is an ordinary differential equation that can be solved for θ as a function of x on a flat plate (zero pressure gradient), provided a reasonable shape is assumed for the boundary layer velocity profile and shear stress is expressed in terms of the other variables. Results for laminar and turbulent flat-plate boundary layer flows are discussed below.

Laminar Boundary Layers. A reasonable approximation to the laminar boundary layer velocity profile is to express u as a polynomial in y. The resulting solutions for δ and τ_w have the same dependence on x as the exact Blasius solution. Numerical results are presented in Table 3.6.1. Comparing the approximate and exact solutions shows remarkable agreement in view of the approximations used in the analysis. The trends are predicted correctly and the approximate values are within 10% of the exact values.

Turbulent Boundary Layers. The turbulent velocity profile may be expressed well using a power law, $u/U = (y/\delta)^{1/n}$, where n is an integer between 6 and 10 (frequently 7 is chosen). For turbulent flow it is

TABLE 3.6.1 Exact and Approximate Solutions for Laminar Boundary Layer Flow over a Flat Plate at Zero Incidence

Velocity Distribution $\frac{u}{U} = f\left(\frac{y}{\delta}\right) = f(\eta)$	$\frac{\theta}{\delta}$	$\frac{\delta^*}{\delta}$	$a = \frac{\delta}{x}\sqrt{\mathrm{Re}_x}$	$b = C_f\sqrt{\mathrm{Re}_x}$
$f(\eta) = 2\eta - \eta^2$	2/15	1/3	5.48	0.730
$f(\eta) = 3/2\ \eta - 1/2\ \eta^3$	39/280	3/8	4.64	0.647
$f(\eta) = \sin(\pi/2\ \eta)$	$(4-\pi)/2\pi$	$(\pi-2)/\pi$	4.80	0.654
Exact	0.133	0.344	5.00	0.664

not possible to express shear stress directly in terms of a simple velocity profile; an empirical correlation is required. Using a pipe flow data correlation gives

$$\frac{\delta}{x} = \frac{0.382}{\mathrm{Re}_x^{1/5}} \tag{3.6.7}$$

and

$$C_f = \frac{\tau_w}{\frac{1}{2}\rho U^2} = \frac{0.0594}{\mathrm{Re}_x^{1/5}} \tag{3.6.8}$$

These results characterize the turbulent boundary layer on a flat plate. They show that turbulent boundary layer thickness varies as $x^{4/5}$ and wall shear stress varies as $1/x^{1/5}$.

Approximate results for laminar and turbulent boundary layers are compared in Table 3.6.2. At a Reynolds number of 1 million, wall shear stress for the turbulent boundary layer is nearly six times as large as for the laminar layer. For a turbulent boundary layer, thickness increases five times faster with distance along the surface than for a laminar layer. These approximate results give a physical feel for relative magnitudes in the two cases.

TABLE 3.6.2 Thickness and Skin Friction Coefficient for Laminar and Turbulent Boundary Layers on a Flat Plate

Reynolds Number	Boundary Layer Thickness/x		Skin Friction Coefficient		Turbulent/Laminar Ratio	
	Laminar BL	Turbulent BL	Laminar BL	Turbulent BL	BL Thickness	Skin Friction
2E + 05	0.0112	0.0333	0.00148	0.00517	2.97	3.48
5E + 05	0.00707	0.0277	0.000939	0.00431	3.92	4.58
1E + 06	0.00500	0.0241	0.000664	0.00375	4.82	5.64
2E + 06	0.00354	0.0210	0.000470	0.00326	5.93	6.95
5E + 06	0.00224	0.0175	0.000297	0.00272	7.81	9.15
1E + 07	0.00158	0.0152	0.000210	0.00236	9.62	11.3
2E + 07	0.00112	0.0132	0.000148	0.00206	11.8	13.9
5E + 07	0.000707	0.0110	0.0000939	0.00171	15.6	18.3

Note: BL = boundary layer.

The MIE cannot be solved in closed form for flows with nonzero pressure gradients. However, the role of the pressure gradient can be understood qualitatively by studying the MIE.

Effect of Pressure Gradient

Boundary layer flow with favorable, zero, and adverse pressure gradients is depicted schematically in Figure 3.6.3. (Assume a thin boundary layer, so flow on the lower surface behaves as external flow on

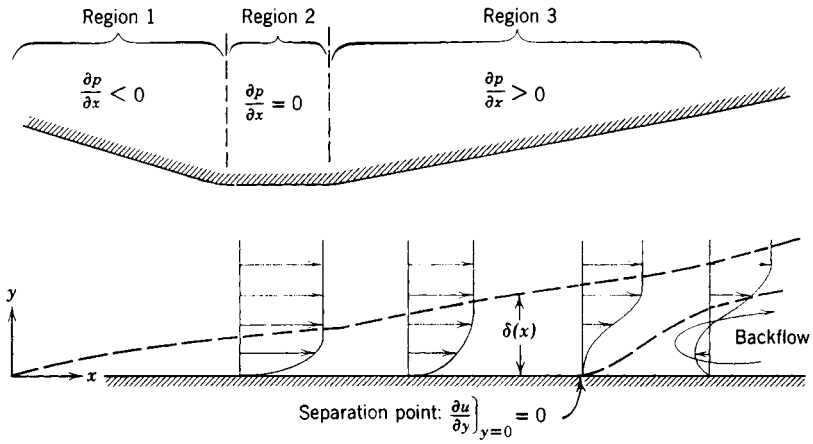


FIGURE 3.6.3 Boundary layer flow with presssure gradient (thickness exaggerated for clarity).

a surface, with the pressure gradient impressed on the boundary layer.) The pressure gradient is favorable when $\partial p/\partial x < 0$, zero when $\partial p/\partial x = 0$, and adverse when $\partial p/\partial x > 0$, as indicated for Regions 1, 2, and 3.

Viscous shear always causes a net retarding force on any fluid particle within the boundary layer. For zero pressure gradient, shear forces alone can never bring the particle to rest. (Recall that for laminar and turbulent boundary layers the shear stress varied as $1/x^{1/2}$ and $1/x^{1/5}$, respectively; shear stress never becomes zero for finite x.) Since shear stress is given by $\tau_w = \mu\ \partial u/\partial y)_{y=0}$, the velocity gradient cannot be zero. Therefore, flow cannot separate in a zero pressure gradient; shear stresses alone can never cause flow separation.

In the favorable pressure gradient of Region 1, pressure forces tend to maintain the motion of the particle, so flow cannot separate. In the adverse pressure gradient of Region 3, pressure forces oppose the motion of a fluid particle. An adverse pressure gradient is a necessary condition for flow separation.

Velocity profiles for laminar and turbulent boundary layers are shown in Figure 3.6.2. It is easy to see that the turbulent velocity profile has much more momentum than the laminar profile. Therefore, the turbulent velocity profile can resist separation in an adverse pressure gradient better than the laminar profile.

The freestream velocity distribution must be known before the MIE can be applied. We obtain a first approximation by applying potential flow theory to calculate the flow field around the object. Much effort has been devoted to calculation of velocity distributions over objects of known shape (the "direct" problem) and to determination of shapes to produce a desired pressure distribution (the "inverse" problem). Detailed discussion of such calculation schemes is beyond the scope of this section; the state of the art continues to progress rapidly.

Drag

Any object immersed in a viscous fluid flow experiences a net force from the shear stresses and pressure differences caused by the fluid motion. *Drag* is the force component parallel to, and *lift* is the force component perpendicular to, the flow direction. *Streamlining* is the art of shaping a body to reduce fluid dynamic drag. Airfoils (hydrofoils) are designed to produce lift in air (water); they are streamlined to reduce drag and thus to attain high lift–drag ratios.

In general, lift and drag cannot be predicted analytically for flows with separation, but progress continues on computational fluid dynamics methods. For many engineering purposes, drag and lift forces are calculated from experimentally derived coefficients, discussed below.

Drag coefficient is defined as

$$C_D = \frac{F_D}{\frac{1}{2}\rho V^2 A} \tag{3.6.9}$$

where $^1/_2\rho V^2$ is dynamic pressure and A is the area upon which the coefficient is based. Common practice is to base drag coefficients on projected *frontal area* (Fox and McDonald, 1992).

Similitude was treated in Section 3.3. In general, the drag coefficient may be expressed as a function of Reynolds number, Mach number, Froude number, relative roughness, submergence divided by length, and so forth. In this section we consider neither high-speed flow nor free-surface effects, so we will consider only Reynolds number and roughness effects on drag coefficient.

Friction Drag

The total friction drag force acting on a plane surface aligned with the flow direction can be found by integrating the shear stress distribution along the surface. The drag coefficient for this case is defined as friction force divided by dynamic pressure and *wetted area* in contact with the fluid. Since shear stress is a function of Reynolds number, so is drag coefficient (see Figure 3.6.4). In Figure 3.6.4, transition occurs at Re_x = 500,000; the dashed line represents the drag coefficient at larger Reynolds numbers. A number of empirical correlations may be used to model the variation in C_D shown in Figure 3.6.4 (Schlichting, 1979).

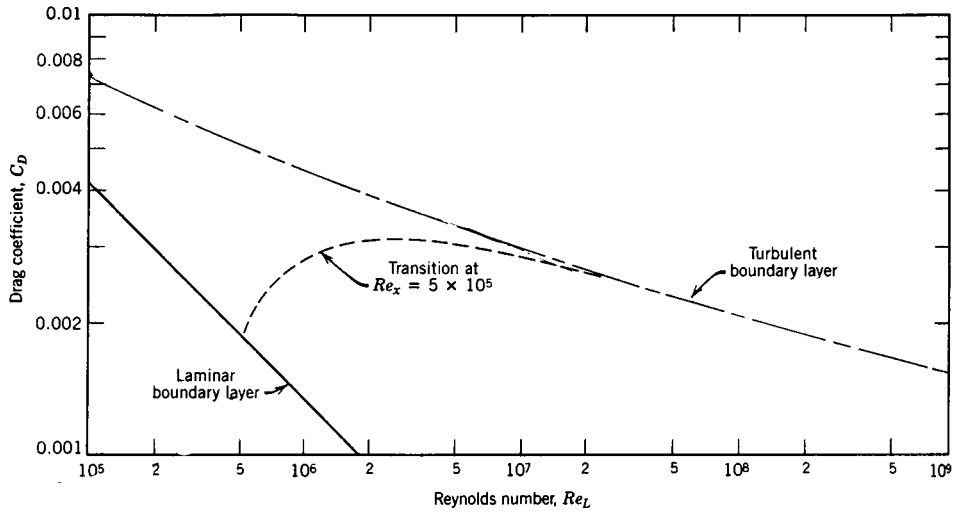


FIGURE 3.6.4 Drag coefficient vs. Reynolds number for a smooth flat plate parallel to the flow.

Extending the laminar boundary layer line to higher Reynolds numbers shows that it is beneficial to delay transition to the highest possible Reynolds number. Some results are presented in Table 3.6.3; drag is reduced more than 50% by extending laminar boundary layer flow to $\mathrm{Re}_L = 10^6$.

Pressure Drag

A thin flat surface normal to the flow has no area parallel to the flow direction. Therefore, there can be no friction force parallel to the flow; all drag is caused by pressure forces. Drag coefficients for objects with sharp edges tend to be independent of Reynolds number (for Re > 1000), because the separation points are fixed by the geometry of the object. Drag coefficients for selected objects are shown in Table 3.6.4.

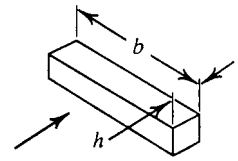
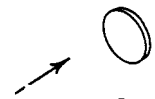
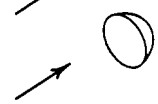
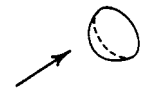
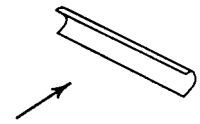
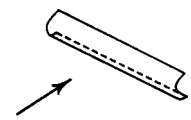
Rounding the edges that face the flow reduces drag markedly. Compare the drag coefficients for the hemisphere and C-section shapes facing into and away from the flow. Also note that the drag coefficient

TABLE 3.6.3 Drag Coefficients for Laminar, Turbulent, and Transition Boundary Layers on a Flat Plate

Reynolds Number	Drag Coefficient			Laminar/ Transition	% Drag Reduction
	Laminar BL	Turbulent BL	Transition		
2E + 05	0.00297	0.00615	—	—	—
5E + 05	0.00188	0.00511	0.00189	—	—
1E + 06	0.00133	0.00447	0.00286	0.464	53.6
2E + 06	0.000939	0.00394	0.00314	0.300	70.0
5E + 06	0.000594	0.00336	0.00304	0.195	80.5
1E + 07	0.000420	0.00300	0.00284	0.148	85.2
2E + 07	0.000297	0.00269	0.00261	0.114	88.6
5E + 07	0.000188	0.00235	0.00232	0.081	9.19

Note: BL = Boundary layer.

TABLE 3.6.4 Drag Coefficient Data for Selected Objects (Re > 1000)

Object	Diagram		$C_D(RE^* \gtrsim 10^3)$
Square prism		$b/h = \infty$	2.05
		$b/h = 1$	1.05
Disk			1.17
Ring			1.20[b]
Hemisphere (open end facing flow)			1.42
Hemisphere (open end facing downstream)			0.38
C-section (open side facing flow)			2.30
C-section (open side facing downstream)			1.20

[a] Data from Hoerner, 1965.
[b] Based on ring area.

for a two-dimensional object (long square cylinder) is about twice that for the corresponding three-dimensional object (square cylinder with $b/h = 1$).

Friction and Pressure Drag: Bluff Bodies

Both friction and pressure forces contribute to the drag of *bluff bodies* (see Shapiro, 1960, for a good discussion of the mechanisms of drag). As an example, consider the drag coefficient for a smooth sphere shown in Figure 3.6.5. Transition from laminar to turbulent flow in the boundary layers on the forward portion of the sphere causes a dramatic dip in drag coefficient at the *critical Reynolds number* ($Re_D \approx 2 \times 10^5$). The turbulent boundary layer is better able to resist the adverse pressure gradient on the rear of the sphere, so separation is delayed and the wake is smaller, causing less pressure drag.

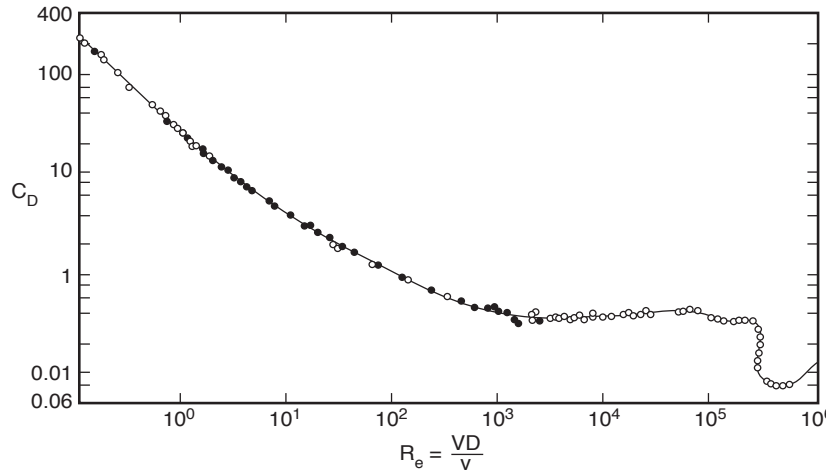


FIGURE 3.6.5 Drag coefficient vs. Reynolds number for a smooth sphere.

Surface roughness (or freestream disturbances) can reduce the critical Reynolds number. Dimples on a golf ball cause the boundary layer to become turbulent and, therefore, lower the drag coefficient in the range of speeds encountered in a drive.

Streamlining

Streamlining is adding a faired tail section to reduce the extent of separated flow on the downstream portion of an object (at high Reynolds number where pressure forces dominate drag). The adverse pressure gradient is taken over a longer distance, delaying separation. However, adding a faired tail increases surface area, causing skin friction drag to increase. Thus, streamlining must be optimized for each shape.

Front contours are of principal importance in road vehicle design; the angle of the back glass also is important (in most cases the entire rear end cannot be made long enough to control separation and reduce drag significantly).

Lift

Lift coefficient is defined as

$$C_L = \frac{F_L}{\frac{1}{2}\rho V^2 A} \tag{3.6.10}$$

Note that lift coefficient is based on projected *planform area*.

Airfoils

Airfoils are shaped to produce lift efficiently by accelerating flow over the upper surface to produce a low-pressure region. Because the flow must again decelerate, inevitably there must be a region of adverse pressure gradient near the rear of the upper surface (pressure distributions are shown clearly in Hazen, 1965).

Lift and drag coefficients for airfoil sections depend on Reynolds number and *angle of attack* between the chord line and the undisturbed flow direction. The *chord line* is the straight line joining the leading and trailing edges of the airfoil (Abbott and von Doenhoff, 1959).

As the angle of attack is increased, the minimum pressure point moves forward on the upper surface and the minimum pressure becomes lower. This increases the adverse pressure gradient. At some angle of attack, the adverse pressure gradient is strong enough to cause the boundary layer to separate completely from the upper surface, causing the airfoil to *stall*. The separated flow alters the pressure distribution, reducing lift sharply.

Increasing the angle of attack also causes the the drag coefficient to increase. At some angle of attack below stall the ratio of lift to drag, the *lift–drag* ratio, reaches a maximum value.

Drag Due to Lift

For wings (airfoils of finite span), lift and drag also are functions of aspect ratio. Lift is reduced and drag increased compared with infinite span, because end effects cause the lift vector to rotate rearward. For a given geometric angle of attack, this reduces effective angle of attack, reducing lift. The additional component of lift acting in the flow direction increases drag; the increase in drag due to lift is called *induced drag*.

The effective aspect ratio includes the effect of planform shape. When written in terms of effective aspect ratio, the drag of a finite-span wing is

$$C_D = C_{D,\infty} + \frac{C_L^2}{\pi ar} \tag{3.6.11}$$

where ar is effective aspect ratio and the subscript ∞ refers to the infinite section drag coefficient at C_L. For further details consult the references.

The lift coefficient must increase to support aircraft weight as speed is reduced. Therefore, induced drag can increase rapidly at low flight speeds. For this reason, minimum allowable flight speeds for commercial aircraft are closely controlled by the FAA.

Boundary Layer Control

The major part of the drag on an airfoil or wing is caused by skin friction. Therefore, it is important to maintain laminar flow in the boundary layers as far aft as possible; laminar flow sections are designed to do this. It also is important to prevent flow separation and to achieve high lift to reduce takeoff and landing speeds. These topics fall under the general heading of boundary layer control.

Profile Shaping

Boundary layer transition on a conventional airfoil section occurs almost immediately after the minimum pressure at about 25% chord aft the leading edge. Transition can be delayed by shaping the profile to maintain a favorable pressure gradient over more of its length. The U.S. National Advisory Committee for Aeronautics (NACA) developed several series of profiles that delayed transition to 60 or 65% of chord, reducing drag coefficients (in the design range) 60% compared with conventional sections of the same thickness ratio (Abbott and von Doenhoff, 1959).

Flaps and Slats

Flaps are movable sections near the trailing edge of a wing. They extend and/or deflect to increase wing area and/or increase wing camber (curvature), to provide higher lift than the clean wing. Many aircraft also are fitted with leading edge slats which open to expose a slot from the pressure side of the wing to the upper surface. The open slat increases the effective radius of the leading edge, improving maximum lift coefficient. The slot allows energized air from the pressure surface to flow into the low-pressure region atop the wing, energizing the boundary layers and delaying separation and stall.

Suction and Blowing

Suction removes low-energy fluid from the boundary layer, reducing the tendency for early separation. Blowing via high-speed jets directed along the surface reenergizes low-speed boundary layer fluid. The objective of both approaches is to delay separation, thus increasing the maximum lift coefficient the wing can achieve. Powered systems add weight and complexity; they also require bleed air from the engine compressor, reducing thrust or power output.

Moving Surfaces

Many schemes have been proposed to utilize moving surfaces for boundary layer control. Motion in the direction of flow reduces skin friction, and thus the tendency to separate; motion against the flow has the opposite effect. The aerodynamic behavior of sports balls — baseballs, golf balls, and tennis balls — depends significantly on aerodynamic side force (lift, down force, or side force) produced by spin. These effects are discussed at length in Fox and McDonald (1992) and its references.

Computation vs. Experiment

Experiments cannot yet be replaced completely by analysis. Progress in modeling, numerical techniques, and computer power continues to be made, but the role of the experimentalist likely will remain important for the foreseeable future.

Computational Fluid Dynamics (CFD)

Computation of fluid flow requires accurate mathematical modeling of flow physics and accurate numerical procedures to solve the equations. The basic equations for laminar boundary layer flow are well known. For turbulent boundary layers generally it is not possible to resolve the solution space into sufficiently small cells to allow direct numerical simulation. Instead, empirical models for the turbulent stresses must be used. Advances in computer memory storage capacity and speed (e.g., through use of massively parallel processing) continue to increase the resolution that can be achieved.

A second source of error in CFD work results from the numerical procedures required to solve the equations. Even if the equations are exact, approximations must be made to discretize and solve them using finite-difference or finite-volume methods. Whichever is chosen, the solver must guard against introducing numerical instability, round-off errors, and numerical diffusion (Hoffman, 1992).

Role of the Wind Tunnel

Traditionally, wind tunnel experiments have been conducted to verify the design and performance of components and complete aircraft. Design verification of a modern aircraft may require expensive scale models, several thousand hours of wind tunnel time at many thousands of dollars an hour, and additional full-scale flight testing.

New wind tunnel facilities continue to be built and old ones refurbished. This indicates a need for continued experimental work in developing and optimizing aircraft configurations.

Many experiments are designed to produce baseline data to validate computer codes. Such systematic experimental data can help to identify the strengths and weaknesses of computational methods.

CFD tends to become only indicative of trends when massive zones of flow separation are present. Takeoff and landing configurations of conventional aircraft, with landing gear, high-lift devices, and

flaps extended, tend to need final experimental confirmation and optimization. Many studies of vertical takeoff and vectored thrust aircraft require testing in wind tunnels.

Defining Terms

Boundary layer: Thin layer of fluid adjacent to a surface where viscous effects are important; viscous effects are negligible outside the boundary layer.

Drag coefficient: Force in the flow direction exerted on an object by the fluid flowing around it, divided by dynamic pressure and area.

Lift coefficient: Force perpendicular to the flow direction exerted on an object by the fluid flowing around it, divided by dynamic pressure and area.

Pressure gradient: Variation in pressure along the surface of an object. For a *favorable* pressure gradient, pressure *decreases* in the flow direction; for an *adverse* pressure gradient, pressure *increases* in the flow direction.

Separation: Phenomenon that occurs when fluid layers adjacent to a solid surface are brought to rest and boundary layers depart from the surface contour, forming a low-pressure *wake* region. Separation can occur only in an *adverse pressure gradient*.

Transition: Change from laminar to turbulent flow within the boundary layer. The location depends on distance over which the boundary layer has developed, pressure gradient, surface roughness, freestream disturbances, and heat transfer.

References

Abbott, I.H. and von Doenhoff, A.E. 1959. *Theory of Wing Sections, Including a Summary of Airfoil Data*. Dover, New York.

Fox, R.W. and McDonald, A.T. 1992. *Introduction to Fluid Mechanics,* 4th ed. John Wiley & Sons, New York.

Hazen, D.C. 1965. *Boundary Layer Control,* film developed by the National Committee for Fluid Mechanics Films (NCFMF) and available on videotape from Encyclopaedia Britannica Educational Corporation, Chicago.

Hoerner, S.F. 1965. *Fluid-Dynamic Drag,* 2nd ed. Published by the author, Midland Park, NJ.

Hoffman, J.D. 1992. *Numerical Methods for Engineers and Scientists*. McGraw-Hill, New York.

Schlichting, H. 1979. *Boundary-Layer Theory*, 7th ed. McGraw-Hill, New York.

Shapiro, A.H. 1960. *The Fluid Dynamics of Drag*, film developed by the National Committee for Fluid Mechanics Film (NCFMF) and available on videotape from Encyclopaedia Britannica Educational Corporation, Chicago.

Further Information

A comprehensive source of basic information is the *Handbook of Fluid Dynamics,* edited by Victor L. Streeter (McGraw-Hill, New York, 1960).

Timely reviews of important topics are published in the *Annual Review of Fluid Mechanics* series (Annual Reviews, Inc., Palo Alto, CA.). Each volume contains a cumulative index.

ASME (American Society of Mechanical Engineers, New York, NY) publishes the *Journal of Fluids Engineering* quarterly. *JFE* contains fluid machinery and other engineering applications of fluid mechanics.

The monthly *AIAA Journal* and bimonthly *Journal of Aircraft* (American Institute for Aeronautics and Astronautics, New York) treat aerospace applications of fluid mechanics.

3.7 Compressible Flow

Ajay Kumar

Introduction

This section deals with compressible flow. Only one- or two-dimensional steady, inviscid flows under perfect gas assumption are considered. Readers are referred to other sources of information for unsteady effects, viscous effects, and three-dimensional flows.

The term *compressible flow* is routinely used to define variable density flow which is in contrast to incompressible flow, where the density is assumed to be constant throughout. In many cases, these density variations are principally caused by the pressure changes from one point to another. Physically, the *compressibility* can be defined as the fractional change in volume of the gas element per unit change in pressure. It is a property of the gas and, in general, can be defined as

$$\tau = \frac{1}{\rho}\frac{d\rho}{dp}$$

where τ is the compressibility of the gas, ρ is the density, and p is the pressure being exerted on the gas. A more precise definition of compressibility is obtained if we take into account the thermal and frictional losses. If during the compression the temperature of the gas is held constant, it is called the isothermal compressibility and can be written as

$$\tau_T = \frac{1}{\rho}\left(\frac{\partial\rho}{\partial p}\right)_T$$

However, if the compression process is reversible, it is called the isentropic compressibility and can be written as

$$\tau_s = \frac{1}{\rho}\left(\frac{\partial\rho}{\partial p}\right)_s$$

Gases in general have high compressibility (τ_T for air is 10^{-5} m^2/N at 1 atm) as compared with liquids (τ_T for water is 5×10^{-10} m^2/N at 1 atm).

Compressibility is a very important parameter in the analysis of compressible flow and is closely related to the *speed of sound*, a, which is the velocity of propagation of small pressure disturbances and is defined as

$$a^2 = \left(\frac{\partial p}{\partial\rho}\right)_s \quad \text{or} \quad a = \sqrt{\left(\frac{\partial p}{\partial\rho}\right)_s}$$

In an isentropic process of a perfect gas, the pressure and density are related as

$$\frac{p}{\rho^\gamma} = \text{constant}$$

Using this relation along with the perfect gas relation $p = \rho RT$, we can show that for a perfect gas

$$a = \sqrt{\gamma RT} = \sqrt{\frac{\gamma p}{\rho}}$$

where γ is the ratio of specific heats at constant pressure and constant volume, R is the gas constant, and T is the temperature. For air under normal conditions, γ is 1.4 and R is 287 m^2/sec^2 K so that the speed of sound for air becomes $a = 20.045\sqrt{T}$ m/sec where T is in kelvin.

Another important parameter in compressible flows is the *Mach number, M,* which is defined as the ratio of the gas velocity to the speed of sound or

$$M = \frac{V}{a}$$

where V is the velocity of gas. Depending upon the Mach number of the flow, we can define the following flow regimes:

$M \ll 1$ Incompressible flow

$M < 1$ Subsonic flow

$M \approx 1$ Transonic flow

$M > 1$ Supersonic flow

$M \gg 1$ Hypersonic flow

Subsonic through hypersonic flows are compressible in nature. In these flows, the velocity is appreciable compared with the speed of sound, and the fractional changes in pressure, temperature, and density are all of significant magnitude. We will restrict ourselves in this section to subsonic through flows only.

Before we move on to study these flows, let us define one more term. Let us consider a gas with static pressure p and temperature T, traveling at some velocity V and corresponding Mach number M. If this gas is brought isentropically to stagnation or zero velocity, the pressure and temperature which the gas achieves are defined as *stagnation pressure* p_0 and *stagnation temperature* T_0 (also called total pressure and total temperature). The speed of sound at stagnation conditions is called the *stagnation speed of sound* and is denoted as a_0.

One-Dimensional Flow

In one-dimensional flow, the flow properties vary only in one coordinate direction. Figure 3.7.1 shows two streamtubes in a flow. In a *truly one-dimensional flow* illustrated in Figure 3.7.1(a), the flow variables are a function of x only and the area of the stream tube is constant. On the other hand, Figure 3.7.1(b) shows a flow where the area of the stream tube is also a function of x but the flow variables are still a function of x only. This flow is defined as the *quasi-one-dimensional flow.* We will first discuss the truly one-dimensional flow.

In a steady, truly one-dimensional flow, conservation of mass, momentum, and energy leads to the following simple algebraic equations.

$$\begin{aligned} \rho u &= \text{constant} \\ p + \rho u^2 &= \text{constant} \\ h + \frac{u^2}{2} + q &= \text{constant} \end{aligned} \tag{3.7.1}$$

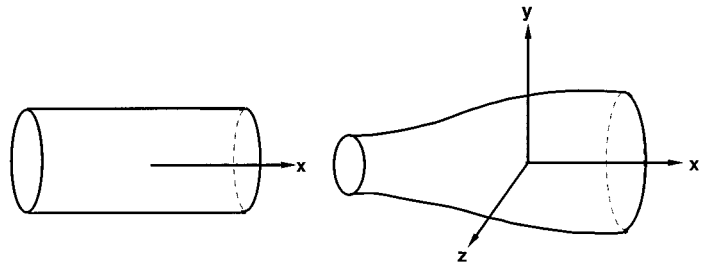


FIGURE 3.7.1 (a) One-dimensional flow; (b) quasi-one-dimensional flow.

where q is the heat added per unit mass of the gas. These equations neglect body forces, viscous stresses, and heat transfer due to thermal conduction and diffusion. These relations given by Equation 3.7.1, when applied at points 1 and 2 in a flow with no heat addition, become

$$\begin{aligned} \rho_1 u_1 &= \rho_2 u_2 \\ p_1 + \rho_1 u_1^2 &= p_2 + \rho_2 u_2^2 \\ h_1 + \frac{u_1^2}{2} &= h_2 + \frac{u_2^2}{2} \end{aligned} \tag{3.7.2}$$

The energy equation for a calorically perfect gas, where $h = c_p T$, becomes

$$c_p T_1 + \frac{u_1^2}{2} = c_p T_2 + \frac{u_2^2}{2}$$

Using $c_p = \gamma R/(\gamma - 1)$ and $a^2 = \gamma RT$, the above equation can be written as

$$\frac{a_1^2}{\gamma - 1} + \frac{u_1^2}{2} = \frac{a_2^2}{\gamma - 1} + \frac{u_2^2}{2} \tag{3.7.3}$$

Since Equation (3.7.3) is written for no heat addition, it holds for an adiabatic flow. If the energy equation is applied to the stagnation conditions, it can be written as

$$c_p T + \frac{u^2}{2} = c_p T_0$$

$$\frac{T_0}{T} = 1 + \frac{\gamma - 1}{2} M^2 \tag{3.7.4}$$

It is worth mentioning that in arriving at Equation (3.7.4), only adiabatic flow condition is used whereas stagnation conditions are defined as those where the gas is brought to rest isentropically. Therefore, the definition of stagnation temperature is less restrictive than the general definition of stagnation conditions. According to the general definition of isentropic flow, it is a reversible adiabatic flow. This definition is needed for the definition of stagnation pressure and density. For an isentropic flow,

$$\frac{p_0}{p} = \left(\frac{\rho_0}{\rho}\right)^{\gamma} = \left(\frac{T_0}{T}\right)^{\gamma/(\gamma-1)} \tag{3.7.5}$$

From Equations 3.7.4 and 3.7.5, we can write

$$\frac{p_0}{p} = \left(1 + \frac{\gamma-1}{2}M^2\right)^{\gamma/(\gamma-1)} \tag{3.7.6}$$

$$\frac{\rho_0}{\rho} = \left(1 + \frac{\gamma-1}{2}M^2\right)^{1/(\gamma-1)} \tag{3.7.7}$$

Values of stagnation conditions are tabulated in Anderson (1982) as a function of M for $\gamma = 1.4$.

Normal Shock Wave

A shock wave is a very thin region (of the order of a few molecular mean free paths) across which the static pressure, temperature, and density increase whereas the velocity decreases. If the shock wave is perpendicular to the flow, it is called a *normal shock wave*. The flow is supersonic ahead of the normal shock wave and subsonic behind it. Figure 3.7.2 shows the flow conditions across a normal shock wave which is treated as a discontinuity. Since there is no heat added or removed, the flow across the shock wave is adiabatic. By using Equations 3.7.2 the normal shock equations can be written as

$$\begin{aligned} \rho_1 u_1 &= \rho_2 u_2 \\ p_1 + \rho_1 u_1^2 &= p_2 + \rho_2 u_2^2 \\ h_1 + \frac{u_1^2}{2} &= h_2 + \frac{u_2^2}{2} \end{aligned} \tag{3.7.8}$$

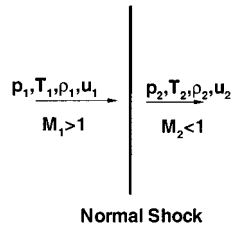


FIGURE 3.7.2 Flow conditions across a normal shock.

Equations (3.7.8) are applicable to a general type of flow; however, for a calorically perfect gas, we can use the relations $p = \rho RT$ and $h = c_p T$ to derive a number of equations relating flow conditions downstream of the normal shock to those at upstream. These equations (also known as Rankine–Hugoniot relations) are

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma+1}\left(M_1^2 - 1\right)$$

$$\frac{\rho_2}{\rho_1} = \frac{u_1}{u_2} = \frac{(\gamma+1)M_1^2}{2 + (\gamma-1)M_1^2} \tag{3.7.9}$$

$$\frac{T_2}{T_1} = \frac{h_2}{h_1} = \left[1 + \frac{2\gamma}{\gamma+1}\left(M_1^2 - 1\right)\right]\left[\frac{2 + (\gamma-1)M_1^2}{(\gamma+1)M_1^2}\right]$$

$$M_2^2 = \frac{1 + \dfrac{\gamma-1}{2}M_1^2}{\gamma M_1^2 - \dfrac{\gamma-1}{2}}$$

Again, the values of p_2/p_1, ρ_2/ρ_1, T_2/T_1, etc. are tabulated in Anderson (1982) as a function of M_1 for $\gamma = 1.4$. Let us examine some limiting cases. As $M_1 \to 1$, Equations 3.7.9 yield $M_2 \to 1$, $p_2/p_1 \to 1$, $\rho_2/\rho_1 \to 1$, and $T_2/T_1 \to 1$. This is the case of an extremely weak normal shock across which no finite changes occur. This is the same as the sound wave. On the other hand, as $M_1 \to \infty$, Equations (3.7.9) yield

$$M_2 \to \sqrt{\frac{\gamma-1}{2\gamma}} = 0.378; \quad \frac{\rho_2}{\rho_1} \to \frac{\gamma+1}{\gamma-1} = 6; \quad \frac{p_2}{p_1} \to \infty; \quad \frac{T_2}{T_1} \to \infty$$

However, the calorically perfect gas assumption no longer remains valid as $M_1 \to \infty$.

Let us now examine why the flow ahead of a normal shock wave must be supersonic even though Equations (3.7.8) hold for $M_1 < 1$ as well as $M_1 > 1$. From the second law of thermodynamics, the entropy change across the normal shock can be written as

$$s_2 - s_1 = c_p \ln\frac{T_2}{T_1} - R\ln\frac{p_2}{p_1}$$

By using Equations (3.7.9) it becomes

$$s_2 - s_1 = c_p \ln\left\{\left[1 + \frac{2\gamma}{\gamma+1}\left(M_1^2 - 1\right)\right]\left[\frac{2 + (\gamma-1)M_1^2}{(\gamma+1)M_1^2}\right]\right\} - R\ln\left[1 + \frac{2\gamma}{\gamma+1}\left(M_1^2 - 1\right)\right] \qquad (3.7.10)$$

Equation (3.7.10) shows that the entropy change across the normal shock is also a function of M_1 only. Using Equation (3.7.10) we see that

$$\begin{aligned} s_2 - s_1 &= 0 \quad \text{for} \quad M_1 = 1 \\ &< 0 \quad \text{for} \quad M_1 < 1 \\ &> 0 \quad \text{for} \quad M_1 > 1 \end{aligned}$$

Since it is necessary that $s_2 - s_1 \geq 0$ from the second law, $M_1 \geq 1$. This, in turn, requires that $p_2/p_1 \geq 1$, $\rho_2/\rho_1 \geq 1$, $T_2/T_1 \geq 1$, and $M_2 \leq 1$.

We now examine how the stagnation conditions change across a normal shock wave. For a calorically perfect gas, the energy equation in Equations (3.7.9) gives

$$c_p T_{01} = c_p T_{02} \quad \text{or} \quad T_{01} = T_{02}$$

In other words, the total temperature remains constant across a stationary normal shock wave.

Let us now apply the entropy change relation across the shock using the stagnation conditions.

$$s_2 - s_1 = c_p \ln \frac{T_{02}}{T_{01}} - R \ln \frac{p_{02}}{p_{01}}$$

Note that entropy at stagnation conditions is the same as at the static conditions since to arrive at stagnation conditions, the gas is brought to rest isentropically. Since $T_{02} = T_{01}$,

$$s_2 - s_1 = -R \ln \frac{p_{02}}{p_{01}}$$

$$\frac{p_{02}}{p_{01}} = e^{-(s_2 - s_1)/R} \tag{3.7.11}$$

Since $s_2 > s_1$ across the normal shockwave, Equation (3.7.11) gives $P_{02} < P_{01}$ or, in other words, the total pressure decreases across a shock wave.

One-Dimensional Flow with Heat Addition

Consider one-dimensional flow through a control volume as shown in Figure 3.7.3. Flow conditions going into this control volume are designated by 1 and coming out by 2. A specified amount of heat per unit mass, q, is added to the control volume. The governing equations relating conditions 1 and 2 can be written as

$$\begin{aligned} \rho_1 u_1 &= \rho_2 u_2 \\ p_1 + \rho_1 u_1^2 &= p_2 + \rho_2 u_2^2 \\ h_1 + \frac{u_1^2}{2} + q &= h_2 + \frac{u_2^2}{2} \end{aligned} \tag{3.7.12}$$

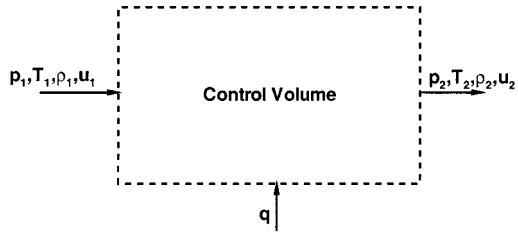


FIGURE 3.7.3 One-dimensional control volume with heat addition.

The following relations can be derived from Equation (3.7.12) for a calorically perfect gas

$$q = c_p \left(T_{02} - T_{01} \right) \tag{3.7.13}$$

$$\frac{p_2}{p_1} = \frac{1 + \gamma M_1^2}{1 + \gamma M_2^2} \tag{3.7.14}$$

$$\frac{T_2}{T_1} = \left(\frac{1+\gamma M_1^2}{1+\gamma M_2^2}\right)^2 \left(\frac{M_2}{M_1}\right)^2 \tag{3.7.15}$$

$$\frac{\rho_2}{\rho_1} = \left(\frac{1+\gamma M_2^2}{1+\gamma M_1^2}\right)^2 \left(\frac{M_1}{M_2}\right)^2 \tag{3.7.16}$$

Equation (3.7.13) indicates that the effect of heat addition is to directly change the stagnation temperature T_0 of the flow. Table 3.7.1 shows some physical trends which can be obtained with heat addition to subsonic and supersonic flow. With heat extraction the trends in Table 3.7.1 are reversed.

TABLE 3.7.1 Effect of Heat Addition on Subsonic and Supersonic Flow

	$M_1 < 1$	$M_1 > 1$
M_2	Increases	Decreases
p_2	Decreases	Increases
T_2	Increases for $M_1 < \gamma^{-1/2}$ and decreases for $M_1 > \gamma^{-1/2}$	Increases
u_2	Increases	Decreases
T_{02}	Increases	Increases
p_{02}	Decreases	Decreases

Figure 3.7.4 shows a plot between enthalpy and entropy, also known as the Mollier diagram, for one-dimensional flow with heat addition. This curve is called the Rayleigh curve and is drawn for a set of given initial conditions. Each point on this curve corresponds to a different amount of heat added or removed. It is seen from this curve that heat addition always drives the Mach numbers toward 1. For a certain amount of heat addition, the flow will become sonic. For this condition, the flow is said to be *choked*. Any further increase in heat addition is not possible without adjustment in initial conditions. For example, if more heat is added in region 1, which is initially supersonic, than allowed for attaining Mach 1 in region 2, then a normal shock will form inside the control volume which will suddenly change the conditions in region 1 to subsonic. Similarly, in case of an initially subsonic flow corresponding to region 1′, any heat addition beyond that is needed to attain Mach 1 in region 2, the conditions in region 1′ will adjust to a lower subsonic Mach number through a series of pressure waves.

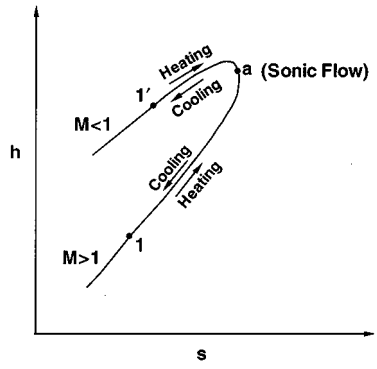


FIGURE 3.7.4 The Rayleigh curve.

Similar to the preceding heat addition or extraction relationships, we can also develop relationships for one-dimensional steady, adiabatic flow but with frictional effects due to viscosity. In this case, the momentum equation gets modified for frictional shear stress. For details, readers are referred to Anderson (1982).

Quasi-One-Dimensional Flow

In quasi-one-dimensional flow, in addition to flow conditions, the area of duct also changes with x. The governing equations for quasi-one-dimensional flow can be written in a differential form as follows using an infinitesimal control volume shown in Figure 3.7.5.

$$d(\rho u A) = 0 \tag{3.7.17}$$

$$dp + \rho u\, du = 0 \tag{3.7.18}$$

$$dh + u\, du = 0 \tag{3.7.19}$$

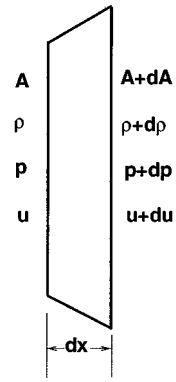


FIGURE 3.7.5 Control volume for quasi-one-dimensional flow.

Equation 3.7.17 can be written as

$$\frac{d\rho}{\rho} + \frac{du}{u} + \frac{dA}{A} = 0 \tag{3.7.20}$$

which can be further written as follows for an isentropic flow:

$$\frac{dA}{A} = \left(M^2 - 1\right)\frac{du}{u} \tag{3.7.21}$$

Some very useful physical insight can be obtained from this area–velocity relation.

- For subsonic flow ($0 \le M < 1$), an increase in area results in decrease in velocity, and vice versa.
- For supersonic flow ($M > 1$), an increase in area results in increase in velocity, and vice versa.
- For sonic flow ($M = 1$), $dA/A = 0$, which corresponds to a minimum or maximum in the area distribution, but it can be shown that a minimum in area is the only physical solution.

Figure 3.7.6 shows the preceding results in a schematic form.

It is obvious from this discussion that for a gas to go isentropically from subsonic to supersonic, and vice versa, it must flow through a convergent–divergent nozzle, also known as the de Laval nozzle. The minimum area of the nozzle at which the flow becomes sonic is called the throat. This physical observation forms the basis of designing supersonic wind tunnels shown schematically in Figure 3.7.7. In general, in a supersonic wind tunnel, a stagnant gas is first expanded to the desired supersonic Mach number. The supersonic flow enters the test section where it passes over a model being tested. The flow

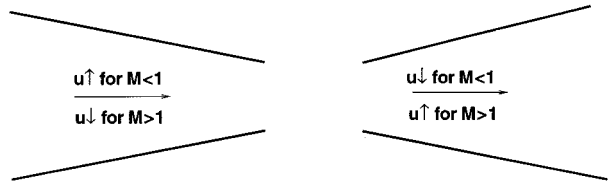


FIGURE 3.7.6 Compressible flow in converging and diverging ducts.

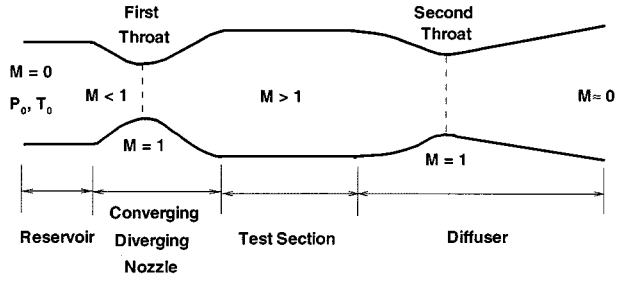


FIGURE 3.7.7 Schematic of a typical supersonic wind tunnel.

then is slowed down by compressing it through a second convergent–divergent nozzle, also known as a diffuser, before it is exhausted to the atmosphere.

Now, using the equations for quasi-one-dimensional flow and the isentropic flow conditions, we can derive a relation for the area ratio that is needed to accelerate or decelerate the gas to sonic conditions. Denoting the sonic conditions by an asterisk, we can write $u^* = a^*$. The area is denoted as A^*, and it is obviously the minimum area for the throat of the nozzle. From Equation (3.7.17) we have

$$\rho u A = \rho^* u^* A^*$$

$$\frac{A}{A^*} = \frac{\rho^* u^*}{\rho u} = \frac{\rho^*}{\rho_0}\frac{\rho_0}{\rho}\frac{u^*}{u} \tag{3.7.22}$$

Under isentropic conditons,

$$\frac{\rho_0}{\rho} = \left(1 + \frac{\gamma - 1}{2} M^2\right)^{1/(\gamma-1)} \tag{3.7.23}$$

$$\frac{\rho_0}{\rho^*} = \left(1 + \frac{\gamma - 1}{2}\right)^{1/(\gamma-1)} = \left(\frac{\gamma + 1}{2}\right)^{1/(\gamma-1)} \tag{3.7.24}$$

Also, $u^*/u = a^*/u$. Let us define a Mach number $M^* = u/a^*$. M^* is known as the *characteristic Mach number* and it is related to the local Mach number by the following relation:

$$M^{*2} = \frac{\dfrac{\gamma + 1}{2} M^2}{1 + \dfrac{\gamma - 1}{2} M^2} \tag{3.7.25}$$

Using Equations (3.7.23) through (3.7.25) in Equation (3.7.22) we can write

$$\left(\frac{A}{A^*}\right)^2 = \frac{1}{M^2}\left[\left(\frac{2}{\gamma+1}\right)\left(1+\frac{\gamma-1}{2}M^2\right)\right]^{(\gamma+1)/(\gamma-1)} \tag{3.7.26}$$

Equation (3.7.26) is called the area Mach number relation. Figure 3.7.8 shows a plot of A/A^* against Mach number. A/A^* is always ≥ 1 for physically viable solutions.

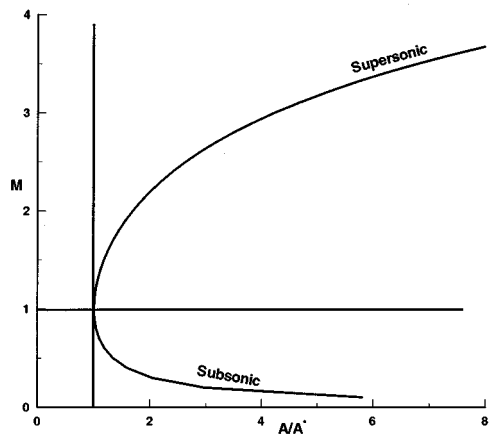


FIGURE 3.7.8 Variation of area ratio A/A^* as a function of Mach number for a quasi-one-dimensional flow.

The area Mach number relation says that for a given Mach number, there is only one area ratio A/A^*. This is a very useful relation and is frequently used to design convergent–divergent nozzles to produce a desired Mach number. Values of A/A^* are tabulated as a function of M in Anderson (1982).

Equation (3.7.26) can also be written in terms of pressure as follows:

$$\frac{A}{A^*} = \frac{\left[1-\left(\frac{p}{p_0}\right)^{(\gamma-1)/\gamma}\right]^{1/2}\left(\frac{p}{p_0}\right)^{1/\gamma}}{\left(\frac{\gamma-1}{2}\right)^{1/2}\left(\frac{2}{\gamma+1}\right)^{(\gamma+1)/2(\gamma-1)}} \tag{3.7.27}$$

Nozzle Flow

Using the area relations, we can now plot the distributions of Mach number and pressure along a nozzle. Figure 3.7.9 shows pressure and Mach number distributions along a given nozzle and the wave configurations for several exit pressures. For curves a and b, the flow stays subsonic throughout and the exit pressure controls the flow in the entire nozzle. On curve c, the throat has just become sonic, and so the pressure at the throat, and upstream of it, can decrease no further. There is another exit pressure corresponding to curve j ($p_j < p_c$) for which a supersonic isentropic solution exists. But if the pressure lies between p_c and p_j, there is no isentropic solution possible. For example, for an exit pressure p_d, a shock will form in the nozzle at location s which will raise the pressure to $p_{d'}$ and turn the flow subsonic. The pressure will then rise to p_d as the subsonic flow goes through an increasing area nozzle. The location, s, depends on the exit pressure. Various possible situations are shown in Figure 3.7.9. It is clear that if the exit pressure is equal to or below p_f, the flow within the nozzle is fully supersonic. This is

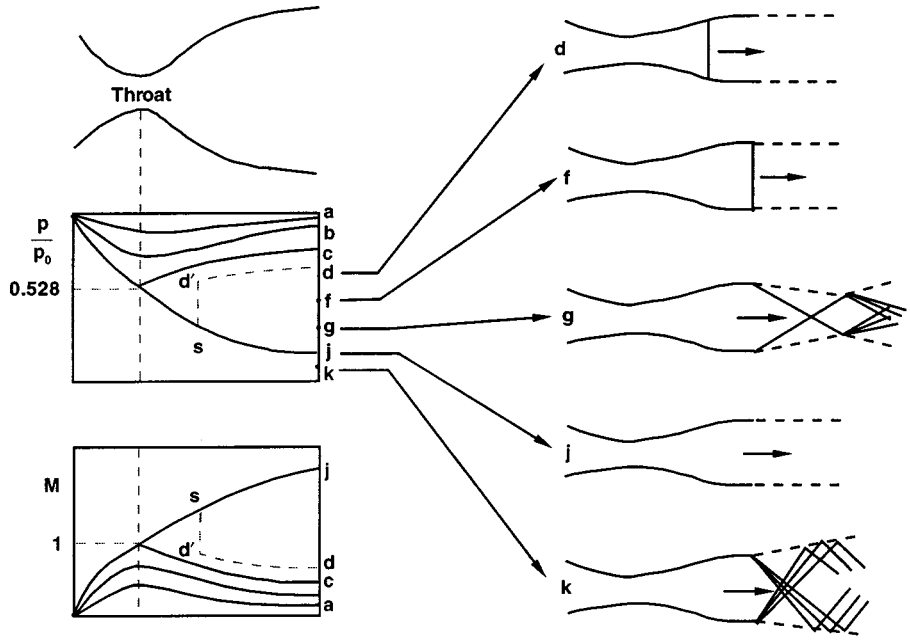


FIGURE 3.7.9 Effect of exit pressure on flow through a nozzle.

the principle used in designing supersonic wind tunnels by operating from a high-pressure reservoir or into a vacuum receiver, or both.

Diffuser

If a nozzle discharges directly into the receiver, the minimum pressure ratio for full supersonic flow in the test section is

$$\left(\frac{p_0}{p_E}\right)_{\min} = \frac{p_0}{p_f}$$

where p_f is the value of p_E at which the normal shock stands right at the nozzle exit. However, by adding an additional diverging section, known as a diffuser, downstream of the test section as shown in Figure 3.7.10 it is possible to operate the tunnel at a lower pressure ratio than p_0/p_f. This happens because the diffuser can now decelerate the subsonic flow downstream of the shock isentropically to a stagnation pressure p_0'. The pressure ratio required then is the ratio of stagnation pressures across a normal shock wave at the test section Mach number. In practice, the diffuser gives lower than expected recovery as a result of viscous losses caused by the interaction of shock wave and the boundary layer which are neglected here.

The operation of supersonic wind tunnels can be made even more efficient; i.e., they can be operated at even lower pressure ratios than p_0/p_0', by using the approach shown in Figure 3.7.7 where the diffuser has a second throat. It can slow down the flow to subsonic Mach numbers isentropically and, ideally, can provide complete recovery, giving $p_0' = p_0$. However, due to other considerations, such as the starting process of the wind tunnel and viscous effects, it is not realized in real life.

Two-Dimensional Supersonic Flow

When supersonic flow goes over a wedge or an expansion corner, it goes through an oblique shock or expansion waves, respectively, to adjust to the change in surface geometry. Figure 3.7.11 shows the two

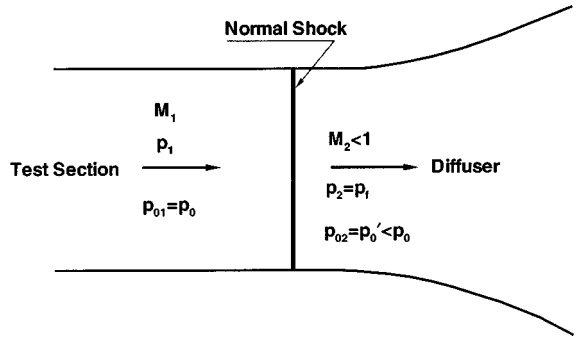


FIGURE 3.7.10 Normal shock diffuser.

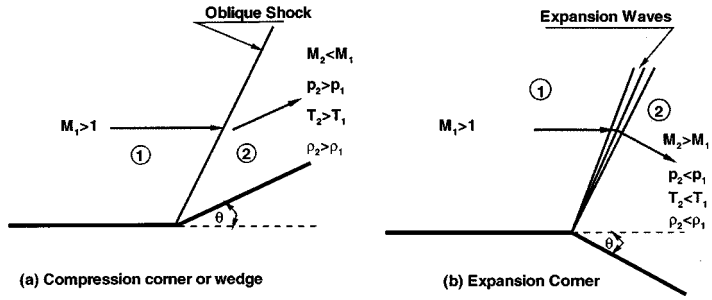


FIGURE 3.7.11 Supersonic flow over a corner.

flow situations. In Figure 3.7.11(a) an oblique shock abruptly turns the flow parallel to the wedge surface. The Mach number behind the shock is less than ahead of it, whereas the pressure, temperature, and density increase. In the case of an expansion corner, oblique expansion waves smoothly turn the flow to become parallel to the surface downstream of the expansion corner. In this case, the Mach number increases, but the pressure, temperature, and density decrease as the flow goes through the expansion corner. Oblique shocks and expansion waves occur in two- and three-dimensional supersonic flows. In this section, we will restrict ourselves to steady, two-dimensional supersonic flows only.

Oblique Shock Waves

The oblique shock can be treated in the same way as the normal shock by accounting for the additional velocity component. If a uniform velocity v is superimposed on the flow field of the normal shock, the resultant velocity ahead of the shock can be adjusted to any flow direction by adjusting the magnitude and direction of v. If v is taken parallel to the shock wave, as shown in Figure 3.7.12, the resultant velocity ahead of the shock is $w_1 = \sqrt{u_1^2 + v_1^2}$ and its direction from the shock is given by $\beta = \tan^{-1}(u_1/v)$. On the downstream side of the shock, since u_2 is less than u_1, the flow always turns toward the shock. The magnitude of u_2 can be determined by the normal shock relations corresponding to velocity u_1 and the magnitude of v is such that the flow downstream of the shock turns parallel to the surface. Since imposition of a uniform velocity does not affect the pressure, temperature, etc., we can use normal shock relations with Mach number replaced in them to correspond to velocity u_1 or u_1/a_1, which is nothing but $M_1 \sin \beta$. Thus, oblique shock relations become

$$\frac{p_2}{p_1} = 1 + \frac{2\gamma}{\gamma + 1}\left(M_1^2 \sin^2 \beta - 1\right) \tag{3.7.28}$$

$$\frac{\rho_2}{\rho_1} = \frac{(\gamma+1)M_1^2 \sin^2\beta}{(\gamma-1)M_1^2 \sin^2\beta + 2} \tag{3.7.29}$$

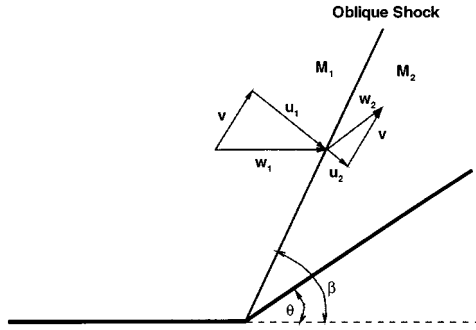


FIGURE 3.7.12 Oblique shock on a wedge.

$$\frac{T_2}{T_1} = \frac{a_2^2}{a_1^2} = \left[1 + \frac{2\gamma}{\gamma+1}\left(M_1^2 \sin^2\beta - 1\right)\right]\left[\frac{2 + (\gamma-1)M_1^2 \sin^2\beta}{(\gamma+1)M_1^2 \sin^2\beta}\right] \tag{3.7.30}$$

The Mach number M_2 (= w_2/a_2) can be obtained by using a Mach number corresponding to velocity u_2 (= $w_2 \sin(\beta - \theta)$) in the normal shock relation for the Mach number. In other words,

$$M_2^2 \sin^2(\beta-\theta) = \frac{1 + \dfrac{\gamma-1}{2}M_1^2 \sin^2\beta}{\gamma M_1^2 \sin^2\beta - \dfrac{\gamma-1}{2}} \tag{3.7.31}$$

To derive a relation between the wedge angle θ and the wave angle β, we have from Figure 3.7.12

$$\tan\beta = \frac{u_1}{v} \quad \text{and} \quad \tan(\beta-\theta) = \frac{u_2}{v}$$

so that

$$\frac{\tan(\beta-\theta)}{\tan\beta} = \frac{u_2}{u_1} = \frac{\rho_1}{\rho_2} = \frac{(\gamma-1)M_1^2 \sin^2\beta + 2}{(\gamma+1)M_1^2 \sin^2\beta}$$

This can be simplified to

$$\tan\theta = 2\cot\beta \frac{M_1^2 \sin^2\beta - 1}{M_1^2(\gamma + \cos 2\beta) + 2} \tag{3.7.32}$$

Dennard and Spencer (1964) have tabulated oblique shock properties as a function of M_1. Let us now make some observations from the preceding relations.

From the normal shock relations, $M_1 \sin\beta \geq 1$. This defines a minimum wave angle for a given Mach number. The maximum wave angle, of course, corresponds to the normal shock or $\beta = \pi/2$. Therefore, the wave angle β has the following range

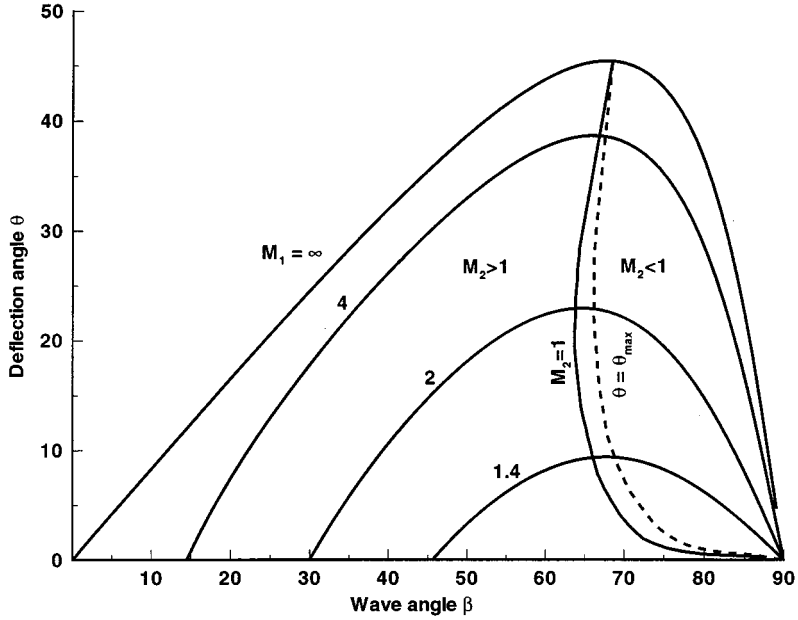


FIGURE 3.7.13 Oblique shock characteristics.

$$\sin^{-1}\frac{1}{M} \le \beta \le \frac{\pi}{2} \tag{3.7.33}$$

Equation 3.7.32 becomes zero at the two limits of β. Figure 3.7.13 shows a plot of θ against β for various values of M_1. For each value of M_1, there is a maximum value of θ. For $\theta < \theta_{max}$, there are two possible solutions having different values of β. The larger value of β gives the stronger shock in which the flow becomes subsonic. A locus of solutions for which $M_2 = 1$ is also shown in the figure. It is seen from the figure that with weak shock solution, the flow remains supersonic except for a small range of θ slightly smaller than θ_{max}.

Let us now consider the limiting case of θ going to zero for the weak shock solution. As θ decreases to zero, β decreases to the limiting value μ, given by

$$M_1^2 \sin^2 \mu - 1 = 0$$

$$\mu = \sin^{-1}\frac{1}{M_1} \tag{3.7.34}$$

For this angle, the oblique shock relations show no jump in flow quantities across the wave or, in other words, there is no disturbance generated in the flow. This angle μ is called the *Mach angle* and the lines at inclination μ are called *Mach lines*.

Thin-Airfoil Theory

For a small deflection angle Δθ, it can be shown that the change in pressure in a flow at Mach M_1 is given approximately by

$$\frac{\Delta p}{p_1} \approx \frac{\gamma M_1^2}{\sqrt{M_1^2 - 1}}\Delta\theta \tag{3.7.35}$$

This expression holds for both compression and expansion. If Δp is measured with respect to the freestream pressure, p_1, and all deflections to the freestream direction, we can write Equation (3.7.35) as

$$\frac{p - p_1}{p_1} = \frac{\gamma M_1^2}{\sqrt{M_1^2 - 1}}\theta \tag{3.7.36}$$

where θ is positive for a compression and negative for expansion. Let us define a pressure coefficient C_p, as

$$C_p = \frac{p - p_1}{q_1}$$

where q_1 is the dynamic pressure and is equal to $\gamma p_1 M_1^2/2$. Equation (3.7.36) then gives

$$C_p = \frac{2\theta}{\sqrt{M_1^2 - 1}} \tag{3.7.37}$$

Equation (3.7.37) states that the pressure coefficient is proportional to the local flow deflection. This relation can be used to develop supersonic thin-airfoil theory. As an example, for a flat plate at angle of attack α_0 (shown in Figure 3.7.14), the pressure coefficients on the upper and lower surfaces are

$$C_p = \mp \frac{2\alpha_0}{\sqrt{M_1^2 - 1}}$$

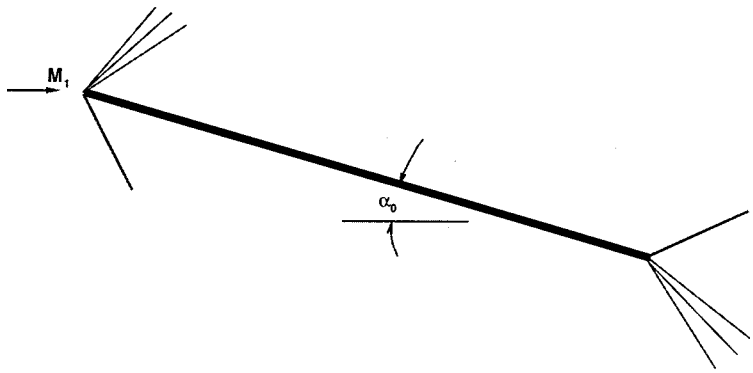


FIGURE 3.7.14 Lifting flat plate.

The lift and drag coefficients can be written as

$$C_L = \frac{(p_L - p_U)c\cos\alpha_0}{q_1 c} = \left(C_{p_L} - C_{p_U}\right)\cos\alpha_0$$

$$C_D = \frac{(p_L - p_U)c\sin\alpha_0}{q_1 c} = \left(C_{p_L} - C_{p_U}\right)\sin\alpha_0$$

where c is the chord length of the plate. Since α_0 is small, we can write

$$C_L = \frac{4\alpha_0}{\sqrt{M_1^2 - 1}}, \qquad C_D = \frac{4\alpha_0^2}{\sqrt{M_1^2 - 1}} \tag{3.7.38}$$

A similar type of expression can be obtained for an arbitrary thin airfoil that has thickness, camber, and angle of attack. Figure 3.7.15 shows such an airfoil. The pressure coefficients on the upper and lower surfaces can be written as

$$C_{p_U} = \frac{2}{\sqrt{M_1^2 - 1}} \frac{dy_U}{dx}, \qquad C_{p_L} = \frac{2}{\sqrt{M_1^2 - 1}} \left(-\frac{dy_L}{dx} \right) \tag{3.7.39}$$

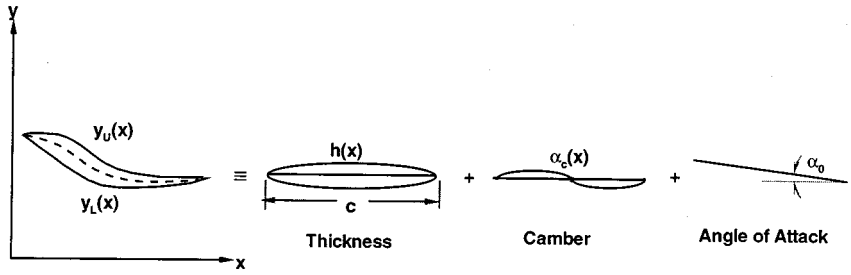


FIGURE 3.7.15 Arbitrary thin airfoil and its components.

For the thin airfoil, the profile may be resolved into three separate components as shown in Figure 3.7.15. The local slope of the airfoil can be obtained by superimposing the local slopes of the three components as

$$\frac{dy_U}{dx} = -\left(\alpha_0 + \alpha_c(x)\right) + \frac{dh}{dx} = -\alpha(x) + \frac{dh}{dx}$$

$$\frac{dy_L}{dx} = -\left(\alpha_0 + \alpha_c(x)\right) - \frac{dh}{dx} = -\alpha(x) - \frac{dh}{dx} \tag{3.7.40}$$

where $\alpha = \alpha_0 + \alpha_c(x)$ is the local total angle of attack of the camber line. The lift and drag for the thin airfoil are given by

$$L = q_1 \int_0^c \left(C_{p_L} - C_{p_U} \right) dx$$

$$D = q_1 \int_0^c \left[C_{p_L} \left(-\frac{dy_L}{dx} \right) + C_{p_U} \left(\frac{dy_U}{dx} \right) \right] dx$$

Let us define an average value of $\alpha(x)$ as

$$\overline{\alpha} = \frac{1}{c} \int_0^c \alpha(x)\, dx$$

Using Equation (3.7.40) and the fact that $\overline{\alpha}_0 = \alpha$ and $\overline{\alpha}_c = 0$ by definition, the lift and drag coefficients for the thin airfoil can be written as

$$C_L = \frac{4\alpha_0}{\sqrt{M_1^2 - 1}}$$

$$C_D = \frac{4}{\sqrt{M_1^2 - 1}}\left[\overline{\left(\frac{dh}{dx}\right)^2} + \overline{\alpha_C^2(x)} + \alpha_0^2\right] \tag{3.7.41}$$

Equations (3.7.41) show that the lift coefficient depends only on the mean angle of attack whereas the drag coefficient is a linear combination of the drag due to thickness, drag due to camber, and drag due to lift (or mean angle of attack).

References

Anderson, J.D. 1982. *Modern Compressible Flow,* McGraw-Hill, New York.

Dennard, J.S. and Spencer, P.B. 1964. Ideal-Gas Tables for Oblique-Shock Flow Parameters in Air at Mach Numbers from 1.05 to 12.0. NASA TN D-2221.

Liepmann, H.W. and Roshko, A. 1966. *Elements of Gas Dynamics,* John Wiley & Sons, New York.

Further Information

As mentioned in the beginning, this section discussed only one- or two-dimensional steady, inviscid compressible flows under perfect gas assumption. Even this discussion was quite brief because of space limitations. For more details on the subject as well as for compressible unsteady viscous flows, readers are referred to Anderson (1982) and Liepmann and Roshko (1966).

3.8 Multiphase Flow

John C. Chen

Introduction

Classic study of fluid mechanics concentrates on the flow of a single homogeneous phase, e.g., water, air, steam. However, many industrially important processes involve simultaneous flow of multiple phases, e.g., gas bubbles in oil, wet steam, dispersed particles in gas or liquid. Examples include vapor–liquid flow in refrigeration systems, steam–water flows in boilers and condensers, vapor–liquid flows in distillation columns, and pneumatic transport of solid particulates. In spite of their importance, multiphase flows are often neglected in standard textbooks. Fundamental understanding and engineering design procedures for multiphase flows are not nearly so well developed as those for single-phase flows. An added complexity is the need to predict the relative concentrations of the different phases in the multiphase flows, a need that doesn't exist for single-phase flows.

Inadequate understanding not withstanding, a significant amount of data have been collected and combinations of theoretical models and empirical correlations are used in engineering calculations. This knowledge base is briefly summarized in this section and references are provided for additional information. While discussions are provided of solid–gas flows and solid–liquid flows, primary emphasis is placed on multiphase flow of gas–liquids since this is the most often encountered class of multiphase flows in industrial applications.

A multiphase flow occurs whenever two or more of the following phases occur simultaneously: gas/vapor, solids, single-liquid phase, multiple (immiscible) liquid phases. Every possible combination has been encountered in some industrial process, the most common being the simultaneous flow of vapor/gas and liquid (as encountered in boilers and condensers). All multiphase flow problems have features which are characteristically different from those found in single-phase problems. First, the relative concentration of different phases is usually a dependent parameter of great importance in multiphase flows, while it is a parameter of no consequence in single-phase flows. Second, the spatial distribution of the various phases in the flow channel strongly affects the flow behavior, again a parameter that is of no concern in single-phase flows. Finally, since the density of various phases can differ by orders of magnitude, the influence of gravitational body force on multiphase flows is of much greater importance than in the case of single-phase flows. In any given flow situation, the possibility exists for the various phases to assume different velocities, leading to the phenomena of slip between phases and consequent interfacial momentum transfer. Of course, the complexity of laminar/turbulent characteristics occurs in multiphase flows as in single-phase flows, with the added complexity of interactions between phases altering the laminar/turbulent flow structures. These complexities increase exponentially with the number of phases encountered in the multiphase problem. Fortunately, a large number of applications occur with just two phase flows, or can be treated as pseudo-two-phase flows.

Two types of analysis are used to deal with two-phase flows. The simpler approach utilizes homogeneous models which assume that the separate phases flow with the same identical local velocity at all points in the fluid. The second approach recognizes the possibility that the two phases can flow at different velocities throughout the fluid, thereby requiring separate conservation equations for mass and momentum for each phase. Brief descriptions of both classes of models are given below.

Fundamentals

Consider n phases in concurrent flow through a duct with cross-sectional area A_c. Fundamental quantities that characterize this flow are

$\dot{m}_i$ = mass flow rate of *i*th phase

u_i = velocity of *i*th phase

α_i = volume fraction of *i*th phase in channel

Basic relationships between these and related parameters are

$$G_i = \text{mass flux of } i\text{th phase} = \frac{\dot{m}_i}{A_c} \tag{3.8.1}$$

$$v_i = \text{superficial velocity of } i\text{th phase} = \frac{G_i}{\rho_i} \tag{3.8.2}$$

$$u_i = \text{actual velocity of } i\text{th phase} = \frac{v_i}{\alpha_i} \tag{3.8.3}$$

$$x_i = \text{flow quality of } i\text{th phase} = \frac{\dot{m}_i}{\sum_i^n m_i} = \frac{G_i}{\sum_{i=1}^n G_i} \tag{3.8.4}$$

$$\alpha_i = \text{volume fraction of } i\text{th phase} = \frac{\left(\frac{x_i}{\rho_i u_i}\right)}{\sum_{i=1}^n \left(\frac{x_i}{\rho_i u_i}\right)} \tag{3.8.5}$$

In most engineering calculations, the above parameters are defined as average quantities across the entire flow area, A_c. It should be noted, however, that details of the multiphase flow could involve local variations across the flow area. In the latter situation, G_i, v_i, and α_i are often defined on a local basis, varying with transverse position across the flow area.

Pressure drop along the flow channel is associated with gravitational body force, acceleration forces, and frictional shear at the channel wall. The total pressure gradient along the flow axis can be represented as

$$\frac{dP}{dz} = \left(\frac{dP}{dz}\right)_g + \left(\frac{dP}{dz}\right)_a + \left(\frac{dP}{dz}\right)_f \tag{3.8.6}$$

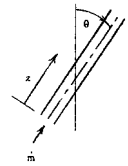
where

$$\left(\frac{dP}{dz}\right)_g = -g\cos\theta \cdot \sum_{i=1}^{n} \alpha_i \rho_i \tag{3.8.7}$$

θ = angle of channel from vertical

and

$$\left(\frac{dP}{dz}\right)_a = -\sum_{i=1}^{n} G_i \frac{du_i}{dz} \tag{3.8.8}$$



$$\left(\frac{dP}{dz}\right)_f = -\frac{\rho u^2}{2D} f \tag{3.8.9}$$

ρ = density of multiphase mixture

$$= \sum_{i=1}^{n} \rho_i \alpha_i \tag{3.8.10}$$

u = an average mixture velocity

$$= \frac{1}{\rho} \sum_{i=1}^{n} G_i \tag{3.8.11}$$

f = equivalent Darcy friction factor for the multiphase flow

In applications, the usual requirement is to determine pressure gradient (dP/dz) and the volume fractions (α_i). The latter quantities are of particular importance since the volume fraction of individual phases affects all three components of the pressure gradient, as indicated in Equations (3.8.7) to (3.8.11). Correlations of various types have been developed for prediction of the volume fractions, all but the simplest of which utilize empirical parameters and functions.

The simplest flow model is known as the homogeneous equilibrium model (HEM), wherein all phases are assumed to be in neutral equilibrium. One consequence of this assumption is that individual phase velocities are equal for all phases everywhere in the flow system:

$$u_i = u \quad \text{for all} \quad i \tag{3.8.12}$$

This assumption permits direct calculation of the volume fractions from known mass qualities:

$$\alpha_i = \frac{x_i}{\rho_i \sum_{i=1}^{n} \left(\frac{x_i}{\rho_i}\right)} \tag{3.8.13}$$

The uniform velocity for all phases is the same as mixture velocity:

$$u = \frac{1}{\rho}\sum_{i=1}^{n} G_i \tag{3.8.14}$$

where

$$\frac{1}{\rho} = \sum_{i=1}^{n}\left(\frac{x_i}{\rho_i}\right) \tag{3.8.15}$$

This homogeneous model permits direct evaluation of all three components of axial pressure gradient, if flow qualities (x_i) are known:

$$\left(\frac{dP}{dz}\right)_g = -\frac{g\cos\theta}{\sum_{i=1}^{n}\left(\frac{x_i}{\rho_i}\right)} \tag{3.8.16}$$

$$\left(\frac{dP}{dz}\right)_a = -\left(\sum_{i=1}^{n} G_i\right)\cdot\frac{du}{dz} \tag{3.8.17}$$

$$\left(\frac{dP}{dz}\right)_f = -\frac{\rho u^2}{2D_f}\cdot f \tag{3.8.18}$$

where u and ρ are given by Equations (3.8.14) and (3.8.15).

Predicting the coefficient of friction (f to clear) remains a problem, even in the homogeneous model. For cases of fully turbulent flows, experience has shown that a value of 0.02 may be used as a first-order approximation for (f to clear). More-accurate estimates require empirical correlations, specific to particular classes of multiphase flows and subcategories of flow regimes.

The following parts of this section consider the more common situations of two-phase flows and describe improved design methodologies specific to individual situations.

Gas–Liquid Two-Phase Flow

The most common case of multiphase flow is two-phase flow of gas and liquid, as encountered in steam generators and refrigeration systems. A great deal has been learned about such flows, including delineation of flow patterns in different flow regimes, methods for estimating volume fractions (gas void fractions), and two-phase pressure drops.

Flow Regimes

A special feature of multiphase flows is their ability to assume different spatial distributions of the phases. These different flow patterns have been classified in flow regimes, which are themselves altered by the direction of flow relative to gravitational acceleration. Figures 3.8.1 and 3.8.2 (Delhaye, 1981) show the flow patterns commonly observed for co-current flow of gas and liquid in vertical and horizontal channels, respectively. For a constant liquid flow rate, the gas phase tends to be distributed as small bubbles at low gas flow rates. Increasing gas flow rate causes agglomeration of bubbles into larger slugs and plugs. Further increasing gas flow rate causes separation of the phases into annular patterns wherein liquid concentrates at the channel wall and gas flows in the central core for vertical ducts. For horizontal ducts, gravitational force tends to drain the liquid annulus toward the bottom of the channel, resulting

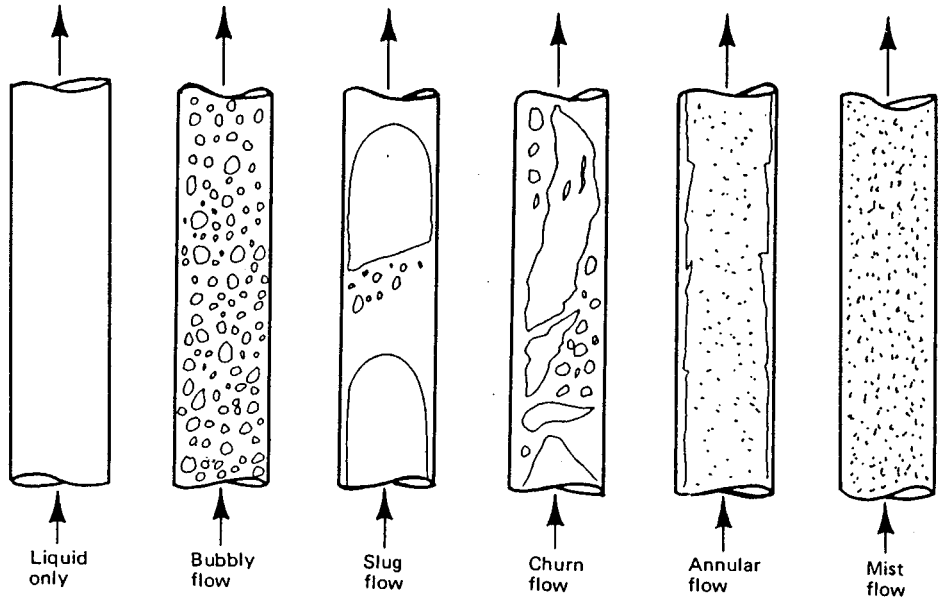


FIGURE 3.8.1 Flow patterns in gas–liquid vertical flow. (From Lahey, R.T., Jr. and Moody, F.I. 1977. *The Thermal Hydraulics of a Boiling Water Nuclear Reactor,* The American Nuclear Society, LaGrange, IL. With permission.)

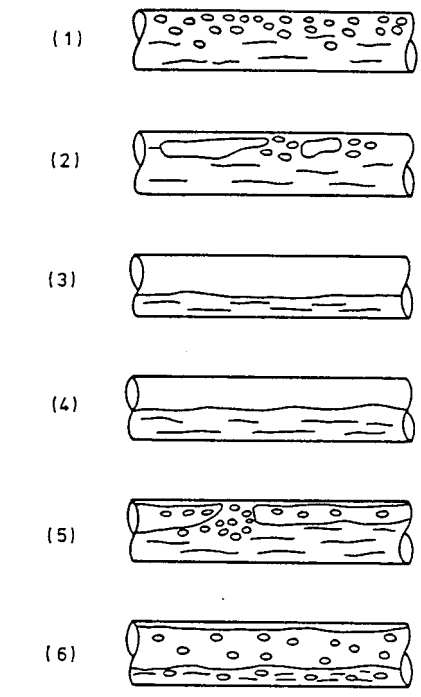


FIGURE 3.8.2 Flow patterns in gas–liquid horizontal flow.

in stratified and stratified wavy flows. This downward segregation of the liquid phase can be overcome by kinetic forces at high flow rates, causing stratified flows to revert to annular flows. At high gas flow rates, more of the liquid tends to be entrained as dispersed drops; in the limit one obtains completely dispersed mist flow.

Flow pattern maps are utilized to predict flow regimes for specific applications. The first generally successful flow map was that of Baker (1954) for horizontal flow, reproduced here in Figure 3.8.3. For vertical flows, the map of Hewitt and Roberts (1969), duplicated in Figure 3.8.4, provides a simple method for determining flow regimes. Parameters used for the axial coordinates of these flow maps are defined as follows:

$$\lambda = \left(\frac{\rho_g \rho_\ell}{\rho_a \rho_w} \right)^{1/2} \tag{3.8.19}$$

$$\psi = \left(\frac{\sigma_w}{\sigma} \right) \left[\left(\frac{\mu_\ell}{\mu_w} \right) \left(\frac{\rho_w}{\rho_\ell} \right)^2 \right]^{1/3} \tag{3.8.20}$$

$$j = \text{volumetric flux}, \frac{G}{\rho} \tag{3.8.21}$$

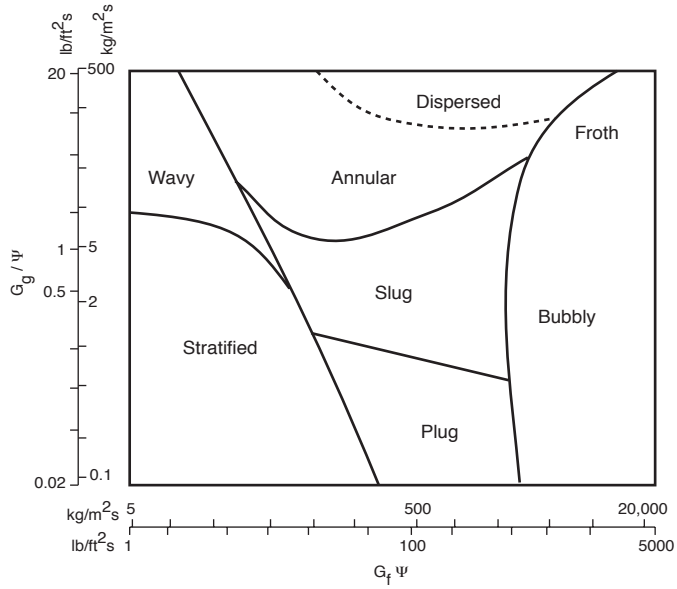


FIGURE 3.8.3 Flow pattern map for horizontal flow (Baker, 1954).

Void Fractions

In applications of gas–liquid flows, the volume fraction of gas (α_g) is commonly called "void fraction" and is of particular interest. The simplest method to estimate void fraction is by the HEM. From Equation (3.8.13), the void fraction can be estimated as

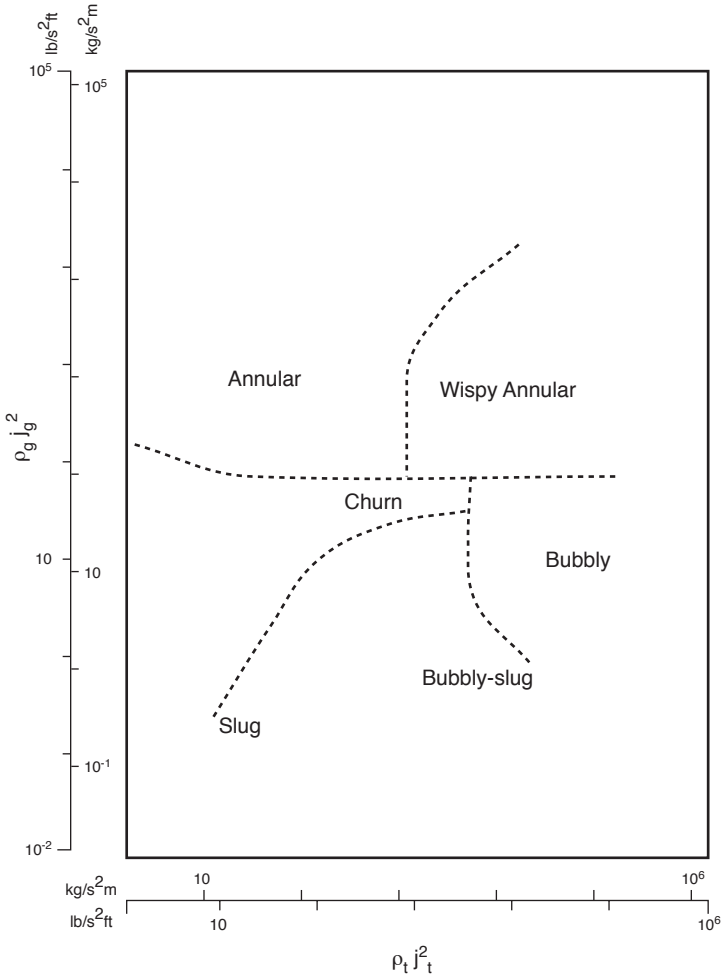


FIGURE 3.8.4 Flow pattern map for vertical flow (Hewitt and Roberts, 1969).

$$\alpha_g = \frac{x_g}{x_g + \left(1 - x_g\right)\dfrac{\rho_g}{\rho_\ell}} \tag{3.8.22}$$

where α_g, x_g, ρ_g, ρ_ℓ are cross-sectional averaged quantities.

In most instances, the homogenous model tends to overestimate the void fraction. Improved estimates are obtained by using separated-phase models which account for the possibility of slip between gas and liquid velocities. A classic separated-phase model is that of Lockhart and Martinelli (1949). The top portion of Figure 3.8.5 reproduces the Lockhart–Martinelli correlation for void fraction (shown as α) as a function of the parameter X which is defined as

$$X = \left[\left(\frac{dP}{dz}\right)_{f\ell} \div \left(\frac{dP}{dz}\right)_{fg}\right]^{1/2} \tag{3.8.23}$$

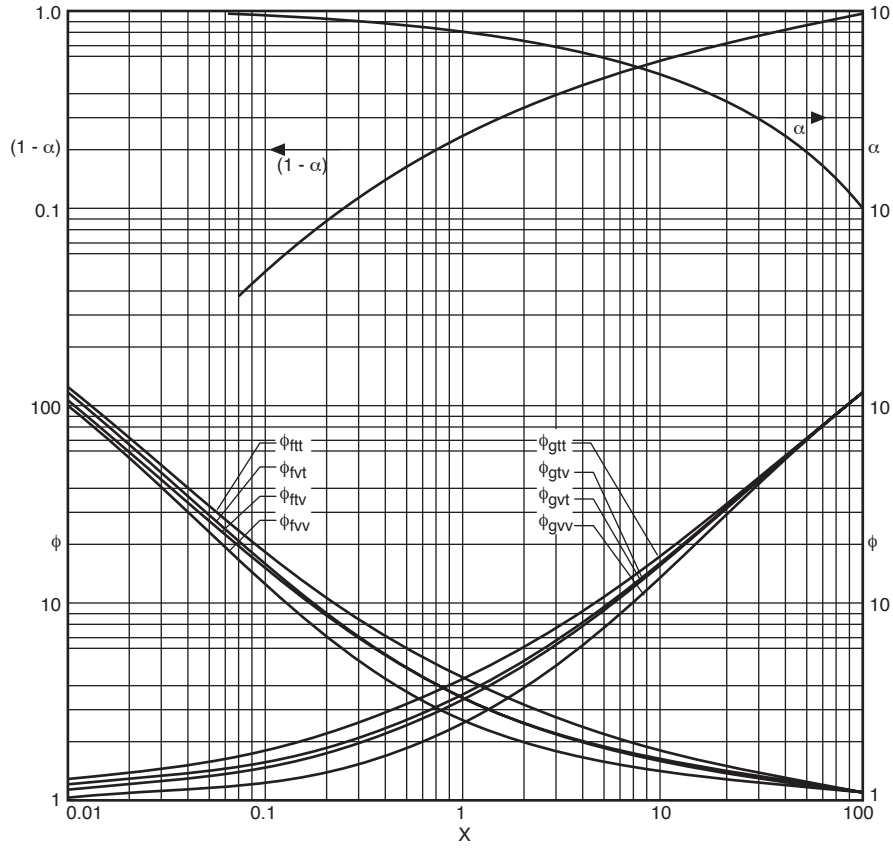


FIGURE 3.8.5 Correlations for void fraction and frictional pressure drop (Lockhart and Martinelli, 1949)

where

$\left(\frac{dP}{dz}\right)_{f\ell}$ = frictional pressure gradient of liquid phase flowing alone in channel

$\left(\frac{dP}{dz}\right)_{fg}$ = frictional pressure gradient of gas phase flowing alone in channel

Often, flow rates are sufficiently high such that each phase if flowing alone in the channel would be turbulent. In this situation the parameter X can be shown to be

$$X_{tt} = \left(\frac{1 - x_g}{x_g}\right)^{0.9} \left(\frac{\rho_g}{\rho_\ell}\right)^{0.5} \left(\frac{\mu_\ell}{\mu_g}\right)^{0.1} \tag{3.8.24}$$

Another type of separated-phase model is the drift-flux formulation of Wallis (1969). This approach focuses attention on relative slip between phases and results in slightly different expressions depending on the flow regime. For co-current upflow in two of the more common regimes, the drift-flux model gives the following relationships between void fraction and flow quality:

Bubbly flow or churn-turbulent flow:

$$\alpha_g = \frac{x_g}{\left(\frac{u_o \rho_g}{G}\right) + C_o\left[x_g + \left(1 - x_g\right)\frac{\rho_g}{\rho_\ell}\right]} \tag{3.8.25}$$

Dispersed drop (mist) flow:

$$x_g = \frac{1 - \left(1 - \alpha_g\right)\left(\frac{u_o \rho_\ell}{G}\alpha_g^2 + 1\right)}{1 - \left(1 - \alpha_g\right)\left(1 - \frac{\rho_\ell}{\rho_g}\right)} \tag{3.8.26}$$

where u_o = terminal rise velocity of bubble, in bubbly flow, or terminal fall velocity of drop in churn-turbulent flow

C_o = an empirical distribution coefficient $\simeq 1.2$

Pressure Drop

Equations 3.8.16 through 3.8.18 permit calculation of two-phase pressure drop by the homogeneous model, if the friction coefficient (f) is known. One useful method for estimating (f) is to treat the entire two-phase flow as if it were all liquid, except flowing at the two-phase mixture velocity. By this approach the frictional component of the two-phase pressure drop becomes

$$\left(\frac{dP}{dz}\right)_f = \left[1 + x_g\left(\frac{\rho_\ell}{\rho_g} - 1\right)\right] \cdot \left(\frac{dP}{dz}\right)_{f\ell G} \tag{3.8.27}$$

where $(dP/dz)_{f\ell G}$ = frictional pressure gradient if entire flow (of total mass flux G) flowed as liquid in the channel.

The equivalent frictional pressure drop for the entire flow as liquid, $(dP/dz)_{f\ell g}$, can be calculated by standard procedures for single-phase flow. In using Equations (3.8.16) through (3.8.18), the void fraction would be calculated with the equivalent homogeneous expression Equation (3.8.13).

A more accurate method to calculate two-phase pressure drop is by the separated-phases model of Lockhart and Martinelli (1949). The bottom half of Figure 3.8.5 shows empirical curves for the Lockhart–Martinelli frictional multiplier, ϕ:

$$\phi_i = \left[\left(\frac{dP}{dz}\right)_f \div \left(\frac{dP}{dz}\right)_{fi}\right]^{1/2} \tag{3.8.28}$$

where (i) denotes either the fluid liquid phase (f) or gas phase (g). The single-phase frictional gradient is based on the ith phase flowing alone in the channel, in either viscous laminar (v) or turbulent (t) modes. The most common case is where each phase flowing alone would be turbulent, whence one could use Figure 3.8.5 to obtain

$$\left(\frac{dP}{dz}\right)_f = \text{frictional pressure gradient for two-phase flow}$$
$$= \phi_{gtt}^2 \cdot \left(\frac{dP}{dz}\right)_{fg} \tag{3.8.29}$$

where $(dP/dz)_{fg}$ is calculated for gas phase flowing alone and $X = X_{tt}$ as given by Equation (3.8.24).

The correlation of Lockhart–Martinelli has been found to be adequate for two-phase flows at low-to-moderate pressures, i.e., with reduced pressures less than 0.3. For applications at higher pressures, the revised models of Martinelli and Nelson (1948) and Thom (1964) are recommended.

Gas–Solid, Liquid–Solid Two-Phase Flows

Two-phase flows can occur with solid particles in gas or liquid. Such flows are found in handling of granular materials and heterogeneous reaction processing. Concurrent flow of solid particulates with a fluid phase can occur with various flow patterns, as summarized below.

Flow Regimes

Consider vertical upflow of a fluid (gas or liquid) with solid particles. Figure 3.8.6 illustrates the major flow regimes that have been identified for such two-phase flows. At low flow rates, the fluid phase percolates between stationary particles; this is termed flow through a fixed bed. At some higher velocity a point is reached when the particles are all suspended by the upward flowing fluid, the drag force between particles and fluid counterbalancing the gravitational force on the particles. This is the point of minimum fluidization, marking the transition from fixed to fluidized beds. Increase of fluid flow rate beyond minimum fluidization causes instabilities in the two-phase mixture, and macroscopic bubbles or channels of fluid are observed in the case of gaseous fluids. In the case of liquid fluids, the two-phase mixture tends to expand, often without discrete bubbles or channels. Further increase of fluid velocity causes transition to turbulent fluidization wherein discrete regions of separated phases (fluid slugs or channels and disperse suspensions of particles) can coexist. Depending on specific operating conditions (e.g., superficial fluid velocity, particle size, particle density, etc.), net transport of solid particles with the flowing fluid can occur at any velocity equal to or greater than that associated with slug flow and turbulent flow. Further increases in fluid velocity increase the net transport of solid particles. This can occur with large-scale clusters of solid particles (as exemplified by the fast fluidization regime) or with dilute dispersions of solid particles (as often utilized in pneumatic conveying). For engineering application of fluid–solid two-phase flows, the important thresholds between flow regimes are marked by the fluid velocity for minimum fluidization, terminal slip, and saltation threshold.

Minimum Fluidization

The transition from flow through packed beds to the fluidization regime is marked by the minimum fluidization velocity of the fluid. On a plot pressure drop vs. superficial fluid velocity, the point of minimum fluidization is marked by a transition from a linearly increasing pressure drop to a relatively constant pressure drop as shown in Figure 3.8.7 for typical data, for two-phase flow of gas with sand particles of 280 μm mean diameter (Chen, 1996). The threshold fluid velocity at minimum fluidization is traditionally derived from the Carman–Kozeny equation,

$$U_{mf} = \frac{\left(\rho_s - \rho_f\right)\left(\phi\, dp\right)^2 g}{150\mu_f} \cdot \frac{\alpha_{mf}^2}{\left(1 - \alpha_{mf}\right)} \tag{3.8.30}$$

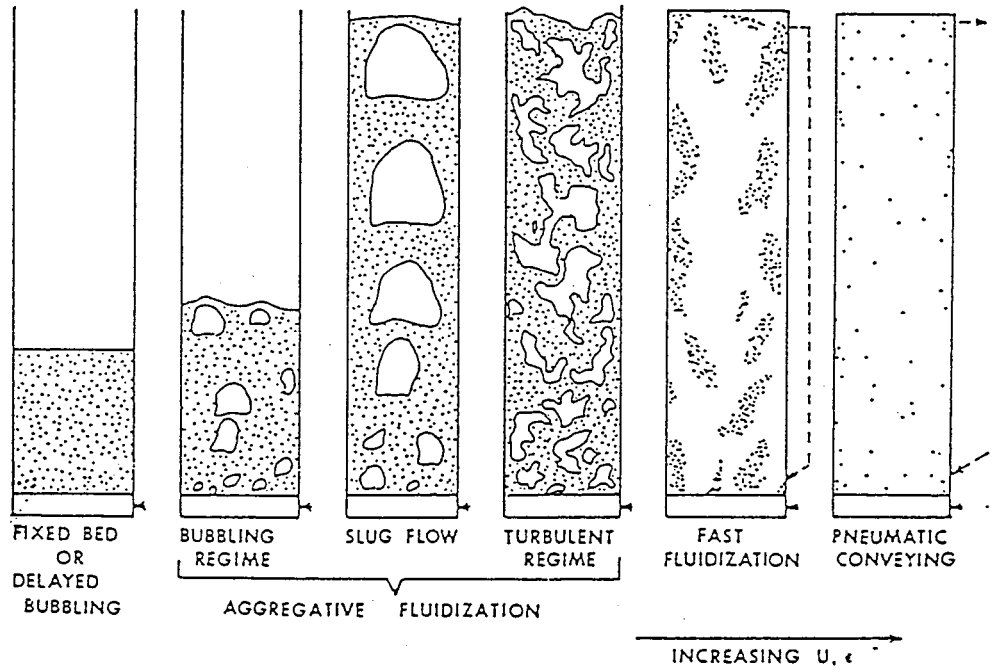


FIGURE 3.8.6 Flow patterns for vertical upflow of solid particles and gas or liquid. (From Chen, J.C. 1994. *Proc. Xth Int. Heat Transfer Conf.,* Brighton, U.K., 1:369–386. With permission.)

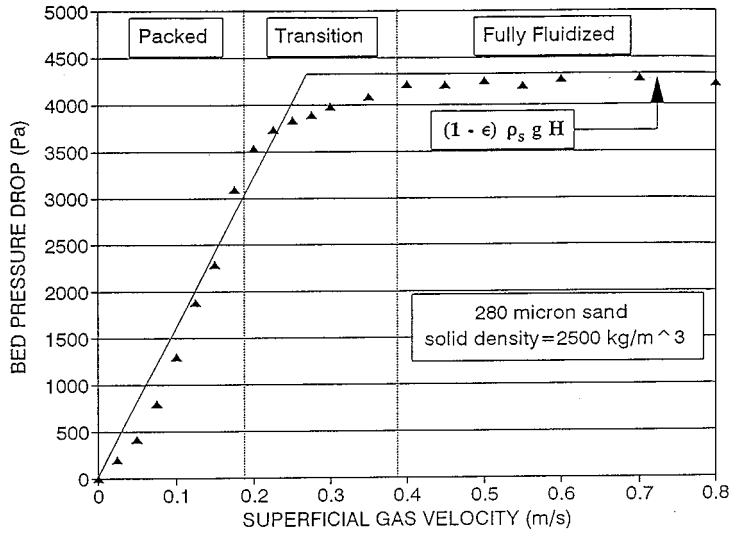


FIGURE 3.8.7 Transition at minimum fluidization. (From Chen, J.C. 1996. In *Annual Review of Heat Transfer,* Vol. VII, Begal House, Washington, D.C. With permission.)

where ϕ = sphericity of particles (unity for spherical particles)

α_{mf} = volumetric fraction of fluid at minimum fluidization

Small, light particles have minimum fluidization voidage (α_{mf}) of the order 0.6, while larger particles such as sand have values closer to 0.4.

An alternative correlation for estimating the point of minimum fluidization is that of Wen and Yu (1966):

$$\frac{U_{mf} d_p \rho_f}{\mu_f} = \left(33.7 + 0.041 Ga\right)^{0.5} - 33.7 \tag{3.8.31}$$

where $Ga = \rho_f d_p^3 (\rho_s - \rho_f) g / \mu_f^2$.

When the fluid velocity exceeds U_{mf}, the two-phase mixture exists in the fluidized state in which the pressure gradient is essentially balanced by the gravitational force on the two-phase mixture:

$$\frac{dP}{dz} = g\left[\alpha_s \rho_s + \alpha_f \rho_f\right] \tag{3.8.32}$$

This fluidized state exists until the fluid velocity reaches a significant fraction of the terminal slip velocity, beyond which significant entrainment and transport of the solid particles occur.

Terminal Slip Velocity

For an isolated single particle the maximum velocity relative to an upflowing fluid is the terminal slip velocity. At this condition, the interfacial drag of the fluid on the particle exactly balances the gravitational body force on the particle:

$$U_t = \left(U_f - U_s\right)_t = \left[\frac{4 d_p \left(\rho_s - \rho_f\right)}{3 \rho_f} \cdot \frac{1}{C_D}\right]^{1/2} \tag{3.8.33}$$

where C_D = coefficient of drag on the particle.

The coefficient of drag on the particle (C_D) depends on the particle Reynolds number:

$$\mathrm{Re}_p = \frac{\rho_f d_p \left(U_f - U_s\right)}{\mu_f} \tag{3.8.34}$$

The following expressions may be used to estimate C_D as appropriate:

$$\begin{aligned} C_D &= \frac{32}{\mathrm{Re}_p}, && \mathrm{Re}_p \le 1 \\ C_D &= \frac{18.5}{\mathrm{Re}_p^{0.67}}, && 1 \le \mathrm{Re}_p \le 10^3 \end{aligned} \tag{3.8.35}$$

Pneumatic Conveying

A desirable mode of pneumatic conveying is two-phase flow with solid particles dispersed in the concurrent flowing fluid. Such dispersed flows can be obtained if the fluid velocity is sufficiently high. For both horizontal and vertical flows, there are minimum fluid velocities below which saltation of the solid particles due to gravitational force occurs, leading to settling of the solid particles in horizontal channels and choking of the particles in vertical channels. Figures 3.8.8 and 3.8.9 for Zenz and Othmer (1960) show these different regimes of pneumatic conveying for horizontal and vertical transport, respectively. Figure 3.8.8 shows that for a given rate of solids flow (W) there is a minimum superficial fluid velocity below which solid particles tend to settle into a dense layer at the bottom of the horizontal channels. Above this saltation threshold, fully dispersed two-phase flow is obtained. In the case of

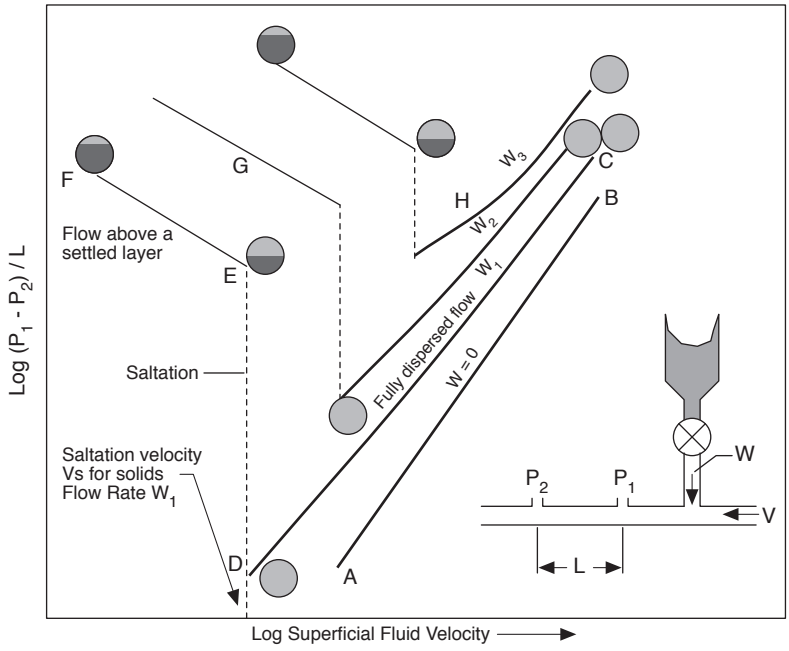


FIGURE 3.8.8 Flow characteristics in horizontal pneumatic conveying.

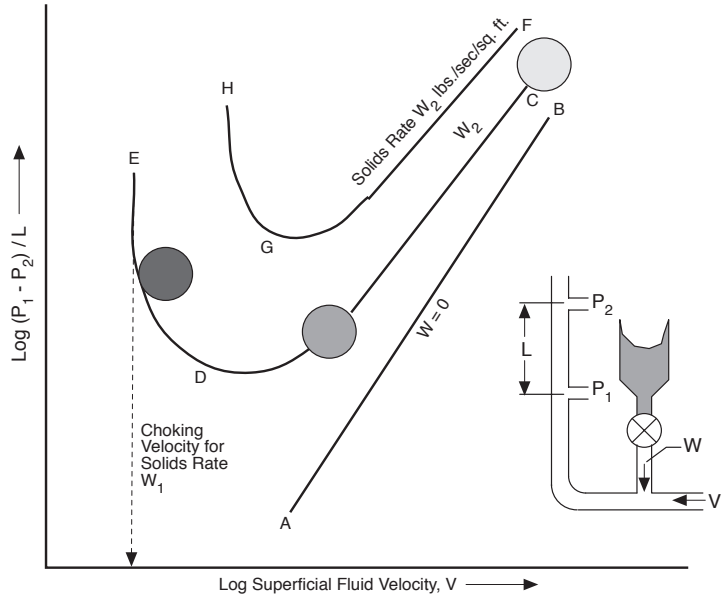


FIGURE 3.8.9 Flow characteristics in vertical pneumatic conveying.

vertical transport illustrated in Figure 3.8.9, there is a minimum fluid velocity below which solid particles tend to detrain from the two-phase suspension. This choking limit varies not only with particle properties but also with the actual rate of particle flow. Well-designed transport systems must operate with superficial fluid velocities greater than these limiting saltation and choking velocities.

Zenz and Othmer (1960) recommend the empirical correlations represented in Figure 3.8.10 estimating limiting superficial fluid velocities at incipient saltation or choking, for liquid or gas transport of uniformly sized particles. Note that these correlations are applicable for either horizontal or vertical concurrent flow. Figure 3.8.10 is duplicated from the original source and is based on parameters in engineering units, as noted in the figure. To operate successfully in dispersed pneumatic conveying of solid particles, the superficial fluid velocity must exceed that determined from the empirical correlations of Figure 3.8.10.

Nomenclature

A_c	cross-sectional flow area of channel
C_o	Wallis' distribution coefficient
d_p	diameter of solid particles
f_D	Darcy friction factor
G	mass flow flux, kg/m^2 · sec
j	volumetric flow flux, m/sec
$\dot{m}$	mass flow rate, kg/sec
P	pressure, N/m^2
u	velocity in axial flow direction, m/sec
v	superficial velocity in axial flow direction, m/sec
x	mass flow quality
z	axial coordinate

Greek Letters

α	volume fraction
λ	parameter in Baker flow map
ϕ	sphericity of solid particles
ϕ_i	frictional multiphase for pressure drag, Equation (3.8.28)
ψ	parameter in Baker flow map
σ	surface tension
θ	angle from vertical

Subscripts

a	air
f	fluid phase
g	gas phase
l	liquid phase
mf	minimum fluidization
p	particle
s	solid phase
t	terminal slip
w	water

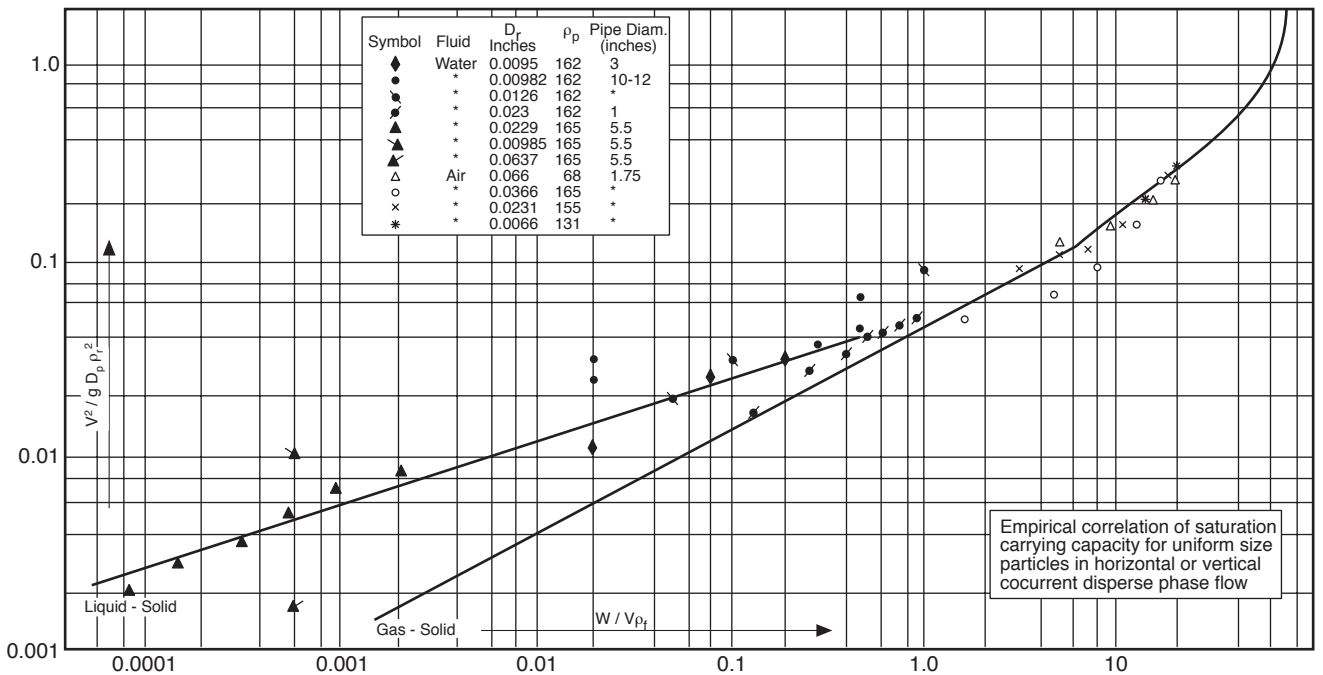


FIGURE 3.8.10 Correlations for limiting velocities in pneumatic conveying. (

References

Baker, O. 1954. Design of pipelines for simultaneous flow of oil and gas, *Oil Gas J.*.

Chen, J.C. 1994. Two-phase flow with and without phase changes: suspension flows. Keynote lecture, *Proc. Xth Int. Heat Transfer Conf.,* Brighton, U.K., 1:369–386.

Chen, J.C. 1996. Heat transfer to immersed surfaces in bubbling fluidized beds, in *Annual Review of Heat Transfer*, Vol. VII, Bengel House, Washington, D.C.

Collier, J.G. 1972. *Convective Boiling and Condensation,* McGraw-Hill, London.

Delhaye, J.M. 1981. Two-phase flow patterns, in *Two-Phase Flow and Heat Transfer,* A.E. Bergles, J.G. Collier, J.M. Delhaye, G.F. Newitt, and F. Mayinger, Eds., Hemisphere Publishing, McGraw-Hill, New York.

Hewitt, G.F. and Roberts, D.N. 1969. Studies of Two-Phase Flow Patterns by Simultaneous X-Ray and Flash Photography, Report AERE-M 2159.

Lahey, R.T., Jr. and Moody, F.I. 1977. *The Thermal Hydraulics of a Boiling Water Nuclear Reactor,* The American Nuclear Society, La Grange, IL.

Lockhart, R.W. and Martinelli, R.C. 1949. Proposed correlation of data for isothermal two-phase two-component flow in pipes, *Chem. Eng. Progr.,* 45:39.

Martinelli, R.C. and Nelson, D.B. 1984. Prediction of pressure drop during forced-circulation boiling of water, *Trans. ASME,* 70:695–702.

Thom, J.R.S. 1964. Prediction of pressure drop during forced circulation boiling of water, *Int. J. Heat Mass Transfer,* 7:709–724.

Wallis, G.B. 1969. *One-Dimensional Two-Phase Flow,* McGraw-Hill, New York.

Wen, C.Y. and Yu, Y.H. 1966. A generalized method of predicting the minimum fluidization velocity, *AIChE J.,* 12:610–612.

Zenz, F.A. and Othmer, D.F. 1960. *Fluidization and Fluid-Particle Systems,* Reinhold, New York.

3.9 New-Newtonian Flows

Thomas F. Irvine Jr. and Massimo Capobianchi

Introduction

An important class of fluids exists which differ from Newtonian fluids in that the relationship between the shear stress and the flow field is more complicated. Such fluids are called non-Newtonian or rheological fluids. Examples include various suspensions such as coal–water or coal–oil slurries, food products, inks, glues, soaps, polymer solutions, etc.

An interesting characteristic of rheological fluids is their large "apparent viscosities". This results in laminar flow situations in many applications, and consequently the engineering literature is concentrated on laminar rather than turbulent flows. It should also be mentioned that knowledge of non-Newtonian fluid mechanics and heat transfer is still in an early stage and many aspects of the field remain to be clarified.

In the following sections, we will discuss the definition and classification of non-Newtonian fluids, the special problems of thermophysical properties, and the prediction of pressure drops in both laminar and turbulent flow in ducts of various cross-sectional shapes for different classes of non-Newtonian fluids.

Classification of Non-Newtonian Fluids

It is useful to first define a Newtonian fluid since all other fluids are non-Newtonian. Newtonian fluids possess a property called viscosity and follow a law analogous to the Hookian relation between the stress applied to a solid and its strain. For a one-dimensional Newtonian fluid flow, the shear stress at a point is proportional to the rate of strain (called in the literature the *shear rate*) which is the velocity gradient at that point. The constant of proportionality is the dynamic viscosity, i.e.,

$$\tau_{y,x} = \mu \frac{du}{dy} = \mu \dot{\gamma} \tag{3.9.1}$$

where x refers to the direction of the shear stress y the direction of the velocity gradient, and $\dot{\gamma}$ is the shear rate. The important characteristic of a Newtonian fluid is that the dynamic viscosity is independent of the shear rate.

Equation (3.9.1) is called a constitutive equation, and if $\tau_{x,y}$ is plotted against $\dot{\gamma}$, the result is a linear relation whose slope is the dynamic viscosity. Such a graph is called a *flow curve* and is a convenient way to illustrate the viscous properties of various types of fluids.

Fluids which do not obey Equation (3.9.1) are called non-Newtonian. Their classifications are illustrated in Figure 3.9.1 where they are separated into various categories of purely viscous time-independent or time-dependent fluids and viscoelastic fluids. Viscoelastic fluids, which from their name possess both viscous and elastic properties (as well as memory), have received considerable attention because of their ability to reduce both drag and heat transfer in channel flows. They will be discussed in a later subsection.

Purely viscous time-dependent fluids are those in which the shear stress in a function only of the shear rate but in a more complicated manner than that described in Equation (3.9.1). Figure 3.9.2 illustrates the characteristics of purely viscous time-independent fluids. In the figure, (a) and (b) are fluids where the shear stress depends only on the shear rate but in a nonlinear way. Fluid (a) is called pseudoplastic (or shear thinning), and fluid (b) is called dilatant (or shear thickening). Curve (c) is one which has an initial yield stress after which it acts as a Newtonian fluid, called Buckingham plastic, and curve (d), called Hershel-Buckley, also has a yield stress after which it becomes pseudoplastic. Curve (e) depicts a Newtonian fluid.

Figure 3.9.3 shows flow curves for two common classes of purely viscous time-dependent non-Newtonian fluids. It is seen that such fluids have a hysteresis loop or memory whose shape depends

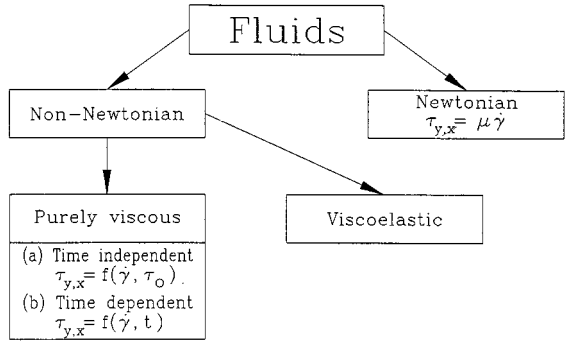


FIGURE 3.9.1 Classification of fluids.

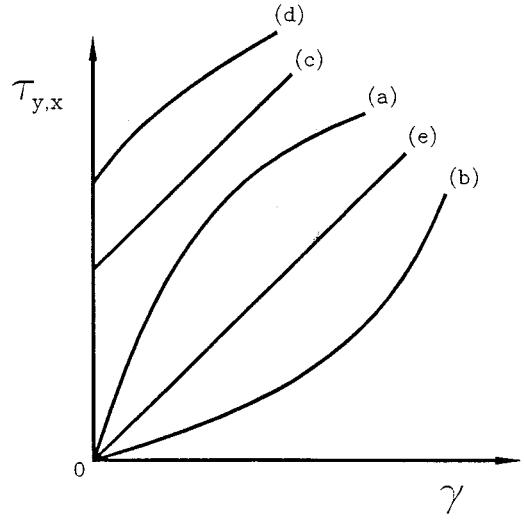


FIGURE 3.9.2 Flow curves of purely viscous, time-independent fluids: (a) pseudoplastic; (b) dilatant; (c) Bingham plastic; (d) Hershel–Buckley; (e) Newtonian.

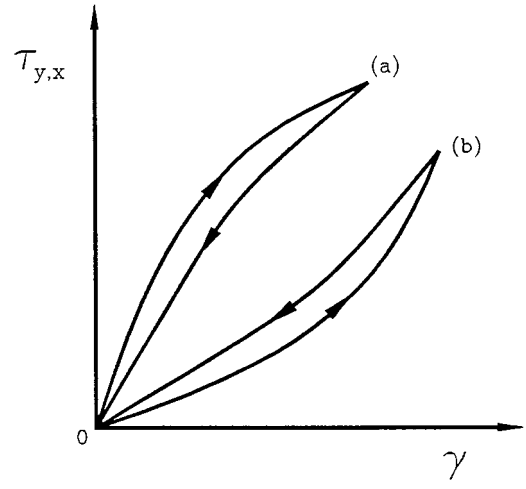


FIGURE 3.9.3 Flow curves for purely viscous, time-dependent fluids: (a) thixotropic; (b) rheopectic.

upon the time-dependent rate at which the shear stress is applied. Curve (a) illustrates a pseudoplastic time-dependent fluid and curve (b) a dilatant time-dependent fluid. They are called, respectively, thixotropic and rheopectic fluids and are complicated by the fact that their flow curves are difficult to characterize for any particular application.

Apparent Viscosity

Although non-Newtonian fluids do not have the property of viscosity, in the Newtonian fluid sense, it is convenient to define an apparent viscosity which is the ratio of the local shear stress to the shear rate at that point.

$$\mu_a = \frac{\tau}{\dot{\gamma}} \tag{3.9.2}$$

The apparent viscosity is not a true property for non-Newtonian fluids because its value depends upon the flow field, or shear rate. Nevertheless, it is a useful quantity and flow curves are often constructed with the apparent viscosity as the ordinate and shear rate as the abscissa. Such a flow curve will be illustrated in a later subsection.

Constitutive Equations

A constitutive equation is one that expresses the relation between the shear stress or apparent viscosity and the shear rate through the rheological properties of the fluid. For example, Equation (3.9.1) is the constitutive equation for a Newtonian fluid.

Many constitutive equations have been developed for non-Newtonian fluids with some of them having as many as five rheological properties. For engineering purposes, simpler equations are normally satisfactory and two of the most popular will be considered here.

Since many of the non-Newtonian fluids in engineering applications are pseudoplastic, such fluids will be used in the following to illustrate typical flow curves and constitutive equations. Figure 3.9.4 is a qualitative flow curve for a typical pseudoplastic fluid plotted with logarithmic coordinates. It is seen in the figure that at low shear rates, region (a), the fluid is Newtonian with a constant apparent viscosity of μ_o (called the *zero shear rate viscosity*). At higher shear rates, region (b), the apparent viscosity begins to decrease until it becomes a straight line, region (c). This region (c) is called the power law region and is an important region in fluid mechanics and heat transfer. At higher shear rates than the power law region, there is another transition region (d) until again the fluid becomes Newtonian in region (e). As discussed below, regions (a), (b), and (c) are where most of the engineering applications occur.

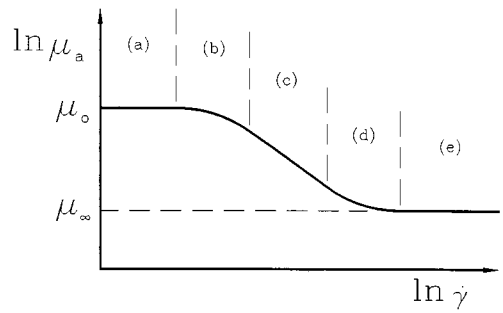


FIGURE 3.9.4 Illustrative flow curve for a pseudoplastic fluid (a) Newtonian region; (b) transition region I: (c) power law region; (d) transition region II; (e) high-shear-rate Newtonian region.

Power Law Constitutive Equation

Region (c) in Figure 3.9.4, which was defined above as the power law region, has a simple constitutive equation:

$$\tau = K\dot{\gamma}^{n} \tag{3.9.3}$$

or, from Equation (3.9.2):

$$\mu_a = K\dot{\gamma}^{n-1} \tag{3.9.4}$$

Here, K is called the fluid consistency and n the flow index. Note that if $n = 1$, the fluid becomes Newtonian and K becomes the dynamic viscosity. Because of its simplicity, the power law constitutive equation has been most often used in rheological studies, but at times it is inappropriate because it has several inherent flaws and anomalies. For example, if one considers the flow of a pseudoplastic fluid ($n < 1$) through a circular duct, because of symmetry at the center of the duct the shear rate (velocity gradient) becomes zero and thus the apparent viscosity from Equation (3.9.4) becomes infinite. This poses conceptual difficulties especially when performing numerical analyses on such systems. Another difficulty arises when the flow field under consideration is not operating in region (c) of Figure 3.9.4 but may have shear rates in region (a) and (b). In this case, the power law equation is not applicable and a more general constitutive equation is needed.

Modified Power Law Constitutive Equation

A generalization of the power law equation which extends the shear rate range to regions (a) and (b) is given by

$$\mu_a = \frac{\mu_o}{1 + \dfrac{\mu_o}{K}\dot{\gamma}^{1-n}} \tag{3.9.5}$$

Examination of Equation (3.9.5) reveals that at low shear rates, the second term in the denominator becomes small compared with unity and the apparent viscosity becomes a constant equal to μ_O. This represents the Newtonian region in Figure 3.9.4. On the other hand, as the second term in the denominator becomes large compared with unity, Equation (3.9.5) becomes Equation (3.9.4) and represents region (c), the power law region. When both denominator terms must be considered, Equation (3.9.5) represents region (b) in Figure 3.9.4.

An important advantage of the modified power law equation is that it retains the rheological properties K and n of the power law model plus the additional property μ_o. Thus, as will be shown later, in the flow and heat transfer equations, the same dimensionless groups as in the power law model will appear plus an additional dimensionless parameter which describes in which of the regions (a), (b), or (c) a particular system is operating. Also, solutions using the modified power law model will have Newtonian and power law solutions as asymptotes.

Equation (3.9.5) describes the flow curve for a pseudoplastic fluid ($n < 1$). For a dilatant fluid, ($n > 1$), an appropriate modified power law model is given by

$$\mu_a = \mu_o\left[1 + \frac{K}{\mu_o}\dot{\gamma}^{n-1}\right] \tag{3.9.6}$$

Many other constitutive equations have been proposed in the literature (Skelland, 1967; Cho and Hartnett, 1982; Irvine and Karni, 1987), but the ones discussed above are sufficient for a large number of engineering applications and agree well with the experimental determinations of rheological properties.

Rheological Property Measurements

For non-Newtonian fluids, specifying the appropriate rheological properties for a particular fluid is formidable because such fluids are usually not pure substances but various kinds of mixtures. This means that the properties are not available in handbooks or other reference materials but must be measured for each particular application. A discussion of the various instruments for measuring rheological properties is outside the scope of the present section, but a number of sources are available which describe different rheological property measurement techniques and instruments: Skelland (1967), Whorlow (1980), Irvine and Karni (1987), and Darby (1988). Figure 3.9.5 is an illustration of experimental flow curves measured with a falling needle viscometer and a square duct viscometer for polymer solutions of different concentrations. Also known in the figure as solid lines is the modified power law equation used to represent the experimental data. It is seen that Equation (3.9.5) fits the experimental data within ±2%. Table 3.9.1 lists the rheological properties used in the modified power law equations in Figure 3.9.5. It must be emphasized that a proper knowledge of these properties is vital to the prediction of fluid mechanics and heat transfer phenomena in rheological fluids.

TABLE 3.9.1 Rheological Properties Used in the Modified Power Law Equations in Figure 3.9.5 for Three Polymer Solutions of CMC-7H4

CMC	K (N · sec^n/m^2)	n	μ_o (N · sec/m^2)n
5000 wppm	2.9040	0.3896	0.21488
2500 wppm	1.0261	0.4791	0.06454
1500 wppm	0.5745	0.5204	0.03673

Source: Park, S. et al., *Proc. Third World Conf. Heat Transfer, Fluid Mechanics, and Thermodynamics,* Vol. 1, Elsevier, New York, 1993, 900–908.

Fully Developed Laminar Pressure Drops for Time-Independent Non-Newtonian Fluids

Modified Power Law Fluids

This important subject will be considered by first discussing modified power law fluids. The reason is that such solutions include both friction factor–Reynolds number relations and a shear rate parameter. The latter allows the designer to determine the shear rate region in which his system is operating and thus the appropriate solution to be used, i.e., regions (a), (b), or (c) in Figure 3.9.4.

For laminar fully developed flow of a modified power law fluid in a circular duct, the product of the friction factor and a certain Reynolds number is a constant depending on the flow index, n, and the shear rate parameter, β.

$$f_D \cdot \mathrm{Re}_m = \text{constant}(n,\beta) \tag{3.9.7}$$

where f_D is the Darcy friction factor and Re_m the modified power law Reynolds number, i.e.,

$$f_D = \frac{2\dfrac{\Delta p}{L} D_H}{\rho \bar{u}^2} \quad \text{(Darcy friction factor)}^*$$

* It should be noted that the Fanning friction factor is also used in the technical literature. The Fanning friction factor is $^1/_4$ of the Darcy friction factor, and will be characterized by the symbol f_F.

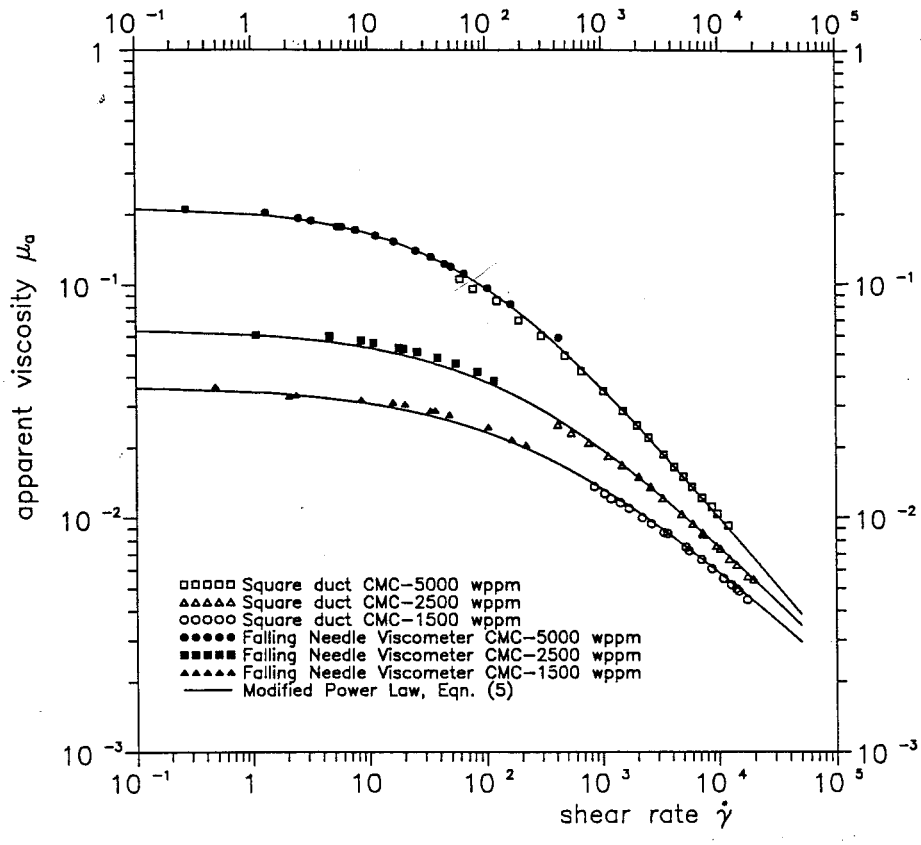


FIGURE 3.9.5 Experimental measurements of apparent viscosity vs. shear rate for polymer solutions (CMC-7H4) at different concentrations. (From Park, S. et al., in *Proc. Third World Conf. Heat Transfer, Fluid Mechanics, and Thermodynamics*, Vol. 1, Elsevier, New York, 1993, 900–908.).

$$\mathrm{Re}_m = \frac{\rho \bar{u} D_H}{\mu^*}$$

$$\mu^* = \frac{\mu_o}{1+\beta}$$

$$\beta = \frac{\mu_o}{K}\left(\frac{\bar{u}}{D_H}\right)^{1-n}$$

where β is the shear rate parameter mentioned previously which can be calculated by the designer for a certain operating duct ($\bar{u}$ and d) and a certain pseudoplastic fluid (μ_o, K, n). The solution for a circular tube has been calculated by Brewster and Irvine (1987) and the results are shown in Figure 3.9.6 and in Table 3.9.2. Referring to 3.9.6, we can see that when the $\log_{10}\beta$ is less than approximately –2, the duct is operating in region (a) of Figure 3.9.4 which is the Newtonian region and therefore classical Newtonian solutions can be used. Note that in the Newtonian region, Re_m reverts to the Newtonian Reynolds number given by

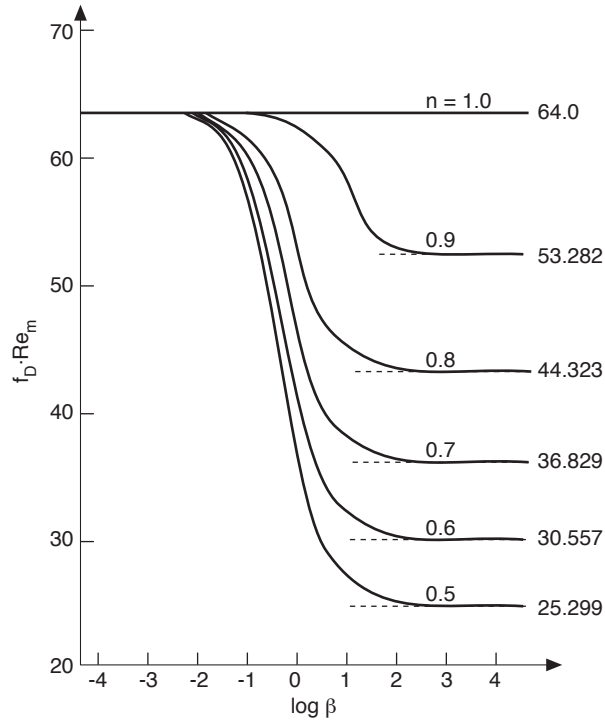


FIGURE 3.9.6 Product of friction factor and modified Reynolds number vs. $\log_{10} \beta$ for a circular duct.

.

TABLE 3.9.2 Summary of Computed Values of $f_D \cdot Re_m$ for Various Values of n and β for a Circular Duct

	$f_D \cdot Re_m$ for Flow Index: n =					
β	**1.0**	**0.9**	**0.8**	**0.7**	**0.6**	**0.5**
10^{-5}	64.000	64.000	64.000	64.000	63.999	63.999
10^{-4}	64.000	63.999	63.997	63.995	63.993	63.990
10^{-3}	64.000	63.987	63.972	63.953	63.930	63.903
10^{-2}	64.000	63.873	63.720	63.537	63.318	63.055
10^{-1}	64.000	62.851	61.519	59.987	58.237	56.243
10^{0}	64.000	58.152	52.377	46.761	41.384	36.299
10^{1}	64.000	54.106	45.597	38.308	32.082	26.771
10^{2}	64.000	53.371	44.458	36.985	30.716	25.451
10^{3}	64.000	53.291	44.336	36.845	30.573	25.314
10^{4}	64.000	53.283	44.324	36.831	30.559	25.300
10^{5}	64.000	53.282	44.323	36.830	30.557	25.299
Exact solution	64.000	53.282	44.323	36.829	30.557	25.298

Source: Brewster, R.A. and Irvine, T.F., Jr., *Wärme und Stoffübertragung*, 21, 83–86, 1987. With permission.

$$\mathrm{Re}_N = \frac{\rho \bar{u} D_H}{\mu_o} \tag{3.9.8}$$

When the value of $\log_{10} \beta$ is approximately in the range $-2 \le \log_{10} \beta \le 2$, the duct is operating in the transition region (b) of Figure 3.9.4 and the values of $f_D \cdot \mathrm{Re}_m$ must be obtained from Figure 3.9.6 or from Table 3.9.2.

When $\log_{10} \beta$ is greater than approximately 2, the duct is operating in the power law region (c) of Figure 3.9.4 and power law friction factor Reynolds number relations can be used. They are also indicated in Figure 3.9.6 and Table 3.9.2. In this region, Re_m becomes the power law Reynolds number given by

$$\mathrm{Re}_g = \frac{\rho \bar{u}^{2-n} D_H^n}{K} \tag{3.9.9}$$

For convenience, Brewster and Irvine (1987) have presented a correlation equation which agrees within 0.1% with the results tabulated in Table 3.9.2.

$$f_D \cdot \mathrm{Re}_m = \frac{1+\beta}{\dfrac{1}{64} + \dfrac{\beta}{2^{3n+3}\left(\dfrac{3n+1}{4n}\right)^n}} \tag{3.9.10}$$

Thus, Equation (3.9.10) contains all of the information required to calculate the circular tube laminar fully developed pressure drop for a pseudoplastic fluid depending upon the shear rate region(s) under consideration, i.e., regions (a), (b), or (c) of Figure 3.9.4. Note that in scaling such non-Newtonian systems, both Re_m and β must be held constant. Modified power law solutions have been reported for two other duct shapes. Park et al. (1993) have presented the friction factor–Reynolds number relations for rectangular ducts and Capobianchi and Irvine (1992) for concentric annular ducts.

Power Law Fluids

Since the power law region of modified power law fluids ($\log_{10} \beta \ge 2$) is often encountered, the friction factor–Reynolds number relations will be discussed in detail in this subsection.

An analysis of power law fluids which is most useful has been presented by Kozicki et al. (1967). Although the method is approximate, its overall accuracy (±5%) is usually sufficient for many engineering calculations. His expression for the friction factor–Reynolds number product is given by

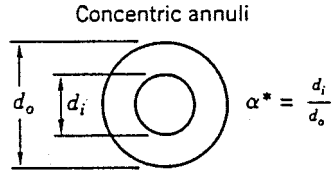
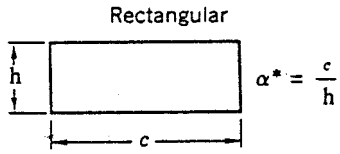
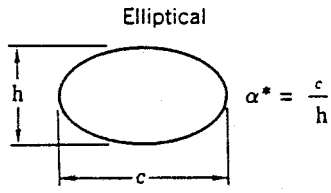
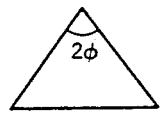
$$f_D \cdot \mathrm{Re}^* = 2^{6n} \tag{3.9.11}$$

where

$$\mathrm{Re}^* = \text{Kozicki Reynolds number}, \quad \mathrm{Re}^* = \frac{\mathrm{Re}_g}{\left[\dfrac{a+bn}{n}\right]^n 8^{n-1}} \tag{3.9.12}$$

and a and b are geometric constants which depend on the cross-sectional shape of the duct. For example, for a circular duct, $a = 0.25$ and $b = 0.75$. Values of a and b for other duct shapes are tabulated in Table 3.9.3. For additional duct shapes in both closed and open channel flows, Kozicki et al. (1967) may be consulted.

TABLE 3.9.3 Constants *a* and *b* for Various Duct Geometrics Used in the Method Due to Kozicki et al. (1967)

Geometry	α^*	*a*	*b*
	0.1	0.4455	0.9510
	0.2	0.4693	0.9739
	0.3	0.4817	0.9847
	0.4	0.4890	0.9911
	0.5	0.4935	0.9946
	0.6	0.4965	0.9972
	0.7	0.4983	0.9987
	0.8	0.4992	0.9994
	0.9	0.4997	1.0000
	1.0[a]	0.5000	1.0000
	0.0	0.5000	1.0000
	0.25	0.3212	0.8482
	0.50	0.2440	0.7276
	0.75	0.2178	0.6866
	1.00	0.2121	0.8766
	0.00	0.3084	0.9253
	0.10	0.3018	0.9053
	0.20	0.2907	0.8720
	0.30	0.2796	0.8389
	0.40	0.2702	0.8107
	0.50	0.2629	0.7886
	0.60	0.2575	0.7725
	0.70	0.2538	0.7614
	0.80	0.2515	0.7546
	0.90	0.2504	0.7510
	1.00[b]	0.2500	0.7500
	2ϕ (deg)		
Isosceles triangular 	10	0.1547	0.6278
	20	0.1693	0.6332
	40	0.1840	0.6422
	60	0.1875	0.6462
	80	0.1849	0.6438
	90	0.1830	0.6395
	N		
Regular polygon (N sides) 	4	0.2121	0.6771
	5	0.2245	0.6966
	6	0.2316	0.7092
	8	0.2391	0.7241

[a] Parallel plates.

[b] Circle.

Source: Irvine, T.F., Jr. and Karni, J., in *Handbook of Single Phase Convective Heat Transfer,* John Wiley and Sons, New York, 1987, pp 20-1–20-57.

Fully Developed Turbulent Flow Pressure Drops

In a number of engineering design calculations for turbulent flow, the shear rate range falls in region (c) of Figure 3.9.4. Thus, power law relations are appropriate for such pressure drop calculations.

Hartnett and Kostic (1990) have investigated the various correlations which have appeared in the literature for circular tubes and have concluded that for a circular tube the relation proposed by Dodge and Metzner (1959) is the most reliable for pseudoplastic fluids. It is given by

$$\frac{1}{f_F^{1/2}} = \frac{4.0}{n^{0.75}} \cdot \log_{10}\left[\mathrm{Re}'_g \left(f_F\right)^{1-(1/2n)}\right] - \frac{0.40}{n^{1.2}} \tag{3.9.13}$$

where f_F is the Fanning friction factor and

$$\mathrm{Re}'_g = \mathrm{Re}_g \left[\frac{8^{1-n}}{\left[\frac{3n+1}{4n}\right]^n}\right] \tag{3.9.14}$$

Figure 3.9.7 is a graphical representation of Equation (3.9.13) which indicates the Dodge and Metzner experimental regions by solid lines, and by dashed lines where the data are extrapolated outside of their experiments.

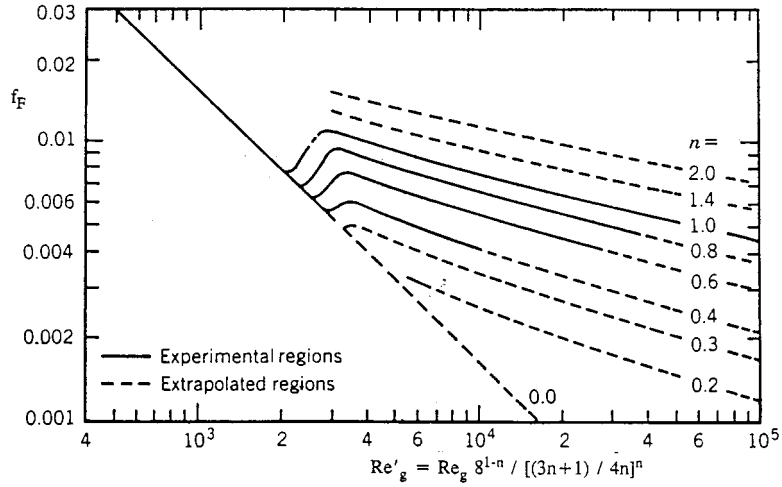


FIGURE 3.9.7 Dodge and Metzner relation between Fanning friction factor and Re'_g. (From Dodge, D.W. and Metzner, A.B., *AIChE J.*, 5, 189–204, 1959.)

For noncircular ducts in turbulent fully developed flow, only a limited amount of experimental data are available. Kostic and Hartnett (1984) suggest the correlation:

$$\frac{1}{f_F^{1/2}} = \frac{4}{n^{0.75}} \cdot \log_{10}\left[\mathrm{Re}^* \left(f_F\right)^{1-(1/2n)}\right] - \frac{0.40}{n^{0.5}} \tag{3.9.15}$$

where f_F is again the Fanning friction factor and Re^* is the Kozicki Reynolds number:

$$\mathrm{Re}^* = \frac{\mathrm{Re}_g}{\left[\frac{a+bn}{n}\right]^n 8^{n-1}} \tag{3.9.16}$$

and a and b are geometric constants given in Table 3.9.3.

Viscoelastic Fluids

Fully Developed Turbulent Flow Pressure Drops

Viscoelastic fluids are of interest in engineering applications because of reductions of pressure drop and heat transfer which occur in turbulent channel flows. Such fluids can be prepared by dissolving small amounts of high-molecular-weight polymers, e.g., polyacrylamide, polyethylene oxide (Polyox), etc., in water. Concentrations as low as 5 parts per million by weight (wppm) result in significant pressure drop reductions. Figure 3.9.8 from Cho and Hartnett (1982) illustrates the reduction in friction factors for Polyox solutions in a small-diameter capillary tube. It is seen that at zero polymer concentration the data agree with the Blasius equation for Newtonian turbulent flow. With the addition of only 7 wppm of Polyox, there is a significant pressure drop reduction and for concentrations of 70 wppm and greater all the data fall on the Virk line which is the maximum drag-reduction asymptote. The correlations for the Blasius and Virk lines as reported by Cho and Hartnett (1982) are

$$f_F = \frac{0.079}{\mathrm{Re}^{1/4}} \quad \text{(Blasius)} \tag{3.9.17}$$

$$f_F = 0.20\,\mathrm{Re}_a^{-0.48} \quad \text{(Virk)} \tag{3.9.18}$$

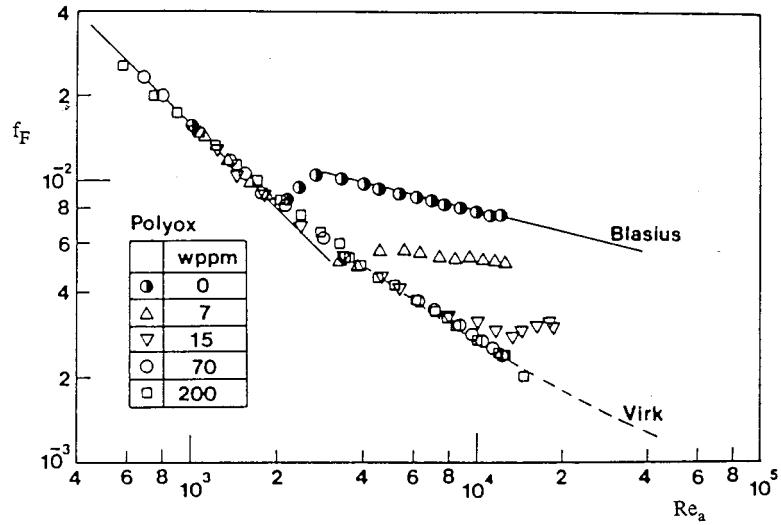


FIGURE 3.9.8 Reduction in friction factors for polyethylene oxide (Polyox) solutions in a small-diameter capillary tube. (From Cho, Y.I. and Harnett, J.P., *Adv. Heat Transfer,* 15, 59–141, 1982. With permission.)

At the present time, no generally accepted method exists to predict the drag reduction between the Blasius and Virk lines. Kwack and Hartnett (1983) have proposed that the amount of drag reduction between those two correlations is a function of the Weissenberg number, defined as

$$w_s = \frac{\lambda \bar{u}}{D_H} \tag{3.9.19}$$

where λ = characteristic time of the viscoelastic fluid. They present correlations which allow the friction factor to be estimated at several Reynolds numbers between the Blasius and Virk lines.

Fully Developed Laminar Flow Pressure Drops

The above discussion on viscoelastic fluids has only considered fully developed turbulent flows. Laminar fully developed flows can be considered as nonviscoelastic but purely viscous non-Newtonian. Therefore, the method of Kozicki et al. (1967) may be applied to such situations once the appropriate rheological properties have been determined.

Nomenclature

a = duct shape geometric constant
b = duct shape geometric constant
c = duct width (see Table 3.9.3) (m)
d_i = concentric annuli inner diameter (see Table 3.9.3) (m)
d_o = concentric annuli outer diameter (see Table 3.9.3) (m)
f_D = Darcy friction factor
f_F = Fanning friction factor
h = duct height (see Table 3.9.3) (m)
K = fluid consistency (Ns^n/m^2)
n = flow index
N = number of sides in polygon (see Table 3.9.3)
Re_g = generalized Reynolds number,

$$\mathrm{Re}_g = \frac{\rho \bar{u}^{2-n} D_H^n}{K}$$

Re_m = modified power law Reynolds number,

$$\mathrm{Re}_m = \frac{\rho \bar{u} D_H}{\mu^*}$$

Re_N = modified power law Reynolds number Newtonian asymptote,

$$\mathrm{Re}_N = \frac{\rho \bar{u} D_H}{\mu_o}$$

Re_a = apparent Reynolds number

$$\mathrm{Re}_a = \frac{\mathrm{Re}_g}{\left(\frac{3n+1}{4n}\right)^{n-1} 8^{n-1}}$$

Re^* = Kozicki Reynolds number

$$\mathrm{Re}^* = \frac{\rho \bar{u}^{2-n} D_H^n}{K\left[\frac{a+bn}{n}\right]^n 8^{n-1}}$$

Re'_g = Metzner Reynolds number

$$\mathrm{Re}'_g = \mathrm{Re}_g \left[\frac{8^{1-n}}{\left[\frac{3n+1}{4n} \right]^n} \right]$$

$\bar{u}$ = average streamwise velocity (m/sec)
t = time (sec)
w_s = Weissenberg number
x = direction of shear stress (m)
y = direction of velocity gradient (m)

Greek

α^* = duct aspect ratio in Table 3.9.3
β = shear rate parameter

$$\beta = \frac{\mu_o}{K} \left(\frac{\bar{u}}{D_H} \right)^{1-n}$$

$\dot{\gamma}$ = shear rate (L/sec)
ΔP = presure drop (N/m^2)
λ = characteristic time of viscoelastic fluid (sec)
μ_a = apparent viscosity (N · sec/m^2)
μ_o = zero shear rate viscosity (N · sec/m^2)
μ_∞ = high shear rate viscosity (N · sec/m^2)
μ^* = reference viscosity

$$\mu^* = \frac{\mu_o}{1+\beta} \quad \left(\mathrm{N \cdot sec/m^2} \right)$$

τ_o = yield stress (N/m^2)
$\tau_{y,x}$ = shear stress (N/m^2)
ϕ = half apex angle (see Table 3.9.3) (°)

References

Brewster, A.A. and Irvine, T.F. Jr. 1987. Similtude considerations in laminar flow of power law fluids in circular ducts, *Wärme und Stoffübertagung,* 21:83–86.

Capobianchi, M. and Irvine, T.F. Jr. 1992. Predictions of pressure drop and heat transfer in concentric annular ducts with modified power law fluids, *Wärme und Stoffübertagung,* 27:209–215.

Cho, Y.I. and Hartnett, J.P. 1982. Non-Newtonian fluids in circular pipe flow, in *Adv. Heat Transfer,* 15:59–141.

Darby, R. 1988. Laminar and turbulent pipe flows of non-Newtonian fluids, in *Encyclopedia of Fluid Mechanics,* Vol. 7, Gulf Publishing, Houston, 7:20–53.

Dodge, D.W. and Metzner, A.B. 1959. Turbulent flow of non-Newtonian systems, *AIChE J.,* 5:189–204.

Harnett, J.P. and Kostic, M. 1990. Turbulent Friction Factor Correlations for Power Law Fluids in Circular and Non-Circular Channels, *Int. Comm. Heat and Mass Transfer,* 17:59–65.

Irvine, T.F. Jr. and Karni, J. 1987. Non-Newtonian fluid flow and heat transfer, in *Handbook of Single Phase Convective Heat Transfer,* pp. 20-1–20-57, John Wiley and Sons, New York.

Kostic, M. and Hartnett, J.P. 1984. Predicting turbulent friction factors of non-Newtonian fluids in non-circular ducts, *Int. Comm. Heat and Mass Transfer,* 11:345–352.

Kozicki, W., Chou, C.H., and Tiu, C. 1967. Non-Newtonian flow in ducts of arbitrary cross-sectional shape, *Can. J. Chem. Eng.,* 45:127–134.

Kwack, E.Y. and Hartnett, J.P. 1983. Empirical correlations of turbulent friction factors and heat transfer coefficients for viscoelastic fluids, *Int. Comm. Heat and Mass Transfer,* 10:451–461.

Park, S., Irvine, T.F. Jr., and Capobianchi, M. 1993. Experimental and numerical study of friction factor for a modified power law fluid in a rectangular duct, *Proc. Third World Conf. Heat Transfer, Fluid Mechanics, and Thermodynamics,* Vol. 1, Elsevier, New York, 1:900–908.

Skelland, A.H.P. 1967. *Non-Newtonian Flow and Heat Transfer,* John Wiley and Sons, New York.

Whorlow, R.W. 1980. *Rheological Techniques,* Halsted Press, New York.

Further Information

It is not possible to include all of the interesting non-Newtonian topics in a section of this scope. Other items which may be of interest and importance are listed below along with appropriate references: hydrodynamic and thermal entrance lengths, Cho and Hartnett (1982); non-Newtonian flow over external surfaces, Irvine and Karni (1987); chemical, solute, and degradation effects in viscoelastic fluids, Cho and Harnett (1982); general references, Skelland (1967), Whorlow (1980), and Darby (1988).

3.10 Tribology, Lubrication, and Bearing Design

Francis E. Kennedy, E. Richard Booser, and Donald F. Wilcock

Introduction

Tribology, the science and technology of contacting surfaces involving friction, wear, and lubrication, is extremely important in nearly all mechanical components. A major focus of the field is on friction, its consequences, especially wear and its reduction through lubrication and material surface engineering. The improper solution of tribological problems is responsible for huge economic losses in our society, including shortened component lives, excessive equipment downtime, and large expenditures of energy. It is particularly important that engineers use appropriate means to reduce friction and wear in mechanical systems through the proper selection of bearings, lubricants, and materials for all contacting surfaces. The aim of this section is to assist in that endeavor.

Sliding Friction and its Consequences

Coefficient of Friction

If two stationary contacting bodies are held together by a normal force W and a tangential force is applied to one of them, the tangential force can be increased until it reaches a magnitude sufficient to initiate sliding. The ratio of the friction force at incipient sliding to the normal force is known as the static coefficient of friction, f_s. After sliding begins, the friction force always acts in the direction opposing motion and the ratio between that friction force and the applied normal force is the kinetic coefficient of friction, f_k.

Generally, f_k is slightly smaller than f_s and both coefficients are independent of the size or shape of the contacting surfaces. Both coefficients are very much dependent on the materials and cleanliness of the two contacting surfaces. For ordinary metallic surfaces, the friction coefficient is not very sensitive to surface roughness. For ultrasmooth or very rough surfaces, however, the friction coefficient can be larger. Typical friction coefficient values are given in Table 3.10.1. Generally, friction coefficients are greatest when the two surfaces are identical metals, slightly lower with dissimilar but mutually soluble metals, still lower for metal against nonmetal, and lowest for dissimilar nonmetals.

TABLE 3.10.1 Some Typical Friction Coefficients[a]

Material Pair	Static Friction Coefficient f_s		Kinetic Friction Coefficient f_k	
	In Air	In Vacuo	In Air, Dry	Oiled
Mild steel vs. mild steel	0.75	—	0.57	0.16
Mild steel vs. copper	0.53	0.5 (oxidized) 2.0 (clean)	0.36	0.18
Copper vs. copper	1.3	21.0	0.8	0.1
Tungsten carbide vs. copper	0.35	—	0.4	—
Tungsten carbide vs. tungsten carbide	0.2	0.4	0.15	—
Mild steel vs. polytetrafluoroethylene	0.04	—	0.05	0.04

[a] The friction coefficient values listed in this table were compiled from several of the references listed at the end of this section.

The kinetic coefficient of friction, f_k, for metallic or ceramic surfaces is relatively independent of sliding velocity at low and moderate velocities, although there is often a slight decrease in f_k at higher velocities. With polymers and soft metals there may be an increase in the friction coefficient with increasing velocity until a peak is reached, after which friction may decrease with further increases in velocity or temperature. The decrease in kinetic friction coefficient with increasing velocity, which may become especially pronounced at higher sliding velocities, can be responsible for friction-induced

vibrations (stick–slip oscillations) of the sliding systems. Such vibrations are an important design consideration for clutches and braking systems, and can also be important in the accurate control and positioning of robotic mechanisms and precision manufacturing systems.

Wear

Wear is the unwanted removal of material from solid surfaces by mechanical means; it is one of the leading reasons for the failure and replacement of manufactured products. It has been estimated that the costs of wear, which include repair and replacement, along with equipment downtime, constitute up to 6% of the U.S. gross national product (Rabinowicz, 1995). Wear can be classified into four primary types: sliding wear, abrasion, erosion, and corrosive wear. Owing to its importance, wear and its control have been the subject of several handbooks (Peterson and Winer, 1980; Blau, 1992), which the interested reader may consult for further information.

Types of Wear. Sliding wear occurs to some degree whenever solid surfaces are in sliding contact. There are two predominant sliding wear mechanisms, adhesion and surface fatigue. *Adhesive wear* is caused by strong adhesive forces between the two surfaces within the real area of contact. It results in the removal of small particles from at least one of the surfaces, usually the softer one. These particles can then transfer to the other surface or mix with other material from both surfaces before being expelled as loose wear debris. Adhesive wear can be particularly severe for surfaces which have a strong affinity for each other, such as those made from identical metals. *Surface fatigue wear* occurs when repeated sliding or rolling/sliding over a wear track results in the initiation of surface or subsurface cracks, and the propagation of those cracks produces wear particles in ductile materials by a process that has been called delamination. With brittle materials, sliding wear often occurs by a *surface fracture* process.

After an initial transition or "running-in" period, sliding wear tends to reach a steady state rate which is approximated by the following Archard (or Holm/Archard) wear equation:

$$V = K * W * s/H \tag{3.10.1}$$

where V = volume of worn material, K = dimensionless wear coefficient, s = sliding distance, W = normal load between the surfaces, and H = hardness of the softer of the two contacting surfaces.

The dimensionless wear coefficient gives an indication of the tendency of a given material combination to wear; relative wear coefficient values are given in Figure 3.10.1. In general, wear coefficients are highest for identical metals sliding without lubrication, and wear is decreased by adding a lubricant and by having material pairs which are dissimilar.

Abrasive wear occurs when a hard, rough surface slides against a softer surface (*two-body abrasion*) or when hard particles slide between softer surfaces (*three-body abrasion*). This process usually results in material removal by plowing or chip formation, especially when the abraded surface is metallic; surface fracture can occur during abrasion of brittle surfaces. In fact, abrasion mechanisms are similar to those of grinding and lapping, which could be considered as intentional abrasion. Consideration of the cutting and plowing processes shows that abrasive wear obeys the same equation (3.10.1) as sliding wear (Archard, 1980; Rabinowicz, 1995). Typical wear coefficients for abrasive wear are given in Figure 3.10.1. Since the relative size, hardness, and sharpness of the abrading particles, or surface asperities, also affect abrasive wear rates, the wear coefficients for abrasion must include recognition of those factors (Rabinowicz, 1995).

Erosion occurs when solid particles or liquid droplets impinge on a solid surface. When impingement is on a ductile metallic surface, the wear process is similar to that caused by abrasion, and is dominated by plastic deformation. Brittle surfaces, on the other hand, tend to erode by surface fracture mechanisms. The material removal rate is dependent on the angle of attack of the particles, with erosion reaching a peak at low angles (about 20°) for ductile surfaces and at high angles (90°) for brittle materials. In either case, the wear rate is proportional to the mass rate of flow of the particles and to their kinetic energy; it is inversely proportional to the hardness of the surface and the energy-absorbing potential (or toughness)

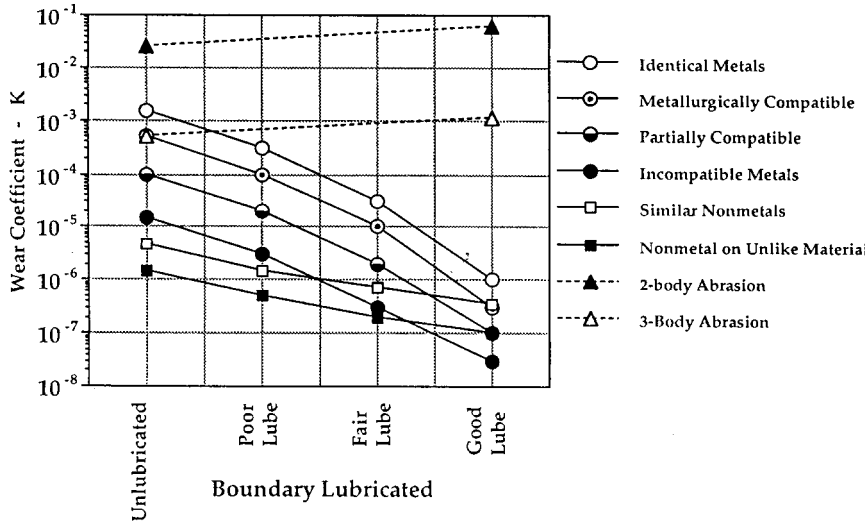


FIGURE 3.10.1 Typical values of wear coefficient for sliding and abrasive wear. (Modified from Rabinowicz, 1980, 1995.)

of the impinged surface (Schmitt, 1980). Although erosion is usually detrimental, it can be used beneficially in such material removal processes as sandblasting and abrasive water jet machining.

Corrosive wear results from a combination of chemical and mechanical action. It involves the synergistic effects of chemical attack (corrosion) of the surface, followed by removal of the corrosion products by a wear mechanism to expose the metallic surface, and then repetition of those processes. Since many corrosion products act to protect the surfaces from further attack, the removal of those films by wear acts to accelerate the rate of material removal. Corrosive wear can become particularly damaging when it acts in a low-amplitude oscillatory contact, which may be vibration induced, in which case it is called *fretting corrosion.*

Means for Wear Reduction.
The following actions can be taken to limit sliding wear:

- Insure that the sliding surfaces are well lubricated. This can best be accomplished by a liquid lubricant (see sub-section on effect of lubrication on friction and wear), but grease, or solid lubricants such as graphite or molybdenum disulfide, can sometimes be effective when liquid lubricants cannot be used.
- Choose dissimilar materials for sliding pairs.
- Use hardened surfaces.
- Add wear-resistant coatings to the contacting surfaces (see the following subsection).
- Reduce normal loads acting on the contact.
- Reduce surface temperatures. This is particularly important for polymer surfaces.

To reduce abrasive wear:

- Use hardened surfaces.
- Add a hard surface coating.
- Reduce the roughness of hard surfaces that are in contact with softer surfaces.
- Provide for the removal of abrasive particles from contacting surfaces. This can be done by flushing surfaces with liquid and/or filtering liquid coolants and lubricants.
- Reduce the size of abrasive particles.

To reduce erosion:

- Modify the angle of impingement of solid particles or liquid droplets.
- Provide for the removal of solid particles from the stream of fluid.
- Use hardened surfaces.
- Use tough materials for surfaces.
- Add protective coating to surfaces.

Surface Engineering for Friction and Wear Reduction

Surface treatments have long been an important remedy for wear problems, and that importance has grown in recent years with the introduction of new techniques to harden surfaces or apply hard surface coatings. Available processes and characteristics for treating a steel substrate are listed in Table 3.10.2.

Thermal transformation hardening processes are used to harden ferrous (primarily steel) surfaces by heating the surface rapidly, transforming it to austenite, and then quenching it to form martensite. The source of heat can be one of the following: an oxyacetylene or oxypropane flame (*flame hardening*), eddy currents induced by a high-frequency electric field (*induction hardening*), a beam from a high-power laser (*laser hardening*), or a focused electron beam (*electron beam hardening*). The depth and uniformity of the hard layer depend on the rate and method of heating. These processes are characterized by a short process time and all except electron beam hardening (which requires a moderate vacuum) can be done in air.

TABLE 3.10.2 Characteristics of Surface Treatment Processes for Steel

Process	Coating or Treated Layer		Substrate Temperature (°C)
	Hardness (*HV*)	Thickness (μm)	
Surface hardening			
Flame or induction hardening	500–700	250–6000	800–1000
Laser or electron beam hardening	500–700	200–1000	950–1050
Carburizing	650–900	50–1500	800–950
Carbonitriding	650–900	25–500	800–900
Nitriding	700–1200	10–200	500–600
Boronizing	1400–1600	50–100	900–1100
Coating			
Chrome plating	850–1250	1–500	25–100
Electroless nickel	500–700	0.1–500	25–100
Hardfacing	800–2000	500–50000	1300–1400
Thermal spraying	400–2000	50–1500	<250
Physical vapor deposition	100–3000	0.05–10	100–300
Chemical vapor deposition	1000–3000	0.5–100	150–2200
Plasma-assisted chemical vapor deposition	1000–5000	0.5–10	<300
Ion implantation	750–1250	0.01–0.25	<200

Thermal diffusion processes involve the diffusion of atoms into surfaces to create a hard layer. In the most widely used of these processes, *carburizing* (or case hardening), carbon diffuses into a low-carbon steel surface to produce a hard, carbon-rich case. The hardness and thickness of the case depend on the temperature, exposure time, and source of carbon (either a hydrocarbon gas, a salt bath, or a packed bed of carbon). *Carbonitriding* is a process similar to carburizing which involves the simultaneous diffusion of carbon and nitrogen atoms into carbon steel surfaces. In the *nitriding* process, nitrogen atoms diffuse into the surface of a steel which contains nitride-forming elements (such as Al, Cr, Mo, V, W, or Ti) and form fine precipitates of nitride compounds in a near-surface layer. The hardness of the surface layer depends on the types of nitrides formed. The source of nitrogen can be a hot gas (usually ammonia) or a plasma. *Nitrocarburizing* and *boronizing* are related processes in which nitrogen or boron atoms diffuse

into steel surfaces and react with the iron to form a hard layer of iron carbonitride or iron boride, respectively.

Thin, hard metallic coatings can be very effective in friction and wear reduction and can be applied most effectively by *electroplating processes* (Weil and Sheppard, 1992). The most common of such coatings are *chromium*, which is plated from a chromic acid bath, and *electroless nickel*, which is deposited without electric current from a solution containing nickel ions and a reducing agent. Chromium coatings generally consist of fine-grained chromium with oxide inclusions, while electroless nickel coatings contain up to 10% of either phosphorus or boron, depending on the reducing agent used.

Thermal spray processes (Kushner and Novinski, 1992) enable a large variety of coating materials, including metals, ceramics and polymers, to be deposited rapidly on a wide range of substrates. Four different thermal spray processes are commercially available: *oxyfuel* (or flame) spraying of metallic wire or metallic or ceramic powder, *electric arc* spraying of metallic wire, *plasma arc* spraying of powder (metallic or ceramic), and *high-velocity oxyfuel* (or detonation gun) powder spray. In each thermal spray process the coating material, in either wire or powder form, is heated to a molten or plastic state, and the heated particles are propelled toward the surface to be coated where they adhere and rapidly solidify to form a coating. The hardness of the coating depends on both the sprayed material and the process parameters.

Weld hardfacing processes (Crook and Farmer, 1992) involve the application of a wear-resistant material to the surface of a part by means of a weld overlay. Weld overlay materials include ferrous alloys (such as martensitic air-hardening steel or high-chromium cast iron), nonferrous alloys (primarily cobalt- or nickel-based alloys containing hard carbide, boride, or intermetallic particles), and cemented carbides (usually tungsten carbide/cobalt cermets). In each case the surface being coated is heated to the same temperature as the molten weld layer, thus posing a limitation to the process. Weld hardfacing is best used when abrasion or sliding wear cannot be avoided (as with earthmoving or mining equipment) and the goal is to limit the wear rate.

Vapor deposition processes for wear-resistant coatings include *physical vapor deposition* (PVD), *chemical vapor deposition* (CVD), and several variants of those basic processes (Bhushan and Gupta, 1991). Each of the processes consists of three steps: (1) creation of a vapor phase of the coating material, (2) transportation of the vapor from source to substrate, and (3) condensation of the vapor phase on the substrate and growth of a thin solid film. In PVD processes the vapor is produced by either evaporation (by heating of the coating source) or sputtering (in which coating material is dislodged and ejected from the source as a result of bombardment by energetic particles). In some PVD processes the vapor becomes ionized or reacts with a gas or plasma en route to the substrate, thus modifying the structure or composition of the deposited film. In CVD processes a gas composed of a volatile component of the coating material is activated either thermally or by other means in the vicinity of the substrate, and it reacts to form a solid deposit on the surface of the hot substrate.

Both PVD and CVD methods can be used to produce a wide variety of coatings, including metals, alloys, and refractory compounds. Among the most popular vapor-deposited hard coatings for wear resistance are titanium nitride and titanium carbide. Deposition rates are relatively low compared with some other coating processes, ranging from <0.1 μm/min for some ion beam–sputtering or ion-plating processes, up to 25 μm/min or more for activated reactive evaporation or CVD processes. Most PVD processes are done in a vacuum, while CVD processes are done in a reaction chamber which may be at atmospheric pressure. *Plasma-assisted chemical vapor deposition* (PACVD) is a hybrid process in which the constituents of the vapor phase react to form a solid film when assisted by a glow discharge plasma. The advantages of PACVD over other CVD processes include lower substrate temperatures, higher deposition rates, and a wider variety of coating possibilities.

Ion implantation (Fenske, 1992) is a process in which charged particles are created in an ion source, accelerated toward the surface at high velocity, and then injected into the substrate surface. The most commonly implanted ions for surface engineering are nitrogen, carbon, boron, and titanium, although virtually any element could be implanted. The microstructure of the near-surface region is changed by

the presence of the implanted ions and the result can be high near-surface hardness and wear resistance. The affected layer is very thin (<1 μm).

Effect of Lubrication on Friction and Wear

Whenever lubricated surfaces slide together at low sliding speeds or with a high applied normal load, the lubricant may not separate the two solid surfaces completely. However, the lubricant can still significantly reduce the friction coefficient by reducing the shear strength of adhesive junctions between the two surfaces. In this so-called boundary lubrication regime, the effectiveness of the lubricant can be improved if the lubricant molecules adhere well to the solid surfaces. This is best accomplished by introducing a lubricant or additive that forms a surface film through adsorption, chemisorption, or chemical reaction with the surface. The ensuing reduced shear strength of the surface film can lower the friction coefficient by as much as an order of magnitude from the dry friction value.

When a good supply of a viscous lubricant is available, the separation between the surfaces will increase as the sliding speed increases or the normal load decreases. As the separation increases, the amount of solid/solid contact between the surfaces will decrease, as will the friction coefficient and wear rate. In this "mixed friction" regime, friction is determined by the amount of plowing deformation on the softer surface by the harder surface asperities and by adhesion within the solid/solid contacts. When the surfaces become completely separated by a self-acting or externally pressurized lubricant film, the lubricating regime is hydrodynamic, wear is reduced to nearly zero, and friction reaches a low value governed by viscous shear of the lubricant. Friction coefficients in such cases can be 0.001 or lower, depending on the surface velocities and the lubricant viscosity. This is the case for most journal or thrust bearings (see subsection on fluid film bearings).

Bearings for Friction Reduction

Most mechanical systems contain moving components, such as shafts, which must be supported and held in position by stationary members. This is best done by appropriate design or selection of bearings to be used wherever the moving member is to be supported. Most bearings may be classified as either fluid film bearings, dry or semilubricated bearings, or rolling element bearings.

Fluid film bearings (see subsection below) have a conformal geometry, with a thin film of fluid separating the two surfaces. The fluid lubricant could be a liquid, such as oil, or a gas, such as air. Fluid film bearings are commonly used to support rotating cylindrical shafts, and the load on such a bearing could be either radial, in which case the bearing is called a journal bearing, or axial, for a thrust bearing. In most cases the fluid film is generated by the motion within the bearing itself, so the bearing is called self-acting or hydrodynamic. Whether or not a self-acting bearing can develop a fluid film sufficient to separate and support the two surfaces is determined by magnitude of the quantity $\mu U/W$, where μ is the (absolute) fluid viscosity, U is the relative sliding velocity, and W is the normal load. If that quantity is too small, the fluid film will be too thin and high friction will occur. This can be a problem during start-up of equipment when sliding velocities are low. That problem can be overcome by pressurizing the fluid film from an external pressure source to create a hydrostatic bearing. Whether the fluid film is externally pressurized (hydrostatic) or self-acting (hydrodynamic), separation of the solid surfaces allows wear to be essentially eliminated and friction to be very low, even when very large loads are carried by the pressurized lubricant.

Dry and semilubricated bearings (see subsection below) have conformal surfaces which are in direct contact with each other. This category includes bearings which run dry (without liquid lubrication) or those which have been impregnated with a lubricant. Dry bearings are made of a material such as a polymer or carbon-graphite which has a low friction coefficient, and they are generally used in low-load and low-speed applications. Semilubricated bearings are made of a porous material, usually metal, and are impregnated with a lubricant which resides within the pores. The lubricant, which could be oil or grease, cannot provide a complete fluid film, but usually acts as a boundary lubricant. Semilubricated bearings can carry greater loads at greater speeds than dry bearings, but not as high as either fluid film or rolling element bearings. The failure mechanism for both dry and semilubricated bearings is wear.

Rolling element bearings (see subsection below) have the advantage that rolling friction is lower than sliding friction. These bearings include rolling elements, either balls or rollers, between hardened and ground rings or plates. Their main advantage over fluid film bearings is that they have low friction both during start-up and at operating velocities, although the friction can be higher than that of fluid film bearings during steady state operation. Ball and roller bearings are most commonly lubricated by either oil or grease. In either case the lubricating film at the concentrated contacts between rolling elements and rings is very thin and the pressures in the film are very high; this is the condition known as elastohydrodynamic lubrication. Rolling element bearings fail by a number of mechanisms, often stemming from improper installation or use or from poor lubrication, but the overriding failure mechanism is rolling contact fatigue.

Each type of bearing has advantages and disadvantages, and these are summarized in Table 3.10.3. The Engineering Sciences Data Unit (ESDU) (1965; 1967) has developed some general guides to the selection of bearing type for different load and speed conditions, and those guides for journal and thrust bearing selection are given in Figures 3.10.2 and 3.10.3.

TABLE 3.10.3 Bearing Characteristics

	Fluid Film Bearings	Dry Bearings	Semilubricated	Rolling Element Bearings
Start-up friction coefficient	0.25	0.15	0.10	0.002
Running friction coefficient	0.001	0.10	0.05	0.001
Velocity limit	High	Low	Low	Medium
Load limit	High	Low	Low	High
Life limit	Unlimited	Wear	Wear	Fatigue
Lubrication requirements	High	None	Low/None	Low
High temperature limit	Lubricant	Material	Lubricant	Lubricant
Low temperature limit	Lubricant	None	None	Lubricant
Vacuum	Not applicable	Good	Lubricant	Lubricant
Damping capacity	High	Low	Low	Low
Noise	Low	Medium	Medium	High
Dirt/dust	Need Seals	Good	Fair	Need seals
Radial space requirement	Small	Small	Small	Large
Cost	High	Low	Low	Medium

Lubricant Properties

Petroleum Oils

The vast majority of lubricants in use today are mineral oils which are obtained through the distillation of crude petroleum. Mineral oils are composed primarily of three types of hydrocarbon structures: paraffinic, aromatic, and alicyclic (naphthenic). The molecular weights of the hydrocarbons range from about 250 for low-viscosity grades, up to nearly 1000 for more-viscous lubricants.

Mineral oils by themselves do not have all of the properties required of modern lubricants. For that reason, almost all current lubricants are fortified with a chemical additive package which consists of some of the following:

Oxidation inhibitors limit oxidation of hydrocarbon molecules by interrupting the hydroperoxide chain reaction.

Rust inhibitors are surface-active additives that preferentially adsorb on iron or steel surfaces and prevent their corrosion by moisture.

Antiwear and extreme pressure agents form low shear strength films on metallic surfaces which limit friction and wear, particularly in concentrated contacts.

Friction modifiers form adsorbed or chemisorbed surface films which are effective in reducing friction of bearings during low-speed operation (boundary lubrication regime).

Detergents and dispersants reduce deposits of oil-insoluble compounds (e.g., sludge) in internal combustion engines.

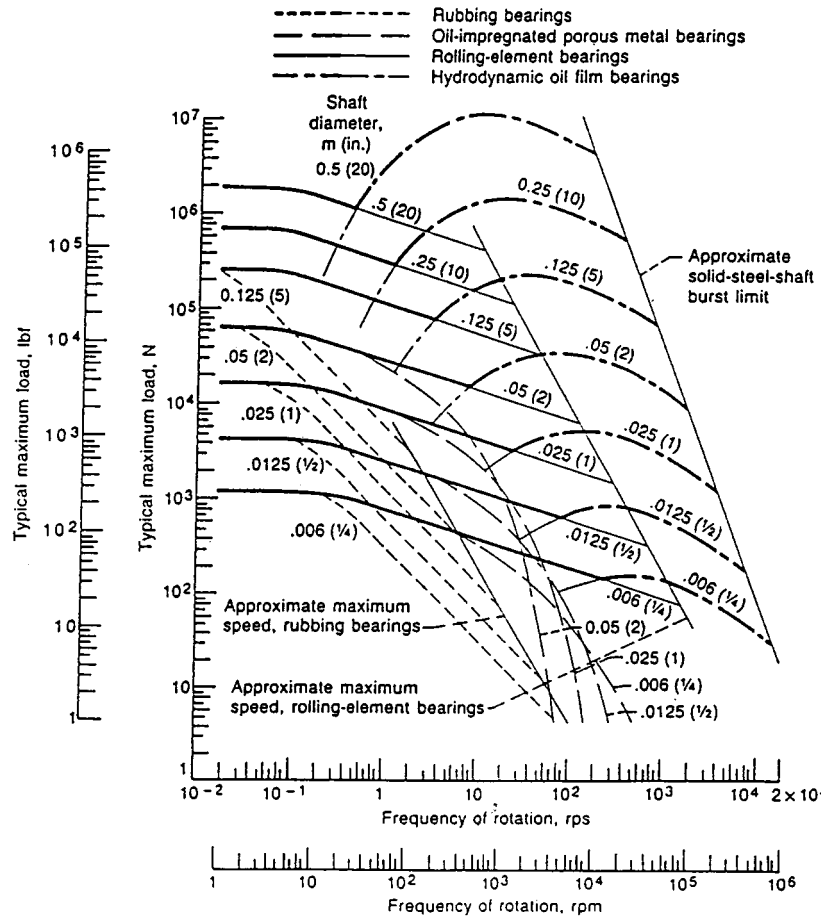


FIGURE 3.10.2 General guide to journal bearing–type selection. Except for rolling element bearings, curves are drawn for bearings with width/diameter = 1. A medium-viscosity mineral oil is assumed for hydrodynamic bearings. From ESDU, *General Guide to the Choice of Journal Bearing Type*, Item 67073, Institution of Mechanical Engineers, London, 1965. With permission.

Pour-point depressants lower the temperature at which petroleum oils become immobilized by crystallized wax.

Foam inhibitors are silicone polymers which enhance the separation of air bubbles from the oil.

Viscosity-index improvers are long-chain polymer molecules which reduce the effect of temperature on viscosity. They are used in multigrade lubricants.

Properties of Petroleum Oils

The lubricating oil property which is of most significance to bearing performance is viscosity. The absolute viscosity, designated as μ, could be given in SI units as pascal second (Pa · sec = N · sec/m^2) or centipoise (1 cP = 0.001 Pa · sec) or in English units as lb · sec/in^2 (or reyn). Kinematic viscosity, designated here as ν, is defined as absolute viscosity divided by density. It is given in SI units as m^2/sec or centistokes (1 cSt = 10^{-6} m^2/sec) and in English units as in.^{2}/sec.

Viscosity data in Table 3.10.4 are representative of typical petroleum "turbine" and "hydraulic" oils which are widely used in industry and closely correspond to properties of most other commercially available petroleum oils. Table 3.10.5 gives equivalent viscosity grades for common automotive (SAE), gear (SAE and AGMA), and reciprocating aircraft engine (SAE) oils (Booser, 1995). Equivalent ISO viscosity grades are listed for the single-graded SAE automotive oils such as SAE 10W and SAE 30.

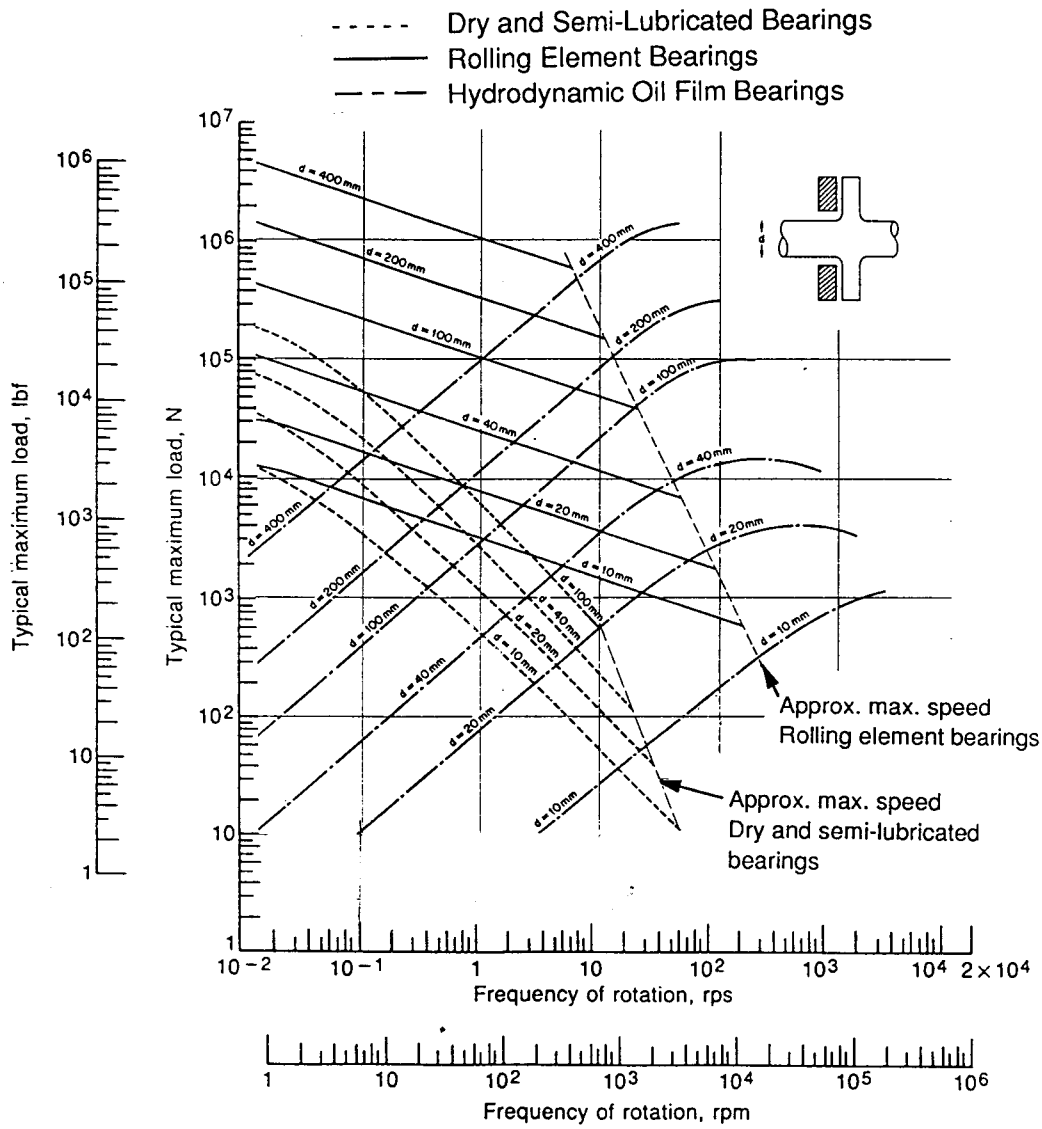


FIGURE 3.10.3 General guide to thrust bearing type selection. Except for rolling element bearings, curves are drawn for a ratio of inside diameter to outside diameter equal to 2 and for a nominal life of 10,000 hr. A medium-viscosity mineral oil is assumed for hydrodynamic bearings. (Based on ESDU, *General Guide to the Choice of Journal Bearing Type*, Item 67073, Institution of Mechanical Engineers, London, 1965, and Neale, M.J., *Bearings*, Butterworth-Heinemann, Oxford, 1993.)

For multigrade oils such as SAE 10W–30, however, the added viscosity-index improvers provide unique viscosity–temperature characteristics. Typical properties of a number of these multigrade SAE oils are included in Table 3.10.4.

ISO viscosity grade 32 and the equivalent SAE 10W are most widely used industrially. Lower-viscosity oils often introduce evaporation and leakage problems, along with diminished load capacity. Higher viscosity may lead to high temperature rise, unnecessary power loss, and start-up problems at low temperature. For low-speed machines, however, higher-viscosity oils ranging up to ISO 150, SAE 40 and sometimes higher are often used to obtain higher load capacity.

TABLE 3.10.4 Representative Oil Properties

	Viscosity				Density	
	Centistokes		10^{-6} reyns(lb·sec/in^2)		gm/cc	lb/in^3
	40°C	100°C	104°F	212°F	40°C	104°F
ISO Grade (Equivalent SAE)						
32 (10W)	32.0	5.36	3.98	0.64	0.857	0.0310
46 (20)	46.0	6.76	5.74	0.81	0.861	0.0311
68 (20W)	68.0	8.73	8.53	1.05	0.865	0.0313
100 (30)	100.0	11.4	12.60	1.38	0.869	0.0314
150 (40)	150.0	15.0	18.97	1.82	0.872	0.0315
220 (50)	220.0	19.4	27.91	2.36	0.875	0.0316
SAE Multigrade						
5W-30	64.2	11.0	8.15	0.99	0.860	0.0311
10W-30	69.0	11.0	8.81	1.08	0.865	0.0312
10W-40	93.5	14.3	11.9	1.45	0.865	0.0312
20W-50	165.5	18.7	21.3	2.74	0.872	0.0315

TABLE 3.10.5 Equivalent Viscosity Grades for Industrial Lubricants

ISO-VG	Viscosity, cSt (at 40°C)		SAE Crankcase	SAE Aircraft	SAE Gear	AGMA Gear Lube Grades	
Grade	Mimimum	Maximum	Oil Grades[a]	Oil Grades[a]	Lube Grades[a]	Regular	EP
2	1.98	2.42	—	—	—	—	—
3	2.88	3.52	—	—	—	—	—
5	4.14	5.06	—	—	—	—	—
7	6.12	7.48	—	—	—	—	—
10	9.00	11.0	—	—	—	—	—
15	13.5	16.5	—	—	—	—	—
22	19.8	24.2	5W	—	—	—	—
32	28.8	35.2	10W	—	—	—	—
46	41.4	50.6	15W	—	75W	1	
68	61.2	74.8	20W	—	—	2	2 EP
100	90.0	110	30	65	80W-90	3	3 EP
150	135	165	40	80	—	4	4 EP
220	198	242	50	100	90	5	5 EP
320	288	352	60	120	—	6	6 EP
460	414	506	—	—	85W-140	7 comp	7 EP
680	612	748	—	—	—	8 comp	8 EP
1000	900	1100	—	—	—	8A comp	8A EP
1500	1350	1650	—	—	250	—	—

[a] Comparisons are nominal since SAE grades are not specified at 40°C viscosity; VI of lubes could change some of the comparisons.

Oil viscosity decreases significantly with increasing temperature as shown in Fig. 3.10.4. While Figure 3.10.4 provides viscosity data suitable for most bearing calculations, oil suppliers commonly provide only the 40°C and 100°C values of kinematic viscosity in centistokes (mm^2/sec). The viscosity at other temperatures can be found by the following ASTM D341 equation relating kinematic viscosity ν in centistokes (mm^2/sec) to temperature T in degrees F:

$$\log\log(\nu + 0.7) = A - B\log(460 + T) \tag{3.10.2}$$

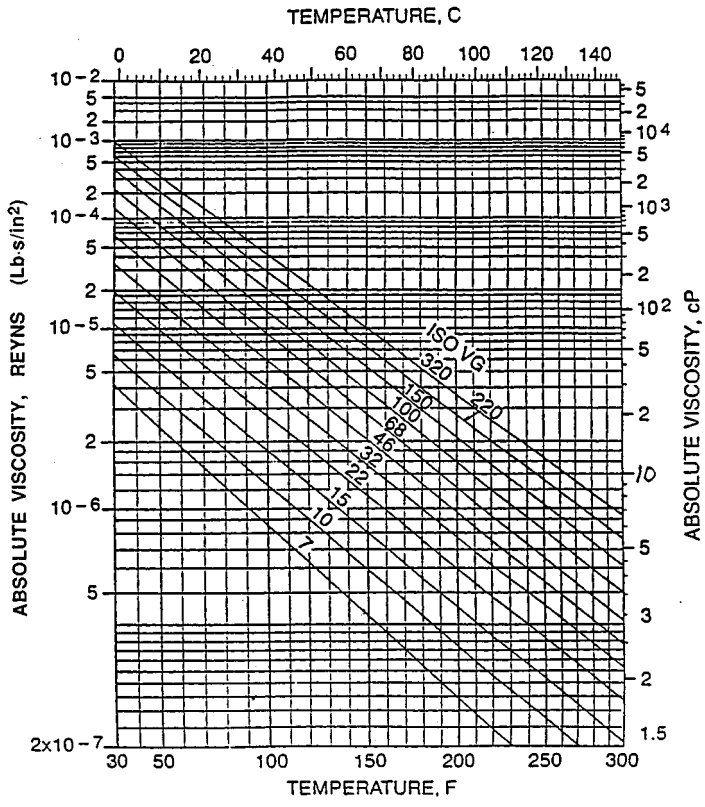


FIGURE 3.10.4 Viscosity–temperature chart for industrial petroleum oils (Ramondi and Szeri, 1984).

where A and B are constants for any particular (petroleum or synthetic) oil. For ISO VG-32 oil in Table 3.10.4, for example, A = 10.54805 and B = 3.76834 based on the 104 and 212°F (40 and 100°C) viscosities. This gives a viscosity of 10.78 cSt at a bearing operating temperature of 160°F.

Conversion of kinematic to absolute viscosity requires use of density, which also varies with temperature. A coefficient of expansion of 0.0004/°F is typical for most petroleum oils, so the variation of density with temperature is in accordance with the following relation:

$$\rho_T = \rho_{104}\left[1 - 0.0004(T - 104)\right] \tag{3.10.3}$$

For the ISO-VG-32 oil with its density of 0.857 g/cc at 100°F (40°C), this equation gives 0.838 g/cc at 160°F. Multiplying 10.78 cSt by 0.838 then gives an absolute viscosity of 9.03 cP. Multiplying by the conversion factor of 1.45×10^{-7} gives 1.31×10^{-6} lb·sec/in.^{2} (reyns).

Heat capacity C_p of petroleum oils used in bearing calculations is given by (Klaus and Tewksburg, 1984)

$$C_p = 3860 + 4.5(T) \quad \text{in.} \cdot \text{lb}/(\text{lb} \cdot °\text{F}) \tag{3.10.4}$$

The viscosity of petroleum oils is also affected significantly by pressure, and that increase can become important in concentrated contacts such as rolling element bearings where elastohydrodynamic lubrication occurs. The following relationship can be used for many oils to find the viscosity at elevated pressure:

$$\mu_p = \mu_o e^{\alpha p} \tag{3.10.5}$$

where μ_o is the viscosity at atmospheric pressure, μ_p is the viscosity at pressure p, and α is the pressure–viscosity exponent. The pressure–viscosity exponent can be obtained from lubricant suppliers.

Synthetic Oils

Synthetic oils of widely varying characteristics are finding increasing use for applications at extreme temperatures and for their unique physical and chemical characteristics. Table 3.10.6 gives a few representative examples of synthetic oils which are commercially available. Cost of the synthetics, ranging up to many times that of equivalent petroleum oils, is a common deterrent to their use where petroleum products give satisfactory performance

TABLE 3.10.6 Properties of Representative Synthetic Oils

Type	Viscosity, cSt at 100°C	Viscosity, cSt at 40°C	Viscosity, cSt at –54°C	Pour Point, °C	Flash Point, °C	Typical Uses
Synthetic hydrocarbons						
Mobil 1, 5W-30[a]	11	58	—	–54	221	Auto engines
SHC 824[a]	6.0	32	—	–54	249	Gas turbines
SHC 629[a]	19	141	—	–54	238	Gears
Organic esters						
MIL-L-7808	3.2	13	12,700	–62	232	Jet engines
MIL-L-23699	5.0	24	65,000	–56	260	Jet engines
Synesstic 68[b]	7.5	65	—	–34	266	Air compressors, hydraulics
Polyglycols						
LB-300-X[c]	11	60	—	–40	254	Rubber seals
50-HB-2000[c]	70	398	—	–32	226	Water solubility
Phosphates						
Fyrquel 150[d]	4.3	29	—	–24	236	Fire-resistant fluids for die casting,
Fyrquel 220[d]	5.0	44	—	–18	236	air compressors and hydraulic systems
Silicones						
SF-96 (50)	16	37	460	–54	316	Hydraulic and damping fluids
SF-95 (1000)	270	650	7,000	–48	316	Hydraulic and damping fluids
F-50	16	49	2,500	–74	288	Aircraft and missiles
Fluorochemicals						
Halocarbon 27[e]	3.7	30	—	–18	None	Oxygen compressors, liquid-oxygen
Krytox 103[f]	5.2	30	—	–45	None	systems

[a] Mobil Oil Corp.
[b] Exxon Corp.
[c] Union Carbide Chemials Co.
[d] Akzo Chemicals
[e] Halocarbon Products Corp.
[f] DuPont Co.

Greases

Grease is essentially a suspension of oil in a thickening agent, along with appropriate additives. The oil generally makes up between 75 and 90% of the weight of a grease, and it is held in place by the gel structure of the thickener to carry out its lubricating function. The simplicity of the lubricant supply system, ease of sealing, and corrosion protection make grease the first choice for many ball-and-roller bearings, small gear drives, and slow-speed sliding applications (Booser, 1995). Consistencies of greases vary from soap-thickened oils that are fluid at room temperature to hard brick-types that may be cut with a knife. Greases of NLGI Grade 2 stiffness (ASTM D217) are most common. Softer greases down to grade 000 provide easier feeding to multiple-row roller bearings and gear mechanisms. Stiffer Grade 3 is used in some prepacked ball bearings to avoid mechanical churning as the seals hold the grease in close proximity with the rotating balls.

Petroleum oils are used in most greases; the oils generally are in the SAE 30 range, with a viscosity of about 100 to 130 cSt at 40°C. Lower-viscosity oil grades are used for some high-speed applications and for temperatures below about –20°C. Higher-viscosity oils are used in greases for high loads and low speeds. Synthetic oils are used only when their higher cost is justified by the need for special properties, such as capability for operation below –20°C or above 125 to 150°C.

The most common gelling agents are the fatty acid metallic soaps of lithium, calcium, sodium, or aluminum in concentrations of 8 to 25%. Of these, the most popular is lithium 12-hydroxystearate; greases based on this lithium thickener are suitable for use at temperatures up to 110°C, where some lithium soaps undergo a phase change. Greases based on calcium or aluminum soaps generally have an upper temperature limit of 65 to 80°C, but this limit can be significantly raised to the 120 to 125°C range through new complex soap formulations. Calcium-complex soaps with improved high-temperature stability, for instance, are prepared by reacting both a high-molecular-weight fatty acid (e.g., stearic acid) and a low-molecular-weight fatty acid (acetic acid) with calcium hydroxide dispersed in mineral oil.

Inorganic thickeners, such as fine particles of bentonite clay, are inexpensively applied by simple mixing with oil to provide nonmelting greases for use up to about 140°C. Polyurea nonmelting organic powders are used in premium petroleum greases for applications up to about 150 to 170°C.

Additives, such as those mentioned in the subsection on petroleum oils, are added to grease to improve oxidation resistance, rust protection, or extreme pressure properties. Because of the incompatibility of oils, thickeners, and additives, greases of different types should be mixed only with caution.

Solid Lubricants

Solid lubricants provide thin solid films on sliding or rolling/sliding surfaces to reduce friction and wear. They are particularly useful for applications involving high operating temperatures, vacuum, nuclear radiation, or other environments which limit the use of oils or greases. Solid lubricant films do not prevent moving surfaces from contacting each other, so they cannot eliminate wear and their lives are limited by wear. The properties of some of the most common solid lubricants are given in Table 3.10.7.

TABLE 3.10.7 Properties of Selected Solid Lubricants

Material	Acceptable Usage Temperature, °C				Average Friction Coefficient, *f*		Remarks
	Minimum		Maximum				
	In Air	In N_2 or Vacuum	In Air	In N_2 or Vacuum	In Air	In N_2 or Vacuum	
Molybdenum disulfide, MoS_2	–240	–240	370	820	0.10–0.25	0.05–0.10	Low *f*, carries high load, good overall lubricant, can promote metal corrosion
Graphite	–240	—	540	Unstable in vacuum	0.10–0.30	0.02–0.45	Low *f* and high load capacity in air, high *f* and wear in vacuum, conducts electricity
PTFE	–70	–70	290	290	0.02–0.15	0.02–0.15	Lowest *f* of solid lubricants, load capacity moderate and decreases at elevated temperature
Calcium flouride–barium fluoride eutectic, CaF_2–BaF_2	430	430	820	820	0.10–0.25 above 540°C 0.25–0.40 below 540°C	Same as in air	Can be used at higher temperature than other solid lubricants, high *f* below 540°C

Modified from Booser, E.R., in *Encyclopedia of Chemical Technology*, 4th ed., Vol. 15, John Wiley & Sons, New York, 1995, 463–517.

The most important inorganic solid lubricants are layer–lattice solids such as molybdenum disulfide (MoS_2) and graphite. These materials are characterized by strong covalent or ionic bonding between atoms in individual layers, but relatively weak van der Waals bonds between layers, enabling the layers

to slide easily relative to one another. Graphite is a very effective lubricant film when moisture or water vapor is present, since adsorbed water vapor lubricates the sliding layers, but it has poor friction properties in vacuum or other low-humidity applications. Molybdenum disulfide does not require the presence of adsorbed water vapor, so it is widely used in vacuum or space applications.

The most commonly used organic solid lubricant is polytetrafluoroethylene (PTFE) which can be used either as a fused surface coating or as a self-lubricating material (see subsection on plastics). Its low friction is attributed to the smooth profile of the PTFE molecule. The chemical inertness of PTFE makes it attractive for chemical and food-processing applications.

New ceramic-based solid lubricants have been developed for high-temperature applications, such as heat engines or space uses. One of the most promising of these is a calcium fluoride–barium fluoride eutectic, which can be used at temperatures exceeding 800°C.

Fluid Film Bearings

Journal Bearings

A journal bearing consists of an approximately cylindrical bearing body or sleeve around a rotating cylindrical shaft. In general, journal bearings are found in motors, pumps, generators, appliances, and internal combustion engines in which a fluid lubricant is used; and in smaller mechanisms such as switches, clocks, small motors, and circuit breakers in which a solid lubricant such as graphite, grease, or certain plastics serves to reduce friction. Air (gas) bearings are designed to utilize both fluid mechanics principles when operating and solid lubricant–surfaced materials for start, stop, and emergency operations.

A hydrodynamic journal bearing maintains separation of shaft from bearing because the lubricant viscosity and the speed of the shaft create pressure in the converging portion of the fluid film which carries load. The governing equations were first developed by Reynolds (1886). Their solution has led to numerous computer solutions, including those used for this section.

Journal Bearing Design. Figure 3.10.5 shows schematics of frequently used types of journal bearing in which one or more lobes of cylindrical shape are positioned around the shaft, their axis being assumed parallel to the shaft axis. The features of each design and applications where it is often found are listed in Table 3.10.8.

Noncontact journal bearings are designed to assure a continuous supply of lubricant to the load-carrying section, and the bearing grooves in Figure 3.10.5 are designed for that purpose. Oil must be resupplied to the bearing because of the continuous loss of lubricant as it is forced from the bearing by the load-carrying pressures generated within it. The subsection on lubricant supply methods describes some of the many systems designed to assure this supply and to cool the lubricant at the same time.

Controlling Variables. Definitions of the variables involved in journal bearing analysis are contained in Table 3.10.9. Because of the large range of many variables, nondimensional quantities are often used which are independent of the dimensional unit system involved. Examples are given in the English system unless otherwise stated.

Calculating Bearing Performance. Journal bearing performance can be calculated directly from dedicated computer codes which take account of load, speed, oil type, and delivery system, as well as bearing dimensions. This subsection presents two approximate solutions: a simple thermal approach and a set of interpolation tables based on computer solutions.

Thermal Approach. It is assumed that the bearing is operating at a constant but elevated temperature. A predicted operating temperature can then be found as a function of an assumed viscosity. A solution is found when the assumed viscosity equals the lubricant viscosity at that temperature. Three equations are used for this method. For radial loads, the power dissipation is

$$H_p = j\pi^3 \mu \left(N/60\right)^2 D^3 L/C \quad \text{in.} \cdot \text{lb/sec} \tag{3.10.6}$$

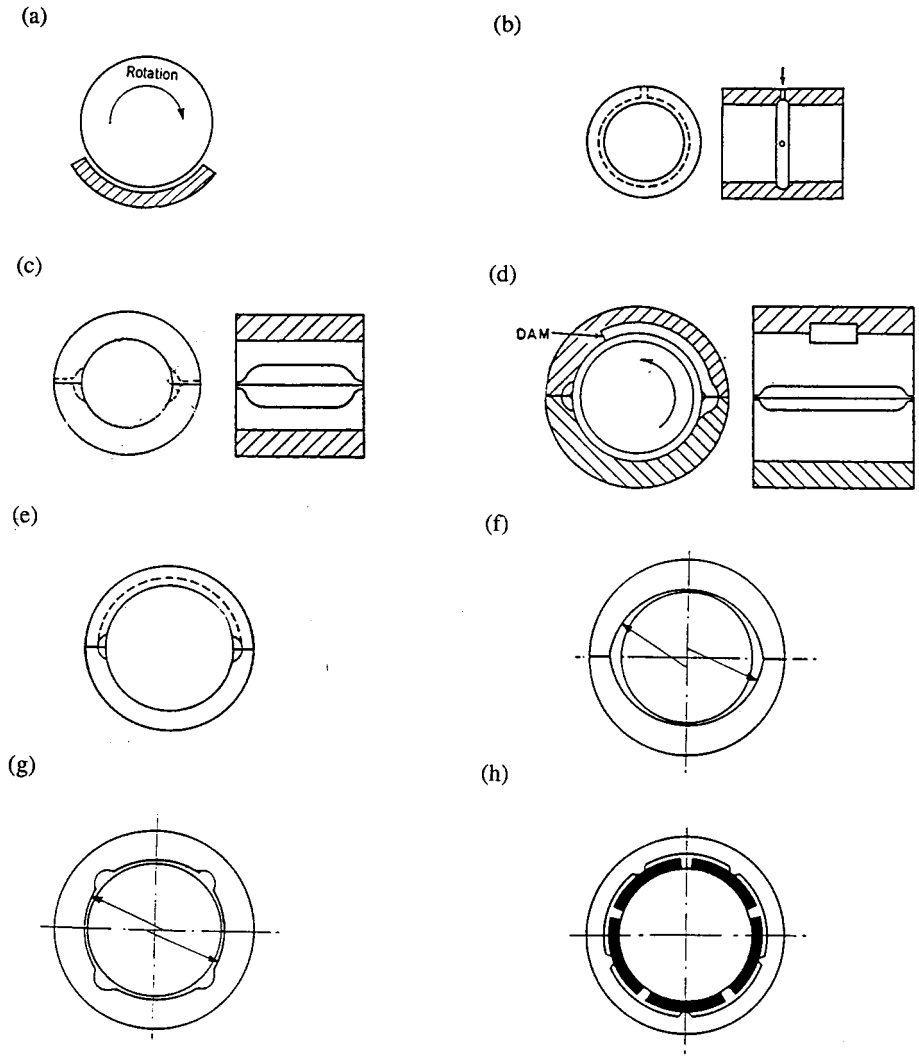


FIGURE 3.10.5 Types of pressure-fed journal bearings: (a) Partial arc. (b) Circumferential groove. (c) Cylindrical bearing–axial groove. (d) Pressure dam. (e) Cylindrical overshot. (f) Elliptical. (g) Multilobe. (h) Tilting pad.

TABLE 3.10.8 Journal Bearing Designs

Type	Typical Loading	Applications
Partial arc	Unidirectional load	Shaft guides, dampers
Circumferential groove	Variable load direction	Internal combustion engines
Axial groove types		
Cylindrical	Medium to heavy unidirectional load	General machinery
Pressure dam	Light loads, unidirectional	High-speed turbines, compressors
Overshot	Light loads, unidirectional	Steam turbines
Multilobe	Light loads, variable direction	Gearing, compressors
Preloaded	Light loads, unidirectional	Minimize vibration
Tilting pad	Moderate variable loads	Minimize vibration

where $j = 1$ for a shaft-centered design. The lubricant flow rate is

TABLE 3.10.9 Journal Bearing Parameters

B	Bearing damping coefficient	lb/in./sec.
C	Radial clearance	in.
C_α	Adiabatic constant	—
C_p	Heat capacity	in.·lb/lb°F
D	Diameter	in.
H_p	Power loss	in.·lb/sec
K	Bearing stiffness	lb/in.
L	Bearing length	in.
N	Shaft velocity	rpm
Q	Lubricant flow rate	$in.^3$/sec
R	Shaft radius	in.
R_e	Reynolds number	—
T_e	Entrance temperature	°F
T_f	Feed temperature	°F
T_q	Torque	in.·lb
ΔT_b	Temperature rise coefficient, bottom half	°F
ΔT_t	Temperature rise coefficient, top half	°F
U	Velocity	in./sec
W	Load	lb
e	Shaft eccentricity	in.
h	Film thickness	in.
j	Ratio: power loss/shaft-centered loss	—
p	Pressure	psi
q	Flow coefficient	—
w	Load coefficient	—
x	Coordinate in direction of load	in.
y	Coordinate orthogonal to load	in.
β	Exponential temperature coefficient of viscosity	—
ε	Shaft eccentricity, nondimensional	—
γ	Angular extent of film	—
ϕ	Attitude angle	—
ρ	Density	$lb/in.^3$
μ	Viscosity	$lb\cdot sec/in.^2$
ω	Shaft velocity	rad/sec
θ	Angle in direction of rotation, from bottom dead center	—
Φ	Energy dissipation	in.·lb/sec

$$Q = Q_o + qCR^2\omega/2 \quad in.^3/sec \tag{3.10.7}$$

where q is the proportion of side flow to circulating flow, and the zero speed flow, Q_o (in^3/sec), represents other flows such as from the ends of the feed grooves which are not related to the load-carrying film itself. Q_o can usually be neglected for rough estimation, and this approximation is useful for eccentricities as high as 0.7. Note that both q and j are functions of specific design as well as load and speed. The average operating temperature for a given viscosity is

$$T_2 = T_f + \frac{\left(H_p - \Phi\right)}{\left(\rho C_p Q\right)} \quad °F \tag{3.10.8}$$

where T_f is the feed temperature and Φ is the energy loss due to conduction and radiation. For diameters of 2″ or more, Φ can usually be assumed to be 0. Plotting T_2 vs. viscosity for several values of μ on a plot of the viscosity vs. T for the lubricant shows the operating temperature for the bearing as the intersection.

Flow Dynamics Solution. A more general solution for journal bearing performance is based on prediction of flow characteristics in the bearing, and of physical behavior of the bearing based on the Reynolds equation. A common two-pad journal bearing with pressurized oil feed will be used to provide specific design information. The Reynolds equation is the differential equation expressing the physical behavior of the fluid film between shaft and bearing, as written for cylindrical bearings:

$$1/R^2\left[\partial/\partial\Theta\left(h^3/\mu\right)\partial p/\partial\Theta\right]+\partial/\partial z\left(h^3/\mu\right)\partial p/\partial z = 6\left(U/R\right)\partial h/\partial\Theta \tag{3.10.9}$$

where z is the axial coordinate and Θ is the angular coordinate.

Bearing Configuration. A cross section through a common type of two-pad cylindrical bearing is shown in Figure 3.10.6. Two pads having a radius of $R + C$ and an angular extent of 150°, and with load applied vertically to the lower half through a shaft, make up the bearing. Lubricant is admitted under moderate feed pressure to the two 30° grooves formed at the split as shown in Figure 3.10.6. The shaft rotates counterclockwise, and lubricant pressures are balanced when the shaft center is displaced down and to the right.

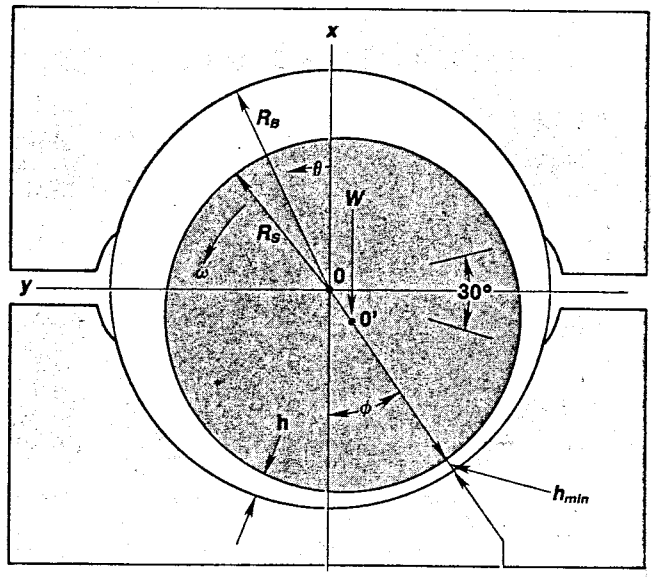


FIGURE 3.10.6 Geometry of split cylindrical journal bearing.

Lubricant Properties. Pressures in the lubricant film are generated as a function of the local shear rates and the local viscosity as described by the Reynolds equation. The local temperature rise is given by the local energy dissipation divided by the local flow rate:

$$\Delta T = \left[2\mu\omega R^2\Delta\theta\right]/\left[h^2\rho C_p\right] \tag{3.10.10}$$

As an alternative to Equation 3.10.2, an exponential relation between viscosity and temperature is used:

$$\mu = \mu_0 e^{-\beta\left(T-T_0\right)} \tag{3.10.11}$$

Assuming an ISO 32 oil, viscosity μ_0 at 104°F is 3.98 reyns, density is 0.0310, and β is 0.0170 for the range from 104 to 212°F. The value for β may be determined from Figure 3.10.4. A nondimensional

coefficient, C_α, the *adiabatic coefficient*, is used as an indicator of the severity of thermal heating in a design. It contains the several factors responsible for temperature rise.

$$C_\alpha = 2\mu_f \beta\omega\left(R/C\right)^2 / \rho C_p \tag{3.10.12}$$

Bearing Performance Tables. Using computer-generated solutions of the Reynolds equation for the pressure field over the bearing surface, the relevant performance properties of the bearing/lubricant system can be determined. A number of programs are commercially available for this purpose. In order to illustrate how the behavior of a journal bearing can be determined, a two-pad split cylindrical bearing with an *L*/*D* of 0.5 has been selected, and a proprietary computer code has been used to develop a group of performance quantities for such a bearing. The code accounts for the internal lubricant circulation, including mixing in the inlet groove of feed lubricant with warm lubricant from the upper half, resulting in the viscosity μ_f. The primary characteristics of the bearing are load, stiffness, and damping. Each of these factors is presented in nondimensional terms in Table 3.10.10, and the corresponding dimensional and other quantities are given as follows:

Film thickness, in. $$h = C\left(1 + e\cos\Theta\right) \tag{3.10.13a}$$

Shaft eccentricity, in. $$e = \varepsilon C \tag{3.10.13b}$$

Load, lb $$W = w\left[6\mu\omega\left(R/C\right)^2 DL\right] \tag{3.10.13c}$$

Flow, $\text{in.}^3/\text{sec}$ $$Q = q\left(CR^2\,\omega/2\right) \tag{3.10.13d}$$

Power loss, in. · lb/sec $$H_p = j\left(2\pi\mu\omega^2 R^3\, L/C\right) \tag{3.10.13e}$$

Stiffness, lb/in. $$S_{xx} = K_{xx} 6\mu\omega R\left(R/C\right)^2 \tag{3.10.13f}$$

Damping, lb · sec/in. $$D_{xx} = B_{xx} 12\mu R\left(R.C\right)^3 \tag{3.10.13g}$$

The axial length/diameter ratio also influences the performance of a journal bearing. To illustrate this, Table 3.10.11 presents the performance of longer bearings (*L*/*D* = 0.75 and 1.00) for comparison to the more common *L*/*D* = 0.5 results in Table 3.10.10.

Comparing Tables 3.10.10 and 3.10.11, the use of longer bearings has several important effects on operation and performance. Comparing key variables, the effects at an eccentricity of ratio of 0.7 are as follows:

Variable	L/D = 0.5	L/D = 0.75	L/D = 1.00
Load, w	0.28	0.69	1.21
Flow, q	0.69	0.82	0.88
Attitude angle, ϕ	36.4	36.1	35.8
Power ratio, j	1.00	1.15	1.17
Stiffness, K_{xx}	1.38	3.06	5.03
Damping, B_{xx}	0.99	2.52	4.31

Effect of Turbulence. Turbulence is a mixing phenomenon that occurs in larger high-speed bearings. When this behavior occurs, the simple viscous drag behavior that has been assumed in the preceding

TABLE 3.10.10 Performance of *L/D* = 0.5 Bearing

Part 1: C_α = 0.0

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	66.5	48.01	36.44	30.07	22.18	16.46
w	0.0246	0.0997	0.2784	0.5649	1.6674	4.4065
q	0.3037	0.6014	0.6927	0.6946	0.6487	0.588
j	0.7779	0.8534	1.1005	1.3905	2.008	3.084
ΔT_b	0	0	0	0	0	0
ΔT_t	0	0	0	0	0	0
K_{xx}	0.041	0.2805	1.379	4.063	22.67	—
K_{xy}	0.1465	0.3745	1.063	2.476	9.390	34.47
K_{yx}	–0.055	–0.072	0.0063	0.193	1.710	8.002
K_{yy}	0.046	0.170	0.4235	0.883	2.622	7.555
B_{xx}	0.142	0.352	0.989	2.311	8.707	32.30
B_{xy}, B_{yx}	0.023	0.094	0.236	0.522	1.547	4.706
B_{yy}	0.056	0.105	0.174	0.302	0.630	1.390

Part 2: C_α = 0.1

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	69.9	50.2	38.7	32.35	24.83	19.8
w	0.022	0.087	0.233	0.451	1.184	2.621
q	0.312	0.620	0.721	0.728	0.692	0.642
j	0.686	0.723	0.863	0.997	1.253	1.545
ΔT_b	0.274	0.403	0.642	0.907	1.519	2.346
ΔT_t	0.243	0.211	0.183	0.168	0.151	0.142
K_{xx}	0.038	0.2365	1.041	2.935	13.66	50.44
K_{xy}	0.126	0.3135	0.870	1.851	3.078	18.30
K_{yx}	–0.047	–0.061	–0.021	0.139	1.068	3.961
K_{yy}	0.037	0.140	0.3585	0.669	1.784	4.327
B_{xx}	0.121	0.286	0.776	1.592	4.97	14.00
B_{xy}, B_{yx}	0.016	0.071	0.195	0.341	0.850	2.10
B_{yy}	0.047	0.086	0.156	0.216	0.394	0.757

Part 3: C_α = 0.2

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	73.4	52.2	40.8	34.55	27.23	22.5
w	0.020	0.077	0.198	0.368	0.890	1.779
q	0.320	0.639	0.747	0.759	0.730	0.760
j	0.613	0.628	0.712	0.791	0.933	1.092
ΔT_b	0.520	0.7520	1.162	1.594	2.521	3.651
ΔT_t	0.472	0.415	0.363	0.333	0.301	0.284
K_{xx}	0.035	0.1925	0.830	2.156	8.86	28.6
K_{xy}	0.11	0.272	0.704	1.477	4.515	11.72
K_{yx}	–0.041	–0.062	–0.018	0.074	0.640	2.371
K_{yy}	0.029	0.125	0.2895	0.551	1.375	2.932
B_{xx}	0.104	0.242	0.596	1.21	3.90	7.830
B_{xy}, B_{yx}	0.011	0.061	0.140	0.212	0.634	1.21
B_{yy}	0.040	0.080	0.121	0.187	0.326	0.501

TABLE 3.10.10 (continued) Performance of *L/D* = 0.5 Bearing

Part 4 C_α = 0.4

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	80.2	56.0	44.5	38.4	31.3	26.7
w	0.016	0.061	0.148	0.260	0.562	1.000
q	0.331	0.6720	0.795	0.815	0.797	0.760
j	0.504	0.498	0.534	0.570	0.637	0.716
ΔT_b	0.946	1.33	1.97	2.61	3.87	5.26
ΔT_t	0.898	0.801	0.712	0.658	0.597	0.562
K_{xx}	0.029	0.137	0.538	1.295	4.56	12.6
K_{xy}	0.085	0.206	0.503	0.985	2.67	6.17
K_{yx}	–0.0315	–0.0548	0.0298	0.0233	0.321	1.136
K_{yy}	0.019	0.094	0.214	0.382	0.860	1.68
B_{xx}	0.079	0.175	0.397	0.734	1.75	3.44
B_{xy}, B_{yx}	0.0041	0.042	0.094	0.166	0.329	0.55
B_{yy}	0.030	0.064	0.092	0.131	0.120	0.276

TABLE 3.10.11 Performance of Long Bearings

Part 1: *L/D* = 0.75, C_α = 0.0

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	64.74	46.54	36.13	30.17	22.64	17.03
w	0.0705	0.2714	0.6947	1.311	3.440	8.241
q	0.392	0.738	0.825	0.811	0.737	0.6545
j	0.777	0.871	1.145	1.450	2.184	3.233
ΔT_b	0	0	0	0	0	0
ΔT_t	0	0	0	0	0	0
K_{xx}	0.121	0.706	3.065	8.506	41.5	—
K_{xy}	0.418	0.992	2.517	5.228	18.1	59.0
K_{yx}	–0.123	–0.189	0.052	0.404	3.18	16.2
K_{yy}	0.113	0.429	1.012	1.891	5.33	13.49
B_{xx}	0.423	0.982	2.52	5.16	17.7	54.4
B_{xy}, B_{yx}	0.057	0.249	0.609	1.10	3.24	7.58
B_{yy}	0.127	0.263	0.444	0.641	1.35	2.32

Part 2: *L/D* = 1.00, C_α = 0.00

ε	0.2	0.5	0.7	0.8	0.9	0.95
ϕ	63.2	45.3	35.8	30.3	22.9	17.4
w	0.138	0.506	1.214	2.18	5.34	12.15
q	0.444	0.800	0.879	0.856	0.769	0.679
j	0.782	0.886	1.174	1.768	2.250	3.323
ΔT_b	0	0	0	0	0	0
ΔT_t	0	0	0	0	0	0
K_{xx}	0.234	1.254	5.026	13.24	60.9	—
K_{xy}	0.818	1.795	4.142	8.12	26.8	83.5
K_{yx}	–0.201	–0.313	–0.075	0.671	4.96	24.9
K_{yy}	0.198	0.732	1.64	2.95	8.04	19.5
B_{xx}	0.82	1.87	4.31	8.27	26.5	75.9
B_{xy}, B_{yx}	0.10	0.45	0.97	1.68	4.78	10.36
B_{yy}	0.21	0.46	0.70	0.98	2.02	3.24

subsections is broken up by numerous eddies which increase the drag. The Reynolds number is a nondimensional quantity that expresses this factor:

$$R_e = hU\rho/\mu \tag{3.10.14}$$

where h is the local film thickness, U is the relative velocity of one surface with respect to the other, ρ is the fluid density, and μ is the local viscosity.

The influence of turbulence on an $L/D = 0.5$ bearing is shown in Table 3.10.12. Examination of Table 3.10.12 shows that the principal effects of operation in the turbulent regime with a Reynolds number above about 1000 are in the greater power required (j) and the maximum bearing temperature. Load capacity and flow are only moderately affected.

TABLE 3.10.12. Influence of Turbulence ($\varepsilon = 0.7$, $C_\alpha = 0.2$, arc = 150°)

R_e	0	1000	2000	4000
ϕ	40.8	43.8	46.4	49.2
w	0.198	0.171	0.197	0.221
q	0.747	0.809	0.862	0.914
j	0.712	0.983	1.459	2.124
ΔT_b	1.162	0.585	0.918	1.404
K_{xx}	0.830	0.627	0.634	0.647
K_{xy}	0.704	0.575	0.577	0.645
K_{yx}	–0.018	–0.034	–0.047	–0.078
K_{yy}	0.289	0.257	0.282	0.330
B_{xx}	0.596	0.483	0.513	0.534
B_{xy}, B_{yx}	0.140	0.125	0.132	0.136
B_{yy}	0.121	—	—	0.104

Example Calculation. The problem is to design a two-pad cylindrical bearing for support of a rotor having a bearing load of 8000 lb, a shaft diameter of 6 in., and a speed of 3600 rpm. Assume use of ISO VG-32 oil fed at a temperature of 120°F. Compute operating characteristics for a 3-in.-long bearing. Assume a radial clearance of 0.0044 in.

Feed viscosity, $\mu_f = 3.98 \times 10^{-6}\, e^{-0.00170(120-104)} = 3.03 \times 10^{-6}$ reyn
Angular velocity, $\omega = 3600 \times 2\pi/60 = 377$ rad/sec
Adiabatic coefficient: $C_\alpha = 2 \times 3.03 \times 10^{-6} \times 0.0170 \times 377 \times (3/0.0044)^2/0.0310/4320 = 0.1345$
Load coefficient (from Equation 3.10.13c: $w = 8000/[6 \times 3.03 \times 10^{-6} \times 377 \times 3 \times 6 \times (33/0.0044)^2] = 0.139$

The desired solution lies between Part 2 and Part 3 of Table 3.10.10. By using linear interpolation between the tabulated values for C_α of 0.1 and 0.2, and values of ε of 0.7 and 0.8, an approximate operating point of $C_\alpha = 0.1345$ yields the following coefficients: $\varepsilon = 0.729$, $w = 0.279$, $q = 0.733$, $j = 0.860$, and $\Delta T_b = 0.915$.

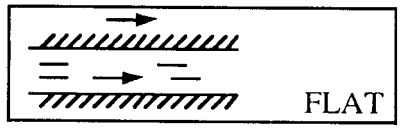
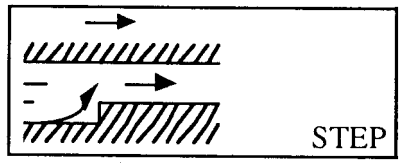
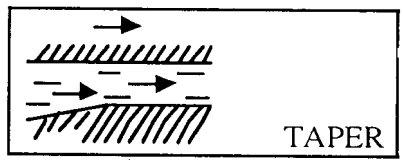
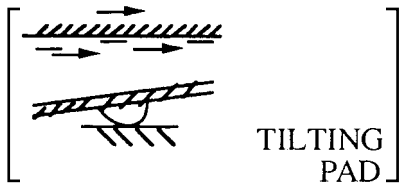
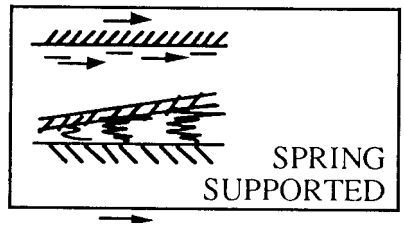
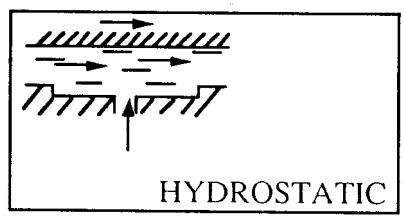
By using Equations 3.10.13, the dimensional operating results are:

Shaft velocity: $\omega = 3600 \times 2\pi/60 = 377$ rad/sec
Flow: $Q = 0.733 \times 0.0044 \times 3^2 \times 377 = 5.47$ in^3/sec
Power loss: $H_p = 0.860 \times 2\pi \times 3.03 \times 10^{-6} \times 377^2 \times 3^3 \times 3/0.0044 = 42.8$ in.·lb/sec
Oil temperature: $T_b = 120 + 0.915/0.0170 = 174$°F

Thrust Bearings

Types of Thrust Bearings. Oil film thrust bearings range from coin-size flat washers to sophisticated assemblies many feet in diameter. Of the six common types of thrust bearings shown in Table 3.10.13, the first five are hydrodynamic. As with journal bearings, each of these generates oil film pressure when

TABLE 3.10.13 Common Thrust Bearings and Their Range of Application

Type	O.D., in.	Unit Load, psi
	0.5–20	20–100
	0.5–10	100–300
	2–35	150–300
	4–120	250–700
	50–120	350–700
	3–50	500–3000

Source: Booser, E.R. and Wilcock, D.F., *Machine Design*, June 20, 1991, 69–72. With permission.

a rotating thrust face pumps oil by shear into a zone of reduced downstream clearance. When thrust load increases, film thickness drops until a new balance is reached between inflow and outflow, raising pressure until the higher bearing load is balanced. The hydrostatic bearing uses a separate oil pump to supply the pressurized flow.

Flat-land bearings, simplest to fabricate and least costly, are the first choice for simple positioning of a rotor and for light loads in electric motors, appliances, pumps, crankshafts, and other machinery. They carry less load than the other types because flat parallel planes do not directly provide the required pumping action. Instead, their action depends on thermal expansion of the oil and warping of the bearing

material induced by heating from the passing oil film. The resulting slight oil wedge then gives a load rating of about 10 to 20% of that for the other types.

Step bearings also offer relatively simple design. With a coined or etched step, they lend themselves to mass production as small-size bearings and thrust washers. Step height for optimum load capacity approximately equals the minimum film thickness, often 0.001 in. or less. Circumferential length beyond the step is ideally 45% of the total bearing segment (Wilcock and Booser, 1956). Step thrust bearings are well suited for low-viscosity fluids such as water, gasoline, fuels, and solvents. Minimum film thickness in these applications is so small that features such as pivots and tapers are usually impractical. Step height must be small enough for good load capacity, yet large enough to accommodate some wear without becoming worn away. Step erosion by contaminants is sometimes a problem.

Tapered-land bearings provide reliable, compact designs for mid- to large-size high-speed machines such as turbines, compressors, and pumps. Taper height normally should be about one to three times the minimum film thickness. For greater load capacity and to minimize wear during starting, stopping, and at low speeds, a flat land is commonly machined at the trailing edge to occupy from 10 to 30% of the circumferential length of each segment. Because operation of these bearings is sensitive to load, speed, and lubricant viscosity, they are typically designed for the rated range of operating conditions for specific machines.

Tilting-pad thrust bearings are used increasingly in turbines, compressors, pumps, and marine drives in much the same range of applications as tapered-land designs. They usually have a central supporting pivot for each of their three to ten bearing segments. Each of these thrust pad segments is free to adjust its position to form a nearly optimum oil wedge for widely varying loads, speeds, and lubricants, and with rotation in both directions. A secondary leveling linkage system is commonly introduced to support the individual pads; this provides a further advantage over tapered-land designs by accommodating some misalignment. Off-the-shelf units are available to match rotor shaft diameters from about 2 to 12 in., and custom designs range up to 120 to 170 in. in outside diameter. Recent trends to increase load capacity have led to offsetting pivots from the circumferential midpoint of a pad to about 60% beyond the leading edge, to substituting copper for steel as the backing for a tin babbitt bearing surface, and to nonflooded lubrication to minimize parasitic power loss from useless churning of oil passing through the bearing housing.

Springs or other flexible supports for thrust segments are employed for bearings ranging up to 10 ft or more in outside diameter and carrying millions of pounds of thrust. This flexible mounting avoids the high load concentration encountered by pivots in supporting large tilting-pads. Rubber backing can provide this flexible mounting for smaller thrust pads.

Hydrostatic thrust bearings are used where sufficient load support cannot be generated by oil film action within the bearing itself. This may be the case with low-viscosity fluids, or for load support on an oil film at standstill and very low speeds. The fluid is first pressurized by an external pump and then introduced into pockets in the bearing surface to float the load. A compact hydrostatic thrust bearing can sometimes be constructed with a single pocket at the end of a rotor. Larger bearings usually use three or more pressurized pockets to resist misalignment or to support off-center loads. Hydraulic flow resistance in the supply line to each pocket, or constant flow to each pocket (as with ganged gear pumps) then provides any asymmetric pressure distribution needed to support an off-center load. Bearing unit load is commonly limited to about 0.5 (0.75 with fixed flow systems) times the hydrostatic fluid supply pressure—up to 5000 psi with conventional lubricating oils.

Design Factors for Thrust Bearings. In preliminary sizing, the inside diameter d of a thrust bearing is made sufficiently larger than the shaft to allow for assembly, and to provide for any required oil flow to the thrust bearing inside diameter (see Figure 3.10.7). This clearance typically ranges from about $^1/_8$ in. for a 2-in. shaft to $^1/_2$ in. for a 10-in. shaft. Bearing outside diameter D is then set to provide bearing

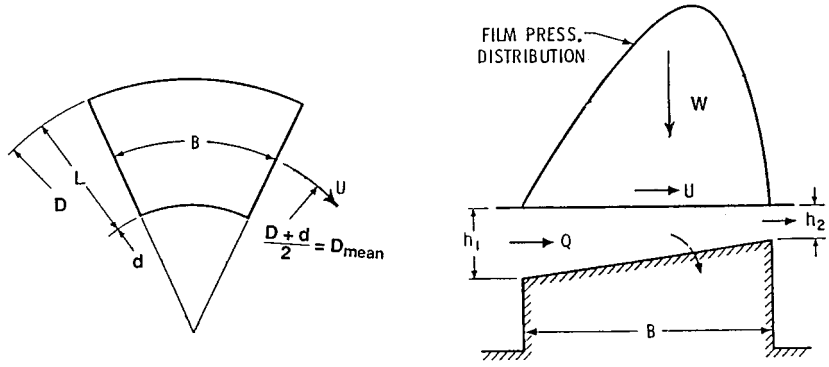


FIGURE 3.10.7 Sector for tapered land thrust bearing.

area sufficient to support total thrust load W (lb or N) with an average loading P (psi or N/m^2), using typical values from Table 3.10.13:

$$D = \left(\frac{4W}{\pi k_g P} + d^2 \right)^{0.5} \tag{3.10.15}$$

where k_g (typically 0.80 to 0.85) is the fraction of area between d and D not occupied by oil-distributing grooves. This bearing area is then divided by radial oil-feed groove passages, usually into "square" sectors with circumferential breadth B at their mean diameter equal to their radial length L.

While square pads usually produce optimum oil film performance, other proportions may be advantageous. With very large bearing diameters, for instance, square sectors may involve such a long circumferential dimension that the oil film overheats before reaching the next oil-feed groove. With a radially narrow thrust face, on the other hand, square sectors may become so numerous as to introduce excessive oil groove area, and their short circumferential length would interfere with hydrodynamic oil film action.

Performance Analysis. Performance analyses for sector thrust bearings using a fixed taper also hold approximately for most thrust bearing shapes (other than flat lands) with the same ratio of inlet to outlet oil film thickness (Wilcock and Booser, 1956; Neale, 1970; Fuller, 1984). Both for centrally pivoted pad thrust bearings and for spring-supported thrust bearings, use of an inlet-to-outlet film thickness ratio of two is usually appropriate in such an analysis.

While computer analyses in polar coordinates and with local oil film viscosity calculated over the whole oil film area gives more-exact solutions, the following constant viscosity approximations are made by relating a rectangular bearing sector (width B, length L) to the circular configuration of Figure 3.10.7. This rectangular representation allows more ready evaluation of a range of variables and gives results which are quite similar to a more accurate polar coordinate analysis.

Employing the nomenclature of Figure 3.10.7, the following relations give minimum film thickness h_2, frictional power loss H, and oil flow into a sector Q. The dimensionless coefficients involved are given in Table 3.10.14 for a range of sector proportions L/B and ratios of inlet to outlet film thicknesses h_1/h_2.

$$h_2 = K_h \left(\mu UB/P \right)^{0.5} \qquad \text{in. (m)} \tag{3.10.16a}$$

$$H = K_f \, \mu U^2 BL/h_2 \qquad \text{lb} \cdot \text{in./sec} \left(\text{N} \cdot \text{m/sec} \right) \tag{3.10.16b}$$

$$Q = K_q U L h_2 \qquad \text{in.}^3/\text{sec}\ \left(\text{m}^3/\text{sec}\right) \tag{3.10.16c}$$

$$\Delta T = H/\left(Q\rho C_p\right) = \frac{K_f}{K_q K_h^2 \rho C_p} P \quad °\text{F}\ (°\text{C}) \tag{3.10.16d}$$

where B = circumferential breadth of sector at mean radius, in. (m)
K_h, K_f, K_q = dimensionless coefficients
L = radial length of sector, in. (m)
P = unit loading on projected area of sector, W/BL, lb/in.^{2} (N/m^2)
U = surface velocity at mean diameter of sector, in./sec (m/sec)
W = load on a sector, lb (N)
C_p = oil specific heat, in.·lb/(lb·°F) (J/(kg·°C))
h_1, h_2 = leading edge and trailing edge film thicknesses, in. (m)
ρ = oil density, lb/in.^{3} (n/m^3)
μ = oil viscosity at the operating temperature, lb·sec/in.^{2} (N·sec/m^2)

Example. The following example involves a bearing for 25,000 lb thrust load in a 1200 rpm compressor whose rotor has a 5-in.-diameter journal. ISO-32 viscosity grade oil is to be fed at 120°F. Allowing $^3/_8$ in. radial clearance along the shaft sets thrust bearing bore $d = 5.75$ in. Taking unit loading $P = 300$ psi allows a margin for uncertainty in expected thrust load, and outside diameter D is given by Equation 3.10.14):

$$D = \left(\frac{4 \times 25000}{\pi(0.85)(300)} + 5.75^2\right)^{0.5} = 12.6 \text{ in.}$$

With 15% of the area used for oil feed passages, $k_g = 0.85$. Thrust bearing segment radial length L = (12.6 – 5.75)/2 = 3.425 in. With mean diameter (5.75 + 12.6)/2 = 9.175 in., total circumferential breadth of all pads at the mean diameter = $\pi D_m k_g$ = 24.5 in. The number of sectors (and grooves) for $B = L$ is then 24.5/3.425 = 7.2. Using seven lands, adjusted circumferential breadth B for each sector = 24.5/7 = 3.5 in. (For simpler fabrication, six or eight sectors should also be considered.) Runner velocity at the mean diameter, $U = \pi(9.175)(1200/60) = 576.5$ in./sec.

For square pads ($L/B = 1$) in a pivoted-paid bearing with $h_1/h_2 = 2$, which represents experience with centrally pivoted pads, Table 3.10.14 gives the following performance coefficients:

$$K_h = 0.261, \quad K_f = 0.727, \quad K_q = 0.849$$

Temperature rise is given by Equation (3.10.16d) which assumes that the total frictional power loss H goes into heating the total oil flow Q passing over the pad.

$$\Delta T = \frac{K_f P}{K_q K_h^2 \rho C_p} = \frac{0.727(300)}{0.849(0.261)^2(0.0313)(4535)} = 27°\text{F}$$

Adding this ΔT to the 120°F feed temperature gives 147°F as the representative oil operating temperature with a viscosity from Figure 3.10.4 of 1.6×10^{-6} lb·sec/in.^{2}. Temperature rise to the maximum oil film temperature would be expected to be about 53°F, twice the 27°F. If this bearing were in a housing fully flooded with oil, feed temperature to each pad would become the housing oil temperature, essentially the same as oil draining from the housing.

TABLE 3.10.14 Thrust Bearing Performance Characteristics

L/B	0.25	0.5	0.75	1.0	1.5	2.0	∞
			h_1/h_2 = 1.2				
K_h	0.064	0.115	0.153	0.180	0.209	0.225	0.266
K_f	0.912	0.913	0.914	0.915	0.916	0.917	0.919
K_q	0.593	0.586	0.579	0.574	0.567	0.562	0.549
			h_1/h_2 = 1.5				
K_h	0.084	0.151	0.200	0.234	0.275	0.296	0.351
K_f	0.813	0.817	0.821	0.825	0.830	0.833	0.842
K_q	0.733	0.714	0.696	0.680	0.659	0.647	0.610
			h_1/h_2 = 2				
K_h	0.096	0.170	0.223	0.261	0.305	0.328	0.387
K_f	0.698	0.708	0.718	0.727	0.739	0.747	0.768
K_q	0.964	0.924	0.884	0.849	0.801	0.772	0.690
			h_1/h_2 = 3				
K_h	0.100	0.173	0.225	0.261	0.304	0.326	0.384
K_f	0.559	0.579	0.600	0.617	0.641	0.655	0.696
K_q	1.426	1.335	1.236	1.148	1.024	0.951	0.738
			h_1/h_2 = 4				
K_h	0.098	0.165	0.212	0.244	0.282	0.302	0.352
K_f	0.476	0.503	0.529	0.551	0.581	0.598	0.647
K_q	1.888	1.745	1.586	1.444	1.242	1.122	0.779
			h_1/h_2 = 6				
K_h	0.091	0.148	0.186	0.211	0.241	0.256	0.294
K_f	0.379	0.412	0.448	0.469	0.502	0.521	0.574
K_q	2.811	2.560	2.273	2.013	1.646	1.431	0.818
			h_1/h_2 = 10				
K_h	0.079	0.121	0.148	0.165	0.185	0.195	0.221
K_f	0.283	0.321	0.353	0.377	0.408	0.426	0.474
K_q	4.657	4.182	3.624	3.118	2.412	2.001	0.834

After Khonsari, M.M., in *Tribology Data Handbook*, CRC Press, Boca Raton, FL, 1997.

Minimum film thickness h_2 becomes, from Equation (3.10.16a)

$$h_2 = 0.261\left[\left(1.6\times10^{-6}\right)(576.5)(3.5)/300\right]^{0.5} = 0.00086 \text{ in.}$$

With a fixed tapered land, rather than a centrally pivoted pad for which it could be assumed that h_1/h_2 = 2, several iterations might be required with different assumed values of the h_1/h_2 ratio in order to determine the performance coefficients in Table 3.10.14. The proper value of h_1/h_2 will be the one that gives the same final calculated value of h_2 from the above equation as was assumed in the preliminary selection of K_h, K_f, and K_q.

After finding the values for h_2 *and* K_f, the power loss H can be determined using Equation (3.10.16b). For this example the power loss would be H = 5510 lb·in./sec.

The total oil feed to the bearing should satisfy two requirements: (1) provide a full oil film over the bearing segment and (2) maintain reasonably cool operation with no more than 30 to 40°F rise in the

oil temperature from feed to drain. Equation (3.10.16c) can be used to find the oil feed Q needed at the sector inlet to form a full film. The oil feed needed for a 40°F rise is given by the following heat balance using typical density and specific heat values for petroleum oil:

$$Q = H \Big/ \left(\rho C_p \Delta T\right) \tag{3.10.17}$$

The required oil feed will be the larger of the values determined by (3.10.16c) and (3.10.17).

The above calculations are for a single sector; power loss and oil feed would be multiplied by the number of sectors (seven) to obtain values for the total bearing. Consideration would normally follow for other pad geometries, and possibly other lubricants and oil flow patterns, in a search for the most-promising design. More-detailed calculations of film thickness, film temperatures, oil flow, and power loss could then be obtained by one of a number of computer codes available from bearing suppliers or other sources.

Oil Film Bearing Materials

Selection of the material for use in a journal or thrust bearing depends on matching its properties to the load, temperature, contamination, lubricant, and required life.

Babbitts. Of the common bearing materials, listed in Table 3.10.15, first consideration for rotating machines is usually a babbitt alloy containing about 85% tin or lead together with suitable alloying elements. With their low hardness, they have excellent ability to embed dirt, conform to shaft misalignment, and rate highest for compatibility with steel journals. Tin babbitts, containing about 3 to 8% copper and 5 to 8% antimony, are usually the first choice for their excellent corrosion resistance. SAE 12 (ASTM Grade 2) tin babbitt is widely used in both automotive and industrial bearings. The much lower cost of lead babbitt, however, with 9 to 16% antimony and as much as 12% tin for improved corrosion resistance, brings SAE 13, 14, and 15 grades into wide use for both general automotive and industrial applications (Booser, 1992).

TABLE 3.10.15 Characteristics of Oil Film Bearing Materials

Material	Brinell Hardness	Load Capacity, psi	Max Operating Temp., °F	Compatibility[a]	Conformability and Embeddability[a]	Corrosion Resistance[a]	Fatigue Strength[a]
Tin babbitt	20–30	800–1500	300	1	1	1	5
Lead babbitt	15–25	800–1200	300	1	1	3	5
Copper lead	20–30	1500–2500	350	2	2	5	4
Leaded bronze	60–65	3000–4500	450	3	4	4	3
Tin bronze	65–80	5000+	500	5	5	2	2
Aluminum alloy	45–65	4000+	300	4	3	1	2
Zinc alloy	90–125	3000	250	4	5	5	3
Silver overplated	—	5000+	300	2	4	2	1
Two-component, babbitt surfaced	—	3000+	300	2	4	2	3
Three-component, babbitt surfaced	—	4000+	300	1	2	2	1

[a] Arbitrary scale: 1 = best, 5 = worst.

To achieve the high fatigue strength needed in reciprocating engines, only a very thin layer (commonly 0.001″) of babbitt is used so that much of the reciprocating load is taken on a stronger backing material (DeHart, 1984; Kingsbury, 1992). For bimetal bushings such as those used in automobile engines, oil

grooves and feed holes are formed in a continuous steel strip coated with babbitt. The strip is then cut to size and the individual segments are rolled into finished bearings.

For heavy-duty reciprocating engines, three-layer bearings are common. By using a steel strip backing, a thin overlay of SAE 19 or 190 lead babbitt is either electroplated or precision cast on an intermediate layer about 0.1 to 0.3″ thick of copper–nickel, copper–lead, leaded bronze, aluminum, or electroplated silver.

Copper Alloys. Copper–lead alloys containing 20 to 50% lead, either cast or sintered on a steel back, provide good fatigue resistance for heavy-duty main and connecting rod bearings for automotive, truck, diesel, and aircraft engines. The 10% lead–10% tin leaded bronze has been a traditional selection for bearings in steel mills, appliances, pumps, automotive piston pins, and trunions. This has been replaced in many applications by CA932 (SAE 660) containing 3% zinc for easier casting. The harder tin bronzes require reliable lubrication, good alignment, and 300 to 400 Brinell minimum shaft hardness. Cast tin bronze bushings are used at high loads and low speeds in farm machinery, earthmoving equipment, rolling mills, and in automotive engines for connecting rod bearings.

Utility of copper alloy bearings is limited to relatively low surface speeds by the tendency to form a copper transfer film on a steel shaft. Above about 1500 to 3000 ft/min, selective plucking of softer copper material from hotter load zones in the bearing may result in welded lumps forming on the cooler, stronger transfer layer on the mating steel shaft.

Zinc Alloys. Zinc alloys containing about 10 to 30% aluminum find some use for lower cost and better wear life in replacing leaded bronzes. They are used for both oscillating and rotating applications involving speeds up to 1400 ft/min and temperatures up to 250°F.

Aluminum Alloys (DeHart, 1984; Shabel et al., 1992). Although finding only minor use in general industrial applications because of their limited compatibility with steel journals, aluminum alloys containing 6.5% tin, 1% copper, and up to 4% silicon are used as solid, bimetal, and trimetal bearings in automotive engines, reciprocating compressors, and aircraft equipment. Good journal finish and shaft hardness of Rockwell B 85 or higher are required. The good fatigue and corrosion resistance of aluminum alloys have led to use of a number of unique alloys containing major additions of silicon, lead, or tin to provide better compatibility characteristics.

Dry and Semilubricated Bearings

Various plastics, porous bronze and porous iron, carbon–graphite, rubber, and wood are widely used for bearings operating dry or with sparse lubrication (Booser, 1992). Unique properties of these materials have led to their broad use in applications once employing oil film and ball and roller bearings. While these materials provide good performance under conditions of poor or nonexistent lubrication at low speeds, performance commonly improves the closer the approach to full film lubrication.

Plastics

Most commercial plastics find some use both dry and lubricated in slow-speed bearings at light loads (Jamison, 1994). The most commonly used thermoplastics for bearings are PTFE, nylon, and acetal resins. Thermosetting plastics used for bearings include phenolics, polyesters, and polyimides. Table 3.10.16 compares characteristics of typical plastic bearings with those of carbon–graphite, wood, and rubber which are used in similar applications.

In addition to the maximum temperature which can be tolerated, three operating limits shown in Table 3.10.16 are normally defined for plastic bearings: (1) maximum load at low speed, which reflects the compressive yield strength, (2) maximum speed for running under very light load, and (3) a $P\nu$ load-speed limit at intermediate speeds, which serves as a measure of the maximum tolerable surface temperature. Since wear volume in dry sliding is approximately proportional to total load and the distance of sliding, $P\nu$ also gives a measure of wear depth d in the modified form of Archard's relation (3.10.1), $d = k(P\nu)t$, where t is the operating time and wear factor k = wear coefficient K/hardness H.

TABLE 3.10.16. Representative Limiting Conditions for Nonmetallic Bearing Materials

Material	Maximum Temperature, °C	Pv Limit, MN/(m·sec)[a]	Maximum Pressure, P, MN/m²[b]	Maximum speed, v, m/sec
Thermoplastics				
Nylon	90	0.90	5	3
Filled	150	0.46	10	—
Acetal	100	0.10	5	3
Filled	—	0.28	—	—
PTFE	250	0.04	3.4	0.3
Filled	250	0.53	17	5
Fabric	—	0.88	400	0.8
Polycarbonate	105	0.03	7	5
Polyurethane	120	—	—	—
Polysulfone	160	—	—	—
Thermosetting				
Phenolics	120	0.18	41	13
Filled	160	0.53	—	—
Polyimides	260	4	—	8
Filled	260	5	—	8
Others				
Carbon–graphite	400	0.53	4.1	13
Wood	70	0.42	14	10
Rubber	65	—	0.3	20

[a] See Table 3.10.18.
[b] To convert MN/m² to psi, multiply by 145.

Typical values of this wear factor k are given in Table 3.10.17. Since k values involve substantial variability, prototype tests are highly desirable for any planned application. Added fillers can reduce the wear factor for the base polymer by a factor of 10 to 1000 and more (Blanchet and Kennedy, 1992). Common fillers include inorganic powders such as clay, glass fibers, graphite, molybdenum disulfide, and powdered metal, and also silicone fluid as an internally available lubricant.

TABLE 3.10.17 Wear Factors for Plastic Bearings[a]

Material	Wear Factor k, m²/N No Filler	Filled[b]
Nylon-6, 6	4.0	0.24
PTFE	400	0.14[c]
Acetal resin	1.3	4.9
Polycarbonate	50	3.6
Polyester	4.2	1.8
Poly(phenylene oxide)	60	4.6
Polysulfone	30	3.2
Polyurethane	6.8	3.6

[a] See Booser (1992).
[b] With 30 wt% glass fiber, unless otherwise noted.
[c] 15% glass fiber.

Porous Metals

Bearings of compressed and sintered bronze, iron, and aluminum alloy powder are produced at the rate of millions per week for shaft sizes ranging from about 1.6 to 150 mm. These sleeve bearings and thrust washers are used in a wide variety of small electric motors, appliances, business machines, machine tools, automotive accessories, and farm and construction equipment (Morgan, 1984; Cusano, 1994). Traditional powder metal bearings consist of 90% copper and 10% tin (Table 3.10.18). The common pore volume of 20 to 30% is usually impregnated with a petroleum oil of SAE 30 viscosity. To promote

formation of an oil film, high porosity with its high oil content is employed for higher speeds, often with an oil wick or grease providing a supplementary lubricant supply. Lower porosity with up to 3.5% added graphite is used for lower speeds and oscillation where oil film formation is difficult.

Porous iron bearings are used for lower cost, often with some copper and graphite added for high load capacity at low speed. Iron with up to 40% of added 90–10 bronze powder provides many of the characteristics of porous bronze bearings while enjoying the lower cost of the iron. Porous aluminum containing 3 to 5% copper, tin, and lead finds limited use for providing cooler operation, better conformability, and lower weight.

Table 3.10.18 gives approximate operating limits for porous metal bearings. Generally, maximum P values for sleeve bearings range up to 50,000 psi·ft/min. $P\nu$ levels for thrust bearings should generally not exceed about 20% of this value.

TABLE 3.10.18 Operating Limits for Porous Metal Bearings

Porous Metal	Nominal Composition, wt%	Pressure limit, P, MN/m²		Speed Limit ν, m/sec	$P\nu$ Limit MN/(m·sec)
		Static	Dynamic		
Bronze	Cu 90, Sn 10	59	28	6.1	1.8[a]
Iron		52	25	2.0	1.3
Iron–copper	Fe 90, Cu 10	140	28	1.1	1.4
Iron–copper–carbon	Fe 96, Cu 3, C 0.7	340	56	0.2	2.6
Bronze-iron	Fe 60, Cu 36, Sn 4	72	17	4.1	1.2
Aluminum		28	14	6.1	1.8

Note: To convert MN/m² to psi, multiply by 145.

[a] Approximately equivalent to 50,000 psi · ft/min limit often quoted by U.S. suppliers.

Rolling Element Bearings

Types of Rolling Element Bearings

Rolling element bearings may be classified according to the type of rolling element, i.e., ball or roller, and the loading direction. Ball and roller bearings can be designed to carry either radial or thrust loads, or a combination of the two. Standard rolling element bearing configurations are shown in Figure 3.10.8, and the capabilities of the different types are summarized in Figure 3.10.9.

Ball bearings usually consist of a number of hardened and precisely ground balls interposed between two grooved and hardened rings or races. A cage or separator is used to keep the balls equally spaced around the groove. The most common *radial ball bearing* is a deep groove, or Conrad, type which is designed to carry large radial loads, with a small thrust load capability. The radial capacity of a deep groove bearing can be increased by inserting more balls in the bearing, by means of either a face-located filling notch (which decreases the thrust capacity) or a split inner or outer ring (which requires a means to hold the ring halves axially together). The thrust capability of a radial ball bearing can be increased by inducing angular contact between ball and rings. A single-row angular contact bearing can carry thrust load in only one direction, with the thrust load capacity being dependent on the contact angle (angle between the line of action of the force and the plane perpendicular to the shaft axis). Duplex angular contact bearings consist of two angular contact bearings mounted together so they can carry thrust loads in either direction with little axial play, or they can be mounted in tandem to double the axial and radial load-carrying capacity. Self-aligning ball bearings are designed to accommodate more shaft misalignment than is possible with other radial ball bearings.

Thrust ball bearings are used primarily in machinery with a vertically oriented shaft which requires a stiff axial support. Many such bearings have a 90° contact angle and, as a result, can carry essentially no radial load; they also have limited high-speed capability. If thrust loads in both directions are to be carried, a second row of balls must be added.

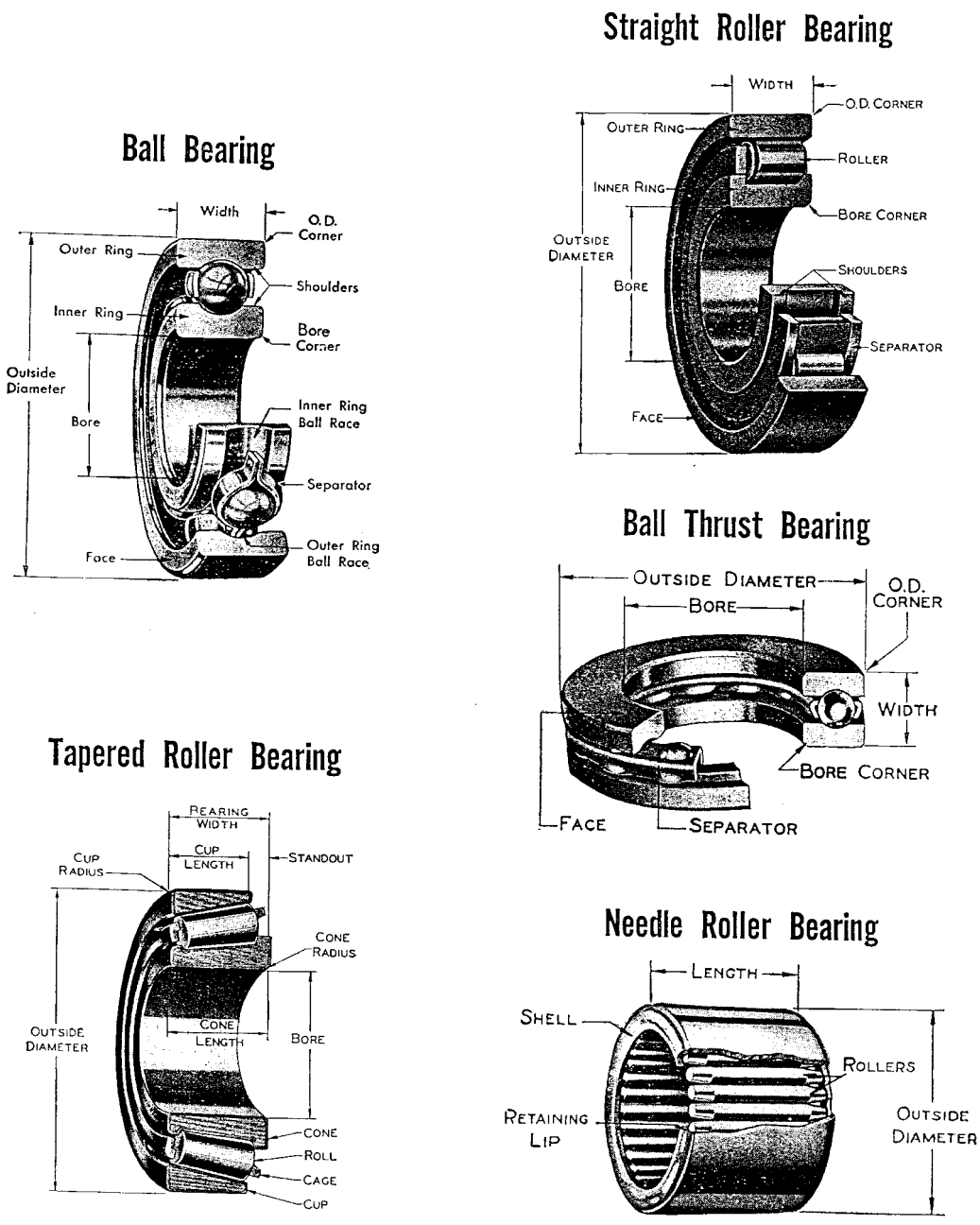


FIGURE 3.10.8 Major types of rolling element bearings.

Roller bearings can be made with cylindrical, tapered, or spherical rollers. As with ball bearings, the rollers are contained between two rings, with a cage or separator used to keep the rollers separated. The cage can be guided by either the rollers or one of the rings. Since roller bearings operate with line contacts, as opposed to the point (or elliptical) contacts that occur in ball bearings, roller bearings are stiffer (less radial displacement per unit load) and have a greater load-carrying capacity than a ball bearing of similar size. Roller bearings are more expensive than ball bearings of comparable size.

CHARACTERISTICS OF STANDARD ROLLING ELEMENT BEARING CONFIGURATIONS

	TYPE	SIZE RANGE IN INCHES		AVERAGE RELATIVE RATINGS				DIMENSIONS	
				Capacity		Limiting Speed	Permissible Misalignment		
		Bore	O.D.	Radial	Thrust			Metric	Inch
BALL BEARINGS	CONRAD TYPE 	.1181 to 41.7323	.3750 to 55.1181	Good	Fair ↔	Conrad is basis for comparison 1.00	± 0° 8′	X	X
	MAXIMUM TYPE 	.6693 to 4.3307	1.5748 to 8.4646	Excellent	Poor ↔	1.00	± 0° 3′	X	
	ANGULAR CONTACT 15°/40°  15° / 40°	.3937 to 7.4803	1.0236 to 15.7480	Good	Good (15°) Excellent (40°) ←	1.00 / 0.70	± 0° 2′	X	
	ANGULAR CONTACT 35° 	.3937 to 4.3307	1.1811 to 9.4488	Excellent	Good ←	0.70	0°	X	
	SELF-ALIGNING 	.1969 to 4.7244	.7480 to 9.4488	Fair	Fair ↔	1.00	± 4°	X	
CYLINDRICAL ROLLER BEARINGS	SEPARABLE INNER RING NON-LOCATING 	.4724 to 19.6850	1.2598 to 28.3465	Excellent	0	1.00	± 0° 4′	X	
	SEPARABLE INNER RING ONE DIR. LOCATING 	.4724 to 12.5984	1.2598 to 22.8346	Excellent	Fair ←	1.00	± 0° 4′	X	
	SELF-CONTAINED TWO DIR. LOCATING 	.4724 to 3.9370	1.4567 to 8.4646	Excellent	Fair ↔	1.00	± 0° 4′	X	
TAPERED ROLLER BEARINGS	SEPARABLE 	.6205 to 6.0000	1.5700 to 10.0000	Good	Good →	0.60	± 0° 2′	X	X
SPHERICAL ROLLER BEARINGS	SELF-ALIGNING 	.9843 to 12.5984	2.0472 to 22.8346	Good	Fair ↔	0.50	± 4°	X	
	SELF-ALIGNING 	.9843 to 35.4331	2.0472 to 46.4567	Excellent	Good ↔	0.75	± 1°	X	
NEEDLE BEARINGS	COMPLETE BEARINGS with or without locating rings & lubricating groove 	.2362 to 14.1732	.6299 to 17.3228	Good	0	0.60	= 0° 2′	X	X
	DRAWN CUP 	.1575 to 2.3622	.3150 to 2.6772	Good	0	0.30	± 0° 2′	X	X
THRUST BEARINGS	SINGLE DIRECTION BALL Grooved Race 	.2540 to 46.4567	.8130 to 57.0866	Poor	Excellent →	0.30	0°	X	X
	SINGLE DIRECTION CYL. ROLLER 	1.1811 to 23.6220	1.8504 to 31.4960	0	Excellent →	0.20	0°	X	
	SELF-ALIGNING SPHERICAL ROLLER 	3.3622 to 14.1732	4.3307 to 22.0472	Poor	Excellent →	0.50	± 3°	X	

FIGURE 3.10.9 Characteristics of standard rolling element bearing configurations.

Radial cylindrical roller bearings are designed to carry primarily radial loads. Cylindrical roller bearings have a high radial load capacity and low friction, so they are suitable for high-speed operation. Their thrust load capacity is limited to that which can be carried by contact (sliding) between the ends of the rollers and the flange or shoulder on the ring which contains them. The rollers in many cylindrical

roller bearings are actually slightly crowned to relieve stress concentrations which would otherwise occur at the ends of the rollers and to compensate for misalignment of the bearing. In order to increase the load-carrying capacity of roller bearings, a second row of rollers is added instead of using longer rollers. This is because long rollers (i.e., length/diameter > 1.7) tend to skew in the roller path, thus limiting their high-speed capability and sometimes shortening their life. Needle bearings have long rollers, however, and they are useful when there are severe radial space limitations and when neither high load capacity nor high speeds are required.

Spherical roller bearings usually have an outer ring with a spherical inside diameter, within which are barrel-shaped rollers. This makes these bearings self-aligning, and also gives them a larger contact area between roller and ring than is the case for other rolling element bearings. Because of this, spherical roller bearings have a very high radial load-carrying capacity, along with some ability to carry thrust loads. They have higher friction between roller and ring, and this limits their high-speed capability.

Tapered roller bearings have tapered rollers, ideally shaped like truncated cones, contained between two mating cones of different angles, the inner cone and the outer cup. The contact angle of the bearing determines its thrust load capability; a steeper angle is chosen for more thrust capacity. If a single row of rollers is used, the bearing is separable and can carry thrust loads in only one direction. If the thrust is double acting, a second bearing can be mounted in a back-to-back configuration or a double row bearing can be selected. In tapered roller bearings there is sliding contact between the ends of the rollers and the guide flange on the inner cone, and this sliding contact requires lubrication to prevent wear and reduce friction.

Thrust roller bearings can be either cylindrical, needle, tapered, or spherical (Figure 3.10.9). In each case there is high load-carrying capacity, but the sliding that occurs between rollers and rings requires lubrication and cooling.

Rolling Element Bearing Materials

Ball and roller bearings require materials with excellent resistance to rolling contact fatigue and wear, as well as good dimensional stability and impact resistance. The rolling elements are subjected to cyclic contact pressures which can range from 70 to 3500 MPa (100 to 500 ksi) or more, and the bearing materials must be hard enough to resist surface fatigue under those conditions. Of the through-hardening steels which meet these requirements, the most popular is AISI 52100, which contains about 1% carbon and 1.5% chromium. In general, balls and rollers made from 52100 are hardened to about Rockwell C60. Standard bearings made from 52100 may suffer from unacceptable dimensional changes resulting from metallurgical transformations at operating temperatures above 140°C (285°F). Special stabilization heat treatments enable operation at higher temperatures, with successful operation at temperatures as high as 200°C (390°F) having been achieved in cases involving low loads. The strength and fatigue resistance of the material diminish if the bearing temperature increases above about 175°C (350°F, however, so above that temperature materials with better hot-hardness, such as M50 tool steel, are required. Carburizing steels such as AISI 8620 have been developed for tapered roller bearings and other heavily loaded types that benefit from the tougher core and case compressive residual stress developed during carburizing. For applications in oxidative or corrosive environments, a hardened martensitic stainless steel such as SAE 440C may be chosen. For the highest operating temperatures, ceramic materials may be used in bearings. The most promising of the ceramics for rolling element bearing applications is silicon nitride. Its major use so far in bearings has been in hybrid bearings with ceramic balls or rollers and metallic rings, but all-ceramic bearings have also been developed.

The temperature limits of these bearing materials are given in Table 3.10.19. For all bearing materials, great care must be taken in the processing and production stages to ensure that no defects or inclusions are present that could serve as an initiation site for fatigue cracks. For most high-performance metallic bearings, this requires a very clean steel production process, such as vacuum arc remelting. Heat treatment of the material is also important to produce the required dimensional stability. The production process for ceramic bearings is even more critical, because a defect in a ceramic bearing element could result in catastrophic fracture of the brittle material.

TABLE 3.10.19 Temperature Limits for Rolling Element Bearing Materials

Material	Maximum Operating Temperature	
	°C	°F
AISI 52100	140–175	285–350
AISI 8620 (carburized)	150	300
440C stainless steel	170	340
M50 tool steel	315	600
Hybrid Si_3N_4–M50	425	800
All-ceramic (Si_3N_4)	650	1200

Bearing cages or retainers have as their primary purpose the separation of the rolling elements. In some cases, they also provide some solid lubrication to the bearing. Low-carbon steel is the most common cage material, but bronze (silicon iron bronze or aluminum bronze) and polymers (particularly nylon 6-6) are used in many applications.

Selection of Rolling Element Bearings

It has been stated that if a rolling element bearing in service is properly lubricated, properly aligned, kept free of abrasive particles, moisture, and corrosive agents, and properly loaded, then all causes of damage will be eliminated except one, contact fatigue (Harris, 1991). The fatigue process results in a spall which may originate on or just beneath the contact surface. Studies of rolling contact fatigue life by Lundberg and Palmgren (1947; 1952) and others showed that most rolling element bearings have fatigue lives which follow a Weibull statistical distribution, wherein the dependence of strength on volume is explained by the dispersion in material strength. Most bearings today are designed according to the Lundberg–Palmgren model, which has been standardized by international (ISO, 1990) and national standards (e.g., ANSI/AFBMA, 1990), although recent work (Ioannides and Harris, 1985) has found that modern bearings have longer lives than those predicted by the standard methods.

The basic rating life of rolling element bearings is the L_{10} life, which is the number of revolutions at constant speed for which there is a 10% probability of failure (or 90% reliability). The basic dynamic load-carrying capacity, or load rating, of a bearing is the constant load C which corresponds to an L_{10} life of one million revolutions. For any other bearing load F, the L_{10} life can be determined by the following relationship:

$$L_{10} = \left(C/F\right)^n \tag{3.10.18}$$

where the load-life exponent $n = 3$ for ball bearings, and $n = 10/3$ for roller bearings.

The equivalent bearing load includes contributions from both radial and thrust loads, and can be calculated by the following expression:

$$F = XF_r + YF_a \tag{3.10.19}$$

where X is a radial load factor, Y is a thrust load factor, F_r is the radial load applied to the bearing, and F_a is the applied thrust (or axial) load.

Values for the dynamic load rating C, as well as the load factors X and Y for any bearing configuration can be found in manufacturers' catalogs, or they can be calculated according to formulas given in bearing texts by Harris (1991) or Eschmann et al. (1985). Those references also give life adjustment factors which can be used to adjust the desired L_{10} life to account for special operating conditions, special material selections, special lubrication conditions, or for a reliability different from 90%.

The bearing user will generally select a commercially available bearing by the following procedure:

1. Determine the axial and thrust loads acting at the bearing location.
2. Determine the required bearing life (L_{10}).

3. Select the most appropriate bearing type from among those given in Figure 3.10.9.
4. Use the X and Y values appropriate to the type of bearing and loading conditions in Equation (3.10.19) to find the equivalent dynamic bearing load F.
5. Determine the required dynamic load capacity C from Equation (3.10.18).
6. Select a bearing with a dynamic load capacity at least as large as the required value from a manufacturer's catalog.
7. Provide an appropriate mounting arrangement for the bearing. Manufacturers' catalogs can be consulted for guidance in designing the mounting and selecting appropriate fits for the bearing. The importance of the fit cannot be overemphasized, since improper fit can result in considerable reduction in bearing life.
8. Provide adequate lubrication for the bearing (see below). Seals and/or shields may be integrated into the bearing to retain or protect the lubricant in the bearing.

Rolling Bearing Lubrication

The primary load-carrying contacts between rolling elements and rings exhibit nearly pure rolling. There are many sliding contacts in rolling element bearings, however, including those where roller ends contact the internal flanges of rings, where rolling elements contact separator/cage, and where the separator contacts the guiding (piloting) ring of the bearing. All of those contacts must be lubricated to limit friction and wear, and either a grease or an oil can be used for that purpose.

Under most normal operating conditions, rolling element bearings can be grease lubricated. *Greases* coat surfaces with a thin boundary lubricant film of oil, thickener, and additive molecules, thus providing protection against sliding wear, and provide oil to lubricate the concentrated rolling contacts (see below). The selection of a grease depends on its effective temperature range, oil viscosity, consistency, and rust-inhibiting properties. For normal applications, a bearing should be filled with grease up to 30 to 50% of its free volume. Overfilling will cause overheating, particularly at high speeds. Grease will deteriorate with time and will leak out. For that reason, there should be a relubrication schedule, with grease being added at intervals which can be estimated by the following expression (Neale, 1993):

$$\text{relubrication interval (hours)} = \left(k/d^{1/2}\right)\left[\left(14 \times 10^6/n\right) - 4d^{1/2}\right] \tag{3.10.20}$$

where k = 10 for radial ball bearings, 5 for cylindrical roller bearings, and 1 for spherical or tapered roller bearings; d = bearing bore diameter (mm); and n = speed (rpm)

Oil lubrication is required when high speed or high operating temperatures preclude the use of grease. It is necessary to choose an oil of proper viscosity and appropriate viscosity–temperature characteristics in order to insure sufficient thickness of oil film in the lubricated concentrated contacts. If viscosity is too low, the film thickness will not prevent metal/metal contact, but if the viscosity is too high, excessive friction will occur. Oil can be applied to a bearing by one of several methods (listed in order of increasing effectiveness at increasing bearing speed): *oil bath*, in which the rolling elements carry the oil through the bearing; *oil circulating system*, in which the oil is pumped from the bearing through an external filter and heat exchanger and back to the bearing; *oil mist*, in which an airstream carries oil droplets to the bearing; and *oil jets*, in which the oil is injected into the bearing through carefully positioned nozzles. The quantity and entry velocity of the oil must be carefully selected and controlled in order to dissipate the heat generated in the bearing.

The lubrication mechanism in the concentrated contacts of rolling element bearings is *elastohydrodynamic lubrication* (EHL). EHL typically occurs in lubricated, nonconforming elastic contacts, such as the elliptical contact that occurs between ball and raceway or the rectangular contact between roller and ring. These lubricated contacts have a very small area and the contact pressures are very high. Because of those high pressures, the contacting surfaces deform and the lubricant viscosity increases, thereby aiding its ability to sustain heavy loading without oil-film breakdown. A diagram of these phenomena is shown in Figure 3.10.10. The most important parameter of the concentrated contact, from

the point of view of rolling element bearing performance, is minimum EHL film thickness, h_o. The following expression can be used to find minimum film thickness in most rolling element bearing contacts (Hamrock and Dowson, 1977):

$$h_o = 3.63R_x U^{0.68} G^{-0.49} W_p^{-0.73}\left(1 - e^{-0.68\kappa}\right) \tag{3.10.21}$$

where $R_x = (R_{x1}\, R_{x2})/(R_{x1} + R_{x2})$, $R_y = (R_{y1}\, R_{y2})/(R_{y1} + R_{y2})$, ellipticity parameter $\kappa = R_x/R_y$, $U = \mu\,(u_1 + u_2)/2E'R_x$, μ = absolute viscosity, u_1 and u_2 are velocities of rolling element and ring, $E' = E/(1–\nu^2)$, E = modulus of elasticity, ν = Poisson's ratio, $G = \alpha E'$, α = pressure–viscosity exponent, $W_p = W/E'\,R_x^2$, and W = radial load.

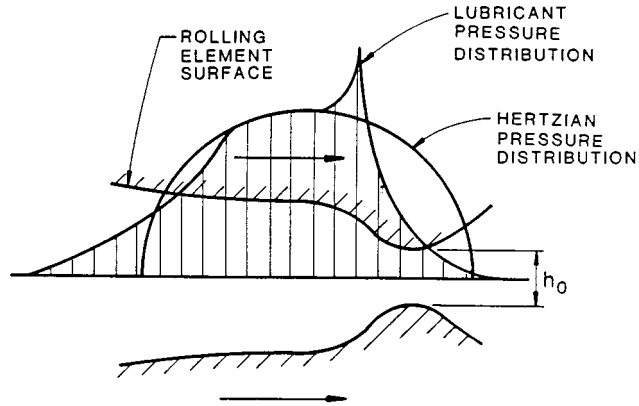


FIGURE 3.10.10 Typical pressure and film thickness distributions in elastohydrodynamic contact.

The minimum film thickness must be large enough to prevent metal/metal contact within the lubricated conjunctions. The criterion for this is stated as

$$h_o \geq 1.5\left(r_{q1}^2 + r_{q2}^2\right)^{0.5} \tag{3.10.22}$$

where r_{q1} and r_{q2} are the rms surface roughness of the rolling element and ring, respectively. If the minimum film thickness is less than this value, complete EHL will not occur, and this could result in wear, surface fatigue, and eventual early bearing failure (i.e., well before the predicted L_{10} life).

An alternative to oil or grease lubrication for rolling element bearings operating under severe conditions is *solid lubrication*. Solid lubricants can be used effectively in high temperatures or vacuum conditions where liquid lubricants are impractical or would provide marginal performance. Solid lubricants do not prevent solid/solid contact, so wear of the lubricant coating can be expected; bearing life is governed by the depletion of the solid lubricant film.

Lubricant Supply Methods

Lubrication systems for oil film bearings can generally be grouped into three classifications: self contained devices for small machines; centralized systems, common in manufacturing plants; and circulating systems dedicated to a single piece of equipment such as a motor, turbine, or compressor. Upper speed limits for common journal bearing lubrication methods are indicated in Figure 3.10.11 (Wilcock and Booser, 1987). Submerging the bearing directly in an oil bath is a common alternative for vertical machines.

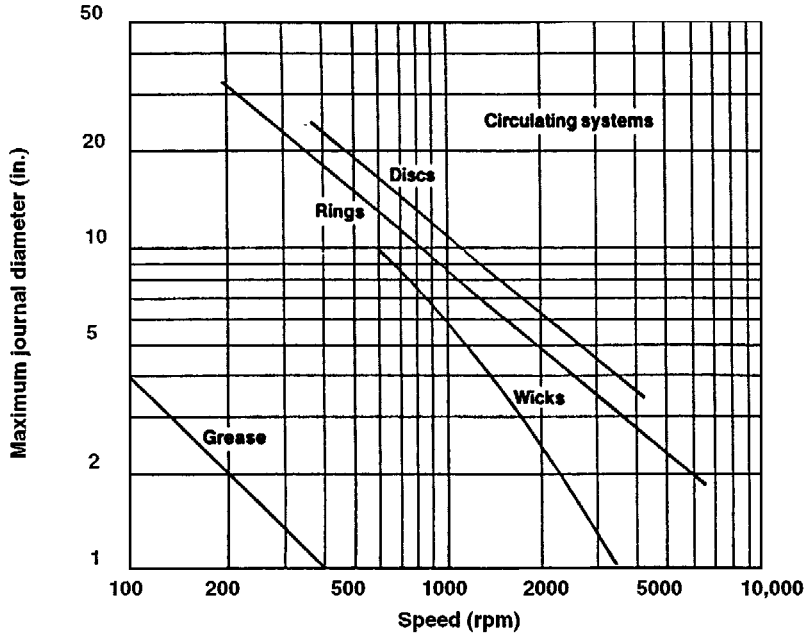


FIGURE 3.10.11 Upper limits for journal bearing lubrication methods. (From Wilcock, D.F. and Booser, E.R., *Mach. Des.*; April 23, 101–107, 1987. With permission.)

Self-Contained Units

Lifting of oil by *capillary action* from a small reservoir is used to feed small bearings in business machines, household appliances, electrical instruments, controls, and timer motors. In capillary tubes, the height h to which oil will rise is (Wilcock and Booser, 1987)

$$h = 2\sigma\cos\theta/(r\rho) \tag{3.10.23}$$

where σ = surface tension, lb/in., r = capillary radius (or spacing between two parallel plates), in.; ρ = oil density, lb/in.^{3}. Because oils wet most surfaces readily, the cosine of the contact angle can be taken as unity. As an example, with $\sigma = 1.7 \times 10^{-4}$ lb/in. for a petroleum oil, the rise in an 0.005-in.-radius capillary will be $h = 2(1.7 \times 10^{-4})(1)/(0.005)(0.0307) = 2.2$ in.

Wick lubrication is applied in millions of fractional horsepower motors annually. Although wicks generally are not efficient at raising oil more than about 2 in., lift may be up to 5 in. in railway journal bearings. By referring to Figure 3.10.12, petroleum oil delivery by a typical wick can be estimated by the following equation (Elwell, 1994):

$$Q = kAF_o(h_u - h)/(\mu L) \quad \text{in.}^3/\text{min} \tag{3.10.24}$$

where the constant k reflects both the capillary spacing in the wick and the surface tension of the oil; A is the wick cross-section area, in.^{2}; F_o is volume fraction of oil in the saturated wick (often about 0.75); h_u is the ultimate wicking height, about 7.5 in. for SAE Grade F-1 felt; h is oil delivery height above the reservoir surface, in.; L is wick length, in.; and μ is viscosity at the wick temperature, lb·sec/in.$^2 \times 10^6$. k is approximately 0.26 for SAE Grade F-1 felt.

Oil rings hanging over a journal, as illustrated in Figure 3.10.13 and Table 3.10.20, are used to lift oil to journal bearings in electric motors, pumps, and medium-size industrial machines (Elwell, 1994).

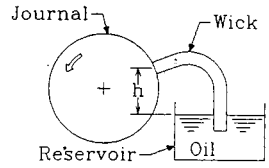


FIGURE 3.10.12 Wick-lubricated journal bearing. (From Elwell, R.C., in *Handbook of Lubrication and Tribology*, Vol. 3, CRC Press, Boca Raton, FL, 1994, 515–533. With permission.)

At very low journal surface speeds below about 2 to 3 ft/sec, the ring will run synchronously with its journal. At higher speeds, increasing viscous drag on the ring in its reservoir will slow the ring surface velocity; oil ring rpm at higher speeds is often in the range of $^1/_{10}$ the journal rpm. Above about 45 ft/sec journal surface velocity, oil delivery drops to an unusably low level as centrifugal throw-off and windage interfere.

These self-contained systems usually supply much less oil to a bearing than needed to form a full hydrodynamic oil film (Elwell, 1994). With the starved oil supply generating an oil wedge of reduced circumferential extent, power loss will be lowered at the expense of reduced load capacity (smaller minimum film thickness).

Centralized Distribution Systems

Limitations with individual localized lubricating devices have led to widespread use of centralized systems for factory production-line equipment, construction and mining machines, and small applications. Oil or soft grease is pumped from a central reservoir in pulses or as metered continuous feed. Oil mist is piped for distances up to 300 ft for machines in petrochemical plants and steel mills. Polymer additives in the 50,000 to 150,000 molecular-weight range greatly reduce the escape of stray oil mist into the atmosphere.

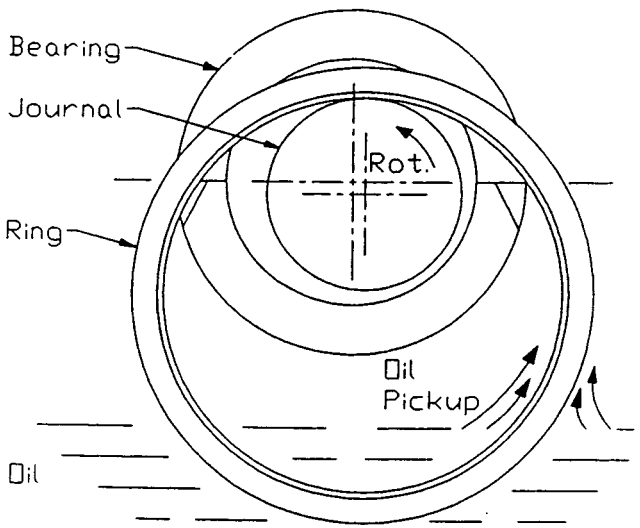
Circulating Systems

Where bearing design, reliability requirements, or equipment considerations preclude use of a simpler oil feed, a circulating system is employed involving an oil reservoir, pump, cooler, and filter (Twidale and Williams, 1984). These systems become quite compact with the space limitations in aircraft, marine, or automobile engines where the reservoir may simply be the machine sump with capacity to hold only a 20- to 60-sec oil supply. Characteristics of typical oil-circulating systems for industrial use are given in Table 3.10.21 (Wilcock and Booser, 1987).

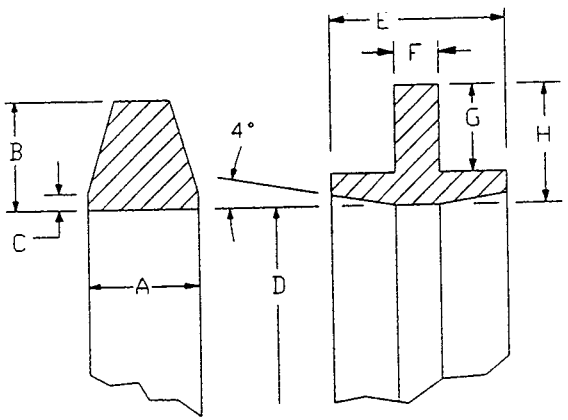
Dynamic Seals

Fluid seals commonly accompany bearings in a wide variety of machinery, both to control leakage of lubricating oil and process fluids and to minimize contamination. Static seals, such as O-rings and gaskets, provide sealing between surfaces which do not move relative to each other. Dynamic seals, which will be the focus of this discussion, restrain flow of fluid between surfaces in relative motion. Most dynamic seals could be classified as either contact or clearance seals. Contact seals are used when the surfaces are in sliding contact, while clearance seals imply that the surfaces are in close proximity to each other but do not contact. The major types of dynamic seals are listed in Table 3.10.22. Further details about the design and analysis of dynamic sealing elements can be found in the handbook article by Stair (1984) and in the book by Lebeck (1991).

As an example of a contact seal, ball bearings are often sealed to retain their lubricant and to keep out contaminants over a long lifetime. For small grease-lubricated bearings, the sealing function is often accomplished by lightly loaded contact between a rubber lip seal component and the bearing ring. For shafts ranging up to about 2 in. (5 cm) in diameter, the lip seal may be replaced by a closely fitted, but noncontacting, shield which helps contain the grease and restricts intrusion of dirt and other contaminants.



Oil-ring bearing elements.



Ring cross-sections.

FIGURE 3.10.13 Oil-ring bearing elements and ring cross-sections. (From Elwell, R.C., in *Handbook of Lubrication and Tribology*, Vol. 3, CRC Press, Boca Raton, FL, 1994, 515–533. With permission.)

TABLE 3.10.20 Typical Oil-Ring Dimensions, mm (in.)

A	B	C	D	E	F	G	H
6 (0.24)	6 (0.24)	1 (0.04)	100 (3.94)				
7 (0.28)	7 (0.28)	1 (0.04)	135 (5.31)				
8 (0.31)	8 (0.31)	2 (0.08)	165 (6.50)				
16 (0.63)	13 (0.51)	2 (0.08)	200 (7.87)	16 (0.63)	5 (0.20)	11 (0.43)	10 (0.39)
			335 (13.2)	21 (0.83)	6 (0.24)	13 (0.51)	14 (0.55)
			585 (23.0)	25 (1.0)	7 (0.28)	14 (0.55)	20 (0.79)
			830 (32.7)	32 (1.3)	8 (0.31)	16 (0.63)	27 (1.06)

TABLE 3.10.21 Typical Oil Circulating Systems

Application	Duty	Oil Viscosity at 40°C (cSt)	Oil Feed (gpm)	Pump Type	Reservoir Dwell Time (min)	Type	Rating (μm)
Electrical machinery	Bearings	32–68	2	Gear	5	Strainer	50
General	Bearings	68	10	Gear	8	Dual cartridge	100
Paper mill dryer section	Bearings, gears	150–220	20	Gear	40	Dual cartridge	120
Steel mill	Bearings	150–460	170	Gear	30	Dual cartridge	150
	Gears	68–680	240	Gear	20	Dual cartridge	
Heavy duty gas turbines	Bearings, controls	32	600	Gear	5	Pleated paper	5
Steam turbine-generators	Bearings	32	1000	Centrifugal	5	Bypass 15%/hr	10

TABLE 3.10.22 Characteristics of Dynamic Seals

Type of Seal	Type of Motion: Rotating	Type of Motion: Reciprocating	Extent of Use	Friction	Leakage	Life
Contact						
Face seals	x		H	L	L	M–H
Lip seals	x		H	L	L	L–M
Piston rings		x	H	H	L	L–M
O-Rings	x	x	M	H	L	L
Packings	x	x	H	M	M	L
Diaphragms		x	L	L	L	H
Controlled clearance						
Hydrodynamic	x		L	L	M	H
Hydrostatic	x		L	L	M	H
Floating bushing	x	x	M	M	M–H	H
Fixed geometry clearance						
Labyrinth	x		H	H	H	H
Bushing	x	x	M	H	H	M–H
Special						
Magnetic fluid	x	x	L	L	L	M
Centrifugal	x		L	M	L	H

H = High, M = Moderate, L = Low.
Modified from Stair, W.K., in *Handbook of Lubrication*, Vol. 2, CRC Press, Boca Raton, FL 1984, 581–622.

For more severe sealing requirements in rolling element bearings, labyrinth clearance seals may be used (Harris, 1991).

The most common seals for rotating shafts passing through fixed housings, such as pumps or gearboxes, are radial lip seals and mechanical face seals. These contact seals can be designed to handle a wide range of sealed fluids, temperatures, velocities, and pressures (Stair, 1984; Lebeck, 1991). Material selection and surface finish considerations are similar to those discussed in the subsections on sliding friction and its consequences and on dry and semilubricated bearings.

When high surface speeds are encountered, wear and frictional heating may prohibit use of rubbing contact seals. For such applications, clearance seals such as close-fitting labyrinth seals or controlled-clearance fluid film seals can be used. Fluid film seals can be either hydrostatic or hydrodynamic; both types have a pressurized film of fluid which prevents contact between the sealed surfaces and use pressure balancing to restrain leakage. The principles governing their operation are similar to those discussed in the subsection on fluid film bearings. Details of fluid film seals can be found in Shapiro (1995).

References

ANSI/AFBMA, 1990. Load Ratings and Fatigue Life for Ball Bearings, ANSI/AFBMA 9–1990, AFBMA, Washington, D.C.

Archard, J.F. 1980. Wear theory and mechanisms, in *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Eds., ASME, New York.

Bhushan, B. and Gupta, B.K. 1991. *Handbook of Tribology*, McGraw-Hill, New York.

Blanchet, T.A. and Kennedy, F.E. 1992. Sliding wear mechanism of polytetrafluoroethylene (PTFE) and PTFE composites, *Wear*, 153:229–243.

Blau, P.J., Ed. 1992. *Friction, Lubrication and Wear Technology, Metals Handbook*, Vol. 18, 10th ed., ASM International, Metals Park, OH.

Booser, E.R. 1992. Bearing materials, in *Encyclopedia of Chemical Technology*, Vol. 4, pp. 1–21, John Wiley & Sons, New York.

Booser, E.R. 1995. Lubricants and lubrication, in *Encyclopedia of Chemical Technology*, 4th ed., Vol. 15, pp. 463–517, John Wiley & Sons, New York.

Booser, E.R. and Wilcock, D.F. 1987. New technique simplifies journal bearing design, *Mach. Des.*, April 23, pp. 101–107.

Booser, E.R. and Wilcock, D.F. 1991. Selecting thrust bearings, *Mach. Des.*, June 20, pp. 69–72.

Crook, P. and Farmer, H.N. 1992. Friction and wear of hardfacing alloys, in *Friction, Lubrication and Wear Technology, Metals Handbook*, Vol. 18, pp. 758–765, ASM International, Metals Park, OH.

Cusano, C. 1994. Porous metal bearings, in *Handbook of Lubrication and Tribology*, Vol. 3, pp. 491–513, CRC Press, Boca Raton, FL.

DeHart, A.O. 1984. Sliding bearing materials, in *Handbook of Lubrication*, Vol. 2, pp. 463–476, CRC Press, Boca Raton, FL.

Derner, W.J. and Pfaffenberger, E.E. 1984. Rolling element bearings, in *Handbook of Lubrication*, Vol. 2, pp. 495, CRC Press, Boca Raton, FL.

Elwell, R.C. 1994. Self-contained bearing lubrication: rings, disks, and wicks, in *Handbook of Lubrication and Tribology*, Vol. 3, pp. 515–533, CRC Press, Boca Raton, FL.

Engineering Sciences Data Unit (ESDU). 1965. *General Guide to the Choice of Journal Bearing Type*, Item 65007, Institution of Mechanical Engineers, London.

Engineering Sciences Data Unit (ESDU). 1967. *General Guide to the Choice of Thrust Bearing Type*, Item 67073, Institution of Mechanical Engineers, London.

Eschmann, P., Hasbargen, L., and Weigand, K. 1985. *Ball and Roller Bearings*, John Wiley & Sons, New York.

Fenske, G.R. 1992. Ion implantation, in *Friction, Lubrication and Wear Technology, Metals Handbook,* Vol. 18, pp. 850–860, ASM International, Metals Park, OH.

Fuller, D.D. 1984. Theory and practice of lubrication for engineers, 2nd ed., John Wiley & Sons, New York.

Hamrock, B. and Dowson, D. 1977. Isothermal elastohydrodynamic lubrication of point contacts, *ASME J. Lubr. Technol.*, 99(2): 264–276.

Harris, T.A. 1991. *Rolling Bearing Analysis*, 3rd ed., John Wiley & Sons, New York.

Ioannides, S. and Harris, T.A. 1985. A new fatigue life model for rolling bearings, *ASME J. Tribology*, 107:367–378.

ISO, 1990. Rolling Bearings Dynamic Load Ratings and Rating Life, International Standard ISO 281.

Jamison, W.E. 1994. Plastics and plastic matrix composites, in *Handbook of Lubrication and Tribology*, Vol. 3, pp. 121–147, CRC Press, Boca Raton, FL.

Khonsari, M.M. 1997. In *Tribology Data Handbook*, CRC Press, Boca Raton, FL.

Kingsbury, G.R. 1992. Friction and wear of sliding bearing materials, in *ASM Handbook*, Vol. 18 pp. 741–757, ASM International, Metals Park, OH.

Klaus, E.E. and Tewksbury, E.J. 1984. Liquid lubricants, in *Handbook of Lubrication*, Vol. 2, pp. 229–254, CRC Press, Boca Raton, FL.

Kushner, B.A. and Novinski, E.R. 1992. Thermal spray coatings, in *Friction, Lubrication and Wear Technology, Metals Handbook*, Vol. 18, pp. 829–833, ASM International, Metals Park, OH.

Lebeck, A.O. 1991. *Principles and Design of Mechanical Face Seals,* John Wiley & Sons, New York.

Lundberg, G. and Palmgren, A. 1947. Dynamic capacity of rolling bearings, *Acta Polytech. Mech. Eng. Ser.*, 1(3):196.

Lundberg, G. and Palmgren, A. 1952. Dynamic capacity of roller bearings, *Acta Polytech. Mech. Eng. Ser.*, 2(4):210.

Morgan, V.T. 1984. *Porous Metal Bearings and Their Application*, MEP-213, Mechanical Engineering Publications, Workington, U.K.

Neale, M.J. 1993. *Bearings*, Butterworth-Heinemann, Oxford.

Neale, P.B. 1970. *J. Mech. Eng. Sci.*, 12:73–84.

Peterson, M.B. and Winer, W.O., Eds., 1980. *Wear Control Handbook*, ASME, New York.

Rabinowicz, E. 1980. Wear coefficients metals, in *Wear Control Handbook*, M.B. Peterson and W.O. Winer, Eds., pp. 475–506, ASME, New York.

Rabinowicz, E. 1995. *Friction and Wear of Materials*, 2nd ed., John Wiley & Sons, New York.

Ramondi, A.A. and Szeri, A.Z. 1984. Journal and thrust bearings, in *Handbook of Lubrication,* Vol. 2, pp. 413–462, CRC Press, Boca Raton, FL.

Reynolds, O. 1886. On the theory of lubrication and its application to Mr. Beauchamp Tower's experiments, *Philos. Trans R. Soc.*, 177:157–234.

Schmitt, G.F. 1980. Liquid and solid particles impact erosion, in *Wear Control Handbook,* M.B. Peterson and W.O. Winer, Eds., pp. 231–282, ASME, New York.

Shabel, B.S. Granger, D.A., and Tuckner, W.G. 1992. Friction and wear of aluminum–silicon alloys, in *ASM Handbook*, Vol. 18, pp. 785–794, ASM International, Metals Park, OH.

Shapiro, W. 1995. Hydrodynamic and hydrostatic seals, in *Handbook of Lubrication and Tribology*, Vol. 3, pp. 445–468, CRC Press, Boca Raton, FL.

Stair, W.K. 1984. Dynamic seals, in *Handbook of Lubrication*, Vol. 2, pp. 581–622, CRC Press, Boca Raton, FL.

Twidale, A.J. and Williams, D.C.J. 1984. Circulating oil systems, in *Handbook of Lubrication*, Vol. 2, pp. 395–409, CRC Press, Boca Raton, FL.

Weil, R. and Sheppard, K. 1992. Electroplated coatings, in *Friction, Lubrication and Wear Technology, Metals Handbook*, Vol. 18, pp. 834–839, ASM International, Ohio.

Wilcock, D.F. and Booser, E.R. 1956. Bearing design and application, McGraw-Hill, New York.

Wilcock, D.F. and Booser, E.R. 1987. Lubrication techniques for journal bearings, *Machine Des.*, April 23, 101–107.

3.11 Pumps and Fans

Robert F. Boehm

Introduction

Pumps are devices that impart a pressure increase to a liquid. Fans are used to increase the velocity of a gas, but this is also accomplished through an increase in pressure. The pressure rise found in pumps can vary tremendously, and this is a very important design parameter along with the liquid flow rate. This pressure rise can range from simply increasing the elevation of the liquid to increasing the pressure hundreds of atmospheres. Fan applications, on the other hand, generally deal with small pressure increases. In spite of this seemingly significant distinction between pumps and fans, there are many similarities in the fundamentals of certain types of these machines as well as with their application and theory of operation.

The appropriate use of pumps and fans depends upon the proper choice of device and the proper design and installation for the application. A check of sources of commercial equipment shows that many varieties of pumps and fans exist. Each of these had special characteristics that must be appreciated for achieving proper function. Preliminary design criteria for choosing between different types is given by Boehm (1987).

As is to be expected, the wise applications of pumps and fans requires knowledge of fluid flow fundamentals. Unless the fluid mechanics of a particular application are understood, the design could be less than desirable.

In this section, pump and fan types are briefly defined. In addition, typical application information is given. Also, some ideas from fluid mechanics that are especially relevant to pump and fan operation are reviewed.

Pumps

Raising of water from wells and cisterns is the earliest form of pumping (a very detailed history of early applications is given by Ewbank, 1842). Modern applications are much broader, and these find a wide variety of machines in use. Modern pumps function on one of two principles. By far the majority of pump installations are of the *velocity head* type. In these devices, the pressure rise is achieved by giving the fluid a movement. At the exit of the machine, this movement is translated into a pressure increase. The other major type of pump is called *positive displacement*. These devices are designed to increase the pressure of the liquid while essentially trying to compress the volume. A categorization of pump types has been given by Krutzsch (1986), and an adaptation of this is shown below.

I. Velocity head
   A. Centrifugal
      1. Axial flow (single or multistage)
      2. Radial flow (single or double suction)
      3. Mixed flow (single or double suction)
      4. Peripheral (single or multistage)
   B. Special Effect
      1. Gas lift
      2. Jet
      3. Hydraulic ram
      4. Electromagnetic
II. Positive displacement
   A. Reciprocating
      1. Piston, plunger
         a. Direct acting (simplex or duplex)

b. Power (single or double acting, simplex, duplex, triplex, multiplex)
2. Diaphragm (mechanically or fluid driven, simplex or multiplex)

B. Rotary
1. Single rotor (vane, piston, screw, flexible member, peristaltic)
2. Multiple rotor (gear, lobe, screw, circumferential piston)

In the next subsection, some of the more common pumps are described.

Centrifugal and Other Velocity Head Pumps

Centrifugal pumps are used in more industrial applications than any other kind of pump. This is primarily because these pumps offer low initial and upkeep costs. Traditionally, pumps of this type have been limited to low-pressure-head applications, but modern pump designs have overcome this problem unless very high pressures are required. Some of the other good characteristics of these types of devices include smooth (nonpulsating) flow and the ability to tolerate nonflow conditions.

The most important parts of the centrifugal pump are the *impeller* and *volute*. An impeller can take on many forms, ranging from essentially a spinning disk to designs with elaborate vanes. The latter is usual. Impeller design tends to be somewhat unique to each manufacturer, as well as finding a variety of designs for a variety of applications. An example of an impeller is shown in Figure 3.11.1. This device imparts a radial velocity to the fluid that has entered the pump perpendicular to the impeller. The volute (there may be one or more) performs the function of slowing the fluid and increasing the pressure. A good discussion of centrifugal pumps is given by Lobanoff and Ross (1992).

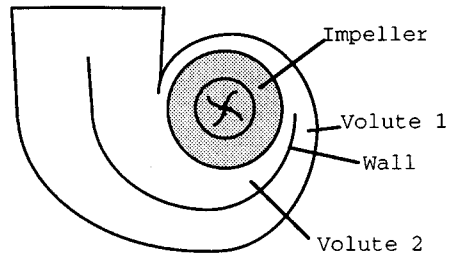


FIGURE 3.11.1. A schematic of a centrifugal pump is shown. The liquid enters perpendicular to the figure, and a radial velocity is imparted by clockwise spin of the impeller.

A very important factor in the specification of a centrifugal pump is the *casing orientation* and *type*. For example, the pump can be oriented vertically or horizontally. Horizontal mounting is most common. Vertical pumps usually offer benefits related to ease of priming and reduction in required net positive suction head (see discussion below). This type also requires less floor space. Submersible and immersible pumps are always of the vertical type. Another factor in the design is the way the casing is split, and this has implications about ease of manufacture and repair. Casings that are split perpendicular to the shaft are called *radially split*, while those split parallel to the shaft axis are denoted as *axially split*. The latter can be *horizontally split* or *vertically split*. The number of *stages* in the pump greatly affects the pump-output characteristics. Several stages can be incorporated into the same casing, with an associated increase in pump output. Multistage pumps are often used for applications with total developed head over 50 atm.

Whether or not a pump is self-priming can be important. If a centrifugal pump is filled with air when it is turned on, the initiation of pumping action may not be sufficient to bring the fluid into the pump. Pumps can be specified with features that can minimize priming problems.

There are other types of velocity head pumps. *Jet pumps* increase pressure by imparting momentum from a high-velocity liquid stream to a low-velocity or stagnant body of liquid. The resulting flow then

goes through a diffuser to achieve an overall pressure increase. *Gas lifts* accomplish a pumping action by a drag on gas bubbles that rise through a liquid.

Positive-Displacement Pumps

Positive-displacement pumps demonstrate high discharge pressures and low flow rates. Usually, this is accomplished by some type of pulsating device. A piston pump is a classic example of positive-displacement machines. Rotary pumps are one type of positive-displacement device that do not impart pulsations to the existing flow (a full description of these types of pumps is given by Turton, 1994). Several techniques are available for dealing with pulsating flows, including use of double-acting pumps (usually of the reciprocating type) and installation of pulsation dampeners.

Positive-displacement pumps usually require special seals to contain the fluid. Costs are higher both initially and for maintenance compared with most pumps that operate on the velocity head basis. Positive-displacement pumps demonstrate an efficiency that is nearly independent of flow rate, in contrast to the velocity head type (see Figure 3.11.2 and the discussion related to it below).

Reciprocating pumps offer very high efficiencies, reaching 90% in larger sizes. These types of pumps are more appropriate for pumping abrasive liquids (e.g., slurries) than are centrifugal pumps.

A characteristic of positive displacement pumps which may be valuable is that the output flow is proportional to pump speed. This allows this type of pump to be used for metering applications. Also a positive aspect of these pumps is that they are self-priming, except at initial start-up.

Very high head pressures (often damaging to the pump) can be developed in positive-displacement pumps if the downstream flow is blocked. For this reason, a pressure-relief-valve bypass must always be used with positive-displacement pumps.

Pump/Flow Considerations

Performance characteristics of the pump must be considered in system design. Simple diagrams of pump applications are shown in Figure 3.11.2. First, consider the left-hand figure. This represents a flow circuit, and the pressure drops related to the piping, fittings, valves, and any other flow devices found in the circuit must be estimated using the laws of fluid mechanics. Usually, these resistances (pressure drops) are found to vary approximately with the square of the liquid flow rate. Typical characteristics are shown in Figure 3.11.3. Most pumps demonstrate a flow vs. pressure rise variation that is a positive value at zero flow and decreases to zero at some larger flow. Positive-displacement pumps, as shown on the right-hand side of Figure 3.11.3, are an exception to this in that these devices usually cannot tolerate a zero flow. An important aspect to note is that a closed system can presumably be pressurized. A contrasting situation and its implications are discussed below.

The piping diagram show on the right-hand side of Figure 3.11.2 is a once-through system, another frequently encountered installation. However, the leg of piping through "pressure drop 1" shown there can have some very important implications related to *net positive suction head*, often denoted as **NPSH**. In simple terms, NPSH indicates the difference between the local pressure and the thermodynamic saturation pressure at the fluid temperature. If NPSH = 0, the liquid can vaporize, and this can result in a variety of outcomes from noisy pump operation to outright failure of components. This condition is called **cavitation**. Cavitation, if it occurs, will first take place at the lowest pressure point within the piping arrangement. Often this point is located at, or inside, the inlet to the pump. Most manufacturers specify how much NPSH is required for satisfactory operation of their pumps. Hence, the actual NPSH (denoted as **NPSHA**) experienced by the pump must be larger than the manufacturer's required NPSH (called **NPSHR**). If a design indicates insufficient NPSH, changes should be made in the system, possibly including alternative piping layout, including elevation and/or size, or use of a pump with smaller NPSH requirements.

The manufacturer should be consulted for a map of operational information for a given pump. A typical form is shown in Figure 3.11.4. This information will allow the designer to select a pump that satisfied the circuit operational requirements while meeting the necessary NPSH and most-efficient-operation criteria.

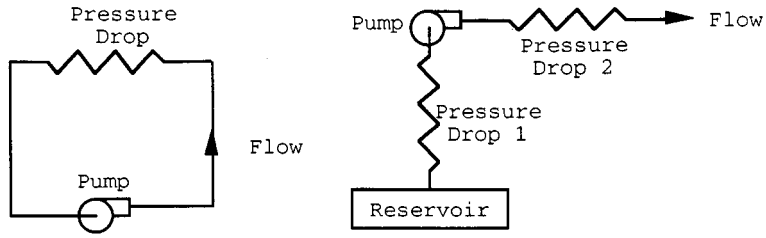


FIGURE 3.11.2. Typical pump applications, either in circuits or once-through arrangements, can be represented as combined fluid resistances as shown. The resistances are determined from fluid mechanics analyses.

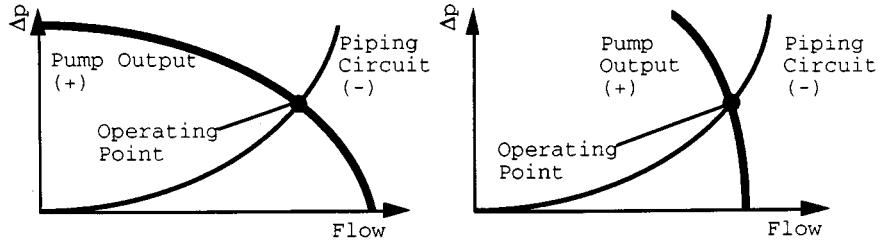


FIGURE 3.11.3. An overlay of the pump flow vs. head curve with the circuit piping characteristics gives the operating state of the circuit. A typical velocity head pump characteristic is shown on the left, while a positive-displacement pump curve is shown on the right.

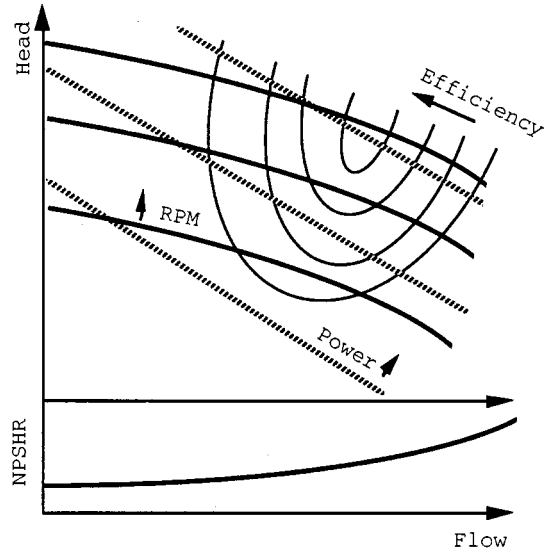


FIGURE 3.11.4. A full range of performance information should be available from the pump manufacturer, and this may include the parameters shown.

Several options are available to the designer for combining pumps in systems. Consider a comparison of the net effect between operating pumps in series or operating the same two pumps in parallel. Examples of this for pumps with characteristics such as centrifugal units are shown in Figure 3.11.5. It is clear that one way to achieve high pumping pressures with centrifugal pumps is to place a number of units in series. This is a related effect to what is found in *multistage* designs.

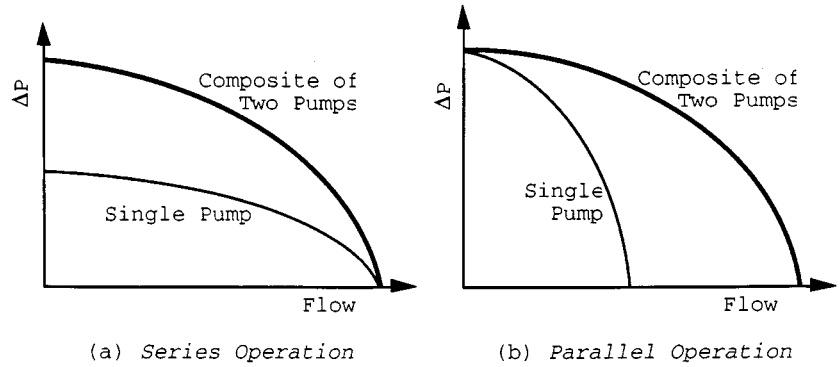


FIGURE 3.11.5. Series (a) and parallel (b) operation of centrifugal pumps are possible. The resultant characteristics for two identical pumps are shown.

Fans

As noted earlier, fans are devices that cause air to move. This definition is broad and can include a flapping palm branch, but the discussion here deals only with devices that impart air movement due to *rotation of an impeller inside a fixed casing*. In spite of this limiting definition, a large variety of commercial designs are included.

Fans find application in many engineering systems. Along with the chillers and boilers, they are the heart of heating, ventilating, and air conditioning (HVAC) systems. When large physical dimensions of a unit are not a design concern (usually the case), centrifugal fans are favored over axial flow units for HVAC applications. Many types of fans are found in *power plants*. Very large fans are used to furnish air to the boiler, as well as to draw or force air through cooling towers and pollution-control equipment. *Electronic cooling* finds applications for small units. Even automobiles have several fans in them. Because of the great engineering importance of fans, several organizations publish rating and testing criteria (see, for example, ASME, 1990).

Generally fans are classified according to how the air flows through the impeller. These flows may be *axial* (essentially a propeller in a duct), *radial* (conceptually much like the centrifugal pumps discussed earlier), *mixed*, and *cross*. While there are many other fan designations, all industrial units fit one of these classifications. Mixed-flow fans are so named because both axial and radial flow occur on the vanes. Casings for these devices are essentially like those for axial-flow machines, but the inlet has a radial-flow component. On cross-flow impellers, the gas traverses the blading twice.

Characteristics of fans are shown in Figure 3.11.6. Since velocities can be high in fans, often both the total and the static pressure increases are considered. While both are not shown on this figure, the curves have similar variations. Of course the total ΔP will be greater than will the static value, the difference being the velocity head. This difference increases as the volume flow increases. At zero flow (the shutoff point), the static and total pressure difference values are the same. Efficiency variation shows a sharp optimum value at the design point. For this reason, it is critical that fan designs be carefully tuned to the required conditions.

A variety of vane type are found on fans, and the type of these is also used for fan classification. Axial fans usually have vanes of airfoil shape or vanes of uniform thickness. Some vane types that might be found on a centrifugal (radial-flow) fan are shown in Figure 3.11.7.

One aspect that is an issue in choosing fans for a particular application is fan efficiency. Typical efficiency comparisons of the effect of blade type on a centrifugal fan are shown in Figure 3.11.8. Since velocities can be high, the value of aerodynamic design is clear. Weighing against this are cost and other factors.

An additional aspect that may be important in the choice of fans is noise generation. This may be most critical in HVAC applications. It is difficult to describe noise characteristics in brief terms because of the frequency-dependent nature of these phenomena. However, a comparison of specific sound power

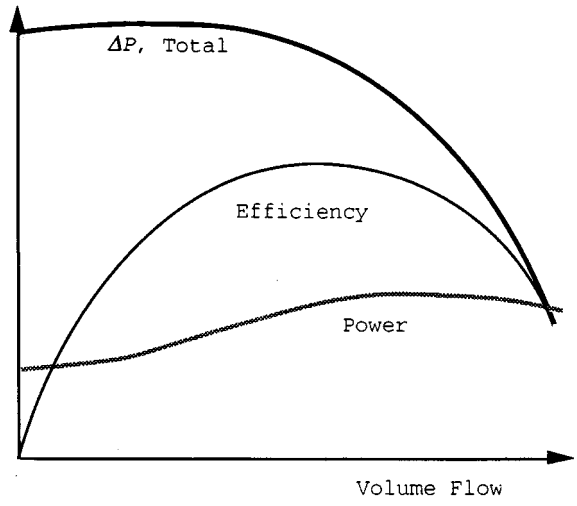


FIGURE 3.11.6. Shown are characteristics of a centrifugal fan. The drawbacks to operating away from optimal conditions are obvious from the efficiency variation.

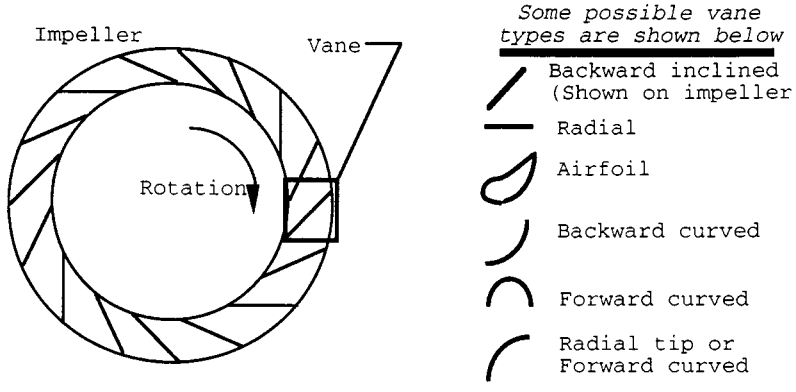


FIGURE 3.11.7. A variety of vane types that might be used on a centrifugal fan are shown.

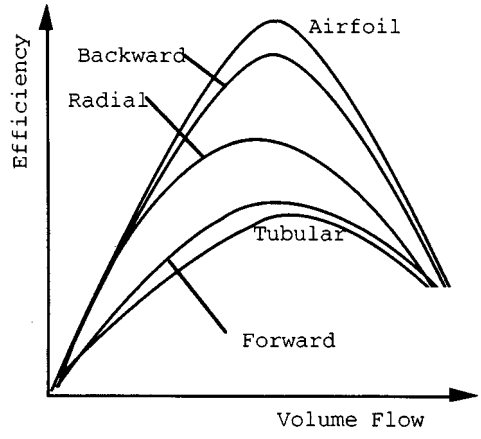


FIGURE 3.11.8. Efficiency variation with volume flow of centrifugal fans for a variety of vane types is shown.

level (usually denoted by K_w) shows backward-curved centrifugal fans with aerodynamic blades perform best among the designs. Details of noise characteristics are given elsewhere (ASHRAE, 1991).

While each type of fan has some specific qualities for certain applications, most installations use centrifugal (radial-flow) fans. A primary exception is for very-high-flow, low-pressure-rise situations where axial (propeller) fans are used.

Similarities exist between fans and pumps because the fluid density essentially does not vary through either type of machine. Of course, in pumps this is because a liquid can be assumed to be incompressible. In fans, a gas (typically air) is moved with little pressure change. As a result, the gas density can be taken to be constant. Since most fans operate near atmospheric pressure, the ideal gas assumptions can be used in determining gas properties.

Flow control in fan applications, where needed, is a very important design concern. Methods for accomplishing this involve use of dampers (either on the inlet or on the outlet of the fan), variable pitch vanes, or variable speed control. Dampers are the least expensive to install, but also the most inefficient in terms of energy use. Modern solid state controls for providing a variable frequency power to the drive motor is becoming the preferred control method, when a combination of initial and operating costs is considered.

Defining Terms

Cavitation: Local liquid conditions allow vapor voids to form (boiling).
NPSH: Net positive suction head is the difference between the local absolute pressure of a liquid and the thermodynamic saturation pressure of the liquid based upon the temperature of the liquid. Applies to the inlet of a pump.
NPSHA: Actual net positive suction head is the NPSH at the given state of operation of a pump.
NPSHR: Required net positive suction head is the amount of NPSH required by a specific pump for a given application.

References

ASHRAE, 1991. *ASHRAE Handbook 1991, HVAC Applications,* American Society of Heating, Refrigerating, and Air Conditioning Engineers, Atlanta, Chapter 42.

ASME, 1990. *ASME Performance Test Codes, Code on Fans*, ASME PTC 11-1984 (reaffirmed 1990), American Society of Mechanical Engineers, New York.

Boehm, R.F. 1987. *Design Analysis of Thermal Systems*, John Wiley and Sons, New York, 17–26.

Ewbank, T. 1842. *A Description and Historical Account of Hydraulic and Other Machines for Raising Water*, 2nd ed., Greeley and McElrath, New York.

Krutzsch, W.C. 1986. Introduction: classification and selection of pumps, in *Pump Handbook*, 2nd ed., I. Karassik et al., Eds., McGraw-Hill, New York, Chapter 1.

Lobanoff, V. and Ross, R. 1992. *Centrifugal Pumps: Design & Application*, 2nd ed., Gulf Publishing Company, Houston.

Turton, R.K. 1994. *Rotodynamic Pump Design*, Cambridge University Press, Cambridge, England.

Further Information

Dickson, C. 1988. *Pumping Manual*, 8th ed., Trade & Technical Press, Morden, England.

Dufour, J. and Nelson, W. 1993. *Centrifugal Pump Sourcebook*, McGraw-Hill, New York.

Fans. 1992. In *1992 ASHRAE Handbook, HVAC Systems and Equipment*, American Society of Heating, Refrigerating, and Air Conditioning Engineers, Atlanta, GA, Chapter 18.

Garay, P.N. 1990. *Pump Application Book*, Fairmont Press, Liburn, GA.

Krivchencko, G.I. 1994. *Hydraulic Machines, Turbines and Pumps*, 2nd ed., Lewis Publishers, Boca Raton, FL.

Stepanoff, A.J. 1993. *Centrifugal and Axial Flow Pumps: Theory, Design, and Application* (Reprint Edition), Krieger Publishing Company, Malabar, FL.

3.12 Liquid Atomization and Spraying

Rolf D. Reitz

Sprays are involved in many practical applications, including in the process industries (e.g., spray drying, spray cooling, powdered metals); in treatment applications (e.g., humidification, gas scrubbing); in coating applications (e.g., surface treatment, spray painting, and crop spraying); in spray combustion (e.g., burners, furnaces, rockets, gas turbines, diesel and port fuel injected engines); and in medicinal and printing applications. To be able to describe sprays it is necessary to obtain a detailed understanding of spray processes.

In the simplest case, the liquid to be sprayed is injected at a high velocity through a small orifice. Atomization is the process whereby the injected liquid is broken up into droplets. Atomization has a strong influence on spray vaporization rates because it increases the total surface area of the injected liquid greatly. Fast vaporization may be desirable in certain applications, but undesirable in others, where the liquid is required to impinge on a target. The trajectories of the spray drops are governed by the injected momentum of the drop, drag forces, and interactions between the drops and the surrounding gas. Control of these and other spray processes can lead to significant improvements in performance and in quality of product, and to reduction of emission of pollutants.

Spray Characterization

Practical atomizers generate sprays with a distribution of drop sizes, with average sizes in the diameter range from a few microns (1 μm = 10^{-6} m) to as large as 0.5 mm. It is important to quantify the details of the distribution depending on the application. For example, the smaller drops in a spray vaporize fast, and this is helpful to control ignition processes in some combustion systems. On the other hand, the large drops carry most of the mass and momentum of the injected liquid and these drops are able to penetrate into the high-pressure gases in engine combustion chambers. Typical average drop sizes for broad classes of sprays are shown schematically in Figure 3.12.1. It should be noted that the terminology used to describe sprays in Figure 3.12.1 is qualitative and is not universally agreed upon.

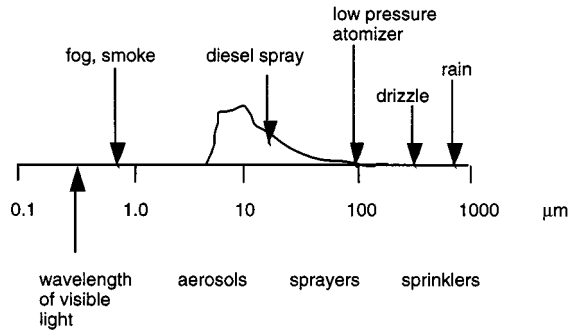


FIGURE 3.12.1 Typical average spray drop sizes for various classes of sprays. A representative size distribution is depicted for the diesel spray.

Methods for characterizing the size distribution of spray drops are discussed in References 1 and 2. A probability distribution function, $F(D)$, is introduced that represents the fraction of drops per unit diameter range about the diameter, D, as shown in Figure 3.12.2. The spray drop sizes span a range from a minimum diameter, D_{min}, to a maximum diameter, D_{max}. It is also convenient to introduce a mean or average drop diameter instead of having to specify the complete drop size distribution. The number median drop diameter (NMD) represents that drop whose diameter is such that 50% of the drops in the spray have sizes less than this size. Spray drop size distribution data can also be represented as a volume

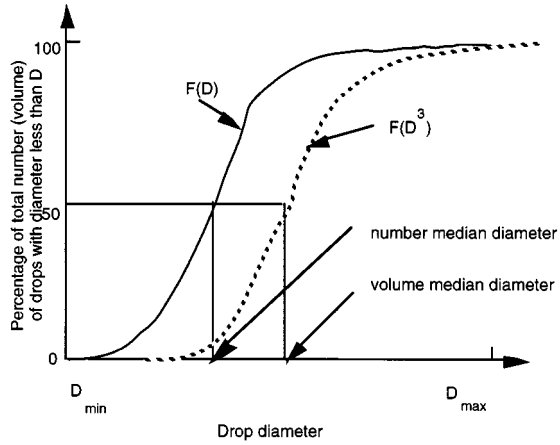


FIGURE 3.12.2 Cumulative spray drop number and volume distributions.

(or mass) distribution function, $F(D^3)$; this gives more weight to the large drops in the distribution. In this case, a volume median diameter (VMD) or a mass median diameter (MMD) can also be defined, as indicated in Figure 3.12.2.

Various other mean diameters are also in common use. These are summarized using the standard notation of Mugele and Evans[2] as

$$\left(D_{jk}\right)^{j-k} = \frac{\int_{D_{\min}}^{D_{\max}} D^j f(D)\, dD}{\int_{D_{\min}}^{D_{\max}} D^k f(D)\, dD} \tag{3.12.1}$$

where $f(D) = dF(D)/dD$ is the drop size probability density function (usually normalized such that $\int_{D_{\min}}^{D_{\max}} f(D)dD = 1$). Commonly used mean diameters are D_{10} (i.e., $j = 1$, $k = 0$, sometimes called the length mean diameter[3] and D_{32} (i.e., $j = 3$, $k = 2$, called the Sauter mean diameter or SMD). The Sauter mean diameter has a useful physical interpretation in combustion applications since drop vaporization rates are proportional to the surface area of the drop. It represents the size of that drop that has the same volume-to-surface area ratio as that of the entire spray.

Several distribution functions have been found to fit experimental data reasonably well. Among these are the Nukiyama–Tanasawa and the Rosin–Rammler distributions which have the general form[3] $f(D) = aD^p \exp(\{-bD\}^q)$, where the constants a, p, b, and q characterize the size distribution. The higher the parameter, q, the more uniform the distribution, and typically $1.5 < q < 4$. Other distributions have been proposed which consist of logarithmic transformations of the normal distribution, such as $f(D) = a \exp(-y^2/2)$, where $y = \delta \ln(\eta D/(D_{\max} - D))$, and a, δ, and η are constants. In this case, the smaller δ, the more uniform the size distribution. It should be noted that there is no theoretical justification for any of these size distributions. Spray drop size distributions can be measured nonintrusively by using optical laser diffraction and phase/Doppler instruments. A discussion of these techniques and their accuracy is reviewed by Chigier.[4]

Atomizer Design Considerations

Atomization is generally achieved by forcing a liquid or a liquid–gas mixture through a small hole or slit under pressure to create thin liquid sheets or jets moving at a high relative velocity with respect to the surrounding ambient gas. Desirable characteristics of atomizers include the ability to atomize the

liquid over a wide range of flow rates, low power requirements, and low susceptibility to blockage or fouling. In addition, atomizers should produce consistent sprays with uniform flow patterns in operation.

Atomizers can be broadly characterized as those producing hollow cone or solid cone sprays, as depicted in Figure 3.12.3. In solid cone (or full cone) sprays the spray liquid is concentrated along the spray axis, Figure 3.12.3(a). These sprays are useful in applications requiring high spray penetration, such as in diesel engines. In hollow cone sprays the axis region is relatively free of drops, giving wide spray dispersal, Figure 3.12.3(b). These sprays are often used in furnaces, gas turbines, and spray-coating applications.

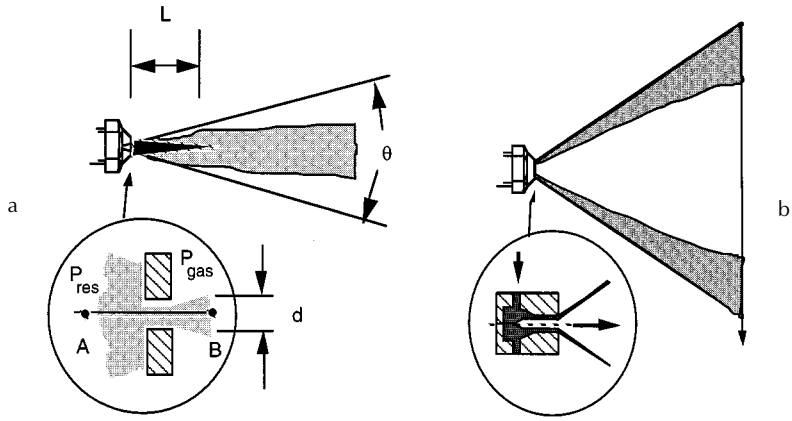


FIGURE 3.12.3 Schematic diagram of (a) solid cone and (b) hollow cone pressure atomizer sprays.

Many different atomizer designs are found in applications. Common atomizer types include pressure, rotary, twin-fluid (air-assist, air-blast, effervescent), flashing, electrostatic, vibratory, and ultrasonic atomizers, as discussed next.

Atomizer Types

In *pressure atomizers* atomization is achieved by means of a pressure difference, $\Delta P = P_{res} - P_{gas}$, between the liquid in the supply reservoir pressure, P_{res}, and the ambient medium pressure, P_{gas}, across a nozzle. The simplest design is the plain orifice nozzle with exit hole diameter, d, depicted in Figure 3.12.3(a). The liquid emerges at the theoretical velocity $U = \sqrt{2\Delta P/\rho_{liquid}}$, the (Bernoulli) velocity along the streamline A–B in Figure 3.12.3(a), where ρ_{liquid} is the density of the liquid. The actual injection velocity is less than the ideal velocity by a factor called the discharge coefficient, C_D, which is between 0.6 and 0.9 for plain hole nozzles. C_D accounts for flow losses in the nozzle.

Four main jet breakup regimes have been identified, corresponding to different combinations of liquid inertia, surface tension, and aerodynamic forces acting on the jet, as shown in Figure 3.12.4. At low injection pressures the low-velocity liquid jet breaks up due to the unstable growth of long-wavelength waves driven by surface tension forces (Rayleigh regime). As the jet velocity is increased, the growth of disturbances on the liquid surface is enhanced because of the interaction between the liquid and the ambient gas (the first and second wind-induced breakup regimes). At high injection pressures the high-velocity jet disintegrates into drops immediately after leaving the nozzle exit (atomization regime). Criteria for the boundaries between the regimes are available.[5] Aerodynamic effects are found to become very important relative to inertial effects when the jet Weber number, $We_j > 40$, where $We_j = \rho_{gas} U^2 d/\sigma$, ρ_{gas} is the gas density, and σ is the liquid surface tension.

Experiments show that the unstable growth of surface waves is aided by high relative velocities between the liquid and the gas, and also by high turbulence and other disturbances in the liquid and gas flows, and by the use of spray liquids with low viscosity and low surface tension.

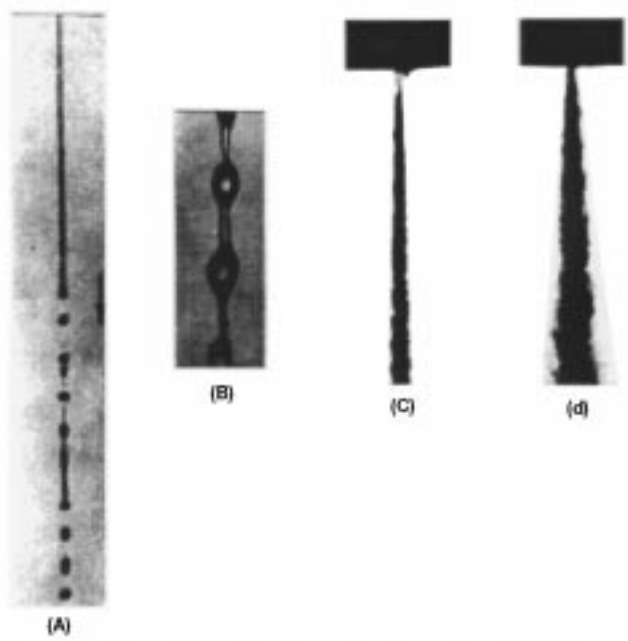


FIGURE 3.12.4 (a) Rayleigh breakup. Drop diameters are larger than the jet diameter. Breakup occurs many nozzle diameters downstream of nozzle. (b) First wind-induced regime. Drops with diameters of the order of jet diameter. Breakup occurs many nozzle diameters downstream of nozzle. (c) Second wind-induced regime. Drop sizes smaller than the jet diameter. Breakup starts some distance downstream of nozzle. (d) Atomization regime. Drop sizes much smaller than the jet diameter. Breakup starts at nozzle exit.

Liquid breakup characteristics such as the spray drop size, the jet breakup length, and the spray angle have been related to the unstable wave growth mechanism. The wavelengths and growth rates of the waves can be predicted using results from a linear stability analysis with[6]

$$\frac{\Lambda}{2} = 9.02\frac{\left(1+0.45Z^{0.5}\right)\left(1+0.4T^{0.7}\right)}{\left(1+0.87\,\mathrm{We}_2^{1.67}\right)^{0.6}} \tag{3.12.2a}$$

$$\Omega\left(\frac{\rho_1 a^3}{\sigma}\right)^{0.5} = \frac{0.34+0.38\,\mathrm{We}_2^{1.5}}{(1+Z)\left(1+1.4T^{0.6}\right)} \tag{3.12.2b}$$

where Λ is the wavelength, Ω is the growth rate of the most unstable surface wave, and a is the liquid jet radius. The maximum wave growth rate increases, and the corresponding wavelength decreases with increasing Weber number, $\mathrm{We}_2 = \rho_{gas}U^2a/\sigma$, where U is the relative velocity between the liquid and the gas. The liquid viscosity appears in the Ohnesorge number, $Z = \mathrm{We}_1^{1/2}/\mathrm{Re}_1$. Here, the Weber number We_1 is based on the liquid density, the Reynolds number is $\mathrm{Re}_1 = Ua/\nu_1$, ν_1 is the liquid viscosity, and the parameter $T = Z\mathrm{We}_2^{1/2}$. The wave growth rate is reduced and the wavelength is increased as the liquid viscosity increases.

The size of the drops formed from the breakup process is often assumed to be proportional to the wavelength of the unstable surface waves in modeling studies.[6] However, the drop sizes in the primary breakup region near the nozzle exist have also been found to be influenced by the length scale of the energy-containing eddies in the turbulent liquid flow.[7] There is uncertainty about atomization mechanisms since spray measurements are complicated by the high optical density of the spray in the breakup region

(e.g., see Figure 3.12.4(d)). As the drops penetrate into the ambient gas, they interact with each other through collisions and coalescence, and the spray drop size changes dynamically within the spray as a result of secondary breakup and vaporization effects. The drop trajectories are determined by complex drop drag, breakup, and vaporization phenomena, and by interactions with the turbulent gas flow.[6]

High-pressure diesel sprays are intermittent and are required to start and stop quickly without dribble between injections. This is accomplished by means of a plunger arrangement that is actuated by a cam and spring system in mechanical "jerk" pump systems (see Figure 3.12.5). Modern electronic injectors include electromagnetic solenoids that permit the duration and injection pressure to be varied independently of each other and of engine speed. Experiments on diesel-type injector nozzles show that the penetration distance, S, of the tip of the spray at time, t, after the beginning of the injection is given by[8]

$$\begin{aligned} S &= 0.39Ut\left(\rho_{\text{liquid}}/\rho_{\text{gas}}\right)^{1/2} && \text{for } t < t_b \\ S &= 2.46\sqrt{U\,dt}\left(\rho_{\text{liquid}}/\rho_{\text{gas}}\right)^{1/4} && \text{for } t > t_b \end{aligned} \tag{3.12.3}$$

where the "breakup time" is $t_b = 40.5d(\rho_{\text{liquid}}/\rho_{\text{gas}})^{1/2}/U$. The jet breakup length (see Figure 3.12.3(a)), $L = Ut_b$ is independent of the injection velocity. On the other hand, for low-speed jets, or for jets injected into a low-gas-density environment, $t_b = 1.04C\,(\rho_{\text{liquid}}\,d^3/\sigma)^{1/2}$, where C is a constant typically between 12 and 16 and σ is the surface tension. In this case L increases with the injection velocity.[9] The functional form of the above jet breakup time and length correlations can be derived for an inviscid liquid in the limits of large and small Weber number, We_2 from the unstable wave growth rate in Equation (3.12.2) with $t_b \sim \Omega^{-1}$.

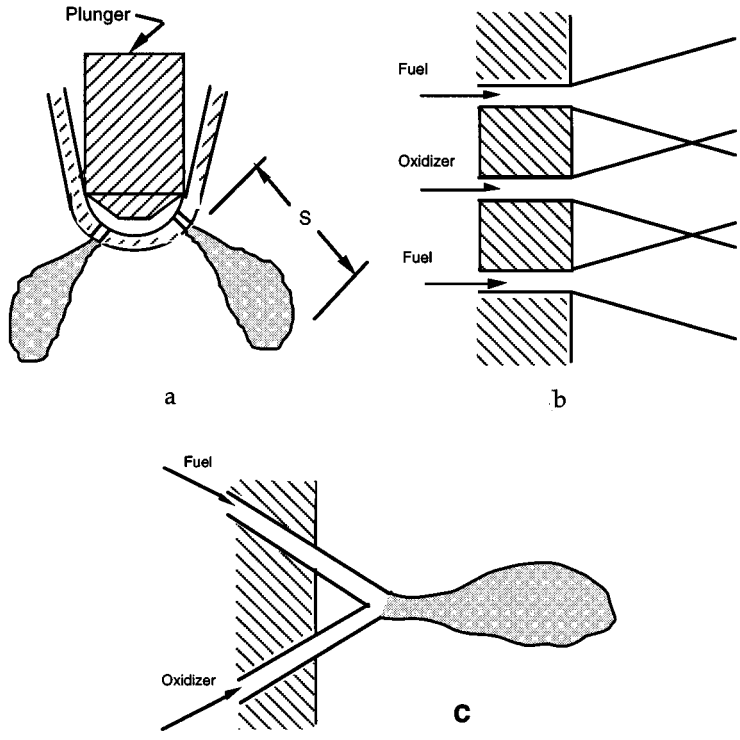


FIGURE 3.12.5 (a) Diesel injector multihole spray nozzle, (b) showerhead, and (c) doublet impingement nozzles.

For high-speed diesel-type jets in the atomization regime the resulting spray diverges in the form of a cone with cone angle, θ, that is usually in the range from 5 to 20°. θ increases with gas density following $\tan\theta = A(\rho_{gas}/\rho_{liquid})^{1/2}$, where A is a constant that depends on the nozzle passage length and (weakly) on the injection velocity.[9] Very high injection pressures are required to produce small drops. In diesel engines ΔP is typically as high as 200 Mpa, and drops are produced with mean diameters of the order of 10 μm (see Figure 3.12.1). Drop size correlations have been proposed for plain-orifice sprays, such as that presented in Table 3.12.1.[3] Note, however, that these correlations do not account for the fact that the spray drop size varies with time, and from place to place in the spray. Moreover, experimental correlations often do not include some parameters that are known to influence spray drop sizes, such as the nozzle passage length and its entrance geometry. Therefore, overall drop size correlations should only be used with caution.

TABLE 3.12.1 Representative Drop Size Correlations for Various Spray Devices (Dimensional quantities are in SI units, kg, m, s)

Device	Correlation	Notes
Plain orifice	$\text{SMD} = 3.08\nu_1^{0.385}\left(\rho_{liquid}\sigma\right)^{0.737}\rho_{gas}^{0.06}\,\Delta P^{-0.54}$	Use SI units
Fan spray	$\text{SMD} = 2.83d_h\left(\sigma\mu_{liquid}^2/\rho_{gas}d_h^3\,\Delta P^2\right)^{0.25}$ $+ 0.26d_h\left(\sigma\rho_{liquid}/\rho_{gas}d_h\,\Delta P\right)^{0.25}$	d_h = nozzle hydraulic diameter
Rotary atomizer	$\text{SMD} = 0.119Q^{0.1}\sigma^{0.5}/Nd^{0.5}\rho_{liquid}^{0.4}\mu_{liquid}^{0.1}$	N = rotational speed (rev/sec), Q = volumetric flow rate, A_{inj} U
Pressure swirl	$\text{SMD} = 4.52\left(\sigma\mu_{liquid}^2/\rho_{gas}\,\Delta P^2\right)^{0.25}\left(t\cos\theta\right)^{0.25}$ $+ 0.39\left(\sigma\rho_{liquid}/\rho_{gas}\,\Delta P\right)^{0.25}\left(t\cos\theta\right)^{0.75}$ $t = 0.0114A_{inj}\rho_{liquid}^{1/2}d\cos\theta$	t = film thickness; θ = cone angle, d = discharge orifice diameter
Twin fluid/air blast	$\text{SMD} = 0.48d\left(\sigma/\rho_{gas}U^2d\right)^{0.4}\left(1+1/\text{ALR}\right)^{0.4}$ $+ 0.15d\left(\mu_{liquid}^2/\sigma\rho_{liquid}d\right)^{0.5}\left(1+1/\text{ALR}\right)$	ALR = air-to-liquid mass ratio
Prefilming air blast	$\text{SMD} = \left(1+1/\text{ALR}\right)\left[0.33d_h\left(\sigma/\rho_{gas}U^2d_p\right)^{0.6}\right.$ $\left.+ 0.068d_h\left(\mu_{liquid}^2/\sigma\rho_{liquid}d_p\right)^{0.5}\right]$	d_h = hydraulic diameter, d_p = prefilmer diameter, Figure 3.12.9
Ultrasonic	$\text{SMD} = \left(4\pi^3\sigma/\rho_{liquid}\omega^2\right)^{1/3}$	ω = vibration frequency

Source: Lefebvre, A.H., *Atomization and Sprays*, Hemisphere Publishing, New York, 1989. With permission.

The plain orifice design is also used in twin-fluid-type liquid rocket engines in showerhead and doublet designs (Figures 3.12.5b and 3.12.5c). In the case of doublet nozzles, shown in Figure 3.12.6c, the impinging jets create unstable liquid sheets which break up to produce the sprays. Drop size correlations are available for liquid sheets such as those formed by discharging the liquid through a rectangular slit (see *fan spray*, Table 3.12.1). Thin liquid sheets or slits lead to the production of small drops. The breakup mechanism of liquid sheets is also thought to involve the unstable growth of surface waves due to surface tension and aerodynamic forces.[5]

In *rotary atomizers* centrifugal forces are used to further enhance the breakup process. In this case the liquid is supplied to the center of a spinning disk and liquid sheets or ligaments are thrown off the edges of the disk. The drop size depends on the rotational speed of the disk, as indicated in Table 3.12.1.

A spinning wheel or cup (turbobell) is used in some spray-painting applications. The spray shape is controlled by supplying a coflowing stream of "shaping-air."

Centrifugal forces also play a role in the breakup mechanism of *pressure swirl* atomizers (*simplex* nozzles). These atomizers give wider spray cone angle than plain orifice nozzles, and are available in hollow cone and solid cone designs. As depicted in Figure 3.12.3(b) the spray liquid enters a swirl chamber tangentially to create a swirling liquid sheet. The air core vortex within the swirl chamber plays an important role in determining the thickness of the liquid sheet or film at the nozzle exit. This type of nozzle produces relatively coarse sprays. A representative SMD correction is listed in Table 3.12.1. The spray cone angle depends on the ratio of the axial and tangential liquid velocity components at the exit of the nozzle. This type of atomizer is not well suited for use in transient applications because it tends to dribble at start-up and to shut down when the air core is not fully formed.

The basic drawback of all pressure atomizers is that the flow rate depends on the square root of ΔP. The volumetric flow rate is $Q = A_{inj}\, U$, where A_{inj} is the liquid flow area at the nozzle exit, so that a factor of 20 increase in flow rate (a typical turndown ratio from idle to full load operation of a gas turbine engine) requires a factor of 400 increase in injection pressure.

This difficulty has led to so-called wide-range atomizer designs such as those shown in Figure 3.12.6. The *duplex* nozzle features two sets of tangential swirl ports; the primary (or pilot) supplies fuel at low flow rates, while the secondary ports become operational at high flow rates. Another variation is the *dual-orifice* nozzle which is conceptually two simplex nozzles arranged concentrically, one supplying the primary flow and the other supplying the secondary flow. The *spill-return* nozzle is a simplex nozzle with a rear passage that returns fuel to the injection pump. In this design the flow rate is controlled by the relative spill amount, and there are no small passages to become plugged. However, the fuel is always supplied at the maximum pressure which increases the demands on the injection pump. But high swirl is always maintained in the swirl chamber and good atomization is achieved even at low flow rates.

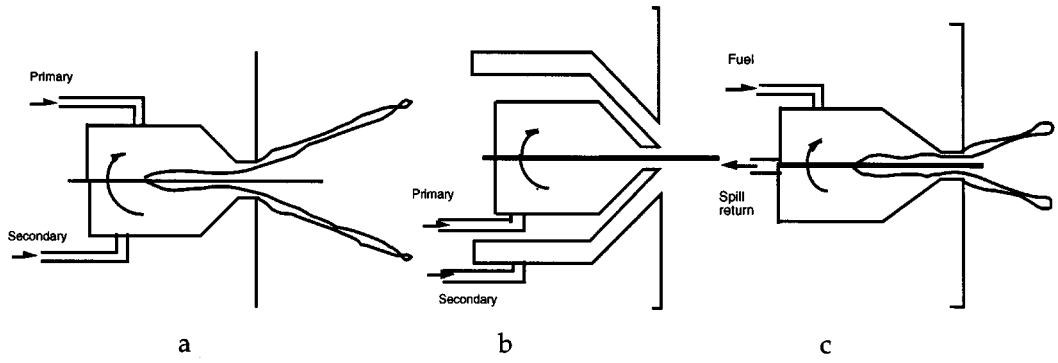


FIGURE 3.12.6 (a) Duplex, (b) dual orifice, and (c) spill-return-type nozzle designs.

In *twin-fluid injectors* atomization is aided by a flow of high-velocity gas through the injector passages. The high-velocity gas stream impinges on a relatively low-velocity liquid either internally (in *internal-mixing* nozzles, Figure 3.12.7) or externally (in *external-mixing* designs, Figure 3.12.8). The liquid and gas flows are typically swirled in opposite directions by means of swirl vanes to improve atomization. *Air-assist* refers to designs that use a relatively small amount of air at high (possibly sonic) velocities. *Air-blast* refers to designs that use large quantities of relatively low-velocity air which often supplies some of the air to help decrease soot formation in combustion systems[3]. (see Figure 3.12.9.)

In *flashing* and *effervescent* atomizers a two-phase flow is passed through the injector nozzle exit. In the former the bubbles are generated by means of a phase change which occurs as the liquid, containing a dissolved propellant gas or vapor, undergoes the pressure drop through the nozzle. This process is exploited in many household spray cans, but has the disadvantage of releasing the propellant gas required for atomization into the atmosphere. In the so-called effervescent atomizer, air bubbles are introduced

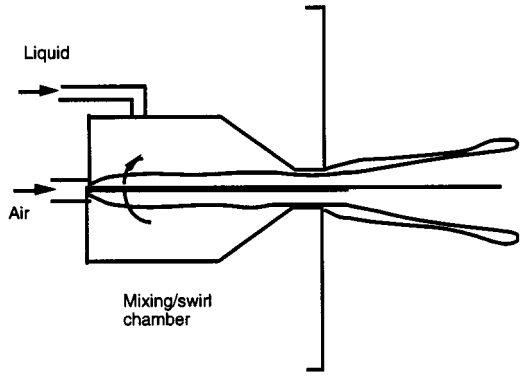


FIGURE 3.12.7 Internal-mixing twin-fluid injector design.

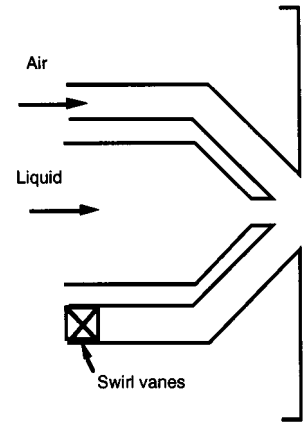


FIGURE 3.12.8 External-mixing twin-fluid injector design.

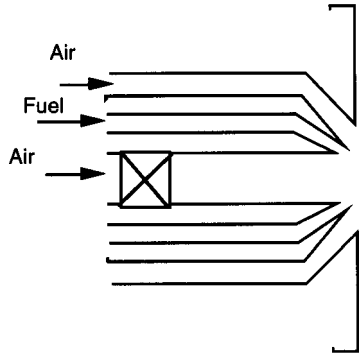


FIGURE 3.12.9 Prefilming air blast atomizer.

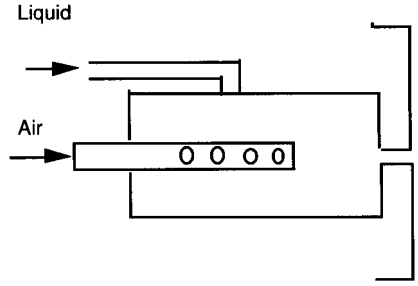


FIGURE 3.12.10 Internal-mixing, effervescent atomizer.

into the liquid upstream of the exit orifice, as depicted in Figure 3.12.10. The spray quality is found to depend weakly on the air bubble size and is independent of the nozzle exit diameter. This makes internal-mixing, air-assist atomizers very attractive for use with high-viscosity fluids and slurries where nozzle plugging would otherwise be a problem.[3]

In *electrostatic* atomizers the spray liquid is charged by applying a high-voltage drop across the nozzle. The dispersion of the spray drops is increased by exploiting electrical repulsive forces between the droplets. An electrostatic charge on the drops is also helpful in spray-coating applications, such as in automotive spray painting using electrostatic turbobell sprayers, since the charged drops are attracted to an oppositely charged target surface.

Other atomizer types include *vibratory* and *ultrasonic* atomizers (or *nebulizers*), where the drops are formed by vibrating the injector nozzle at high frequencies and at large amplitudes to produce short-wavelength disturbances to the liquid flow. Ultrasonic atomizers are used in inhalation therapy where very fine sprays (submicron sizes) are required, and an available representative drop size correlation is also listed in Table 3.12.1.

References

1. American Society for Testing and Materials (ASTM) Standard E799. 1988. Data Criteria and Processing for Liquid Drop Size Analysis.
2. Mugele, R. and Evans, H.D. 1951. Droplet size distributions in sprays, *Ind. Eng. Chem.*, 43, 1317–1324.
3. Lefebvre, A.H. 1989. *Atomization and Sprays*, Hemisphere Publishing, New York.
4. Chigier, N.A. 1983. Drop size and velocity instrumentation, *Prog. Energ. Combust. Sci.*, 9, 155–177.
5. Chigier, N. and Reitz, R.D. 1996. Regimes of jet breakup, in *Progress in Astronautics and Aeronautics Series*, K. Kuo, Ed., AIAA, New York, Chapter 4, pp. 109–135.
6. Reitz, R.D. 1988. Modeling atomization processes in high-pressure vaporizing sprays, *Atomisation Spray Technol.,* 3, 309–337.
7. Wu, P-K., Miranda, R.F., and Faeth, G.M. 1995. Effects of initial flow conditions on primary breakup of nonturbulent and turbulent round liquid jets, *Atomization Sprays*, 5, 175–196.
8. Hiroyasu, H. and Arai, M. 1978. Fuel spray penetration and spray angle in diesel engines, *Trans. JSAE*, 34, 3208.
9. Reitz, R.D. and Bracco, F.V. 1986. Mechanisms of breakup of round liquid jets, in *The Encyclopedia of Fluid Mechanics,* Vol. 3, N. Cheremisnoff, Ed., Gulf Publishing, Houston, TX, Chapter 10, 233–249.

Further Information

Information about recent work in the field of atomization and sprays can be obtained through participation in the Institutes for Liquid Atomization and Spraying Systems (ILASS-Americas, -Europe, -Japan, -Korea). These regional ILASS sections hold annual meetings. An international congress (ICLASS) is also held biennially. More information is available on the ILASS-Americas homepage at http://ucicl.eng.uci.edu/ilass. Affiliated with the ILASS organizations is the Institute's Journal publication *Atomization and Sprays* published by Begell House, Inc., New York.

3.13 Flow Measurement*

Alan T. McDonald and Sherif A. Sherif

This section deals with the measurement of mass flow rate or volume flow rate of a fluid in an enclosed pipe or duct system. Flow measurement in open channels is treated in Section 3.10 of this book.

The choice of a flow meter type and size depends on the required accuracy, range, cost, ease of reading or data reduction, and service life. Always select the simplest and cheapest device that gives the desired accuracy.

Direct Methods

Tanks can be used to determine the flow rate for steady liquid flows by measuring the volume or mass of liquid collected during a known time interval. If the time interval is long enough, flow rates may be determined precisely using tanks. Compressibility must be considered in gas volume measurements. It is not practical to measure the mass of gas, but a volume sample can be collected by placing an inverted "bell" over water and holding the pressure constant by counterweights. No calibration is required when volume measurements are set up carefully; this is a great advantage of direct methods.

Positive-displacement **flow meters** may be used in specialized applications, particularly for remote or recording uses. For example, household water and natural gas meters are calibrated to read directly in units of product. Gasoline metering pumps measure total flow and automatically compute the cost. Many positive-displacement meters are available commercially. Consult manufacturers' literature or Reference 7 for design and installation details.

Restriction Flow Meters for Flow in Ducts

Most restriction flow meters for internal flow (except the laminar flow element) are based on acceleration of a fluid stream through some form of nozzle, shown schematically in Figure 3.13.1. Flow separating from the sharp edge of the nozzle throat forms a recirculation zone shown by the dashed lines downstream from the nozzle. The main flow stream continues to accelerate from the nozzle throat to form a *vena contracta* at section (2) and then decelerates again to fill the duct. At the vena contracta, the flow area is a minimum, the flow streamlines are essentially straight, and the pressure is uniform across the channel section. The theoretical flow rate is

$$\dot{m}_{\text{theoretical}} = \frac{A_2}{\sqrt{1-\left(A_2/A_1\right)^2}}\sqrt{2\rho\left(p_1 - p_2\right)} \tag{3.13.1}$$

Equation (3.13.1) shows the general relationship for a **restriction flow meter**: Mass flow rate is proportional to the square root of the pressure differential across the meter taps. This relationship limits the flow rates that can be measured accurately to approximately a 4:1 range.

Several factors limit the utility of Equation (3.13.1) for calculating the actual mass flow rate through a meter. The actual flow area at section (2) is unknown when the vena contracta is pronounced (e.g., for orifice plates when D_t is a small fraction of D_1). The velocity profiles approach uniform flow only at large Reynolds number. Frictional effects can become important (especially downstream from the meter) when the meter contours are abrupt. Finally, the location of the pressure taps influences the differential pressure reading, $p_1 - p_2$.

The actual mass flow rate is given by

* The contributors of this chapter were asked to conform as much as possible to the nomenclature presented below, but define new terms pertaining to their particular area of specialty in the context of the chapter.

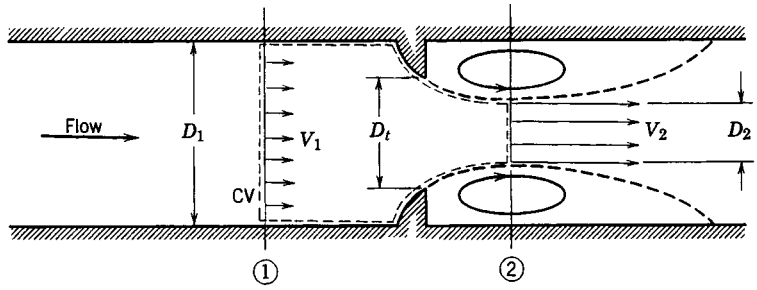


FIGURE 3.13.1 Internal flow through a generalized nozzle, showing control volume used for analysis.

$$\dot{m}_{\text{actual}} = \frac{CA_t}{\sqrt{1-\left(A_t/A_1\right)^2}}\sqrt{2\rho\left(p_1 - p_2\right)} \tag{3.13.2}$$

where C is an empirical *discharge coefficient*.

If $\beta = D_t/D_1$, then $(A_t/A_1)^2 = (D_t/D_1)^4 = \beta^4$, and

$$\dot{m}_{\text{actual}} = \frac{CA_t}{\sqrt{1-\beta^4}}\sqrt{2\rho\left(p_1 - p_2\right)} \tag{3.13.3}$$

where $1/(1 - \beta^4)^{1/2}$ is the *velocity correction factor*. Combining the discharge coefficient and velocity correction factor into a single *flow coefficient*,

$$K \equiv \frac{C}{\sqrt{1-\beta^4}} \tag{3.13.4}$$

yields the mass flow rate in the form:

$$\dot{m}_{\text{actual}} = KA_t\sqrt{2\rho\left(p_1 - p_2\right)} \tag{3.13.5}$$

Test data can be used to develop empirical equations to predict flow coefficients vs. pipe diameter and Reynolds number for standard metering systems. The accuracy of the equations (within specified ranges) is often adequate to use the meter without calibration. Otherwise, the coefficients must be measured experimentally.

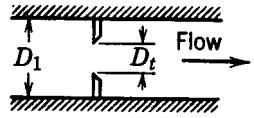
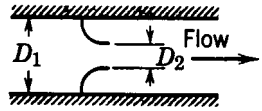
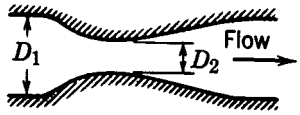
For the turbulent flow regime ($\text{Re}_D > 4000$), the flow coefficient may be expressed by an equation of the form:[7]

$$K = K_\infty + \frac{1}{\sqrt{1-\beta^4}}\frac{b}{\text{Re}_{D_1}^n} \tag{3.13.6}$$

where subscript ∞ denotes the flow coefficient at infinite Reynolds number and constants b and n allow for scaling to finite Reynolds numbers. Correlating equations and curves of flow coefficients vs. Reynolds number are given for specific metering elements in the next three subsections following the general comparison of the characteristics of orifice plate, flow nozzle, and venturi meters in Table 3.13.1 (see Reference 4).

Flow meter coefficients reported in the literature have been measured with fully developed turbulent velocity distributions at the meter inlet (Section 3.1). When a flow meter is installed downstream from

TABLE 3.13.1.Characteristics of Orifice, Flow Nozzle, and Venturi Flow Meters

Flow Meter Type	Diagram	Head Loss	Cost
Orifice		High	Low
Flow nozzle		Intermediate	Intermediate
Venturi		Low	High

a valve, elbow, or other disturbance, a straight section of pipe must be placed in front of the meter. Approximately 10 diameters of straight pipe upstream are required for venturi meters, and up to 40 diameters for orifice plate or flow nozzle meters. Some design data for incompressible flow are given below. The same basic methods can be extended to compressible flows.[7]

Orifice Plates

The orifice plate (Figure 3.13.2) may be clamped between pipe flanges. Since its geometry is simple, it is low in cost and easy to install or replace. The sharp edge of the orifice will not foul with scale or suspended matter. However, suspended matter can build up at the inlet side of a concentric orifice in a horizontal pipe; an eccentric orifice may be placed flush with the bottom of the pipe to avoid this difficulty. The primary disadvantages of the orifice are its limited capacity and the high permanent head loss caused by uncontrolled expansion downstream from the metering element.

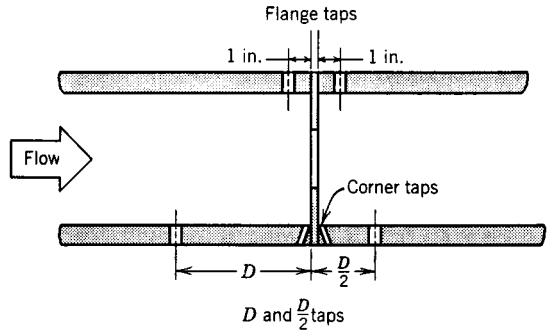


FIGURE 3.13.2 Orifice geometry and pressure tap locations.

Pressure taps for orifices may be placed in several locations as shown in Figure 3.13.2 (see Reference 7 for additional details). Since the location of the pressure taps influences the empirically determined flow coefficient, one must select handbook values of K consistent with the pressure tap locations.

The correlating equation recommended for a concentric orifice with corner taps is

$$C = 0.5959 + 0.0312\beta^{2.1} - 0.184\beta^{8} + \frac{91.71\beta^{2.5}}{\mathrm{Re}_{D_1}^{0.75}} \tag{3.13.7}$$

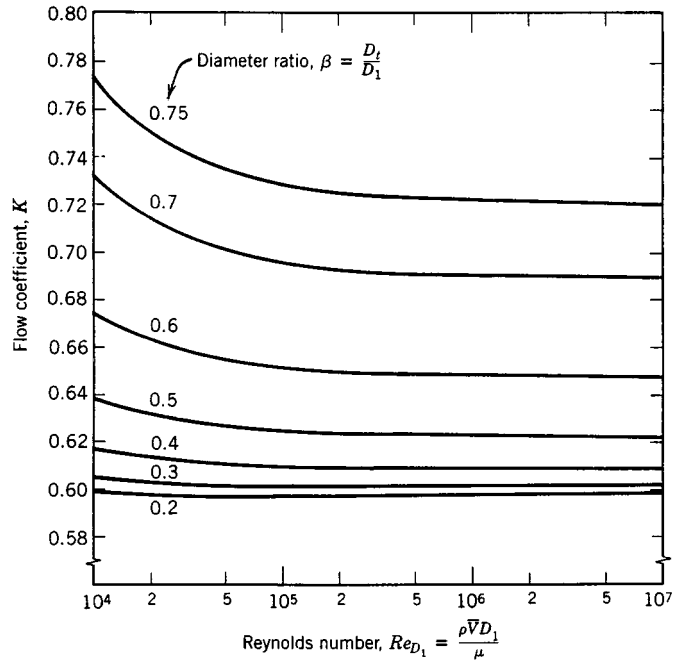


FIGURE 3.13.3 Flow coefficients for concentric orifices with corner taps.

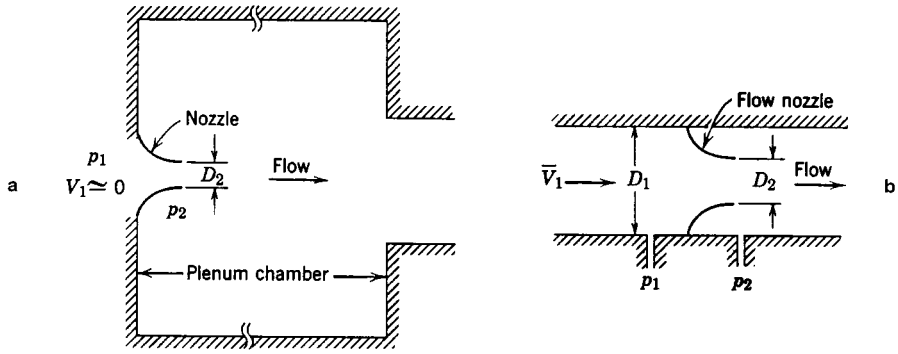


FIGURE 3.13.4 Typical installations of nozzle flow meters. (a) In plenum, (b) In duct.

Equation (3.13.7) predicts orifice discharge coefficients within ± 0.6% for $0.2 < \beta < 0.75$ and for $10^4 < \mathrm{Re}_{D_1} < 10^7$. Some flow coefficients calculated from Equation (3.13.7) are presented in Figure 3.13.3. Flow coefficients are relatively insensitive to Reynolds number for $\mathrm{Re}_{D_1} > 10^5$ when $\beta > 0.5$.

A similar correlating equation is available for orifice plates with D and $D/2$ taps. Flange taps require a different correlation for every line size. Pipe taps, located at $2^1/_2\,D$ and $8D$, no longer are recommended.

Flow Nozzles

Flow nozzles may be used as metering elements in either plenums or ducts, as shown in Figure 3.13.4; the nozzle section is approximately a quarter ellipse. Design details and recommended locations for pressure taps are given in Reference 7.

The correlating equation recommended for an ASME long-radius flow nozzles[7] is

$$C = 0.9975 - \frac{6.53\beta^{0.5}}{\mathrm{Re}_{D_1}^{0.5}} \tag{3.13.8}$$

Equation (3.13.8) predicts discharge coefficients for flow nozzles within ± 2.0% for $0.25 < \beta < 0.75$ for $10^4 < \mathrm{Re}_{D_1} < 10^7$. Some flow coefficients calculated from Equation 3.13.8 are presented in Figure 3.13.5. (K can be greater than 1 when the velocity correction factor exceeds 1.) For plenum installation, nozzles may be fabricated from spun aluminum, molded fiberglass, or other inexpensive materials. Typical flow coefficients are in the range $0.95 < K < 0.99$; the larger values apply at high Reynolds numbers. Thus, the mass flow rate can be computed within approximately ± 2% using $K = 0.97$.

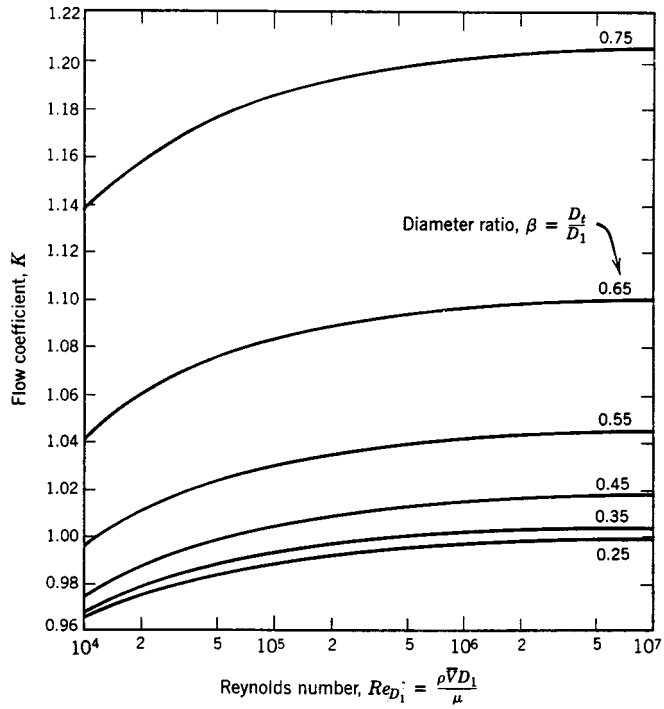


FIGURE 3.13.5 Flow coefficients for ASME long-radius flow nozzles.

Venturis

Venturi meters are generally made from castings machined to close tolerances to duplicate the performance of the standard design, so they are heavy, bulky, and expensive. The conical diffuser section downstream from the throat gives excellent pressure recovery; overall head loss is low. Venturi meters are self-cleaning because of their smooth internal contours.

Experimentally measured discharge coefficients for venturi meters range from 0.980 to 0.995 at high Reynolds numbers ($\mathrm{Re}_{D_1} > 2 \times 10^5$). Thus, $C = 0.99$ can be used to calculate mass flow rate within about ±1% at high Reynolds number.[7] Consult manufacturers' literature for specific information at Reynolds numbers below 10^5.

Orifice plates, flow nozzles, and venturis all produce pressure drops proportional to flow rate squared, according to Equation 3.13.4. In practice, a meter must be sized to accommodate the largest flow rate expected. Because the pressure drop vs. flow rate relationship is nonlinear, a limited range of flow rate

can be measured accurately. Flow meters with single throats usually are considered for flow rates over a 4:1 range.[7]

Unrecoverable head loss across a metering element may be expressed as a fraction of the differential pressure across the element. Unrecoverable head losses are shown in Figure 3.13.6.[7]

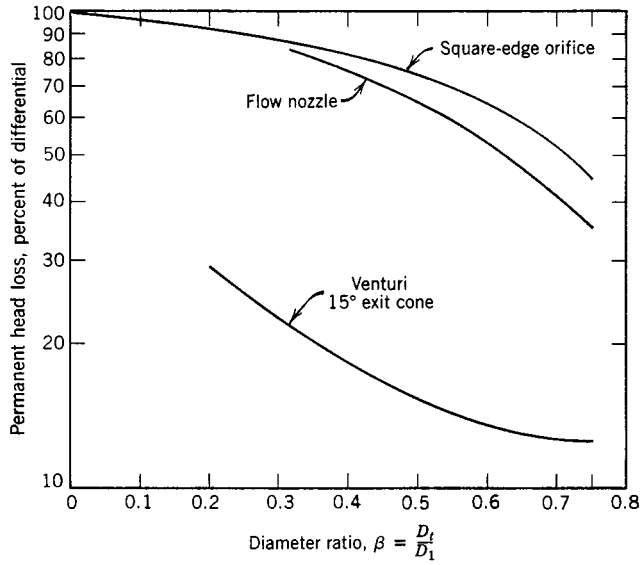


FIGURE 3.13.6 Permanent head loss produced by various flow metering elements.

Laminar Flow Elements

The laminar flow element (LFE)* produces a pressure differential proportional to flow rate. The LFE contains a metering section subdivided into many passages, each of small enough diameter to assure fully developed laminar flow. Because the pressure drop in laminar duct flow is directly proportional to flow rate, the pressure drop vs. flow rate relationship is linear. The LFE may be used with reasonable accuracy over a 10:1 flow rate range. Because the relationship between pressure drop and flow rate for laminar flow depends on viscosity, which is a strong function of temperature, fluid temperature must be known in order to obtain accurate metering.

An LFE costs approximately as much as a venturi, but is much lighter and smaller. Thus, the LFE is widely used in applications where compactness and extended range are important.

Linear Flow Meters

Several flow meter types produce outputs proportional to flow rate. Some meters produce signals without the need to measure differential pressure. The most common linear flow meters are discussed briefly below.

Float meters indicate flow rate directly for liquids or gases. An example is shown in Figure 3.13.7. The float is carried upward in the tapered clear tube by the flowing fluid until drag force and float weight are in equilibrium. Float meters (often called rotameters) are available with factory calibration for a number of common fluids and flow rate ranges.

A *turbine flow meter* is a free-running, vaned impeller mounted in a cylindrical section of tube (Figure 3.13.8). With proper design, the rate of rotation of the impeller may be made proportional to volume

* Patented and manufactured by Meriam Instrument Co., 10920 Madison Ave., Cleveland, OH.

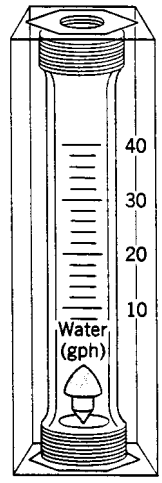


FIGURE 3.13.7 Float-type variable-area flow meter. (Courtesy of Dwyer Instrument Co., Michigan City, IN.)

flow rate over as much as a 100:1 flow rate range. Rotational speed of the turbine element can be sensed using a magnetic or modulated carrier pickup external to the meter. This sensing method therefore requires no penetrations or seals in the duct. Thus, turbine flow meters can be used safely to measure flow rates in corrosive or toxic fluids. The electrical signal can be displayed, recorded, or integrated to provide total flow information.

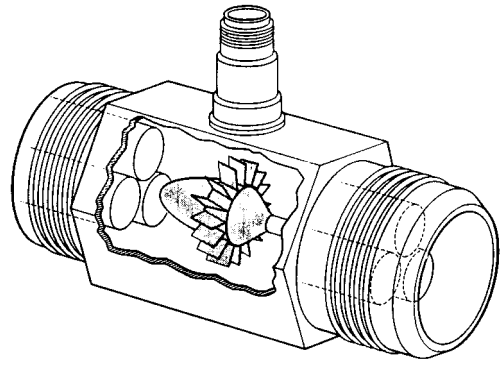


FIGURE 3.13.8 Turbine flow meter. (Courtesy of Potter Aeronautical Corp., Union, NJ.)

Vortex shedding from a bluff obstruction may be used to meter flow. Since Strouhal number, $St = fL/V$, is approximately constant ($St \cong 0.21$), vortex shedding frequency f is proportional to flow velocity. Vortex shedding causes velocity and pressure changes. Pressure, thermal, or ultrasonic sensors may be used to detect the vortex shedding frequency, and thus to infer the fluid velocity. (The velocity profile does affect the constancy of the shedding frequency.) Vortex flow meters can be used over a 20:1 flow rate range.[7]

Electromagnetic flow meters create a magnetic field across a pipe. When a conductive fluid passes through the field, a voltage is generated at right angles to the field and velocity vectors. Electrodes placed on a pipe diameter detect the resulting signal voltage, which is proportional to the average axial velocity when the profile is axisymmetric. The minimum flow speed should be above about 0.3 m/sec, but there are no restrictions on Reynolds number. The flow rate range normally quoted is 10:1.[7]

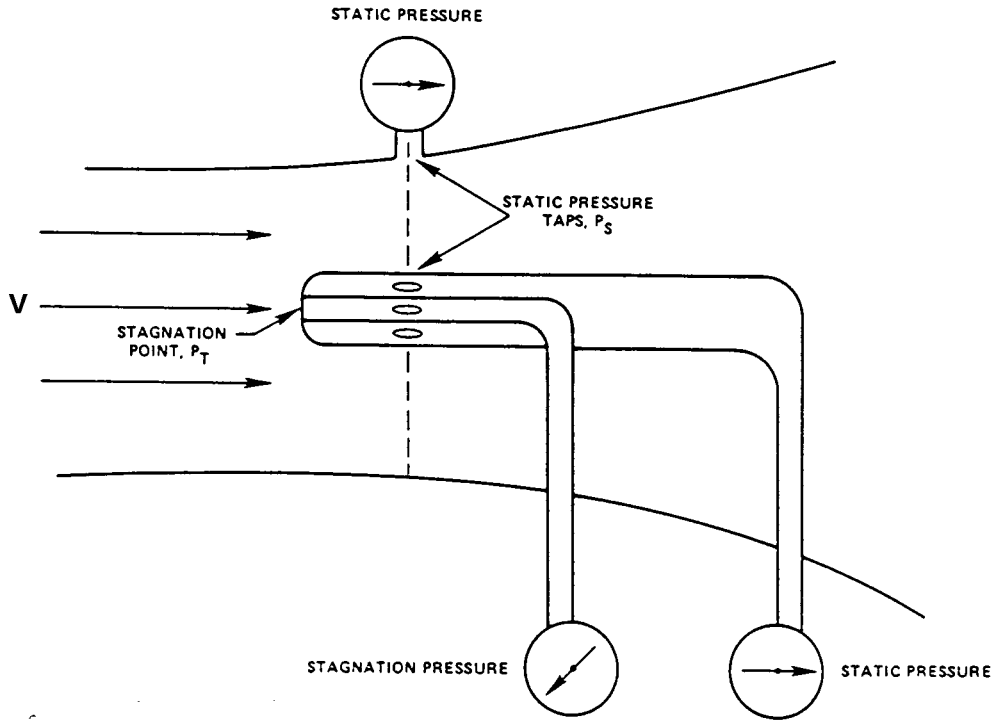
Ultrasonic flow meters also respond to average velocity at a pipe cross section. Two principal types of ultrasonic meters measure propagation time for clean liquids, and reflection frequency shift (Doppler effect) for flows carrying particulates. The speed of an acoustic wave increases in the flow direction and decreases when transmitted against the flow. For clean liquids, an acoustic path inclined to the pipe axis is used to infer flow velocity. Multiple paths are used to estimate volume flow rate accurately.

Doppler effect ultrasonic flow meters depend on reflection of sonic waves (in the megahertz range) from scattering particles in the fluid. When the particles move at flow speed, the frequency shift is proportional to flow speed; for a suitably chosen path, output is proportional to volume flow rate; ultrasonic meters may require calibration in place. One or two transducers may be used; the meter may be clamped to the outside of the pipe. Flow rate range is 10:1.[7]

A relatively new type of true mass flow meter is based on the effect of *Coriolis acceleration* on natural frequency of a bent tube carrying fluid. The bent tube is excited and its vibration amplitude measured. The instrument measures mass flow rate directly, and thus is ideal for two-phase or liquid–solid flow measurements. Pressure drop of the Coriolis meter may be high, but its useful flow rate range is 100:1.

Traversing Methods

In situations such as in air handling or refrigeration equipment, it may be impractical or impossible to install a fixed flow meter, but it may be possible to measure flow rate using a traversing technique.[1] To measure flow rate by **traverse**, the duct cross section is subdivided into segments of equal area. The fluid velocity is measured at the center of each area segment using a pitot tube (see Figure 3.13.9), a total head tube, or a suitable anemometer. The volume flow rate for each segment is approximated by the product of the measured velocity and segment area. Flow rate through the entire duct is the sum of these segmental flow rates. For details of recommended procedures see Reference 1 or 6.



$$V = \sqrt{2(P_T - P_S)/\rho}$$

FIGURE 3.13.9 Pitot tube.

Use of probes for traverse measurements requires direct access to the flow field. Pitot tubes give uncertain results when pressure gradients or streamline curvature are present, and they respond slowly. Two types of anemometers — **thermal anemometers** and **laser Doppler anemometers** — partially overcome these difficulties, although they introduce new complications.

Thermal anemometers use electrically heated tiny elements (either hot-wire or hot-film elements). Sophisticated feedback circuits are used to maintain the temperature of the element constant and to sense the input heating rate. The heating rate is related to the local flow velocity by calibration. The primary advantage of thermal anemometers is the small size of the sensing element. Sensors as small as 0.002 mm in diameter and 0.1 mm long are available commercially. Because the thermal mass of such tiny elements is extremely small, their response to fluctuations in flow velocity is rapid. Frequency responses to the 50-kHz range have been quoted.[3] Thus, thermal anemometers are ideal for measuring turbulence quantities. Insulating coatings may be applied to permit their use in conductive or corrosive gases or liquids.

Because of their fast response and small size, thermal anemometers are used extensively for research. Numerous schemes for treating the resulting data have been published. Digital processing techniques, including fast Fourier transforms, can be used to obtain mean values and moments, and to analyze signal frequency content and correlations.

Laser Doppler anemometers (LDAs) can be used for specialized applications where direct physical access to the flow field is difficult or impossible.[5] Laser beam(s) are focused to a small volume in the flow at the location of interest; laser light is scattered from particles present in the flow or introduced for this purpose. A frequency shift is caused by the local flow speed (Doppler effect). Scattered light and a reference beam are collected by receiving optics. The frequency shift is proportional to the flow speed; this relationship may be calculated, so there is no need for calibration. Since velocity is measured directly, the signal is unaffected by changes in temperature, density, or composition in the flow field. The primary disadvantages of LDAs are the expensive and fragile optical equipment and the need for careful alignment.

Viscosity Measurements

Viscometry is the technique of measuring the viscosity of a fluid. Viscometers are classified as rotational, capillary, or miscellaneous, depending on the technique employed. Rotational viscometers use the principle that a rotating body immersed in a liquid experiences a viscous drag which is a function of the viscosity of the liquid, the shape and size of the body, and the speed of its rotation. Rotational viscometers are widely used because measurements can be carried out for extended periods of time. Several types of viscometers are classified as rotational and Figure 3.13.10 is a schematic diagram illustrating a typical instrument of this type.

Capillary viscometry uses the principle that when a liquid passes in laminar flow through a tube, the viscosity of the liquid can be determined from measurements of the volume flow rate, the applied pressure, and the tube dimensions. Viscometers that cannot be classified either as rotational or capillary include the falling ball viscometer. Its method of operation is based on Stokes' law which relates the viscosity of a Newtonian fluid to the velocity of a sphere falling in it. Falling ball viscometers are often employed for reasonably viscous fluids. Rising bubble viscometers utilize the principle that the rise of an air bubble through a liquid medium gives a visual measurement of liquid viscosity. Because of their simplicity, rising bubble viscometers are commonly used to estimate the viscosity of varnish, lacquer, and other similar media.

Defining Terms

Flow meter: Device used to measure mass flow rate or volume flow rate of fluid flowing in a duct.

Restriction flow meter: Flow meter that causes flowing fluid to accelerate in a nozzle, creating a pressure change that can be measured and related to flow rate.

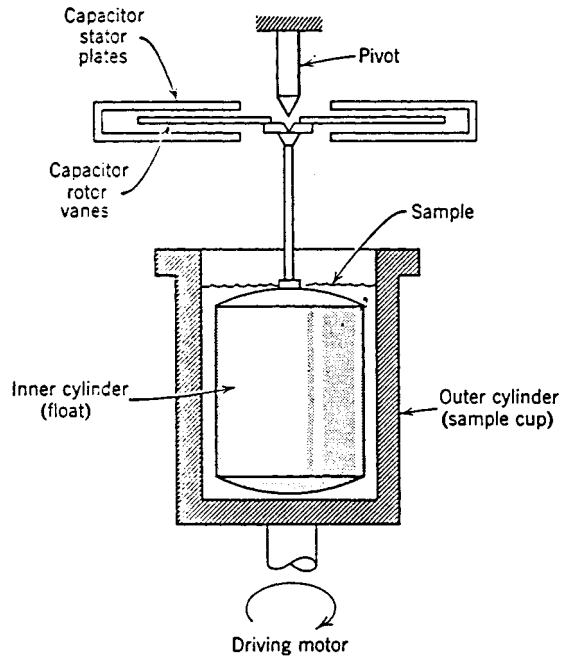


FIGURE 3.13.10 Rotational viscometer.

Thermal anemometer: Heated sensor used to infer local fluid velocity by sensing changes in heat transfer from a small electrically heated surface exposed to the fluid flow.

Traverse: Systematic procedure used to traverse a probe across a duct cross-section to measure flow rate through the duct.

References

1. *ASHRAE Handbook Fundamentals*. 1981. American Society of Heating, Refrigerating, and Air Conditioning Engineers, Atlanta, GA.
2. Baker, R.C. *An Introductory Guide to Flow Measurement*. 1989. Institution of Mechanical Engineers, London.
3. Bruun, H.H. 1995. *Hot-Wire Anemometry: Principles and Signal Analysis*. Oxford University Press, New York.
4. Fox, R.W. and McDonald, A.T. 1992. *Introduction to Fluid Mechanics*, 4th ed., John Wiley & Sons, New York.
5. Goldstein, R.J., Ed. 1996. *Fluid Mechanics Measurements*, 2nd ed., Taylor and Francis, Bristol, PA.
6. ISO 7145, *Determination of Flowrate of Fluids in Closed Conduits or Circular Cross Sections Method of Velocity Determination at One Point in the Cross Section*, ISO UDC 532.57.082.25:532.542. International Standards Organization, Geneva, 1982.
7. Miller, R.W. 1996. *Flow Measurement Engineering Handbook*, 3rd ed., McGraw-Hill, New York.
8. Spitzer, R.W., Ed. 1991. *Flow Measurement: A Practical Guide for Measurement and Control*. Instrument Society of America, Research Triangle Park, NC.
9. White, F.M. 1994. *Fluid Mechanics*, 3rd ed., McGraw-Hill, New York.

Further Information

This section presents only a survey of flow measurement methods. The references contain a wealth of further information. Baker[2] surveys the field and discusses precision, calibration, probe and tracer methods, and likely developments. Miller[7] is the most comprehensive and current reference available for information on restriction flow meters. Goldstein[5] treats a variety of measurement methods in his recently revised and updated book. Spitzer[8] presents an excellent practical discussion of flow measurement. Measurement of viscosity is extensively treated in *Viscosity and Flow Measurement: A Laboratory Handbook of Rheology*, by Van Wazer, J.R., Lyons, J.W., Kim, K.Y., and Colwell, R.E., Interscience Publishers, John Wiley & Sons, New York, 1963.

3.14 Micro/Nanotribology

Bharat Bhushan

Introduction

The emerging field of micro/nanotribology is concerned with processes ranging from atomic and molecular scales to microscale, occurring during adhesion, friction, wear, and thin-film lubrication at sliding surfaces (Bhushan, 1995, 1997; Bhushan et al., 1995). The differences between conventional tribology or macrotribology and micro/nanotribology are contrasted in Figure 3.14.1. In macrotribology, tests are conducted on components with relatively large mass under heavily loaded conditions. In these tests, wear is inevitable and the bulk properties of mating components dominate the tribological performance. In micro/nanotribology, measurements are made on at least one of the mating components, with relatively small mass under lightly loaded conditions. In this situation, negligible wear occurs and the surface properties dominate the tribological performance.

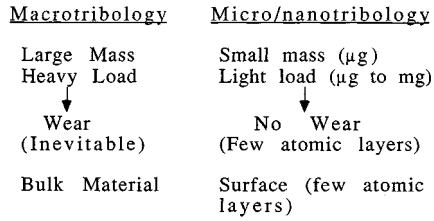


FIGURE 3.14.1 Comparison between macrotribology and microtribology.

Micro/nanotribological investigations are needed to develop fundamental understanding of interfacial phenomena on a small scale and to study interfacial phenomena in the micro- and nanostructures used in magnetic storage systems, microelectromechanical systems (MEMS), and other industrial applications (Bhushan, 1995, 1996, 1997). Friction and wear of lightly loaded micro/nanocomponents are highly dependent on the surface interactions (few atomic layers). These structures are generally lubricated with molecularly thin films. Micro- and nanotribological studies are also valuable in fundamental understanding of interfacial phenomena in macrostructures to provide a bridge between science and engineering (Bowden and Tabor, 1950, 1964; Bhushan and Gupta, 1997; Bhushan, 1996).

In 1985, Binnig et al. (1986) developed an "atomic force microscope" (AFM) to measure ultrasmall forces (less than 1 μN) present between the AFM tip surface and the sample surface. AFMs can be used for measurement of *all engineering surfaces* of any surface roughness, which may be either electrically conducting or insulating. AFM has become a popular surface profiler for topographic measurements on micro- to nanoscale. These are also used for scratching, wear, and nanofabrication purposes. AFMs have been modified in order to measure both normal and friction forces and this instrument is generally called a friction force microscope (FFM) or a lateral force microscope (LFM). New transducers in conjunction with an AFM can be used for measurements of elastic/plastic mechanical properties (such as load-displacement curves, indentation hardness, and modulus of elasticity) (Bhushan et al., 1996). A surface force apparatus (SFA) was first developed in 1969 (Tabor and Winterton, 1969) to study both static and dynamic properties of molecularly thin liquid films sandwiched between two molecularly smooth surfaces. SFAs are being used to study rheology of molecularly thin liquid films; however, the liquid under study has to be confined between molecularly smooth surfaces with radii of curvature on the order of 1 mm (leading to poorer lateral resolution as compared with AFMs) (Bhushan, 1995). Only AFMs/FFMs can be used to study *engineering surfaces* in the *dry and wet conditions* with *atomic resolution*. The scope of this section is limited to the applications of AFMs/FFMs.

At most solid–solid interfaces of technological relevance, contact occurs at numerous asperities with a range of radii; a sharp AFM/FFM tip sliding on a surface simulates just one such contact. Surface roughness, adhesion, friction, wear, and lubrication at the interface between two solids with and without liquid films have been studied using the AFM and FFM. The status of current understanding of micro/nanotribology of engineering interfaces follows.

Experimental Techniques

An AFM relies on a scanning technique to produce very high resolution, three-dimensional images of sample surfaces. The AFM measures ultrasmall forces (less than 1 nN) present between the AFM tip surface and a sample surface. These small forces are measured by measuring the motion of a very flexible cantilever beam having an ultrasmall mass. The deflection can be measured to with ±0.02 nm, so for a typical cantilever force constant of 10 N/m, a force as low as 0.2 nN can be detected. An AFM is capable of investigating surfaces of both conductors and insulators on an atomic scale. In the operation of a high-resolution AFM, the sample is generally scanned; however, AFMs are also available where the tip is scanned and the sample is stationary. To obtain atomic resolution with an AFM, the spring constant of the cantilever should be weaker than the equivalent spring between atoms. A cantilever beam with a spring constant of about 1 N/m or lower is desirable. Tips have to be sharp as possible. Tips with a radius ranging from 10 to 100 nm are commonly available.

In the AFM/FFM shown in Figure 3.14.2, the sample is mounted on a PZT tube scanner which consists of separate electrodes to precisely scan the sample in the *X*–*Y* plane in a raster pattern and to move the sample in the vertical (*Z*) direction. A sharp tip at the end of a flexible cantilever is brought in contact with the sample. Normal and frictional forces being applied at the tip–sample interface are measured simultaneously, using a laser beam deflection technique.

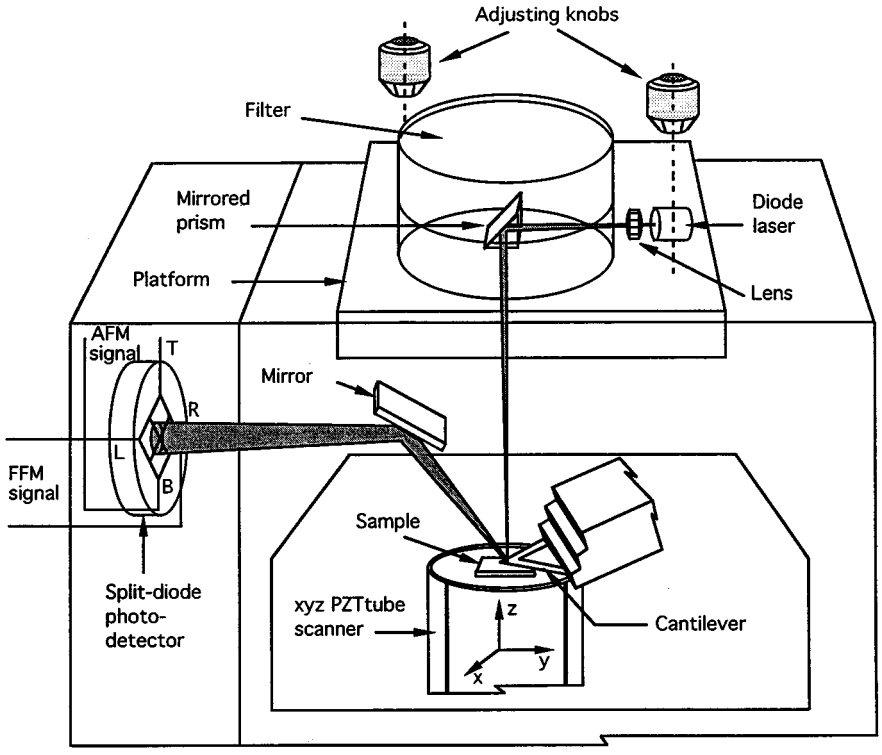


FIGURE 3.14.2 Schematic of a commercial AFM/FFM using laser beam deflection method.

Topographic measurements are typically made using a sharp tip on a cantilever beam with normal stiffness on the order of 0.5 N/m at a normal load of about 10 nN, and friction measurements are carried out in the load range of 10 to 150 nN. The tip is scanned in such a way that its trajectory on the sample forms a triangular pattern. Scanning speeds in the fast and slow scan directions depend on the scan area and scan frequency. A maximum scan size of 125 × 125 μm and scan rate of 122 Hz typically can be used. Higher scan rates are used for small scan lengths.

For nanoscale boundary lubrication studies, the samples are typically scanned over an area of 1 × 1 μm at a normal force of about 300 nN, in a direction orthogonal to the long axis of the cantilever beam (Bhushan, 1997). The samples are generally scanned with a scan rate of 1 Hz and the scanning speed of 2 μm/sec. The coefficient of friction is monitored during scanning for a desired number of cycles. After the scanning test, a larger area of 2 × 2 μm is scanned at a normal force of 40 nN to observe for any wear scar.

For microscale scratching, microscale wear, and nano-scale indentation hardness measurements, a sharp single-crystal natural diamond tip mounted on a stainless steel cantilever beam with a normal stiffness on the order of 25 N/m is used at relatively higher loads (1 to 150 μN). For wear studies, typically an area of 2 × 2 μm is scanned at various normal loads (ranging from 1 to 100 μN) for a selected number of cycles. For nanoindentation hardness measurements the scan size is set to zero and then normal load is applied to make the indents. During this procedure, the diamond tip is continuously pressed against the sample surface for about 2 sec at various indentation loads. The sample surface is scanned before and after the scratching, wear, or indentation to obtain the initial and the final surface topography, at a low normal load of about 0.3 μN using the same diamond tip. An area larger than the indentation region is scanned to observe the indentation marks. Nanohardness is calculated by dividing the indentation load by the projected residual area of the indents.

In measurements using conventional AFMs, the hardness value is based on the projected residual area after imaging the incident. Identification of the boundary of the indentation mark is difficult to accomplish with great accuracy, which makes the direct measurement of contact area somewhat inaccurate. A capacitive transducer with the dual capability of depth sensing as well as *in situ* imaging is used in conjunction with an AFM (Bhushan et al., 1996). This indentation system, called nano/picoindention, is used to make load-displacement measurements and subsequently carry out *in situ* imaging of the indent, if necessary. Indenter displacement at a given load is used to calculate the projected indent area for calculation of the hardness value. Young's modulus of elasticity is obtained from the slope of the unloading portion of the load-displacement curve.

Surface Roughness, Adhesion, and Friction

Solid surfaces, irrespective of the method of formation, contain surface irregularities or deviations from the prescribed geometrical form. When two nominally flat surfaces are placed in contact, surface roughness causes contact to occur at discrete contact points. Deformation occurs in these points and may be either elastic or plastic, depending on the nominal stress, surface roughness, and material properties. The sum of the areas of all the contact points constitutes the real area that would be in contact, and for most materials at normal loads this will be only a small fraction of the area of contact if the surfaces were perfectly smooth. In general, real area of contact must be minimized to minimize adhesion, friction, and wear (Bhushan and Gupta, 1997; Bhushan, 1996). Characterizing surface roughness is therefore important for predicting and understanding the tribological properties of solids in contact.

Surface roughness most commonly refers to the variations in the height of the surface relative to a reference plane (Bowden and Tabor, 1950; Bhushan, 1996). Commonly measured roughness parameters, such as standard deviation of surface heights (rms), are found to be scale dependent and a function of the measuring instrument, for any given surface, Figure 3.14.3 (Poon and Bhushan, 1995). The topography of most engineering surfaces is fractal, possessing a self-similar structure over a range of scales. By using fractal analysis one can characterize the roughness of surfaces with two scale-independent fractal parameters D and C which provide information about roughness at all length scales (Ganti and

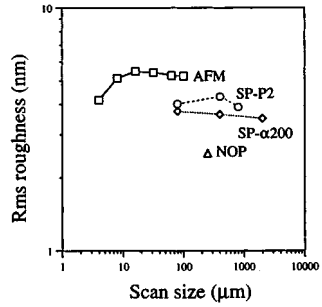


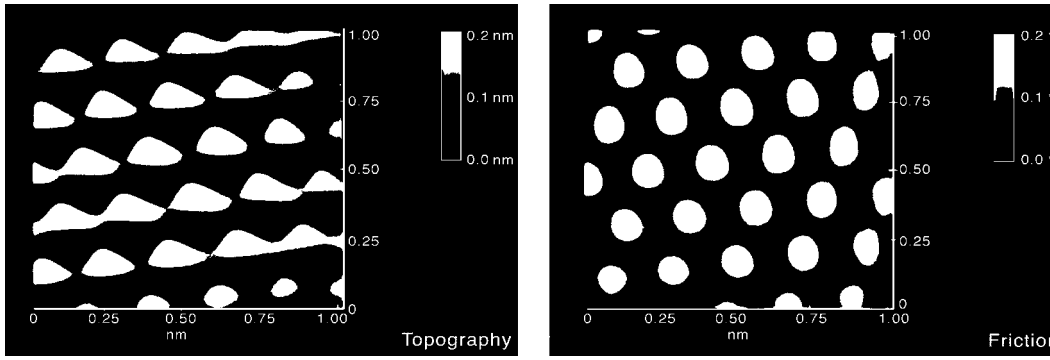
FIGURE 3.14.3 Scale dependence of standard deviation of surface heights (rms) for a glass–ceramic substrate, measured using an AFM, a stylus profiler (SP-P2 and SP-α 200), and a noncontact optical profiler (NOP).

Bhushan, 1995; Bhushan, 1995). These two parameters are instrument independent and are unique for each surface. *D* (generally ranging from 1 to 2) primarily relates to distribution of different frequencies in the surface profile, and *C* to the amplitude of the surface height variations at all frequencies. A fractal model of elastic plastic contact has been used to predict whether contacts experience elastic or plastic deformation and to predict the statistical distribution of contact points.

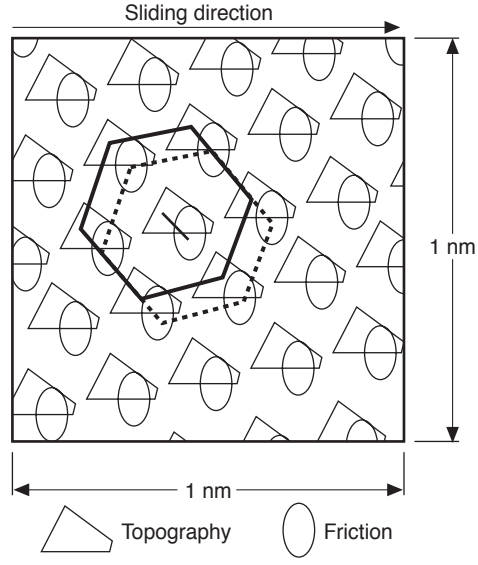
Based on atomic-scale friction measurements of a well-characterized freshly cleaved surface of highly oriented pyrolytic graphite (HOPG), the atomic-scale friction force of HOPG exhibits the same periodicity as that of corresponding topography (Figure 3.14.4(a)), but the peaks in friction and those in topography were displaced relative to each other (Figure 3.14.4(b)). A Fourier expansion of the interatomic potential has been used to calculate the conservative interatomic forces between atoms of the FFM tip and those of the graphite surface. Maxima in the interatomic forces in the normal and lateral directions do not occur at the same location, which explains the observed shift between the peaks in the lateral force and those in the corresponding topography. Furthermore, the observed local variations in friction force were explained by variation in the intrinsic lateral force between the sample and the FFM tip, and these variations may not necessarily occur as a result of an atomic-scale stick–slip process.

Friction forces of HOPG have also been studied. Local variations in the microscale friction of cleaved graphite are observed, which arise from structural changes that occur during the cleaving process. The cleaved HOPG surface is largely atomically smooth, but exhibits line-shaped regions in which the coefficient of friction is more than an order of magnitude larger. Transmission electron microscopy indicates that the line-shaped regions consist of graphite planes of different orientation, as well as of amorphous carbon. Differences in friction can also be seen for organic mono- and multilayer films, which again seem to be the result of structural variations in the films. These measurements suggest that the FFM can be used for structural mapping of the surfaces. FFM measurements can be used to map chemical variations, as indicated by the use of the FFM with a modified probe tip to map the spatial arrangement of chemical functional groups in mixed organic monolayer films. Here, sample regions that had stronger interactions with the functionalized probe tip exhibited larger friction. For further details, see Bhushan (1995).

Local variations in the microscale friction of scratched surfaces can be significant and are seen to depend on the local surface slope rather than on the surface height distribution (Bhushan, 1995). Directionality in friction is sometimes observed on the macroscale; on the microscale this is the norm (Bhushan, 1995). This is because most “engineering” surfaces have asymmetric surface asperities so that the interaction of the FFM tip with the surface is dependent on the direction of the tip motion. Moreover, during surface-finishing processes material can be transferred preferentially onto one side of the asperities, which also causes asymmetry and directional dependence. Reduction in local variations and in the directionality of frictional properties therefore requires careful optimization of surface roughness distributions and of surface-finishing processes.



a



b

FIGURE 3.14.4 (a) Gray-scale plots of surface topography (left) and friction profiles (right) of a 1 × 1 nm area of freshly cleaved HOPG, showing the atomic-scale variation of topography and friction, (b) diagram of superimposed topography and friction profiles from (a); the symbols correspond to maxima. Note the spatial shift between the two profiles.

Table 3.14.1 shows the coefficient of friction measured for two surfaces on micro- and macroscales. The coefficient of friction is defined as the ratio of friction force to the normal load. The values on the microscale are much lower than those on the macroscale. When measured for the small contact areas and very low loads used in microscale studies, indentation hardness and modulus of elasticity are higher than at the macroscale. This reduces the degree of wear. In addition, the small apparent areas of contact reduce the number of particles trapped at the interface, and thus minimize the "ploughing" contribution to the friction force.

At higher loads (with contact stresses exceeding the hardness of the softer material), however, the coefficient of friction for microscale measurements increases toward values comparable with those obtained from macroscale measurements, and surface damage also increases (Bhushan et al., 1995; Bhushan and Kulkarni, 1996). Thus, Amontons' law of friction, which states that the coefficient of

TABLE 3.14.1 Surface Roughness and Micro- and Macroscale Coefficients of Friction of Various Samples

Material	rms Roughness, nm	Microscale Coefficient of Friction vs. Si_3N_4 Tip[a]	Macroscale Coefficient of Friction vs. Alumina Ball[b]
Si(111)	0.11	0.03	0.18
C^+-implanted Si	0.33	0.02	0.18

[a] Tip radius of about 50 nm in the load range of 10 to 150 nN (2.5 to 6.1 GPa), a scanning speed of 5 m/sec and scan area of 1 × 1 μm.

[b] Ball radius of 3 mm at a normal load of 0.1 N (0.3 GPa) and average sliding speed of 0.8 mm/sec.

friction is independent of apparent contact area and normal load, does not hold for microscale measurements. These findings suggest that microcomponents sliding under lightly loaded conditions should experience very low friction and near zero wear.

Scratching, Wear, and Indentation

The AFM can be used to investigate how surface materials can be moved or removed on micro- to nanoscales, for example, in scratching and wear (Bhushan, 1995) (where these things are undesirable) and, in nanomachining/nanofabrication (where they are desirable). The AFM can also be used for measurements of mechanical properties on micro- to nanoscales. Figure 3.14.5 shows microscratches made on Si(111) at various loads after 10 cycles. As expected, the depth of scratch increases with load. Such microscratching measurements can be used to study failure mechanisms on the microscale and to evaluate the mechanical integrity (scratch resistance) of ultrathin films at low loads.

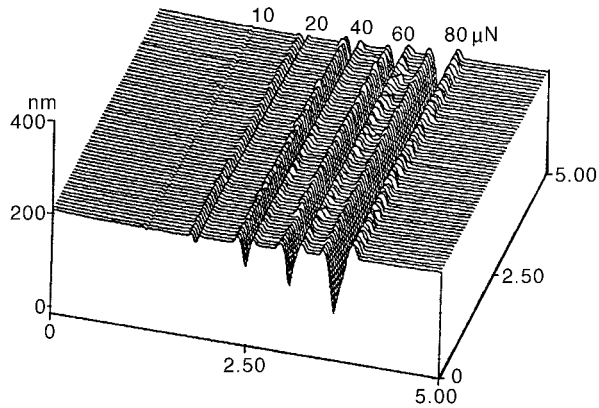


FIGURE 3.14.5 Surface profiles of Si(111) scratched at various loads. Note that the *x* and *y* axes are in micrometers and the *z* axis is in nanometers.

By scanning the sample in two dimensions with the AFM, wear scars are generated on the surface. The evolution of wear of a diamond-like carbon coating on a polished aluminum substrate is showing in Figure 3.14.6 which illustrates how the microwear profile for a load of 20 μN develops as a function of the number of scanning cycles. Wear is not uniform, but is initiated at the nanoscratches indicating that surface defects (with high surface energy) act as initiation sites. Thus, scratch-free surfaces will be relatively resistant to wear.

Mechanical properties, such as load-displacement curves, hardness, and modulus of elasticity can be determined on micro- to picoscales using an AFM and its modifications (Bhushan, 1995; Bhushan et al., 1995, 1996). Indentability on the scale of picometers can be studied by monitoring the slope of cantilever deflection as a function of sample traveling distance after the tip is engaged and the sample

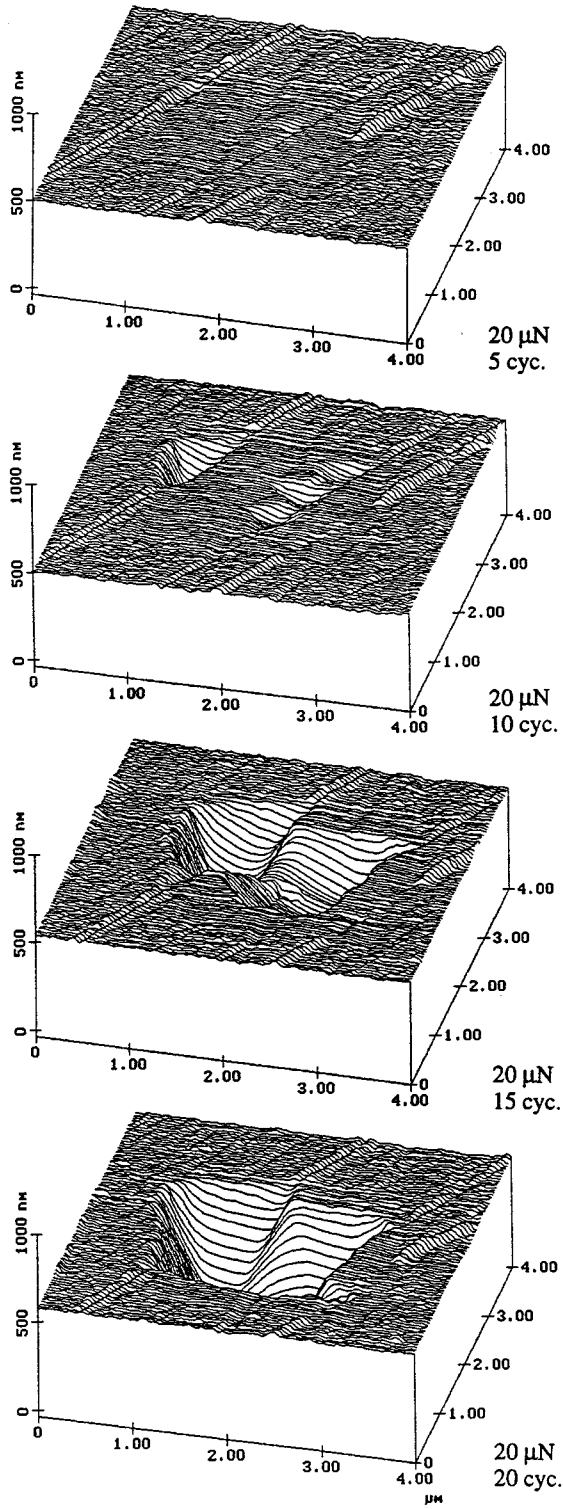


FIGURE 3.14.6 Surface profiles of diamond-like carbon-coated thin-film disk showing the worn region; the normal load and number of test cycles are indicated. (From Bhushan, B., *Handbook of Micro/Nanotribology*, CRC Press, Boca Raton, FL, 1995. With permission.)

is pushed against the tip. For a rigid sample, cantilever deflection equals the sample traveling distance; but the former quantity is smaller if the tip indents the sample. The indentation hardness on nanoscale of bulk materials and surface films with an indentation depth as small as 1 nm can be measured. An example of hardness data as a function of indentation depth is shown in Figure 3.14.7. A decrease in hardness with an increase in indentation depth can be rationalized on the basis that, as the volume of deformed materials increases, there is a higher probability of encountering material defects. AFM measurements on ion-implanted silicon surfaces show that ion implantation increases their hardness and, thus, their wear resistance (Bhushan, 1995). Formation of surface alloy films with improved mechanical properties by ion implantation is growing in technological importance as a means of improving the mechanical properties of materials.

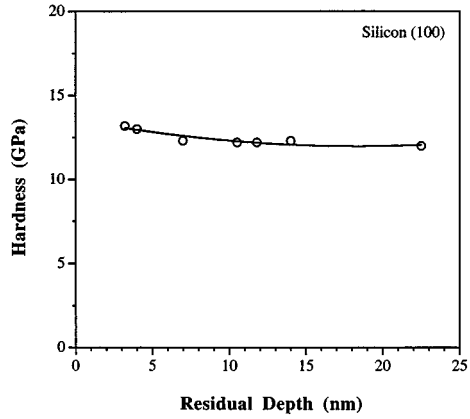


FIGURE 3.14.7 Indentation hardness as a function of residual indentation depth for Si(100).

Young's modulus of elasticity is calculated from the slope of the indentation curve during unloading (Bhushan, 1995; Bhushan et al., 1996). AFM can be used in a *force modulation mode* to measure surface elasticities: an AFM tip is scanned over the modulated sample surface with the feedback loop keeping the average force constant. For the same applied force, a soft area deforms more, and thus causes less cantilever deflection, than a hard area. The ratio of modulation amplitude to the local tip deflection is then used to create a *force modulation image*. The force modulation mode makes it easier to identify soft areas on hard substrates.

Detection of the transfer of material on a nanoscale is possible with the AFM. Indentation of C_{60}-rich fullerene films with an AFM tip has been shown to result in the transfer of fullerene molecules to the AFM tip, as indicated by discontinuities in the cantilever deflection as a function of sample traveling distance in subsequent indentation studies (Bhushan, 1995).

Boundary Lubrication

The "classical" approach to lubrication uses freely supported multimolecular layers of liquid lubricants (Bowden and Tabor, 1950, 1964; Bhushan, 1996). The liquid lubricants are chemically bonded to improve their wear resistance (Bhushan, 1995, 1996). To study depletion of boundary layers, the microscale friction measurements are made as a function of the number of cycles. For an example of the data of virgin Si(100) surface and silicon surface lubricated with about 2-nm-thick Z-15 and Z-Dol perfluoropolyether (PEPE) lubricants, see Figure 3.14.8. Z-Dol is PFPE lubricant with hydroxyl end groups. Its lubricant film was thermally bonded. In Figure 3.14.8, the unlubricated silicon sample shows a slight increase in friction force followed by a drop to a lower steady state value after some cycles. Depletion of native oxide and possible roughening of the silicon sample are responsible for the decrease in this

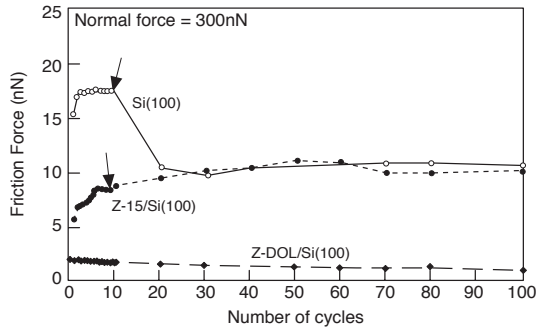


FIGURE 3.14.8 Friction force as a function of number of cycles using a silicon nitride tip at a normal force of 300 nN for the unlubricated and lubricated silicon samples.

friction force. The initial friction force for the Z-15-lubricated sample is lower than that of the unlubricated silicon and increases gradually to a friction force value comparable with that of the silicon after some cycles. This suggests the depletion of the Z-15 lubricant in the wear track. In the case of the Z-Dol-coated silicon sample, the friction force starts out to be low and remains low during the entire test. It suggests that Z-Dol does not get displaced/depleted as readily as Z-15. Additional studies of freely supported liquid lubricants showed that either increasing the film thickness or chemically bonding the molecules to the substrate with a mobile fraction improves the lubrication performance (Bhushan, 1997).

For lubrication of microdevices, a more effect approach involves the deposition of organized, dense molecular layers of long-chain molecules on the surface contact. Such monolayers and thin films are commonly produced by Langmuir–Blodgett (LB) deposition and by chemical grafting of molecules into self-assembled monolayers (SAMs). Based on the measurements, SAMs of octodecyl (C_{18}) compounds based on aminosilanes on a oxidized silicon exhibited a lower coefficient of friction of (0.018) and greater durability than LB films of zinc arachidate adsorbed on a gold surface coated with octadecylthiol (ODT) (coefficient of friction of 0.03) (Bhushan et al., 1995). LB films are bonded to the substrate by weak van der Waals attraction, whereas SAMs are chemically bound via covalent bonds. Because of the choice of chain length and terminal linking group that SAMs offer, they hold great promise for boundary lubrication of microdevices.

Measurement of ultrathin lubricant films with nanometer lateral resolution can be made with the AFM (Bhushan, 1995). The lubricant thickness is obtained by measuring the force on the tip as it approaches, contacts, and pushes through the liquid film and ultimately contacts the substrate. The distance between the sharp "snap-in" (owing to the formation of a liquid of meniscus between the film and the tip) at the liquid surface and the hard repulsion at the substrate surface is a measure of the liquid film thickness. This technique is now used routinely in the information-storage industry for thickness measurements (with nanoscale spatial resolution) of lubricant films, a few nanometers thick, in rigid magnetic disks.

References

Bhushan, B. 1995. *Handbook of Micro/Nanotribology*, CRC Press, Boca Raton, FL.

Bhushan, B. 1996. *Tribology and Mechanics of Magnetic Storage Devices*, 2nd ed., Springer, New York.

Bhushan, B. 1997. *Micro/Nanotribiology and Its Applications*, NATO ASI Series E: Applied Sciences, Kluwer, Dordrecht, Netherlands.

Bhushan, B. and Gupta, B.K. 1997. *Handbook of Tribology: Materials, Coatings and Surface Treatments*, McGraw-Hill, New York (1991); Reprint with corrections, Kreiger, Malabar, FL.

Bhushan, B. and Kulkarni, A.V. 1996. Effect of normal load on microscale friction measurements, *Thin Solid Films*, 278, 49–56.

Bhushan, B., Israelachvili, J.N., and Landman, U. 1995. Nanotribology: friction, wear and lubrication at the atomic scale, *Nature*, 374, 607–616.

Bhushan, B., Kulkarni, A.V., Bonin, W., and Wyrobek, J.T. 1996. Nano-indentation and pico-indentation measurements using capacitive transducer system in atomic force microscopy, *Philos. Mag.*, A74, 1117–1128.

Binning, G., Quate, C.F., and Gerber, Ch. 1986. Atomic force microscopy, *Phys. Rev. Lett.*, 56, 930–933.

Bowden, F.P. and Tabor, D. 1950; 1964. *The Friction and Lubrication of Solids*, Parts I and II, Clarendon, Oxford.

Ganti, S. and Bhushan, B. 1995. Generalized fractal analysis and its applications to engineering surfaces, *Wear*, 180, 17–34.

Poon, C.Y. and Bhushan, B. 1995. Comparison of surface roughness measurements by stylus profiler, AFM and non-contact optical profiler, *Wear*, 190, 76–88.

Tabor, D. and Winterton, R.H.S. 1969. The direct measurement of normal and retarded van der Walls forces, *Proc. R. Soc. London*, A312, 435–450.

Nomenclature for Fluid Mechanics

Symbol	Quantity	Unit		Dimensions (*MLtT*)
		SI	English	
a	Velocity of sound	m/sec	ft/sec	Lt^{-1}
a	Acceleration	m/sec^2	ft/sec^2	Lt^{-2}
A	Area	m^2	ft^2	L^2
b	Distance, width	m	ft	L
c_p	Specific heat, constant pressure	J/kg·K	ft·lb/lb$_m$·°R	$L^2t^{-2}T^{-1}$
c_v	Specific heat, constant volume	J/kg·K	ft·lb/lb$_m$·°R	$L^2t^{-2}T^{-1}$
C	Concentration	No./m^3	No./ft^3	L^{-3}
C	Coefficient	—	—	—
C	Empirical constant	—	—	—
D	Diameter	m	ft	L
D_H	Hydraulic diameter	m	ft	L
e	Total energy per unit mass	J/kg	ft·lb/lb$_m$	L^2t^{-2}
E	Total energy	J	ft·lb or Btu	ML^2t^{-2}
E	Modulus of leasticity	Pa	lb/ft^2	ML^1t^{-2}
Eu	Euler number	—	—	—
f	Friction factor	—	—	—
F	Force	N	lb	MLt^{-2}
Fr	Froude number	—	—	—
F_B	Buoyant force	N	lb	MLt^{-2}
g	Acceleration of gravity	m/sec^2	ft/sec^2	Lt^{-2}
g_0	Gravitation constant	kg·m/N·sec^2	lb$_m$·ft/lb·sec^2	—
G	Mass flow rate per unit area	kg/sec·m^2	lb$_m$/sec·ft^2	$ML^{-2}t^{-1}$
h	Head, vertical distance	m	ft	L
h	Enthalpy per unit mass	J/kg	ft·lb/lb$_m$	L^2t^{-2}
H	Head, elevation of hydraulic grade line	m	ft	L
I	Moment of inertia	m^4	ft^4	L^4
k	Specific heat ratio	—	—	—
K	Bulk modulus of elasticity	Pa	lb/ft^2	ML^1t^{-2}
K	Minor loss coefficient	—	—	—
L	Length	m	ft	L
L	Lift	N	lb	MLt^{-2}
l	Length, mixing length	m	ft	L
ln	Natural logarithm	—	—	—
m	Mass	kg	lb$_m$	M
m	Strength of source	m^3/sec	ft^3/sec	L^3t^{-1}
m	Mass flow rate	kg/sec	lb$_m$/sec	Mt^{-1}

Symbol	Quantity	Unit		Dimensions (*MLtT*)
		SI	**English**	
M	Molecular weight	—	—	—
$M\mathcal{V}$	Momentum per unit time	N	lb	MLt^{-2}
M	Mach number	—	—	—
n	Exponent, constant	—	—	—
n	Normal direction	m	ft	L
n	Manning roughness factor	—	—	—
n	Number of moles	—	—	—
N	Rotation speed	1/sec	1/sec	t^{-1}
NPSH	Net positive suction head	m	ft	L
p	Pressure	Pa	lb/ft^2	$ML^{-1}t^{-2}$
P	Height of weir	m	ft	L
P	Wetted perimeter	m	ft	L
q	Discharge per unit width	m^2/sec	ft^2/sec	L^2t^{-1}
q	Heat transfer per unit time	J/sec	Btu	ML^2t^{-3}
r	Radial distance	m	ft	L
R	Gas constant	J/kg·K	ft·lb/lb$_m$·°R	$L^2t^{-2}T^{-1}$
Re	Reynolds number	—	—	—
s	Distance	m	ft	L
s	Entropy per unit mass	J/kg·K	ft·lb/lb$_m$·°R	$L^2t^{-2}T^{-1}$
S	Entropy	J/K	ft·lb/°R	$ML^2t^{-2}T^{-1}$
S	Specific gravity, slope	—	—	—
t	Time	sec	sec	t
t	Distance, thickness	m	ft	L
T	Temperature	K	°R	T
T	Torque	N·m	lb·ft	ML^2t^{-2}
u	Velocity, Velocity component	m/sec	ft/sec	Lt^{-1}
u	Peripheral speed	m/sec	ft/sec	Lt^{-1}
u	Internal energy per unit mass	J/kg	ft·lb/lb$_m$	L^2t^{-2}
u_*	Shear stress velocity	m/sec	ft/sec	Lt^{-1}
U	Internal energy	J	Btu	ML^2t^{-2}
v	Velocity, velocity component	m/sec	ft/sec	Lt^{-1}
v_s	Specific volume	m^3/kg	ft^3/lb$_m$	$M^{-1}L^3$
V	Volume	m^3	ft^3	L^3
V	Volumetric flow rate	m^3/sec	ft^3/sec	L^3t^{-1}
V	Velocity	m/sec	ft/sec	Lt^{-1}
w	Velocity component	m/sec	ft/sec	Lt^{-1}
w	Work per unit mass	J/kg	ft·lb/lb$_m$	L^2t^{-2}
W	Work per unit time	J/sec	ft·lb/sec	ML^2t^{-3}
W_s	Shaft work	m·N	ft·lb	ML^2t^{-2}
W	Weight	N	lb	MLt^{-2}
We	Weber number	—	—	—
x	Distance	m	ft	L
y	Distance, depth	m	ft	L
Y	Weir height	m	ft	L
z	Vertical distance	m	ft	L

Greek Symbols

α	Angle, coefficient	—	—	—
β	Blade angle	—	—	—
Γ	Circulation	m^2	ft^2	L^2t^{-1}
ύ	Vector operator	1/m	1/ft	L^{-1}
γ	Specific weight	N/m^3	lb/ft^3	$ML^{-2}t^{-2}$
δ	Boundary layer thickness	m	ft	L
ε	Kinematic eddy viscosity	m^2/sec	ft^2/sec	L^2t^{-1}
ε	Roughness height	m	ft	L
η	Eddy viscosity	$N{\cdot}sec/m^2$	$lb{\cdot}sec/ft^2$	$ML^{-1}t^{-1}$
η	Head ratio	—	—	—
η	Efficiency	—	—	—
Θ	Angle	—	—	—
κ	Universal constant	—	—	—
λ	Scale ratio, undetermined multiplier	—	—	—
μ	Viscosity	$N{\cdot}sec/m^2$	$lb{\cdot}sec/ft^2$	$ML^{-1}t^{-1}$
ν	Kinematic viscosity (= μ/ρ)	m^2/sec	ft^2/sec	L^2t^{-1}
Φ	Velocity potential	m^2/sec	ft^2/sec	L^2t^{-1}
Φ	Function	—	—	—
π	Constant	—	—	—
Π	Dimensionless constant	—	—	—
ρ	Density	kg/m^3	lb_m/ft^3	ML^{-3}
σ	Surface tension	N/m	lb/ft	Mt^{-2}
σ	Cavitation index	—	—	—
τ	Shear stress	Pa	lb/ft^2	ML^1t^2
ψ	Stream function, two dimensions	m/sec	ft/sec	L^2t^{-1}
ψ	Stokes' stream function	m^3/sec	ft^3/sec	L^3t^{-1}
ω	Angular velocity	rad/sec	rad/sec	t^{-1}

Subscripts

c	Critical condition
u	Unit quantities
c.s.	Control surface
c.v.	Control volume
o	Stagnation or standard state condition
1, 2	Inlet and outlet, respectively, of control volume or machine rotor
∞	Upstream condition far away from body, free stream
T	Total pressure
J	Static pressure